

# Basel Committee on Banking Supervision

MAR

Calculation of RWA for  
market risk

MAR99

Guidance on use of the  
internal models approach

**Version effective as of  
01 Jan 2022**

Reflects revisions to internal models approach.



BANK FOR INTERNATIONAL SETTLEMENTS

© Bank for International Settlements 2021. All rights reserved.

## Trading desk-level backtesting

- 99.1** An additional consideration in specifying the appropriate risk measures and trading outcomes for profit and loss (P&L) attribution test and backtesting arises because the internally modelled 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.
- 99.2** While this is straightforward in theory, in practice it complicates the issue of backtesting. For instance, it is often argued that neither expected shortfall nor value-at-risk measures can be compared against actual trading outcomes, since the actual outcomes will reflect 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.
- 99.3** This argument is persuasive with regard to the use of risk measures based on price shocks calibrated to longer holding periods. That is, comparing the liquidity-adjusted time horizon 99th percentile risk measures from the internal models capital requirement with actual liquidity-adjusted time horizon trading outcomes would probably not be a meaningful exercise. In particular, in any given multi-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 paper, the test would be based on how banks model risk internally.
- 99.4** 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 programme. 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.

- 99.5** On the one hand, intraday 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 measure 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. 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.
- 99.6** To the extent that backtesting programmes are viewed purely as a statistical test of the integrity of the calculation of the risk measures, it is appropriate to employ a definition of daily trading outcome that allows for an uncontaminated test. To meet this standard, banks must have the capability to perform the tests based on the hypothetical changes in portfolio value that would occur were end-of-day positions to remain unchanged.
- 99.7** Backtesting using actual daily P&Ls 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.
- 99.8** For these reasons, the Committee requires banks to develop the capability to perform these tests 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. The total number of backtesting exceptions for the purpose of the thresholds in [MAR32.9] must be calculated as the maximum of the exceptions generated under hypothetical or actual trading outcomes.

### **Bank-wide backtesting**

- 99.9** To place the definitions of three zones of the bank-wide backtesting 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.
- 99.10** 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.

**99.11** 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 (ie these are binomial probabilities). For example, the left-hand portion of Table 1 sets out 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.

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

**Notes to Table 1:** 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 on 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 type 2 columns 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%.

- 
- 99.12** The right-hand portion of the table 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.
- 99.13** 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.
- 99.14** Under the assumptions that the model's true level of coverage is not 99%, 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 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.

**99.15** 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. It is for this reason that the Committee has rejected an approach that contains only a single threshold.

**99.16** Given these limitations, the Committee has classified outcomes for the backtesting of the bank-wide model 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 (ie backtesting "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 (ie backtesting "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 supervisor should encourage a bank to present additional information about its model before taking action (ie backtesting "amber zone").

**99.17** Table 2 sets out the Committee's agreed 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 backtesting amber 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 backtesting amber zone begins at five exceptions. Similarly, the beginning of the backtesting 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 10 exceptions.



Backtesting zone	Number of exceptions	Backtesting-dependent multiplier (to be added to any qualitative add-on per [MAR33.44])	Cumulative probability
Green	0	1.50	8.11%
	1	1.50	28.58%
	2	1.50	54.32%
	3	1.50	75.81%
	4	1.50	89.22%
Amber	5	1.70	95.88%
	6	1.76	98.63%
	7	1.83	99.60%
	8	1.88	99.89%
	9	1.92	99.97%
Red	10 or more	2.00	99.99%

**Notes to Table 2:** The table defines the backtesting green, amber 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 amber 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.

- 99.18** The backtesting 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.
- 99.19** The range from five to nine exceptions constitutes the backtesting amber 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.
- 99.20** Table 2 sets out the Committee's agreed guidelines for increases in the multiplication factor applicable to the internal models capital requirement, resulting from backtesting results in the backtesting amber zone.
- 99.21** 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 1.13 for a multiplier of 1. 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.

## **Examples of the application of the principles for risk factor modellability**

- 99.22** Although supervisors may use discretion regarding the types of evidence required of banks to provide risk factor modellability, the following are examples of the types of evidence that banks may be required to provide.

- (1) Regression diagnostics for multi-factor beta models. In addition to showing that indices or other regressors are appropriate for the region, asset class and credit quality (if applicable) of an instrument, banks must be prepared to demonstrate that the coefficients used in multi-factor models are adequate to capture both general market risk and idiosyncratic risk. If the bank assumes that the residuals from the multi-factor model are uncorrelated with each other, the bank should be prepared to demonstrate that the modellable residuals are uncorrelated. Further, the factors in the multi-factor model must be appropriate for the region and asset class of the instrument and must explain the general market risk of the instrument. This must be demonstrated through goodness-of-fit statistics (eg an adjusted-R2 coefficient) and other diagnostics on the coefficients. Most importantly, where the estimated coefficients are not used (ie the parameters are judgment-based), the bank must describe how the coefficients are chosen and why they cannot be estimated, and demonstrate that the choice does not underestimate risk. In general, risk factors are not considered modellable in cases where parameters are set by judgment.
- (2) Recovery of price from risk factors. The bank must periodically demonstrate and document that the risk factors used in its risk model can be fed into front office pricing models and recover the actual prices of the assets. If the recovered prices substantially deviate from the actual prices, this can indicate a problem with prices used to derive the risk factors and call into question the validity of data inputs for risk purposes. In such cases, supervisors may determine that the risk factor is non-modellable.
- (3) Risk pricing is periodically reconciled with front office and back office prices. While banks are free to use price data from external sources, these external prices should periodically be reconciled with internal prices (from both front office and back office) to ensure they do not deviate substantially, and that they are not consistently biased in any fashion. Results of these reconciliations should be made available to supervisors, including statistics on the differences of the risk price from front office and back office prices. It is standard practice for banks to conduct reconciliation of front office and back office prices; the risk prices must be included as part of the reconciliation of the front office and whenever there is a potential for discrepancy. If the discrepancy is large, supervisors may determine that the risk factor is non-modellable.

- (4) Risk factor backtesting. Banks must periodically demonstrate the appropriateness of their modelling methodology by comparing the risk factor returns forecast produced by the risk management model with actual returns produced by front office prices. Alternatively, a bank could backtest hypothetical portfolios that are substantively dependent on key risk factors (or combinations thereof). This risk factor backtesting is intended to confirm that risk factors accurately reflect the volatility and correlations of the instruments in the risk model. Hypothetical backtesting can be effective in identifying whether risk factors in question adequately reflect volatility and correlations when the portfolio of instruments is chosen to highlight specific products.
- (5) Risk factors generated from parameterised models. For options, implied volatility surfaces are often built using a parameterised model based on single-name underlyings and/or option index RPOs and/or market quotes. Liquid options at moneyness, tenor and option expiry points may be used to calibrate level, volatility, drift and correlation parameters for a single-name or benchmark volatility surface. Once these parameters are set, they are derived risk factors in their own right that must be updated and recalibrated periodically as new data arrive and trades occur. In the event that these risk factors are used to proxy for other single-name option surface points, there must be an additional-basis non-modellable risk factor overlay for any potential deviations.