Appendix CA-15

Supervisory Framework for the Use of "Backtesting" in Conjunction with the Internal Models Approach to Market Risk Capital Requirements

I.   Introduction

1. This Appendix presents the framework for incorporating backtesting into the internal models approach to market risk capital requirements.

2. Banks that have adopted an internal model-based approach to market risk measurement must routinely compare daily profits and losses with model-generated risk measures to gauge the quality and accuracy of their risk measurement systems.

3. The essence of backtesting is the comparison of actual trading results with model-generated risk measures. If this comparison is close enough, the backtest raises no issues regarding the quality of the risk measurement model. In some cases, however, the comparison uncovers sufficient differences that problems almost certainly must exist, either with the model or with the assumptions of the backtest. In between these two cases is a grey area where the test results are, on their own, inconclusive.

4. Backtesting offers the best opportunity for incorporating suitable incentives into the internal models approach in a manner that is consistent and that will cover a variety of circumstances. The framework outlined in this appendix strikes an appropriate balance between recognition of the potential limitations of backtesting and the need to put in place appropriate incentives. At the same time, the CBB recognises that the techniques for risk measurement and backtesting are still evolving, and the CBB is committed to incorporating important new developments in these areas into these regulations.

5. The remainder of this appendix describes the backtesting framework that is to accompany the internal models capital requirement. The aim of this framework is the promotion of more rigorous approaches to backtesting and the supervisory interpretation of backtesting results. The next section deals with the nature of the backtests themselves, while the section that follows concerns the supervisory interpretation of the results and sets out the agreed standards of the CBB in this regard.

II.   Description of the backtesting framework

6. The backtesting programs typically consist of a periodic comparison of the bank's daily value-at-risk measures with the subsequent daily profit or loss ("trading outcome"). The value-at-risk measures are intended to be larger than all but a certain fraction of the trading outcomes, where that fraction is determined by the confidence level of the value-at-risk measure. Comparing the risk measures with the trading outcomes simply means that the bank counts the number of times that the risk measures were larger than the trading outcome. The fraction actually covered can then be compared with the intended level of coverage to gauge the performance of the bank's risk model. In some cases, this last step is relatively informal, although there are a number of statistical tests that may also be applied.

7. The framework for backtesting in this appendix involves all of the steps identified in the previous paragraph, and attempts to set out as consistent an interpretation of each step as is feasible without imposing unnecessary burdens. Under the value-at-risk framework, the risk measure is an estimate of the amount that could be lost on a set of positions due to general market movements over a given holding period, measured using a specified confidence level.

8. The backtests to be applied compare whether the observed percentage of outcomes covered by the risk measure is consistent with a 99% level of confidence. That is, they attempt to determine if a bank's 99th percentile risk measures truly cover 99% of the firm's trading outcomes. While it can be argued that the extreme-value nature of the 99th percentile makes it more difficult to estimate reliably than other, lower percentiles, the CBB has concluded that it is important to align the test with the confidence level specified in the chapter CA-14.

9. An additional consideration in specifying the appropriate risk measures and trading outcomes for backtesting arises because the value-at-risk approach to risk measurement is generally based on the sensitivity of a static portfolio to instantaneous price shocks. That is, end-of-day trading positions are input into the risk measurement model, which assesses the possible change in the value of this static portfolio due to price and rate movements over the assumed holding period.

10. While this is straightforward in theory, in practice it complicates the issue of backtesting. For instance, it is often argued that value-at-risk measures cannot be compared against actual trading outcomes, since the actual outcomes will inevitably be "contaminated" by changes in portfolio composition during the holding period. According to this view, the inclusion of fee income together with trading gains and losses resulting from changes in the composition of the portfolio should not be included in the definition of the trading outcome because they do not relate to the risk inherent in the static portfolio that was assumed in constructing the value-at-risk measure.

11. This argument is persuasive with regard to the use of value-at-risk measures based on price shocks calibrated to longer holding periods. That is, comparing the ten-day, 99th percentile risk measures from the internal models capital requirement with actual ten-day trading outcomes would probably not be a meaningful exercise. In particular, in any given ten day period, significant changes in portfolio composition relative to the initial positions are common at major trading institutions. For this reason, the backtesting framework described here involves the use of risk measures calibrated to a one-day holding period. Other than the restrictions mentioned in this appendix, the test would be based on how banks model risk internally.

12. Given the use of one-day risk measures, it is appropriate to employ one-day trading outcomes as the benchmark to use in the backtesting program. The same concerns about "contamination" of the trading outcomes discussed above continue to be relevant, however, even for one-day trading outcomes. That is, there is a concern that the overall one-day trading outcome is not a suitable point of comparison, because it reflects the effects of intraday trading, possibly including fee income that is booked in connection with the sale of new products.

13. On the one hand, intra-day trading will tend to increase the volatility of trading outcomes, and may result in cases where the overall trading outcome exceeds the risk measure. This event clearly does not imply a problem with the methods used to calculate the risk measure; rather, it is simply outside the scope of what the value-at-risk method is intended to capture. On the other hand, including fee income may similarly distort the backtest, but in the other direction, since fee income often has annuity-like characteristics.

14. Since this fee income is not typically included in the calculation of the risk measure, problems with the risk measurement model could be masked by including fee income in the definition of the trading outcome used for backtesting purposes.

15. Some have argued that the actual trading outcomes experienced by the bank are the most important and relevant figures for risk management purposes, and that the risk measures should be benchmarked against this reality, even if the assumptions behind their calculations are limited in this regard. Others have also argued that the issue of fee income can be addressed sufficiently, albeit crudely, by simply removing the mean of the trading outcomes from their time series before performing the backtests. A more sophisticated approach would involve a detailed attribution of income by source, including fees, spreads, market movements, and intra-day trading results.

16. To the extent that the backtesting program is viewed purely as a statistical test of the integrity of the calculation of the value-at-risk measure, it is clearly most appropriate to employ a definition of daily trading outcome that allows for an "uncontaminated" test. To meet this standard, banks should develop the capability to perform backtests based on the hypothetical changes in portfolio value that would occur were end-of-day positions to remain unchanged.

17. Backtesting using actual daily profits and losses is also a useful exercise since it can uncover cases where the risk measures are not accurately capturing trading volatility in spite of being calculated with integrity.

18. For these reasons, the CBB requires banks to develop the capability to perform backtests using both hypothetical and actual trading outcomes. In combination, the two approaches are likely to provide a strong understanding of the relation between calculated risk measures and trading outcomes.

19. The next step in specifying the backtesting program concerns the nature of the backtest itself, and the frequency with which it is to be performed. The framework adopted by the CBB, which is also the most straightforward procedure for comparing the risk measures with the trading outcomes, is simply to calculate the number of times that the trading outcomes are not covered by the risk measures ("exceptions"). For example, over 200 trading days, a 99% daily risk measure should cover, on average, 198 of the 200 trading outcomes, leaving two exceptions.

20. With regard to the frequency of the backtest, the desire to base the backtest on as many observations as possible must be balanced against the desire to perform the test on a regular basis. The backtesting framework to be applied entails a formal testing and accounting of exceptions on a quarterly basis using the most recent twelve months of data.

21. Using the most recent twelve months of data yields approximately 250 daily observations for the purposes of backtesting. The national supervisor will use the number of exceptions (out of 250) generated by the bank's model as the basis for a supervisory response. In many cases, there will be no response. In other cases, the CBB will initiate a dialogue with the bank to determine if there is a problem with a bank's model. In the most serious cases, the CBB may impose an increase in a bank's capital requirement or disallow use of the model.

22. The appeal of using the number of exceptions as the primary reference point in the backtesting process is the simplicity and straightforwardness of this approach. From a statistical point of view, using the number of exceptions as the basis for appraising a bank's model requires relatively few strong assumptions. In particular, the primary assumption is that each day's test (exception/no exception) is independent of the outcome of any of the others.

23. The CBB of course recognises that tests of this type are limited in their power to distinguish an accurate model from an inaccurate model. To a statistician, this means that it is not possible to calibrate the test so that it correctly signals all the problematic models without giving false signals of trouble at many others. This limitation has been a prominent consideration in the design of the framework presented here, and is also prominent among the considerations of CBB in interpreting the results of a bank's backtesting program. However, the CBB does not view this limitation as a decisive objection to the use of backtesting. Rather, conditioning supervisory standards on a clear framework, though limited and imperfect, is seen as preferable to a purely judgmental standard or one with no incentive features whatsoever.

III. Supervisory framework for the interpretation of backtesting results

A. Description of three-zone approach

24. It is with the statistical limitations of backtesting in mind that the CBB is introducing a framework for the supervisory interpretation of backtesting results that encompasses a range of possible responses, depending on the strength of the signal generated from the backtest. These responses are classified into three zones, distinguished by colours into a hierarchy of responses. The green zone corresponds to backtesting results that do not themselves suggest a problem with the quality or accuracy of a bank's model. The yellow zone encompasses results that do raise questions in this regard, but where such a conclusion is not definitive. The red zone indicates a backtesting result that almost certainly indicates a problem with a bank's risk model.

25. The CBB has agreed to standards regarding the definitions of these zones in respect of the number of exceptions generated in the backtesting program, and these are set forth below. To place these definitions in proper perspective, however, it is useful to examine the probabilities of obtaining various numbers of exceptions under different assumptions about the accuracy of a bank's risk measurement model.

B. Statistical considerations in defining the zones

26. Three zones have been delineated and their boundaries chosen in order to balance two types of statistical error: (1) the possibility that an accurate risk model would be classified as inaccurate on the basis of its backtesting result, and (2) the possibility that an inaccurate model would not be classified that way based on its backtesting result.

27. Table 1 reports the probabilities of obtaining a particular number of exceptions from a sample of 250 independent observations under several assumptions about the actual percentage of outcomes that the model captures (that is, these are binomial probabilities). For example, the left-hand portion of Table 1 reports probabilities associated with an accurate model (that is, a true coverage level of 99%). Under these assumptions, the column labelled "exact" reports that exactly five exceptions can be expected in 6.7% of the samples.

28. The right-hand portion of Table 1 reports probabilities associated with several possible inaccurate models, namely models whose true levels of coverage are 98%, 97%, 96%, and 95%, respectively. Thus, the column labelled "exact" under an assumed coverage level of 97% shows that five exceptions would then be expected in 10.9% of the samples.

29. Table 1 also reports several important error probabilities. For the assumption that the model covers 99% of outcomes (the desired level of coverage), the table reports the probability that selecting a given number of exceptions as a threshold for rejecting the accuracy of the model will result in an erroneous rejection of an accurate model ("type 1" error). For example, if the threshold is set as low as one exception, then accurate models will be rejected fully 91.9% of the time, because they will escape rejection only in the 8.1% of cases where they generate zero exceptions. As the threshold number of exceptions is increased, the probability of making this type of error declines.

30. Under the assumptions that the model's true level of coverage is not 99%, Table 1 reports the probability that selecting a given number of exceptions as a threshold for rejecting the accuracy of the model will result in an erroneous acceptance of a model with the assumed (inaccurate) level of coverage ("type 2" error). For example, if the model's actual level of coverage is 97%, and the threshold for rejection is set at seven or more exceptions, the table indicates that this model would be erroneously accepted 37.5% of the time.

31. In interpreting the information in Table 1, it is also important to understand that although the alternative models appear close to the desired standard in probability terms (97% is close to 99%), the difference between these models in terms of the size of the risk measures generated can be substantial. That is, a bank's risk measure could be substantially less than that of an accurate model and still cover 97% of the trading outcomes. For example, in the case of normally distributed trading outcomes, the 97th percentile corresponds to 1.88 standard deviations, while the 99th percentile corresponds to 2.33 standard deviations, an increase of nearly 25%. Thus, the supervisory desire to distinguish between models providing 99% coverage, and those providing say, 97% coverage, is a very real one.

C.   Definition of the green, yellow, and red zones

32. The results in Table 1 also demonstrate some of the statistical limitations of backtesting. In particular, there is no threshold number of exceptions that yields both a low probability of erroneously rejecting an accurate model and a low probability of erroneously accepting all of the relevant inaccurate models.

33. Given these limitations, the CBB has classified outcomes into three categories. In the first category, the test results are consistent with an accurate model, and the possibility of erroneously accepting an inaccurate model is low (green zone). At the other extreme, the test results are extremely unlikely to have resulted from an accurate model, and the probability of erroneously rejecting an accurate model on this basis is remote (red zone). In between these two cases, however, is a zone where the backtesting results could be consistent with either accurate or inaccurate models, and the banks are encouraged to present additional information about its model before taking action (yellow zone).

34. Table 2 sets out the boundaries for these zones and the presumptive supervisory response for each backtesting outcome, based on a sample of 250 observations. For other sample sizes, the boundaries should be deduced by calculating the binomial probabilities associated with true coverage of 99%, as in Table 1. The yellow zone begins at the point such that the probability of obtaining that number or fewer exceptions equals or exceeds 95%. Table 2 reports these cumulative probabilities for each number of exceptions. For 250 observations, it can be seen that five or fewer exceptions will be obtained 95.88% of the time when the true level of coverage is 99%. Thus, the yellow zone begins at five exceptions.

35 Similarly, the beginning of the red zone is defined as the point such that the probability of obtaining that number or fewer exceptions equals or exceeds 99.99%. Table 2 shows that for a sample of 250 observations and a true coverage level of 99%, this occurs with ten exceptions.

D.   The green zone

36. The green zone needs little explanation. Since a model that truly provides 99% coverage would be quite likely to produce as many as four exceptions in a sample of 250 outcomes, there is little reason for concern raised by backtesting results that fall in this range. This is reinforced by the results in Table 1, which indicate that accepting outcomes in this range leads to only a small chance of erroneously accepting an inaccurate model.

E.   The yellow zone

37. The range from five to nine exceptions constitutes the yellow zone. Outcomes in this range are plausible for both accurate and inaccurate models, although Table 1 suggests that they are generally more likely for inaccurate models than for accurate models. Moreover, the results in Table 1 indicate that the presumption that the model is inaccurate should grow as the number of exceptions increases in the range from five to nine.

38. Within the yellow zone, the number of exceptions will guide the size of potential supervisory increases in a firm's capital requirement. Table 2 sets out the CBB agreed guidelines for increases in the multiplication factor applicable to the internal models capital requirement, resulting from backtesting results in the yellow zone.

39. These guidelines help in maintaining the appropriate structure of incentives applicable to the internal models approach. In particular, the potential supervisory penalty increases with the number of exceptions. The results in Table 1 generally support the notion that nine exceptions is a more troubling result than five exceptions, and these steps are meant to reflect that.

40. These particular values reflect the general idea that the increase in the multiplication factor should be sufficient to return the model to a 99th percentile standard. For example, five exceptions in a sample of 250 imply only 98% coverage. Thus, the increase in the multiplication factor should be sufficient to transform a model with 98% coverage into one with 99% coverage. Needless to say, precise calculations of this sort require additional statistical assumptions that are not likely to hold in all cases. For example, if the distribution of trading outcomes is assumed to be normal, then the ratio of the 99th percentile to the 98th percentile is approximately 1.14, and the increase needed in the multiplication factor is therefore approximately 0.40 for a scaling factor of 3. If the actual distribution is not normal, but instead has "fat tails", then larger increases may be required to reach the 99th percentile standard. The concern about fat tails was also an important factor in the choice of the specific increments set out in Table 2.

41. It is important to stress, however, that these increases are not meant to be purely automatic. The results in Table 1 indicate that results in the yellow zone do not always imply an inaccurate model, and the CBB has no interest in penalising banks solely for bad luck. Nevertheless, to keep the incentives aligned properly, backtesting results in the yellow zone should generally be presumed to imply an increase in the multiplication factor unless the bank can demonstrate to CBB that such an increase is not warranted.

42. In other words, the burden of proof in these situations is not on the CBB to prove that a problem exists, but rather is on the bank to prove that their model is fundamentally sound. In such a situation, there are many different types of additional information that might be relevant to an assessment of the bank's model.

43. For example, it would then be particularly valuable to see the results of backtests covering disaggregated subsets of the bank's overall trading activities. Many banks that engage in regular backtesting programs break up their overall trading portfolio into trading units organised around risk factors or product categories. Disaggregating in this fashion could allow the tracking of a problem that surfaced at the aggregate level back to its source at the level of a specific trading unit or risk model.

44. Banks should also document all of the exceptions generated from their ongoing backtesting program, including an explanation for the exception. This documentation is important to determining an appropriate supervisory response to a backtesting result in the yellow zone. Banks may also implement backtesting for confidence intervals other than the 99th percentile, or may perform other statistical tests not considered here. Naturally, this information could also prove very helpful in assessing their model.

45. In practice, there are several possible explanations for a backtesting exception, some of which go to the basic integrity of the model, some of which suggest an underspecified or low-quality model, and some of which suggest either bad luck or poor intra-day trading results. Classifying the exceptions generated by a bank's model into these categories can be a very useful exercise.

Basic integrity of the model

(1) The bank's systems simply are not capturing the risk of the positions themselves (e.g. the positions of an overseas office are being reported incorrectly).

(2) Model volatilities and/or correlations were calculated incorrectly (e.g. the computer is dividing by 250 when it should be dividing by 225).

Model's accuracy could be improved

(3) The risk measurement model is not assessing the risk of some instruments with sufficient precision (e.g. too few maturity buckets or an omitted spread).

Bad luck or markets moved in fashion unanticipated by the model

(4) Random chance (a very low probability event).

(5) Markets moved by more than the model predicted was likely (i.e. volatility was significantly higher than expected).

(6) Markets did not move together as expected (i.e. correlations were significantly different than what was assumed by the model).

Intra-day trading

(7) There was a large (and money-losing) change in the bank's positions or some other income event between the end of the first day (when the risk estimate was calculated) and the end of the second day (when trading results were tabulated).

46. In general, problems relating to the basic integrity of the risk measurement model are potentially the most serious. If there are exceptions attributed to this category for a particular trading unit, the plus must apply. In addition, the model may be in need of substantial review and/or adjustment, and the CBB will take appropriate action to ensure that this occurs.

47. The second category of problem (lack of model precision) is one that can be expected to occur at least part of the time with most risk measurement models. No model can hope to achieve infinite precision, and thus all models involve some amount of approximation. If, however, a particular bank's model appears more prone to this type of problem than others, the CBB will impose the plus factor and also consider what other incentives are needed to spur improvements.

48. The third category of problems (markets moved in a fashion unanticipated by the model) should also be expected to occur at least some of the time with value-at-risk models. In particular, even an accurate model is not expected to cover 100% of trading outcomes. Some exceptions are surely the random 1% that the model can be expected not to cover. In other cases, the behaviour of the markets may shift so that previous estimates of volatility and correlation are less appropriate. No value-at-risk model will be immune from this type of problem; it is inherent in the reliance on past market behaviour as a means of gauging the risk of future market movements.

49. Finally, depending on the definition of trading outcomes employed for the purpose of backtesting, exceptions could also be generated by intra-day trading results or an unusual event in trading income other than from positioning. Although exceptions for these reasons would not necessarily suggest a problem with the bank's value-at-risk model, they could still be cause for CBB concern and the plus will be imposed.

50. The extent to which a trading outcome exceeds the risk measure is another relevant piece of information. All else equal, exceptions generated by trading outcomes far in excess of the risk measure are a matter of greater concern than are outcomes only slightly larger than the risk measure.

51. In deciding whether or not to apply increases in a bank's capital requirement, it is envisioned that the CBB can weigh these factors as well as others, including an appraisal of the bank's compliance with applicable qualitative standards of risk management. Based on the additional information provided by the bank, the CBB will decide on the appropriate course of action.

52. In general, the imposition of a higher capital requirement for outcomes in the yellow zone is an appropriate response when the CBB believes the reason for being in the yellow zone is a correctable problem in a bank's model. This can be contrasted with the case of an unexpected bout of high market volatility, which nearly all models may fail to predict. While these episodes may be stressful, they do not necessarily indicate that a bank's risk model is in need of redesign. Finally, in the case of severe problems with the basic integrity of the model, the CBB will consider whether to disallow the use of the model for capital purposes altogether.

F.   The red zone

53. Finally, in contrast to the yellow zone where the CBB may exercise judgement in interpreting the backtesting results, outcomes in the red zone (ten or more exceptions) will generally lead to an automatic presumption that a problem exists with a bank's model. This is because it is extremely unlikely that an accurate model would independently generate ten or more exceptions from a sample of 250 trading outcomes.

54. In general, therefore, if a bank's model falls into the red zone, the CBB will automatically increase the multiplication factor applicable to a firm's model by one (from three to four). The CBB will also begin investigating the reasons why the bank's model produced such a large number of misses, and will require the bank to begin work on improving its model immediately.

55. Although ten exceptions is a very high number for 250 observations, there will on very rare occasions be a valid reason why an accurate model will produce so many exceptions. In particular, when financial markets are subjected to a major regime shift, many volatilities and correlations can be expected to shift as well, perhaps substantially. Unless a bank is prepared to update its volatility and correlation estimates instantaneously, such a regime shift could generate a number of exceptions in a short period of time. In essence, however, these exceptions would all be occurring for the same reason, and therefore the appropriate supervisory reaction might not be the same as if there were ten exceptions, but each from a separate incident. For example, one possible supervisory response in this instance would be to simply require the bank's model to take account of the regime shift as quickly as it can while maintaining the integrity of its procedures for updating the model.

56. This exception will be allowed only under the most extraordinary circumstances.

IV.   Conclusion

57. The above framework is intended to set out a consistent approach for incorporating backtesting into the internal models approach to market risk capital requirements. The goals of this effort have been to build appropriate and necessary incentives into a framework that relies heavily on the efforts of banks themselves to calculate the risks they face, to do so in a way that respects the inherent limitations of the available tools, and to keep the burdens and costs of the imposed procedures to a minimum.

58. The CBB believes that the framework described above strikes the right balance in this regard. Perhaps more importantly, however, the CBB believes that this approach represents the first, and therefore critical, step toward a tighter integration of supervisory guidelines with verifiable measures of bank performance.

Table 1

	Model is accurate			Model is inaccurate: Possible alternative levels of coverage								
Exceptions (out of 250)	Coverage = 99%		Exceptions (out of 250)	Coverage = 98%		Coverage = 97%		Coverage = 96%		Coverage = 95%	
	exact	type 1		exact	type 2	exact	type 2	exact	type 2	exact	type 2
0	8.1%	100.0%	0	0.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
1	20.5%	91.9%	1	3.3%	0.6%	0.4%	0.0%	0.0%	0.0%	0.0%	0.0%
2	25.7%	71.4%	2	8.3%	3.9%	1.5%	0.4%	0.2%	0.0%	0.0%	0.0%
3	21.5%	45.7%	3	14.0%	12.2%	3.8%	1.9%	0.7%	0.2%	0.1%	0.0%
4	13.4%	24.2%	4	17.7%	26.2%	7.2%	5.7%	1.8%	0.9%	0.3%	0.1%
5	6.7%	10.8%	5	17.7%	43.9%	10.9%	12.8%	3.6%	2.7%	0.9%	0.5%
6	2.7%	4.1%	6	14.8%	61.6%	13.8%	23.7%	6.2%	6.3%	1.8%	1.3%
7	1.0%	1.4%	7	10.5%	76.4%	14.9%	37.5%	9.0%	12.5%	3.4%	3.1%
8	0.3%	0.4%	8	6.5%	86.9%	14.0%	52.4%	11.3%	21.5%	5.4%	6.5%
9	0.1%	0.1%	9	3.6%	93.4%	11.6%	66.3%	12.7%	32.8%	7.6%	11.9%
10	0.0%	0.0%	10	1.8%	97.0%	8.6%	77.9%	12.8%	45.5%	9.6%	19.5%
11	0.0%	0.0%	11	0.8%	98.7%	5.8%	86.6%	11.6%	58.3%	11.1%	29.1%
12	0.0%	0.0%	12	0.3%	99.5%	3.6%	92.4%	9.6%	69.9%	11.6%	40.2%
13	0.0%	0.0%	13	0.1%	99.8%	2.0%	96.0%	7.3%	79.5%	11.2%	51.8%
14	0.0%	0.0%	14	0.0%	99.9%	1.1%	98.0%	5.2%	86.9%	10.0%	62.9%
15	0.0%	0.0%	15	0.0%	100.0%	0.5%	99.1%	3.4%	92.1%	8.2%	72.9%

Notes: The table reports both exact probabilities of obtaining a certain number of exceptions from a sample of 250 independent observations under several assumptions about the true level of coverage, as well as type 1 or type 2 error probabilities derived from these exact probabilities.

The left-hand portion of the table pertains to the case where the model is accurate and its true level of coverage is 99%. Thus, the probability of any given observation being an exception is 1% (100% - 99% = 1%). The column labelled "exact" reports the probability of obtaining exactly the number of exceptions shown under this assumption in a sample of 250 independent observations. The column labelled "type 1" reports the probability that using a given number of exceptions as the cut-off for rejecting a model will imply erroneous rejection of an accurate model using a sample of 250 independent observations. For example, if the cut-off level is set at five or more exceptions, the type 1 column reports the probability of falsely rejecting an accurate model with 250 independent observations is 10.8%.

The right-hand portion of the table pertains to models that are inaccurate. In particular, the table concentrates of four specific inaccurate models, namely models whose true levels of coverage are 98%, 97%, 96% and 95% respectively. For each inaccurate model, the "exact" column reports the probability of obtaining exactly the number of exceptions shown under this assumption in a sample of 250 independent observations. The columns labelled "type 2" report the probability that using a given number of exceptions as the cut-off for rejecting a model will imply erroneous acceptance of an inaccurate model with the assumed level of coverage using a sample of 250 independent observations. For example, if the cut-off level is set at five or more exceptions, the type 2 column for an assumed coverage level of 97% reports the probability of falsely accepting a model with only 97% coverage with 250 independent observations is 12.8%.

Table 2

Zone	Number of exceptions	Increase in scaling factor	Cumulative probability
Green Zone	0	0.00	8.11%
	1	0.00	28.58%
	2	0.00	54.32%
	3	0.00	75.81%
	4	0.00	89.22%
Yellow Zone	5	0.40	95.88%
	6	0.50	98.63%
	7	0.65	99.60%
	8	0.75	99.89%
	9	0.85	99.97%
Red Zone	10 or more	1.00	99.99%

Notes: The table defines the green, yellow and red zones that supervisors will use to assess backtesting results in conjunction with the internal models approach to market risk capital requirements. The boundaries shown in the table are based on a sample of 250 observations. For other sample sizes, the yellow zone begins at the point where the cumulative probability equals or exceeds 95%, and the red zone begins at the point where the cumulative probability equals or exceeds 99.99%.

The cumulative probability is simply the probability of obtaining a given number or fewer exceptions in a sample of 250 observations when the true coverage level is 99%. For example, the cumulative probability shown for four exceptions is the probability of obtaining between zero and four exceptions.

Note that these cumulative probabilities and the type 1 error probabilities reported in Table 1 do not sum to one because the cumulative probability for a given number of exceptions includes the possibility of obtaining exactly that number of exceptions, as does the type 1 error probability. Thus, the sum of these two probabilities exceeds one by the amount of the probability of obtaining exactly that number of exceptions.