



AlphaIQ

CERTIFYIQ · SAMPLE VALIDATION REPORT

Model validation memo — Desk A VaR Model v2.4

Market risk — VaR back-testing (FRTB)

FRTB back-testing and P&L attribution

MODEL

Desk A VaR Model v2.4

ENGAGEMENT

Sample — trading desk VaR back-test review

PREPARED

2026-10-07

PREPARED BY

AlphaIQ CertifyIQ, computed by the product

RECOMMENDATION



May be relied upon, with conditions

10 sections · 10 traced figures · 2 finding(s) · 6 test(s) named as not performed

SYNTHETIC SAMPLE

Every figure in this report was computed by CertifyIQ from a synthetic book — A desk's daily VaR forecasts against actual and hypothetical P&L. No real institution, model or customer is described. The book is market-backtest-desk-a.csv; how it was generated, and the defects planted in it, are documented in market-backtest-dgp.json.

AT A GLANCE · SYNTHETIC SAMPLE

Desk A VaR Model v2.4

Market risk — VaR back-testing (FRTB) · FRTB back-testing and P&L attribution

May be relied upon, with conditions

2 material finding(s) require remediation or an accepted compensating control before unqualified reliance. Each is listed in section 8 with the evidence that produced it. 5 of 10 rows were not judged (no threshold declared) and are outside this conclusion.

10

SECTIONS

10TRACED
FIGURES**2**

MATERIAL

0

NOTABLE

0

OBSERVATIONS

6NOT
PERFORMED

Supervisory frame: this memo is built to "FRTB Desk Model Approval", published under Basel · Trading book. No supervisory frame is declared for this engagement, so the frame above was chosen by the model's family rather than by its supervisor. That is not a statement that the two agree: nobody has recorded who supervises this model, and a reader in another jurisdiction should treat the section structure, the citations and the recommendation vocabulary as Basel's.

Evidence manifest: none sealed. The documents behind each figure are named where the figure is; the set of documents on record when this memo was written has not been attested by anyone, so it cannot be cited here.

How to read this sample. Every number below was computed by CertifyIQ from a synthetic book (market-backtest-desk-a.csv), through the same resolver, metric pack and report builder a signed-in engagement uses. Each figure carries its own trace — where it came from and what it was judged against. Tests that could not be performed are named with the reason, never left out. Where no threshold was declared, a metric is reported but not judged, because this report does not invent an institution's tolerances.

CONTENTS

- | | |
|---|--|
| 1 Purpose and scope | 6 P&L attribution test |
| 2 Trading desk, observation window and coverage | 7 Expected shortfall |
| 3 Back-testing outcome: the supervisory zone | 8 Findings |
| 4 Desk-level back-testing against the FRTB limits | 9 Third-party dependencies and residual risk |
| 5 Statistical adequacy of the exception sequence | 10 Not performed |

1 Purpose and scope

This memo validates Desk A VaR Model v2.4 against the one-day VaR back-test of the Desk A — rates and FX desk, extracted on 2026-04-21: 600 line(s) read, 600 observation day(s) kept from 2024-01-02 to 2026-04-20, of which the most recent 250 form the window every figure below is computed over.

The book was declared, not inferred. The DESK is declared rather than read from a column, because a book covering two desks is two series and this pack reads one: a two-desk book presented as one desk would compute MAR32's per-desk limits over a mixture, and MAR32's consequence lands on a desk. The VaR

CONVENTION is declared as a positive loss magnitude, because a signed VaR read as a magnitude turns every day into an exception; a negative value in a VaR column refuses the book rather than being reinterpreted. And NO COLUMN IS BOUND BY NAME: each of the book's columns was bound to a declared field by a named person with a written reason, and the bindings are stored with the book and pinned into the same hash, because the same bytes read through a different mapping are a different book.

This validation carries no replica of the risk model under test and reads no ongoing-monitoring series: the figures are measured directly from the declared book against the institution's own declared claims, so the recommendation is not gated on a replication step this family's evidence chain has no equivalent of. It does not validate the pricing models the revaluations came from, the risk-factor eligibility test, the stress-testing programme or the trading-desk boundary — each is named in section 9 with the reason, and the documentation obligations among them are counted in the engagement's completeness checklist, not here.

2 Trading desk, observation window and coverage

Desk: Desk A — rates and FX, declared. Window: the most recent 250 observation days, which is what Basel's back-test and MAR32's desk limits are both defined over. This book holds 600 observation day(s) — 2024-01-02 to 2026-04-20, with 240 missing calendar day(s) across 120 gap(s) in the 840-day span.

The window is complete: 250 observation days are available and every figure below is computed over them.

The gaps are reported and never filled: 2024-01-05 to 2024-01-08 (2 day(s)); 2024-01-12 to 2024-01-15 (2 day(s)); 2024-01-19 to 2024-01-22 (2 day(s)); 2024-01-26 to 2024-01-29 (2 day(s)); 2024-02-02 to 2024-02-05 (2 day(s)); 2024-02-09 to 2024-02-12 (2 day(s)); 2024-02-16 to 2024-02-19 (2 day(s)); 2024-02-23 to 2024-02-26 (2 day(s)); and 112 further gap(s). Weekends and holidays are gaps in the calendar and not in the evidence; a gap that is neither is a book with days missing, and only the desk can say which this is. Nothing here interpolates a missing day, because an interpolated day would enter Christoffersen's independence test as an observation nobody made.

600 physical line(s) were read and 600 loaded, with none excluded. The order of this series is recoverable from its content — the date is in every record and a duplicate date refuses the book — so a re-export in another row order is the same evidence and is filed as the same run.

3 Back-testing outcome: the supervisory zone

The loss exceeded the desk's own 99% one-day VaR on 14 of the 250 days in the window. Against BCBS 22's Table 2 that count falls in the RED zone (green ≤ 4 · amber 5–9 · red ≥ 10 (250 days, 99%)).

An amber count is not a defect and is not recorded as a finding. BCBS 22 anticipates it: the framework's response to four to nine exceptions is an increase in the capital multiplier, which is a consequence with a number attached, not a verdict that the model is wrong. Only a red count — ten or more — raises a finding here, because that is the count at which the framework itself stops treating the exceedances as consistent with a correctly specified 99% model.

The add-on to the capital multiplier implied by that count is 1.00, read off BCBS 22 Table 2. It is REPORTED AND NOT JUDGED, deliberately: the add-on is a function of the exception count above and carries no information the count does not already carry, so grading it would report one fact twice and let this memo raise two findings where one defect exists. What it is for is the reader who needs the capital consequence in the same paragraph as the count that produced it.

Exceptions at 99% in the most recent 250 days (supervisory zone)

MEASURED

14

Exceptions at 99% in the most recent 250 days (supervisory zone): 14 measured on 250 observation days carrying 14 exceptions, 2025-05-06 to 2026-04-20 — a 350-day span with 100 missing day(s) across 50 gap(s). Benchmark source: BCBS 22, Supervisory framework for the use of backtesting in conjunction with the internal models approach (1996), Table 2 — green zone 0–4 exceptions, yellow 5–9, red 10 or more, in 250 observations at 99%; Basel Framework MAR33 applies the resulting add-on to the multiplier.

Plus factor to the capital multiplier implied by the exception count

MEASURED

1.00

Plus factor to the capital multiplier implied by the exception count: 1.00 measured on 250 observation days carrying 14 exceptions, 2025-05-06 to 2026-04-20 — a 350-day span with 100 missing day(s) across 50 gap(s). Benchmark source: BCBS 22, Table 2 — 0.00 for 0–4 exceptions, 0.40 / 0.50 / 0.65 / 0.75 / 0.85 for 5–9, 1.00 for 10 or more. Reported, never judged: the zone above is the judgement and this is its consequence.

MATERIAL F-BACKTEST-EXCEPTIONS-99 Exceptions at 99% in the most recent 250 days (supervisory zone) does not meet its declared threshold

Measured 14 against green ≤ 4 · amber 5–9 · red ≥ 10 (250 days, 99%). Threshold source: BCBS 22, Supervisory framework for the use of backtesting in conjunction with the internal models approach (1996), Table 2 — green zone 0–4 exceptions, yellow 5–9, red 10 or more, in 250 observations at 99%; Basel Framework MAR33 applies the resulting add-on to the multiplier.

Evidence to re-run: Metric pack market on 250 observation day(s) carrying 14 exception(s) at 99%; thresholds from AlphaIQ market-risk pack — each threshold cited to its own instrument.

4 Desk-level back-testing against the FRTB limits

This section asks a different question from section 3 with the same arithmetic, and the difference is the consequence. Section 3 reads the 99% exception count against BCBS 22's zones, which set the ADD-ON to the capital multiplier for the model. MAR32 sets two hard limits on THE DESK — more than 12 exceptions at 99%, or more than 30 at 97.5%, over the same most recent 250 days — and breaching either removes that desk from the internal models approach for its capital requirement. A desk can sit in section 3's amber zone and be inside both limits here, and it can sit in green and breach the 97.5% limit; neither is a contradiction, and folding the two into one heading would let a reader carry the wrong consequence away.

At 99%: 14 exception(s) against a limit of 12 — the limit is breached, and this desk falls out of the internal models approach until it is back inside.

At 97.5%: 19 exception(s) against a limit of 30 — inside the limit.

Desk-level exceptions at 99% in the most recent 250 days (FRTB desk limit)

MEASURED

14

Desk-level exceptions at 99% in the most recent 250 days (FRTB desk limit): 14 measured on 250 observation days carrying 14 exceptions, 2025-05-06 to 2026-04-20 — a 350-day span with 100 missing day(s) across 50 gap(s). Benchmark source: Basel Framework MAR32 — trading-desk backtesting: a desk with more than 12 exceptions at the 99th percentile in the most recent 12-month period must be capitalised under the standardised approach.

Desk-level exceptions at 97.5% in the most recent 250 days (FRTB desk limit)

MEASURED

19

Desk-level exceptions at 97.5% in the most recent 250 days (FRTB desk limit): 19 measured on 250 observation days carrying 14 exceptions, 2025-05-06 to 2026-04-20 — a 350-day span with 100 missing day(s) across 50 gap(s). Benchmark source: Basel Framework MAR32 — trading-desk backtesting: a desk with more than 30 exceptions at the 97.5th percentile in the most recent 12-month period must be capitalised under the standardised approach.

MATERIAL **F-DESK-EXCEPTIONS-99** **Desk-level exceptions at 99% in the most recent 250 days (FRTB desk limit) does not meet its declared threshold**

Measured 14 against more than 12 exceptions removes the desk from the internal models approach. Threshold source: Basel Framework MAR32 — trading-desk backtesting: a desk with more than 12 exceptions at the 99th percentile in the most recent 12-month period must be capitalised under the standardised approach.

Evidence to re-run: Metric pack market on 250 observation day(s) carrying 14 exception(s) at 99%; thresholds from AlphaIQ market-risk pack — each threshold cited to its own instrument.

5 Statistical adequacy of the exception sequence

Three tests, asking three different things of the same sequence of exceptions. Kupiec's proportion-of-failures test asks whether the RATE of exceptions is consistent with a 99% model — too many and the model understates risk, too few and it overstates it. Christoffersen's independence test asks whether exceptions arrive independently of one another, or cluster: a model that is right on average and wrong in bursts is the one that fails when it matters. The conditional-coverage test asks whether both hold jointly, and is the sum of the two statistics against two degrees of freedom.

AN IMPORTANT PROPERTY OF THE SECOND TEST, because a reader who does not know it will over-read a high p-value: Christoffersen's independence test is a first-order test, comparing the chance of an exception the day after an exception with the chance the day after a quiet day. Exceptions that cluster inside a stressed REGIME rather than on consecutive days can leave it comfortably un-rejected while the conditional-coverage test rejects firmly. A high independence p-value is therefore evidence of nothing on its own, and 'not rejected' is never 'independent'.

None of the three is judged in this run. A p-value is judged against a SIGNIFICANCE LEVEL, and no supervisory instrument publishes one for these tests — 0.05 is a convention of the literature, not a rule of any authority, and this platform ships a threshold only when a named published instrument sets it and it does not vary by institution. So the three p-values are reported for the reader to apply the institution's own level to, and the memo passes no judgement it cannot cite. Declaring that level as a claim on the engagement is what turns these rows into graded ones.

Kupiec 0.000 (reported, not judged); Christoffersen independence 0.806 (reported, not judged); conditional coverage 0.000 (reported, not judged).

Kupiec proportion-of-failures test — p-value of the unconditional coverage LR at 99% **MEASURED** **0.000**

Kupiec proportion-of-failures test — p-value of the unconditional coverage LR at 99%: 0.000 measured on 250 observation days carrying 14 exceptions, 2025-05-06 to 2026-04-20 — a 350-day span with 100 missing day(s) across 50 gap(s). Benchmark source: Kupiec (1995), Techniques for verifying the accuracy of risk measurement models — LR_{uc} against chi-square(1). The significance level is a convention, not a rule: declared, or the p-value is reported unjudged.

Christoffersen independence test — p-value of the first-order Markov LR on the exception sequence **MEASURED** **0.806**

Christoffersen independence test — p-value of the first-order Markov LR on the exception sequence: 0.806 measured on 250 observation days carrying 14 exceptions, 2025-05-06 to 2026-04-20 — a 350-day span with 100 missing day(s) across 50 gap(s). Benchmark source: Christoffersen (1998), Evaluating interval forecasts — LR_{ind} against chi-square(1). A first-order test: it sees consecutive exceptions, not a regime in which exceptions merely cluster, so 'not rejected' is not 'independent'. Declared level, or reported unjudged.

Christoffersen conditional coverage test — p-value of the joint LR (coverage and independence)

MEASURED

0.000

Christoffersen conditional coverage test — p-value of the joint LR (coverage and independence): 0.000 measured on 250 observation days carrying 14 exceptions, 2025-05-06 to 2026-04-20 — a 350-day span with 100 missing day(s) across 50 gap(s). Benchmark source: Christoffersen (1998) — $LR_{cc} = LR_{uc} + LR_{ind}$ against $\chi^2(2)$. Declared level, or reported unjudged.

6 P&L attribution test

The P&L attribution test asks whether the risk model's own revaluation of the desk's positions tracks what the front office's pricing models produced on the same positions with no new trades — the risk-theoretical P&L against the hypothetical P&L. It is not a test of whether the desk made money; it is a test of whether the risk model and the pricing models are describing the same book. MAR32 publishes two measures and their bands, and both are read here.

Spearman correlation between the two series: 0.763 (green ≥ 0.80 · amber 0.70–0.80 · red < 0.70) — marginal. It measures whether the two series move together in rank order, which is the property that survives the fact that the two models disagree about levels.

Kolmogorov-Smirnov statistic between the two distributions: 0.072 (green ≤ 0.09 · amber 0.09–0.12 · red > 0.12) — pass. It measures the largest gap between the two empirical distributions, which catches a risk model that tracks direction well and gets the tail wrong. The two measures are reported together because a desk can pass either alone: rank agreement with a mis-shaped tail, or a well-shaped distribution whose days do not line up.

P&L attribution — Spearman correlation between risk-theoretical and hypothetical P&L

MEASURED

0.763

P&L attribution — Spearman correlation between risk-theoretical and hypothetical P&L: 0.763 measured on 250 observation days carrying 14 exceptions, 2025-05-06 to 2026-04-20 — a 350-day span with 100 missing day(s) across 50 gap(s). Benchmark source: Basel Framework MAR32 — P&L attribution test metrics: Spearman correlation between RTPL and HPL over the most recent 250 days; a desk is in the red zone below 0.70 and in the green zone at 0.80 and above.

P&L attribution — Kolmogorov-Smirnov statistic between risk-theoretical and hypothetical P&L

MEASURED

0.072

P&L attribution — Kolmogorov-Smirnov statistic between risk-theoretical and hypothetical P&L: 0.072 measured on 250 observation days carrying 14 exceptions, 2025-05-06 to 2026-04-20 — a 350-day span with 100 missing day(s) across 50 gap(s). Benchmark source: Basel Framework MAR32 — P&L attribution test metrics: the Kolmogorov-Smirnov test statistic between the RTPL and HPL distributions over the most recent 250 days; red above 0.12, green at 0.09 and below.

7 Expected shortfall

The desk's realised loss averaged over its worst 2.5% of days in the window is 4197823.24. This is what the book actually did in its tail, computed from the realised P&L alone and not from any model's forecast.

IT IS REPORTED AND NOT GRADED, and the reason is worth stating plainly because its absence from the graded rows will otherwise look like an omission. The FRTB capital measure is expected shortfall, but Basel BACK-TESTS VaR: the back-testing and attribution requirements in MAR32 are written about VaR exceptions and P&L attribution, and no supervisory back-test of expected shortfall is published to grade this figure against. Inventing a pass line here would be this platform publishing a supervisory-looking threshold that no supervisor wrote. The

institution's own expected -shortfall calculation, with its liquidity horizons, is its document to supply and is named in section 9.

Realised expected shortfall at 97.5% — mean loss on the worst 2.5% of days in the window

MEASURED 4197823.24

Realised expected shortfall at 97.5% — mean loss on the worst 2.5% of days in the window: 4197823.24 measured on 250 observation days carrying 14 exceptions, 2025-05-06 to 2026-04-20 — a 350-day span with 100 missing day(s) across 50 gap(s). Benchmark source: Basel Framework MAR33 specifies expected shortfall as the capital measure and prescribes no back-test of it; ES back-testing is an open research question (Acerbi and Székely, 2014, and after). Reported in the book's currency, judged against nothing — a pack that graded ES against a number would be inventing a supervisory expectation.

8 Findings

2 finding(s) were raised, 2 of them material. Each carries the evidence a reviewer would re-run to check it.

Against the findings register: 2 finding(s) open, 2 at grade 1, 2 carrying a grade the engine PROPOSED and nobody has owned, 2 newly raised. Oldest open finding is 0 day(s) old.

2 live finding(s) have no owner assigned. An unowned finding has no one accountable for closing it and will age without anyone noticing.

MATERIAL F-BACKTEST-EXCEPTIONS-99 Exceptions at 99% in the most recent 250 days (supervisory zone) does not meet its declared threshold

Measured 14 against green ≤ 4 · amber 5–9 · red ≥ 10 (250 days, 99%). Threshold source: BCBS 22, Supervisory framework for the use of backtesting in conjunction with the internal models approach (1996), Table 2 — green zone 0–4 exceptions, yellow 5–9, red 10 or more, in 250 observations at 99%; Basel Framework MAR33 applies the resulting add-on to the multiplier.

Evidence to re-run: Metric pack market on 250 observation day(s) carrying 14 exception(s) at 99%; thresholds from AlphaIQ market-risk pack — each threshold cited to its own instrument.

MATERIAL F-DESK-EXCEPTIONS-99 Desk-level exceptions at 99% in the most recent 250 days (FRTB desk limit) does not meet its declared threshold

Measured 14 against more than 12 exceptions removes the desk from the internal models approach. Threshold source: Basel Framework MAR32 — trading-desk backtesting: a desk with more than 12 exceptions at the 99th percentile in the most recent 12-month period must be capitalised under the standardised approach.

Evidence to re-run: Metric pack market on 250 observation day(s) carrying 14 exception(s) at 99%; thresholds from AlphaIQ market-risk pack — each threshold cited to its own instrument.

9 Third-party dependencies and residual risk

No third-party dependency is declared for this engagement. That is a statement by the institution that the model embeds no vendor or open-source component it cannot validate inside — not an absence of the question.

10 Not performed

The following checks were not performed, each with the reason. A validation report's most dangerous property is the test it quietly omitted, so this register is part of the memo rather than an appendix to it. Entries that name a documentation obligation are established by the institution and recorded in the engagement's completeness checklist; they are listed here so the reader of this memo knows they exist and knows this memo did not establish them.

NOT PERFORMED — AND WHY

Revaluation reproducibility and pricing-benchmark deviation

Replicating a VaR engine means re-running its pricing library over the desk's positions in a sealed runtime, which this platform does not have. The demo's 'revaluation reproducibility 100%' and 'pricing benchmark deviation 11 bp' are figures nothing here can produce; the back-test reads what the forecasts DID, not how they were made.

Risk factor eligibility test and non-modellable risk factors (MAR31)

Modellability is evidenced by real price observations per risk factor over twelve months (24 observations, no 90-day gap) — a third evidence shape, an observation register per risk factor, that no back-test book carries. Recorded by the institution as an obligation; not computed here.

A supervisory back-test of expected shortfall

None exists to run. Basel Framework MAR32 back-tests VaR at 99% and 97.5%; ES is the capital measure (MAR33) and is not back-tested in the framework. The realised ES is reported with no threshold, and an Acerbi-Székely test would be a research method presented as a supervisory one.

The stress-testing programme over the desks' positions (MAR30)

A stress-testing programme is the institution's own, over its positions and risk factors, reviewed by senior management and reflected in limits; this pack holds one desk's back-test book and no positions. Recorded as an obligation.

Expected shortfall across the prescribed liquidity horizons, calibrated to a stress period (MAR33)

The regulatory ES is computed by the institution's engine over risk-factor histories at 10, 20, 40, 60 and 120-day horizons with a stressed calibration; it is a capital calculation over positions, not a statistic over a back-test book. The institution's calculation is recorded as an obligation.

Trading-desk boundary appropriateness

Whether the desk is the right unit — its trader ownership, its strategy, its limits — is a documentary judgement over the institution's desk structure (MAR30/MAR31), not a figure over a series.



AlphaIQ CertifyIQ — independent validation of models and the decisions they drive. This sample was produced by the product itself; nothing in it was written by hand.

Memo content hash ad9a53b22715de255b6ccc41128d47b1574487b722530e24ae8776c82e3fe3ac

This document is a recommendation, not a certification. A certificate is issued only on attested evidence, which a synthetic sample does not have.