



ECB TRIM / IRB review — Corporate PD Rating System v3.2

IRB — ECB TRIM rating system review

ECB TRIM / CRR IRB validation

MODEL

Corporate PD Rating System v3.2

ENGAGEMENT

Sample — corporate PD rating system, annual
TRIM review

PREPARED

2026-10-07

PREPARED BY

AlphaIQ CertifyIQ, computed by the product

RECOMMENDATION



May be relied upon, with conditions

10 sections · 6 traced figures · 1 finding(s) · 15 test(s) named as not performed

SYNTHETIC SAMPLE

Every figure in this report was computed by CertifyIQ from a synthetic book — A graded corporate population read against its declared master rating scale, grade by grade. No real institution, model or customer is described. The book is `corporate-pd-graded-population.csv`; how it was generated, and the defects planted in it, are documented in `corporate-pd-dgp.json`.

AT A GLANCE · SYNTHETIC SAMPLE

Corporate PD Rating System v3.2

IRB — ECB TRIM rating system review · ECB TRIM / CRR IRB validation

May be relied upon, with conditions

1 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. 9 of 10 rows were not judged (no threshold declared) and are outside this conclusion.

10

SECTIONS

6TRACED
FIGURES**1**

MATERIAL

0

NOTABLE

0

OBSERVATIONS

15NOT
PERFORMED

Supervisory frame: this memo is built to "ECB TRIM / IRB Review Pack", published under EU · ECB. 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 EU'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 (corporate-pd-graded-population.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, scope and the rating system reviewed	6 Rating grade concentration
2 Master scale and grade assignment	7 Rating migration and LGD predictive ability
3 Discriminatory power	8 Tests not performed, and what their absence costs
4 Stability of discriminatory power (§2.5.4.1)	9 Conclusion and findings
5 Predictive ability and calibration by grade	10 Third-party dependencies and residual risk

1 Purpose, scope and the rating system reviewed

This memo reviews the rating system behind Corporate PD Rating System v3.2 against a graded population of 4,000 obligor(s), of which 66 defaulted, assigned across the 7 grade(s) of the declared master scale. Every figure below is measured on that population; none is transcribed from the institution's own reporting.

The master scale was DECLARED, not inferred: each grade carries an assigned PD and an upper bound read from the institution's own scale document. A grade's obligors are those the file assigns to it, and this review does not re-grade anybody, because re-grading would make the figures below statements about AlphaIQ's

assignment rather than about the bank's rating system. 1 declared grade(s) hold no obligors in this population (CC-C); each is printed with NO observed rate rather than a rate of zero, because 0 of 0 is not 0 and a grade nobody was assigned to is not a grade that never defaulted.

This review does not compute a margin of conservatism and does not assess representativeness. Both are named in section 7 with the reason: the MoC is a documentary decomposition per risk parameter, and representativeness compares two populations this platform is given one of. It also carries no replica of the rating system and reads no ongoing-monitoring series, so its conclusion is not gated on a replication step this evidence chain has no equivalent of.

2 Master scale and grade assignment

The declared scale, grade by grade — AAA-AA: assigned 0.030% (upper 0.080%), 7 obligor(s), 0 default(s); A: assigned 0.150% (upper 0.250%), 205 obligor(s), 0 default(s); BBB: assigned 0.500% (upper 0.800%), 1,442 obligor(s), 6 default(s); BB: assigned 1.500% (upper 2.500%), 1,986 obligor(s), 36 default(s); B: assigned 4.500% (upper 8.000%), 350 obligor(s), 23 default(s); CCC: assigned 12.000% (upper 20.000%), 10 obligor(s), 1 default(s); CC-C: assigned 30.000% (upper 100.000%), 0 obligor(s), 0 default(s).

The assigned PD is the institution's own figure for the grade and the upper bound is the scale's own boundary. Neither is derived here; a scale this platform reconstructed from the population would be a scale nobody had approved.

Lowest assigned PD on the master scale

MEASURED 0.030%

Lowest assigned PD on the master scale: 0.030% measured on 4,000 graded exposures carrying 66 realised defaults. Benchmark source: CRR Article 160(1), as amended by Regulation (EU) 2024/1623 (CRR3) — PD input floor of 0.05% for exposures to corporates and institutions, applying since 1 January 2025.

MATERIAL F-PD-FLOOR The master scale assigns a PD below the regulatory floor

The lowest assigned PD on the declared scale is 0.030%, below the 0.050% floor. Unlike every other bar in this review this one is not the institution's: it is in the Regulation, so the breach is a matter of fact rather than of the bank's own tolerance.

Evidence to re-run: Master scale as declared, 7 grade(s). CRR Article 160(1), as amended by Regulation (EU) 2024/1623 (CRR3) — PD input floor of 0.05% for exposures to corporates and institutions, applying since 1 January 2025

3 Discriminatory power

Discriminatory power is measured on this population as the area under the ROC curve of the assigned score against realised default. It answers whether the system ranks defaulters above non-defaulters, and says nothing about whether the PD levels themselves are right — that is section 5.

Rating system AUC on this population

MEASURED 0.764

Rating system AUC on this population: 0.764 measured on 4,000 graded exposures carrying 66 realised defaults. Benchmark source: ECB validation reporting instructions §2.5.4.1 compares the current AUC with the AUC recorded at initial validation. The initial figure is the owner's own documented number and no floor is published; the ECB's S-statistic itself is declared not performed, because it needs the initial AUC's variance as well as its value.

NOT PERFORMED — AND WHY

Comparison against the AUC recorded at initial validation

The institution has declared no initial-validation AUC for this engagement, so the current figure is reported and not judged. The ECB compares the two; no supervisor publishes a floor for either, and a bar this platform chose would be AlphaIQ deciding when a bank's rating system has lost discriminatory power.

4 Stability of discriminatory power (§2.5.4.1)

$S = (\text{AUC}_{\text{init}} \text{ less } \text{AUC}_{\text{curr}}) \text{ divided by } s$, with $p = 1 \text{ less } \Phi(S)$, is the ECB's own test of whether discriminatory power has DETERIORATED since initial validation. It is one-sided by construction: a two-sided p would report an improvement as a problem and halve the significance of a real decline.

The p -value is REPORTED AND NOT SCORED. Section 2.5.4.1 defines the statistic and publishes no significance level, and a level is a convention of practice — shipping 0.05 here would be AlphaIQ deciding when an ECB test fails, printed under the ECB's section number. An institution that holds its own level declares it and this becomes a judged row.

NOT PERFORMED — AND WHY

ECB AUC stability S -statistic and its p -value (section 2.5.4.1)

The ECB AUC stability statistic needs the AUC recorded at initial validation ($\text{auc}_{\text{initial}}$) and the standard deviation of the difference ($\text{auc}_{\text{initial_sd}}$), which are not declared for this engagement. They are the institution's own figures from the initial validation report and cannot be recovered from this population: the standard deviation is a property of the sample the original AUC was measured on. Declare them as an input against this engagement and this row computes.

5 Predictive ability and calibration by grade

Calibration is judged PER GRADE, which is the point of a TRIM review. A portfolio-level default rate says nothing about a rating system: a model can be well calibrated in aggregate while every grade is wrong in alternating directions, and the aggregate would hide exactly that.

Grade by grade — AAA-AA: assigned 0.030%, observed 0.000% on 7 obligor(s) (0 default(s)), deviation -3 bp, Jeffreys p 0.053; A: assigned 0.150%, observed 0.000% on 205 obligor(s) (0 default(s)), deviation -15 bp, Jeffreys p 0.568; BBB: assigned 0.500%, observed 0.416% on 1,442 obligor(s) (6 default(s)), deviation -8 bp, Jeffreys p 0.656; BB: assigned 1.500%, observed 1.813% on 1,986 obligor(s) (36 default(s)), deviation 31 bp, Jeffreys p 0.127; B: assigned 4.500%, observed 6.571% on 350 obligor(s) (23 default(s)), deviation 207 bp, Jeffreys p 0.037; CCC: assigned 12.000%, observed 10.000% on 10 obligor(s) (1 default(s)), deviation -200 bp, Jeffreys p 0.523; CC-C: no obligors, so no observed rate — NOT a rate of zero.

The Jeffreys binomial test is the one ECB section 2.5.3 names, applied to the weakest grade; the deviation figure is the largest single grade's, in basis points. Both are reported against the institution's own declared tolerance, because no supervisor publishes one.

Jeffreys back-test — weakest grade p -value

MEASURED 0.037

Jeffreys back-test — weakest grade p -value: 0.037 measured on 4,000 graded exposures carrying 66 realised defaults.

Benchmark source: ECB, Instructions for reporting the validation results of internal models — credit risk, §2.5.3.1: one-sided Jeffreys binomial back-test, $p = \text{Beta}(D+\frac{1}{2}, N-D+\frac{1}{2})$ evaluated at the grade's assigned PD. The ECB prescribes the test and publishes no significance level; the level is the institution's own.

Predicted vs observed default rate — largest grade deviation**MEASURED 207 bp**

Predicted vs observed default rate — largest grade deviation: 207 bp measured on 4,000 graded exposures carrying 66 realised defaults. Benchmark source: ECB validation reporting instructions §2.5.3: the PD assigned to each grade is reported against the observed default rate for that grade. No supervisor publishes a tolerance; the deviation a grade may show before it is a finding is institution policy.

Calibration bias (mean predicted PD – observed default rate)**MEASURED -4049.459%**

Calibration bias (mean predicted PD – observed default rate): -4049.459% measured on 4,000 graded exposures carrying 66 realised defaults. Benchmark source: Internal model risk policy — the over-prediction a portfolio must show, in basis points. This is a calibration bias and NOT the EBA/GL/2017/16 margin of conservatism, which is a documentary decomposition into categories A, B and C and is declared not performed by this pack.

6 Rating grade concentration

The ECB's normalised Herfindahl index over the distribution of obligors across grades. A concentrated distribution is not itself a defect — a portfolio may genuinely sit in two grades — but it bounds what the per-grade tests in section 5 can establish, because a grade with few obligors supports no conclusion whatever its observed rate.

Herfindahl index of rating grade concentration**MEASURED 0.512**

Herfindahl index of rating grade concentration: 0.512 measured on 4,000 graded exposures carrying 66 realised defaults. Benchmark source: ECB validation reporting instructions §2.5.5.3: $HI = 1 + \log((CV^2 + 1)/K)/\log K$ over the relative frequencies of the K rating grades. No limit is published; the ECB's own test compares the index against its value at initial validation, which only the institution holds.

7 Rating migration and LGD predictive ability

Two evidence artefacts beyond the graded population, each optional and each independently so: a rating migration book read against the institution's own declared transition matrix, and a facility-level book of estimated against realised loss given default. An engagement may hold either, both or neither, and a row without its artefact says which artefact and where to import it rather than shortening the table.

NOT PERFORMED — AND WHY**Rating migration stability — weakest cell p-value**

No migration book is imported for this engagement. Rating migration stability compares the SAME obligors at TWO dates against the transition matrix the rating system declares, which is a second evidence artefact and not a column of the graded population — one photograph has no 'from' grade. Import one at /irb-intake and this row computes.

LGD predictive ability — paired t-test p-value

No facility book is imported for this engagement. LGD predictive ability is FACILITY-level evidence — estimated against realised loss given default, per facility — and one obligor holds several facilities with different collateral and different seniority, so it is not the obligor book with a column added. Import one at /irb-intake and this row computes.

LGD discriminatory power — gAUC over the ordinal outcome

No facility book is imported for this engagement. The gAUC measures discriminatory power over an ORDINAL realised outcome — does a facility this model rated worse actually lose more — which the obligor book cannot answer because it carries no realised loss. Import a facility book at /irb-intake and this row computes.

8 Tests not performed, and what their absence costs

A review's most dangerous property is the test it quietly omitted, so this register is part of the memo rather than an appendix to it. Two entries in it are permanent properties of this pack rather than gaps in this engagement: the margin-of-conservatism decomposition and representativeness are not computed at all, and no evidence supplied would change that.

NOT PERFORMED — AND WHY

Comparison against the AUC recorded at initial validation

The institution has declared no initial-validation AUC for this engagement, so the current figure is reported and not judged. The ECB compares the two; no supervisor publishes a floor for either, and a bar this platform chose would be AlphaIQ deciding when a bank's rating system has lost discriminatory power.

ECB AUC stability S-statistic and its p-value (section 2.5.4.1)

The ECB AUC stability statistic needs the AUC recorded at initial validation (`auc_initial`) and the standard deviation of the difference (`auc_initial_sd`), which are not declared for this engagement. They are the institution's own figures from the initial validation report and cannot be recovered from this population: the standard deviation is a property of the sample the original AUC was measured on. Declare them as an input against this engagement and this row computes.

Rating migration stability — weakest cell p-value

No migration book is imported for this engagement. Rating migration stability compares the SAME obligors at TWO dates against the transition matrix the rating system declares, which is a second evidence artefact and not a column of the graded population — one photograph has no 'from' grade. Import one at `/irb-intake` and this row computes.

LGD predictive ability — paired t-test p-value

No facility book is imported for this engagement. LGD predictive ability is FACILITY-level evidence — estimated against realised loss given default, per facility — and one obligor holds several facilities with different collateral and different seniority, so it is not the obligor book with a column added. Import one at `/irb-intake` and this row computes.

LGD discriminatory power — gAUC over the ordinal outcome

No facility book is imported for this engagement. The gAUC measures discriminatory power over an ORDINAL realised outcome — does a facility this model rated worse actually lose more — which the obligor book cannot answer because it carries no realised loss. Import a facility book at `/irb-intake` and this row computes.

Margin of conservatism, categories A, B and C (EBA/GL/2017/16)

The margin of conservatism is a decomposition into deficiencies in data and methods (A), relevant changes to standards or processes (B) and general estimation error (C), applied per risk parameter and substantially documentary. It is not a portfolio-level over-prediction and cannot be derived from a scored population; `calbias` above is the statistic this pack does compute, under its own name.

Representativeness of the calibration sample for the current portfolio

A judgement about whether the data the model was estimated on still describes the book it is applied to. It compares two populations this platform is given one of.

Appropriateness of the long-run average default rate

Requires the full historical observation period and a judgement about where in the economic cycle it sits. A one-period population cannot evidence a long-run average, and a figure computed from one would read as though it had.

Downturn LGD adequacy

Out of scope twice over: LGD evidence is not read by this pack at all, and the downturn adjustment is a judgement about which historical period is the relevant downturn.

The use test, and the governance of rating overrides

Whether the rating system is actually used in credit decisions, and how overrides are proposed, approved and monitored, are properties of an organisation rather than of a model. Nothing in a scored population evidences either.

9 Conclusion and findings

1 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. 9 of 10 rows were not judged (no threshold declared) and are outside this conclusion.

1 finding(s) were raised, 1 of them material. Each carries the evidence a reviewer would re-run to check it.

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

1 live finding(s) have no owner assigned. An unowned finding has nobody accountable for closing it and will age without anyone noticing.

MATERIAL F-PD-FL00R The master scale assigns a PD below the regulatory floor

The lowest assigned PD on the declared scale is 0.030%, below the 0.050% floor. Unlike every other bar in this review this one is not the institution's: it is in the Regulation, so the breach is a matter of fact rather than of the bank's own tolerance.

Evidence to re-run: Master scale as declared, 7 grade(s). CRR Article 160(1), as amended by Regulation (EU) 2024/1623 (CRR3) — PD input floor of 0.05% for exposures to corporates and institutions, applying since 1 January 2025

10 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.



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 dc733e0d7676b2bf26d3f88cd13542555f7654c6b94c5a1bc8bc64d035fed30a

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