Basel Committee
on Banking Supervision

Consultative document

Sound practices for backtesting counterparty credit risk models

Issued for comment by 31 May 2010

April 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 92-9131-824-8

ISBN 92-9197-824-8

Contents

Sound practices for backtesting counterparty credit risk models

Introduction

1. Firms with permission to use internal model methods to calculate counterparty credit risk (CCR) regulatory capital – hereafter referred to as "IMM firms" – are required to carry out on-going validation of their CCR exposure models. The *International Convergence of Capital Measurement and Capital Standards* (ie Basel II) specifies that IMM firms backtest their expected positive exposure (EPE) models, where backtesting is defined as the comparison of the IMM model's output against realised values. Backtesting is only one element of the validation process and recent experience with IMM firms 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 firms' ability to develop their own validation techniques. Instead, this paper identifies issues with the current text of Basel II based on observed industry practice and proposes alternate text for backtesting.

2. An EPE model is designed to produce a distribution of possible exposure values at future time horizons for a particular counterparty. This distribution of exposures is currently used to determine the regulatory capital through application of the regulatory capital metric "effective expected positive exposure" (EEPE). The IMM model itself can consist of a number of component models. Risk Factor models are used to generate paths of possible future market risk factors. Pricing models are used to translate those risk factors on a trade by trade basis into exposure values. Netting and collateral models are then used to aggregate the trade exposures into a measure of counterparty exposure.

3. Backtesting is the comparison of forecasts to realised outcomes. This comparison is either the comparison of a distribution with a single realised value at a point in time, as for market risk factor or exposure distribution backtesting, or the comparison of a single predicted value against some realised value at a point in time, as for backtesting EPE or pricing models. VaR backtesting is a particular example of the former comparison of testing forecast distributions against realised outcomes. This paper argues that the VaR approach is inappropriate for backtesting the internal models used for counterparty credit risk calculations and suggests approaches that are more suitable.

4. The validation requirements for firms' CCR EPE models are set out in Basel II's Annex 4, Section E, paragraphs 42-46. This document revisits these requirements in light of backtesting practice observed in IMM firms and is set out as follows. The general problem of evaluating probabilistic forecast distributions in the context of CCR backtesting is discussed in the section "Evaluating Probabilistic Forecasts". The section on Backtesting then sets out the principle terminology used in IMM backtesting, discusses backtesting and presents guidance on IMM backtesting requirements.

Evaluating Probabilistic Forecasts

5. An IMM model produces a distribution of exposures, with each single exposure forecast assumed to be just as likely as any other. Ideally the forecast distribution of exposure would be a probability forecast, with the forecasting system (ie the IMM model)

being **reliable**. A reliable forecasting system is one for which events occur with an observed relative frequency that is consistent with the forecasted values. For example, for a series of forecast distributions from a reliable forecasting system, 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 CCR 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). Similarly, for a risk factor model the relative frequency of the observed values that lie between the p^{th} percentile and the q^{th} percentile of the risk factor distribution should be consistent with the forecast probability (q-p).

6. If a forecasting system, or IMM model, is not reliable, the frequency of realised events is not consistent with the predicted probability. Consequently, it is misleading to interpret the predicted probability forecast as a probability forecast. For example, given a utility function on the exposure distribution, theoretically optimal decisions can be made by maximising expected utility. When maximising expected utility using forecast distribution from a forecasting system that is not reliable, the resulting decisions are unlikely to be optimal. Throughout this document the term *probabilistic forecast, or probabilistic forecast distribution* is used to denote forecast distributions generated by forecasting systems that have not yet been demonstrated to be reliable.

7. Reliability is a property of the forecast system (the EPE model) and is defined in terms of the observed realisations (the observed exposures and the market risk factors). Reliability does not require that the data used to test the model are independent. Independent data makes it easier to determine the methodologies used to determine reliability, but independent data are not required to test for reliability. If the model is reliable, events should occur as frequently as expected, the relative frequency of events lying beyond the 95^{th} percentile should be consistent with 5% forecast probability, the relative frequency of events lying between the 47^{th} and the 62^{nd} percentile should be consistent with 15% forecast probability. Consequently, overlapping data windows can be used to determine the reliability of EPE models. This issue is discussed further in paragraph 21.

8. Value-at-Risk backtesting, based on counting the exceptions beyond the 99^{th} percentile, is a test of reliability for one part of the probabilistic forecast, that part lying beyond the 99^{th} percentile. If significantly more, or less, than 1 percent of observations lies outside this region, the hypothesis that the probabilistic forecasting system is reliable can be rejected. Note, however, that a probabilistic forecast might not fail a test of reliability at the 99^{th} -100^{th} percentile, but yet still not be reliable at other percentiles. Regardless of the metric used to determine CCR capital, the IMM model produces a distribution of exposure and the model should be evaluated on its ability to produce a distribution. For this reason, it is important to test the quality of the whole distribution and not just one percentile.

Guidance: Backtesting of forecast distributions produced by EPE models and market risk factor models needs to be performed on the entire forecast distribution.

Guidance: The validation requirements as set out in Basel II for EPE Models should not make reference to VaR requirements and instead the qualitative standards set out in paragraph 718 (LXXIV) should be transposed into the validation requirements for EPE models and the language adapted where required.

9. The tests by which firms are able to reject the IMM model as a reliable forecasting system is determined by the firms themselves. Supervisors will require that firms are able to justify the criteria by which model performance is judged and, in particular, where thresholds are

used to define acceptable or unacceptable performance those thresholds must be demonstrably conservative so that exposure is not under-estimated.

Backtesting

10. This section defines the terminology and discusses a number of issues around backtesting. Particular attention is given to those terms used to describe the data sets on which backtesting is carried out.

Definitions

11. **Backtesting** is used to refer to any 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.

12. 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. In order to calculate EEPE daily, IMM firms must initialise their models daily.

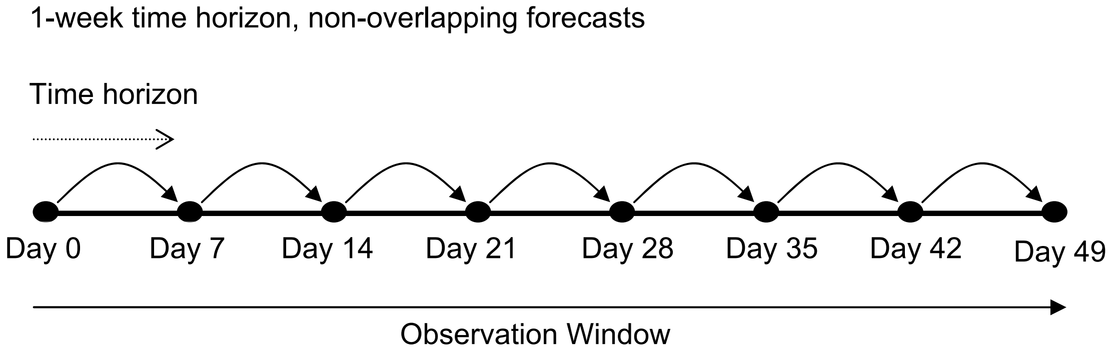
13. Each forecast distribution has a **time horizon**, the time between initialisation and the realisation of the forecast. A forecast initialised on January 1st that realises on January 15th has a 14 day time horizon, a two week forecast. Note that forecasts with different time horizons can have the 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.

14. Backtesting is a statistical test with the significance of any result depending on the amount 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 can be put together in a number of ways. Backtesting using data from a single counterparty over a short period of time may not produce a meaningful conclusion about the quality of the EPE models and its sub-components used to generate that exposure. Firms with advanced model permission have addressed the data requirement problem by aggregating backtesting data across a number of dimensions. The possible dimensions are discussed below.

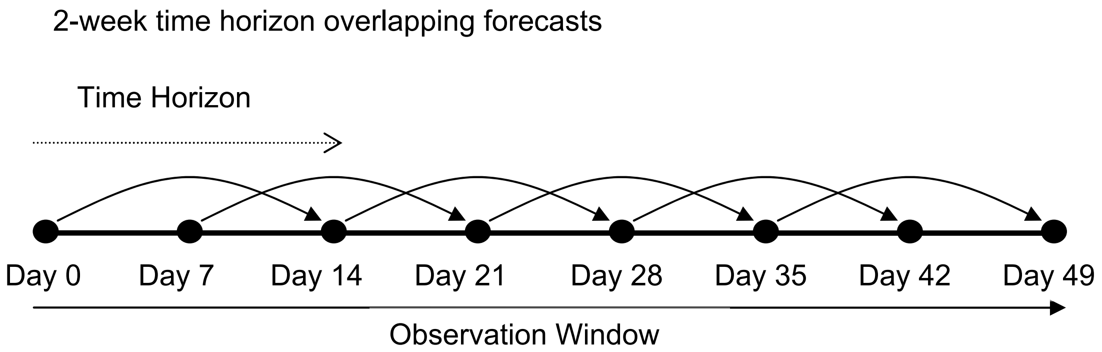
15. The backtesting data set can be aggregated over time, over trades/risk factors or over both time and trades/risk factors. The time period over which data is aggregated is referred to as the **observation window**. There are a number of methodologies for generating a backtesting data set over a given observation window. A selection of frequently used methodologies are set out below.

16. **Non-overlapping, fixed time horizon, aggregating over initialisation dates** data set. In the example below the backtesting data set is constructed by taking, say, 1-week time horizon forecasts initialised on dates 1 week apart. Note that as the time horizon increases to the levels required by IMM backtesting, the effective period of the observation window must increase in order to maintain the same number of data points. Large observation windows are required in order to achieve 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.

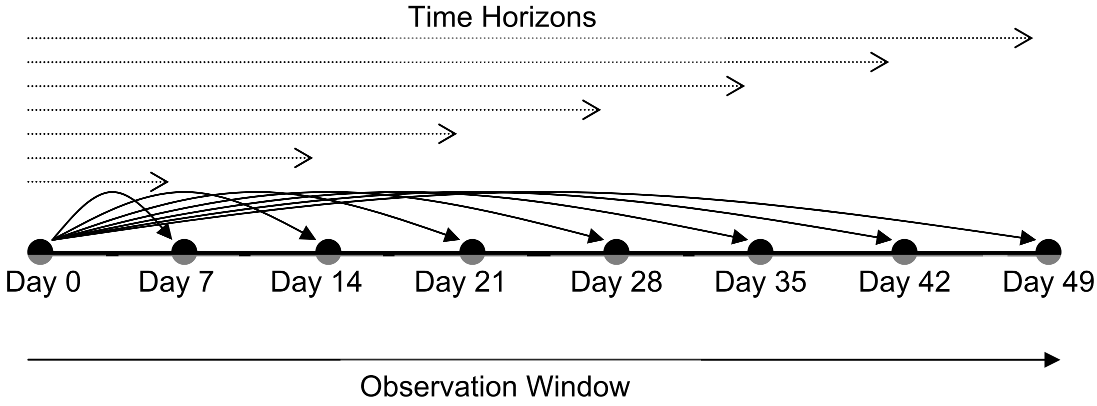


17. **Overlapping, fixed time horizon aggregating over initialisation dates** data set. In the example below the backtesting data set is constructed by taking 2 week ahead forecasts initialised on dates 1 week apart. The data are no longer independent and appropriate tests need to be designed. Note that although the non-independence of data is expected to result in exceptions on day 14 being likely to be followed by an exception on day 21, the converse is also true, with non-exceptions on day 14 likely to be followed by a non-exception on day 21. Regardless of non-independence of the data, the forecasts are still expected to be such that exceptions occur as frequently as expected and while apparently poor performance may be explained by the impact of non-independent data in the analysis, an IMM firm's backtesting programme should not disregard exceptions because of expected non-independence effects.



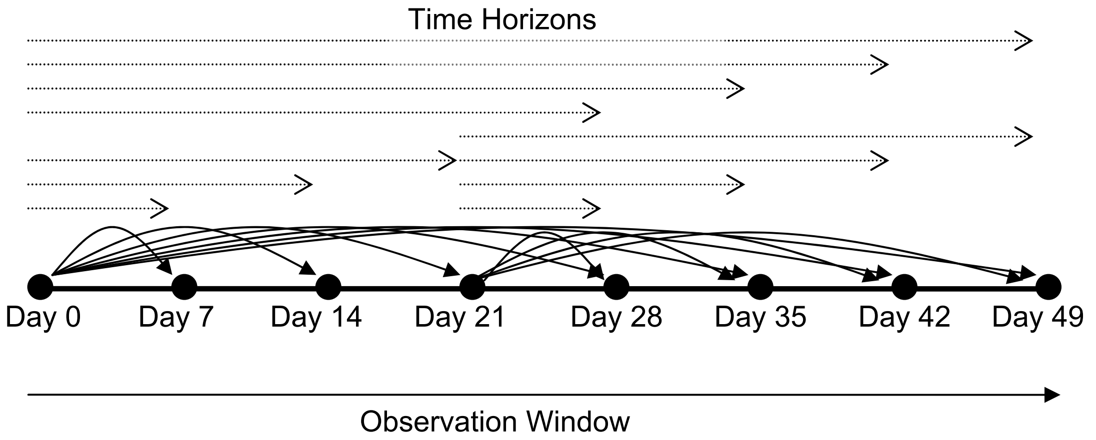
18. **Aggregation over time horizons.** In the example below the backtesting data set is constructed by taking forecasts with different time horizons, initialised on the same date. The assessment of model performance is carried out on the model as parameterised on Day 0, ie any performance benefits (or detriment) resulting from re-parameterisations of the model can not be determined. It is usual for firms taking this approach to aggregate over all time horizons, ie to take forecasts, initialised on the same day whose realisation dates are 1 day apart. This approach to data aggregation increases the effects of non-independence.

Aggregation over time horizons



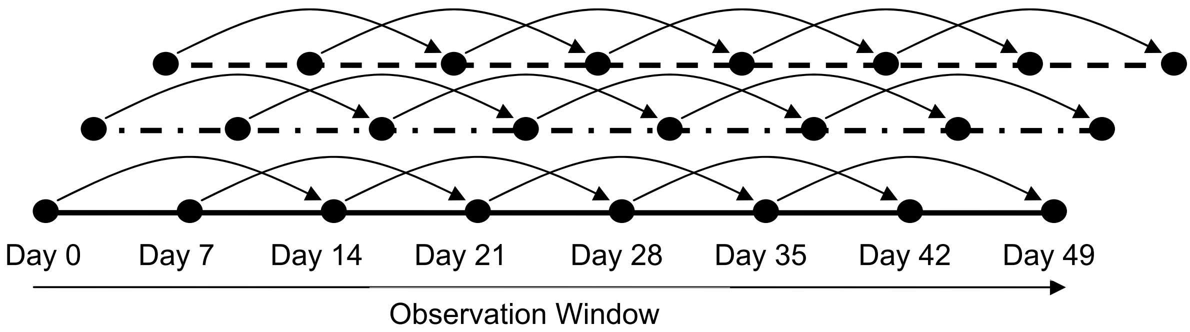
19. **Aggregation over time horizons and initialisation dates.** The backtesting data is constructed by taking forecasts with different time horizons initialised on a number of dates.

Aggregation over time horizons and initialisation dates



20. **Aggregation over trades/risk factors/counterparties.** The data set is constructed by taking data from a number of risk factors (eg USD GBP, USD EUR, USD CHF FX rates), or trades (equity forward, equity option) or counterparty over a given observation window. The methodology for aggregating over time can vary, using either aggregation over initialisation dates or time horizons or both. Note that if the observation window is too small the assessment of model performance depends very much on the date that the data is collected. The resulting assessment is unlikely to be robust.

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



Time Horizons

21. Market risk factor simulation models often assume a fixed volatility parameter, which may not be appropriate in periods of market turbulence. Aggregation Over-Time Horizon backtesting can not distinguish performance at different time horizons and would not enable a firm to identify to what degree assumptions regarding a fixed volatility parameter impact the quality of risk factor forecasts. In order to be able to assess the performance of exposure calculations out to different time horizons IMM firms have to validate for a distinct number of time horizons.

Guidance: The Validation of EPE models and all the relevant models that input into the calculation of EPE must be performed separately for a number of distinct time horizons.

Aggregation of Data

22. Basel II is unclear on the level of data aggregation that is acceptable. Poor performance at the aggregate level will likely be driven by poor performance at the individual risk factor level. However, aggregation over all trades or market risk factors may result in the poor performance of particular trade type or market factor model being masked by acceptable performance elsewhere. Basel II states that the “performance of the forecasting model for market risk factors must be validated over a long time horizon”, but it is not clear what constitutes a market risk factor model. A market risk factor model can be taken to consist of both the model structure (ie a 1-factor model or a 2-factor model) and the methodology used to calibrate that model. Under this definition, the model (structure and calibration) that simulates all Equity prices, for example, is taken to be a market risk factor model. Another definition of a market risk factor model can be taken to be the model structure and the particular parameter values used to simulate a particular risk factor. This more refined definition may well result in data constraint issues. The definition of a market risk factor model for backtesting is taken to be the structure and calibration methodology in the validation requirements of firms with advanced model permissions. As a consequence, firms 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 firms must not be blind to systematic poor performance in a particular risk factor. Firms should assess the appropriateness of the risk factor model on an aggregate basis but still be able to identify poor performance in individual risk factors.

Guidance: 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.

23. In order to achieve meaningful results across-trade aggregation of backtesting data requires that a number of initialisation dates are included. At the extreme, if an assessment of model performance using across trade aggregation is carried out based on forecasts initialised on a single day then the results may likely depend on the particular market condition on either the date that the forecasts were initialised, or the date that the forecasts were 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 must be made using forecasts initialised on a number of historical dates.

Frequency of Backtesting Analysis

24. The frequency with which a firm must carry out backtesting for the purposes of on-going validation is determined by its supervisor, as set out in Basel II's Annex 4 (paragraph 43). Basel II's text requires regular backtesting of representative portfolios but does not require, for example, that market risk factor models be backtested regularly. Basel II should require representative counterparty portfolios and risk factor historical backtesting to be conducted at regular intervals by its supervisor.

Guidance: 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.

25. A number of IMM firms based their backtesting on observation windows spanning a number of years. This assessment of performance allows firms to net performance over benign and turbulent market periods, can mask current poor performance and runs the risk that the backtesting framework is not sufficiently responsive to the model's performance.

26. Paragraph 44 of Annex 4 of Basel II states that:

"These realised exposures would then be compared with the model's forecast distribution at various time horizons. The above must be repeated for several historical dates covering a wide range of market conditions (eg rising rates, falling rates, quiet markets, volatile markets)."

27. It is unclear whether the current Basel text refers to backtesting requirements for initial validation or backtesting requirements for on-going validation. Backtesting as an on-going validation should be an assessment of the current performance of the EPE model and as such should be based on a comparison of recent forecasts to their realised values. Firms are free to carry out an assessment of their counterparty credit risk models using large observation windows for their own validation purposes, but supervisors need to assess the recent performance of the EPE model.

Guidance: Backtesting of EPE and all the relevant models that input into the calculation of EPE must be based on recent performance.

28. 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.

29. The frequency with which IMM models are re-parameterised is a feature of an IMM models that should be assessed as part of the on-going validation. IMM firms update the parameters of their models with varying frequency. During the recent market turbulence some firms might well have benefitted from more frequent updates of parameters. IMM firms need to demonstrate that their schedules for re-parameterisation are appropriate.

Guidance: The frequency with which the parameters of an EPE model are updated needs be assessed as part of the on-going validation process.

Backtesting Criteria

30. The criteria for passing or failing VaR backtesting are well defined and embedded in firms' and regulatory practices. The criteria for passing or failing EPE model backtesting are not currently prescribed and given the range of tests that firms can apply it would not be practical to prescribe pass/fail criteria. Basel II states that "Significant differences between the realised exposures and the model's forecast distribution could indicate a problem with the model or the underlying data that the supervisor would require the bank to correct." As currently worded, Basel II does not empower supervisors to require action given significant differences between realised market risk factors and market risk factor model's forecast distribution.

31. The requirements for passing/failing backtesting are undefined given the range of possible methods. At the same time, firms should have criteria in place for their own backtesting programmes that determines whether the performance of their models is acceptable or not. Moreover, these criteria need to be developed independently of observed results and assessed by the firm's supervisors as appropriate.

Guidance: 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.

32. IMM firms have faced considerable difficulties in specifying appropriate criteria to define acceptable and unacceptable performance when backtesting their IMM models. Those firms 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 firms have used this to justify increasing the thresholds that define acceptable performance. While it is acceptable to take into account the number of data points in determining performance criteria, the non-independence of the data should not be used to justify relaxing the criteria for acceptable models, which in turn can mask poor performance. Serial dependence of the data impacts exceptions at a particular percentile both positively and negatively. Given poor performance, firms 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.

Representative Portfolios

33. Basel II states that "Static historical backtesting on representative counterparty portfolios must be part of the model validation process." Firm's with advanced EPE model

method permission 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 clear idea of what constitutes good practice for determining representative portfolios. It is important that the trades chosen for a firm's representative counterparty portfolio are pre-defined and consistently chosen. To ensure consistency of representative counterparty portfolios firms should set out what constitutes a representative portfolio for the purposes of EPE model backtesting.

Guidance: Firms need to define what constitutes a representative counterparty portfolio for the purposes of carrying out EPE model backtesting.

34. Basel II states that "… representative portfolios must be chosen based on their sensitivity to the material risk factor and correlations to which the bank is exposed." The backtesting of portfolios is the principal way in which firms test their ability to model the relationship between risk factors and the different tenors of the same risk factor. The correlation benefits of IMM models are significant and in order to justify this benefit IMM firms need to be able to demonstrate, through backtesting, that their models appropriately capture the relationship between risk factors and between tenors of the same risk factor. To test these relationships, IMM firms need to construct hypothetical portfolios designed to test risk factor model assumptions and the relationships between risk factors.

Guidance: IMM firms need to conduct hypothetical portfolio backtesting that is designed to test risk factor model assumptions, eg the relationship between tenors of the same risk factor, and the modelled relationships between risk factors.

Materiality of Results

35. 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. Basel II states that "Significant differences between the realised exposures and the model's forecast distribution could indicate a problem with the model or the underlying data that the supervisor would require the bank to correct. Under such circumstances, supervisors may require additional capital."

36. The definition of what constitutes a "significant difference" between realised exposures and the model's forecast distribution is not well defined. It is unclear whether the assessment should be based on a whole portfolio basis or a counterparty basis. Since CCR capital is computed on a netting set basis it is expected that firms should be assessing whether or not the firm level and netting set level exposures are appropriate.

Guidance: Firms need to assess whether or not the firm level and netting set level exposure calculations are appropriate.

Real Trade Backtesting vs Hypothetical Trade Backtesting

37. Firms with advanced model method permission to calculate counterparty credit risk capital have adopted one of two general approaches in the data used to carry out backtesting. Some firms use real trade data, while others have developed stand-alone 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, hypothetical portfolios.

38. Real-trade backtesting is subject to a dynamic portfolio. As a result the number of trades that persist in the backtesting sample decreases with increasing time horizon. Regardless, firms need to demonstrate that their EPE models are capable of estimating counterparty credit exposure out to long time horizons.

Time Horizons

39. Basel II requires that the forecasting model for market risk factors must be validated over a long time horizon, but does not specify the minimum time horizons for validating other IMM model components. The precise definition of what constitutes a long time horizon is not specified. Basel II should specify that backtesting be carried out to time horizons of at least one year, commensurate with the time horizon of the exposure calculation used to determine EEPE. This minimum requirement is for both margined and un-margined exposures since margining reduces exposures but does not reduce the tenor of exposure. Furthermore, for margined portfolios with optionality the near term exposure might be materially different from the long term exposure.

Guidance: 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.

40. IMM firms also use their exposure models to risk trades out to time horizons beyond 1 year. By requiring firms to validate their EPE models and relevant components to one year time horizons, at a minimum, runs the risk that firms will optimise model performance to that time horizon to the detriment of longer time horizon performance. Basel II does not currently specify whether or not the maturity of trades covered by an IMM waiver needs to be restricted based on the time horizons over which the firm can produce a robust assessment of model performance.

Guidance: 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.

Initial Validation

41. Basel II's Annex 4, Section E is concerned with Model Validation in its entirety. Paragraph 42, for example can be interpreted as both requirements for initial validation and for on-going validation. Explicit text should be put in place to address the validation of new EPE models or new models that input into the calculation of EPE. This should include new products for which a new pricing model has been developed. It is proposed that the following requirements are made, regarding backtesting, for initial validation.

Guidance: 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.

Non-Simulation Models

42. A number of firms with advanced model methods for CCR have developed non-simulation models that are more conservative than Effective EPE. Basel II currently states that " … The degree of relative conservatism will be assessed upon initial supervisory

approval and subject to periodic validation." The validation of alternative EPE models needs to be combined with that of the simulation model.

Guidance: Under the internal model method, a measure that is more conservative than Effective EPE (eg a measure based on peak rather than average exposure) for every counterparty may be used in place of alpha times EEPE with the prior approval of the supervisor. The degree of relative conservatism will be assessed upon initial supervisory approval and at regular intervals in conjunction with other EPE models. The assessment needs to cover all counterparties. The firm must have an unambiguous definition of what constitutes acceptable performance for these models and a documented process in place for remediating poor performance.

Pricing Models

43. Basel II currently requires "pricing models used to calculate counterparty exposure for a given scenario of future shocks to market risk factors must be tested as part of the model validation process. These pricing models may be different from those used to calculate VaR over a short horizon. Pricing models for options must account for the nonlinearity of option value with respect to market risk factors."

44. The text does not distinguish between the validation required prior to implementation and the on-going validation required to monitor that the model assumptions remain valid and that the data feeds are correct. It is proposed that additional text is provided that requires IMM firms to carry out an ongoing validation, on a trade by trade basis, of their counterparty credit risk pricing models.

45. Basel II does not currently specify how frequently the validation of pricing models should be. Those pricing models where there is a mismatch between front office and credit risk recalibration frequencies should be monitored closely.

Guidance: Firms are required to carry out an ongoing validation of their counterparty credit risk pricing models against an appropriate benchmark at an appropriate frequency.