Just in Time

Systematic Backtesting of Probability of Default Models with Regulatory Data

July 2026

iason ESSENTIAL SERVICES FOR FINANCIAL INSTITUTIONS

Executive Summary

The paper[1] examines methodologies for systematic backtesting of Probability of Default (PD) models used by EU banks under the Internal Ratings-Based (IRB) framework. It addresses regulatory requirements under the Capital Requirements Regulation (CRR), which mandate regular validation of credit risk models, combining qualitative and quantitative techniques.

The paper identifies limitations in traditional backtesting methodologies since asset and serial correlations are unaccounted for. Those relies heavily on resource-intensive, bank-specific inspections and lacks transparency and scalability. To address this, it proposes a backtesting framework using a proprietary dataset collected by the EBA from 2017–2024 which includes also asset and serial correlation. This new systematic framework is more reliable because it accounts for more realistic conditions and it works both at individual bank level and at aggregate level across EU banks.

(1) *EBA Staff Paper Series No. 24, "Systematic Backtesting of Probability of Default Models with Regulatory Data", April 2026*

Copyright © 2026 - All rights reserved

At a Glance

Keywords: Credit Risk, IRB, PD

Copyright © 2026 - All rights reserved

01

Introduction

IRB Model Validation – Backtesting

Main Updates Implemented

iason

Copyright © 2026 - All rights reserved

Introduction 1/2

IRB Model Validation – Backtesting

Within the **Internal Ratings-Based (IRB) framework**, model validation is a key regulatory requirement aimed at ensuring that risk estimates remain reliable, robust and appropriately calibrated over time. In accordance with the **Capital Requirements Regulation** (CRR, Article 185), institutions are required to perform a **comprehensive validation process** covering both **qualitative** (e.g., governance, documentation, design) and **quantitative** (e.g., backtesting and benchmarking) aspects of rating systems and risk parameters.

Traditional supervisory validation relies heavily on on-site inspections, which are thorough but present **three structural limitations**:

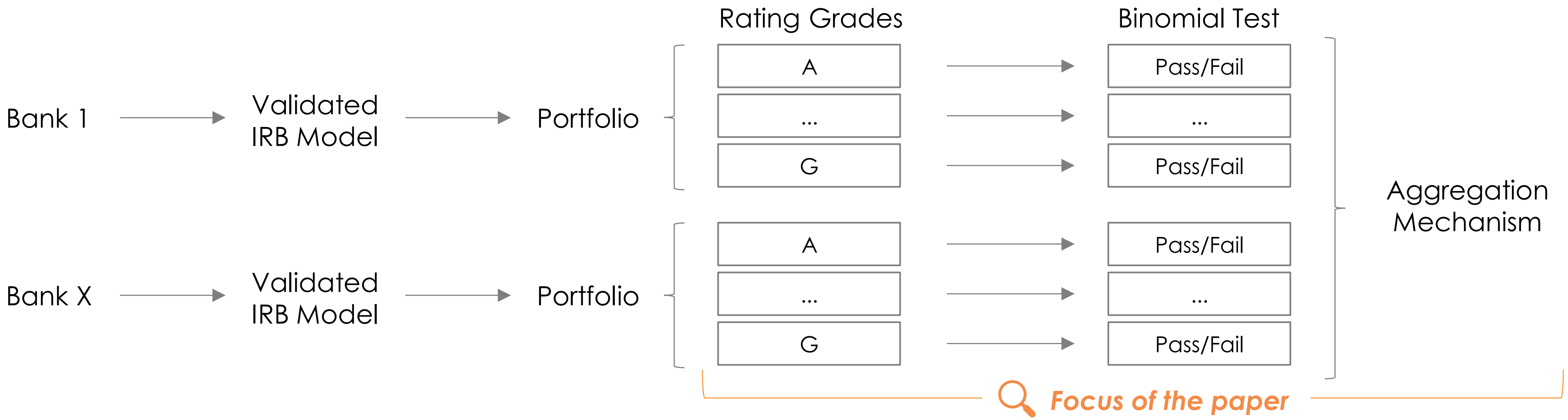
High resource intensity

Limited scalability

Low transparency, as results are rarely disclosed publicly

Backtesting

Backtesting, defined as the **comparison** between **predicted PDs** and **observed default rates (DRs)**, is a core tool but is typically **applied at individual bank and model levels**, limiting its usefulness for system-wide analysis. This creates a **gap** between **microprudential supervision** and **macroprudential oversight**, as aggregate insights into model performance are lacking.



Copyright © 2026 - All rights reserved

Introduction 2/2

Main Updates Implemented

The paper proposes a **comprehensive framework** for **systematic PD backtesting**, contributing to supervisory model validation through **novel data**, **empirical evidence** and **methodological innovation**.

Data

Data Perspective

A **novel proprietary dataset** collected by the **EBA between 2017 and 2024** is used.

This dataset includes **granular information on PDs**, **default rates**, **exposures**, and rating grades across EU banks, enabling both **cross-sectional and time-series analyses**.

Modelling Approach

Empirical Perspective

A **systematic framework** is introduced to:

- conduct **bottom-up backtesting** both at the **individual bank level** and at the **aggregate level** across EU banks. A robust **aggregation method** to transform granular backtesting results into an informative, **system-wide representation** of model calibration performance is proposed.
- quantify the **economic impact** of **miscalibrated corporate SME PD models** by assessing their impact on risk-weighted assets (RWAs) and Tier 1 capital ratios.

Methodological Perspective

A key innovation is the introduction of a **generalised correction to the binomial test**, addressing:

- **Asset correlation (cross-sectional dependence)**
- **Serial correlation (time dependence)**

The classical binomial test assumes independence of defaults, leading to **over-rejection of correctly calibrated models** (inflated Type I error). The proposed correction provides a more realistic framework supported by simulation evidence.

Copyright © 2026 - All rights reserved

02

Data

Perimeter of the Analysis

iason

Copyright © 2026 - All rights reserved

Data

Perimeter of the Analysis

The paper is based on proprietary regulatory **data** collected by the EBA **from 2017 to 2024** for **corporate SME** exposures from **two main data sources**:

Template C 08.05 COREP

Data is explicitly collected for the **backtesting of Probability of Default** (PD) in **Internal Ratings-Based** (IRB) **systems** used by **EU banks**, it includes: number of obligors; PDs; number of defaults; annual DRs per asset class and rating grade; long-run DRs per asset class and rating grade.

Exposures are **mapped** to a **common master scale** with fixed PD ranges. This allows data from different banks (which may have different internal rating scales) **to be compared** uniformly.

Data collected in the supervisory benchmarking exercise

In order to **expand the dataset** to cover the 2017–2020 period, **data** related to the **supervisory benchmarking exercise** for this period is added.

The information obtained from this dataset is the same as in the new COREP template in accordance with Article 78 of the 2013/36/EU Capital Requirements Directive (CRD).

	Mean	SD	P5	P25	P50	P75	P95	RG
A. Bank Level								
RWA	183,613	185,430	3,413	52,556	99,071	314,149	605,244	267
Tier 1 Ratio	16.5	2.4	13.2	14.8	16.0	18.0	21.1	267
B. Rating Grade Level								
PD	8.0	10.9	0.1	0.5	3.4	12.0	33.4	2,250
DR	4.8	8.5	0.0	0.2	1.5	5.7	20.5	2,250
Defaults	63	121	0	2	16	71	274	2,250
Obligors	3,670	6,990	34	260	1,140	4,314	14,738	2,250
EAD	2,582	4,733	16	155	734	2,927	11,402	2,250
RWA	1,204	1,885	18	116	489	1,431	5,149	2,250
Maturity	2.6	0.7	1.7	2.5	2.5	2.8	3.6	2,250

Tier 1 Ratio, PD, and DR are expressed in percentages; Defaults and Obligors are expressed in frequencies; Maturity is expressed in years; and RG indicates the number of observed rating grades.

iason
Copyright © 2026 - All rights reserved

03

Modelling Approach

Traditional Backtesting

Generalized Backtesting Approach

Distributional Properties

Monte Carlo Simulation

Multiyear Backtesting

Copyright © 2026 - All rights reserved

Modelling Approach 1/5

Traditional Backtesting

Backtesting procedures are used to assess the **reliability of the calibration of statistical models**, e.g., credit risk models used to estimate PDs.
The study proposes to compare the EAD weighted average PDs reported by different banks for a given asset class with the corresponding EAD weighted average annual DRs. Generally, **a well calibrated model should not exhibit neither an upward or downward bias**.

Binomial Test

For each bank and each rating grade, a **one-sided binomial test** is applied.
The null hypothesis $\mathbf{H_0}$ is that the model is **prudently calibrated** (PD ≥ DR in expectation).
H_0 is rejected when the observed DR exceeds the critical threshold.

$$p^*_{binomial} \approx \Phi^{-1}(q)\sqrt{\frac{PD\,(1-PD)}{n}} + PD$$

Aggregation Mechanism

$$MCB_b = \frac{\sum_{g=1}^{G} \mathcal{M}_{g,b} \cdot \mathrm{EAD}_{g,b}}{\sum_{g=1}^{G} \mathrm{EAD}_{g,b}}$$

To aggregate the results an **indicator variable $\mathcal{M}_g$ is introduced**, it is equal to 0 if the rating grade g is properly calibrated, 1 otherwise. This is applied to all the rating grades of all the banks (b) in the sample. This is then **aggregated by computing the weighted average percentage of miscalibrations using EAD**.

Recalibration

Since the binomial test provides only a pass/fail outcome, **assessing the impact of a prudent model recalibration on RWAs** and capital requirements helps evaluate the materiality of a failed test result. Hence, it is computed **the minimum PD (p^+) that would just pass the binomial test**. Using p^+ instead of the reported PD, hypothetical RWAs are computed with the Basel IRB formula: $\mathrm{RWA} = K \cdot 12.5 \cdot \mathrm{EAD}$.

$$p^+ - \Phi^{-1}(q)\sqrt{\frac{p^+(1-p^+)}{n}} \leq DR$$

Tier 1 Capital Ratio Impact

$$\mathrm{T1CR}_{\mathrm{impact}} = \left(\frac{\mathrm{T1C}_b}{\mathrm{RWA}_b^{\mathrm{adj}}} - \frac{\mathrm{T1C}_b}{\mathrm{RWA}_b}\right) \cdot 10^4$$

Finally, the **impact on the Tier 1 capital ratio** is calculated; the $\mathrm{RWA}_b^{\mathrm{adj}}$ augments reported RWAs by the additional capital implied by recalibrating failing grades.

Copyright © 2025 - All rights reserved

Modelling Approach 2/5

Generalized Backtesting Approach

Since in real world data the **assumption of independent defaults is rarely true** it is introduced a **generalized corrected binomial test** that accounts for both **asset and serial correlation**. The objective is to show that **once the correlation assumption** is **relaxed** the **distribution of Default Rate becomes right-skewed**.

Creditworthiness

It is assumed that creditworthiness is a random variable $\boldsymbol{z}$ that is unique for each obligor and **that $\boldsymbol{\rho}$ is the underlying asset correlation** (based on a one-factor Vasicek credit risk model). The systemic factor follows an autoregressive process of order one (AR(1)), where $\boldsymbol{\psi}$ denotes the **persistence parameter**.

$$z_{i,t} = \sqrt{\rho}\, y_t + \sqrt{1-\rho}\, u_{i,t} \qquad z_{i,t} \sim \mathcal{N}\left(0, \frac{1-\psi^2(1-p)}{1-\psi^2}\right)$$

$$y_t = \psi y_{t-1} + e_t \qquad y_t \sim \mathcal{N}\left(0, \frac{1}{1-\psi^2}\right)$$

Conditional PD

By using those specifications, it is possible to obtain the distributions of the creditworthiness and of the systemic factor y_t. Since the common systemic factor is fixed, the **conditional probability of default** can be calculated.

$$\mathbb{P}(z_{i,t} < c \mid y_t = y) = \mathbb{P}(\sqrt{\rho}\, y_t + \sqrt{1-\rho}\, u_{i,t} < c) =$$
$$= \mathbb{P}\left(u_{i,t} < \frac{c-\sqrt{\rho}y}{\sqrt{1-\rho}}\right) = \phi\left(\frac{c-\sqrt{\rho}y}{\sqrt{1-\rho}}\right) = f(y)$$

Portfolio Defaults

Then it is possible to determine the probability that the number of defaults in the portfolio is greater than or equal to a critical value k. The number of defaults observed in a portfolio is: $D_{n,t} = \sum_{i=1}^{n} D_{i,t}$. It is possible to calculate the **cumulative distribution function** of the random variable $D_{n,t}$

$$\mathbb{P}(\mathcal{D}_{n,t} \le k) =$$
$$\int_{-\infty}^{\infty} \sum_{j=0}^{k} \binom{n}{j} f(y)^j (1-f(y))^{n-j} \phi\left(\frac{y}{\sqrt{1-\psi^2}}\right) dy$$

Approximation

Since computing the exact CDF is **computationally expensive**, an **approximation** can be employed to determine the **critical value k** by numerically solving the expression for a specified **confidence level** (e.g., 95%).

$$\mathbb{P}(\mathcal{D}_{n,t} \le k) = \mathbb{P}\left(DR_t \le \frac{k}{n}\right) \to \mathbb{P}\left(\phi\left(\frac{c-\sqrt{\rho}y}{\sqrt{1-\rho}}\right) \le \frac{k}{n}\right) =$$
$$= \mathbb{P}\left(y \le \frac{\sqrt{1-\rho}\,\phi^{-1}\left(\frac{k}{n}\right)-c}{\sqrt{\rho}}\right) = \phi\left(\frac{\sqrt{1-\rho}\,\phi^{-1}\left(\frac{k}{n}\right)-c}{\sqrt{\rho(1-\psi^2)}}\right)$$

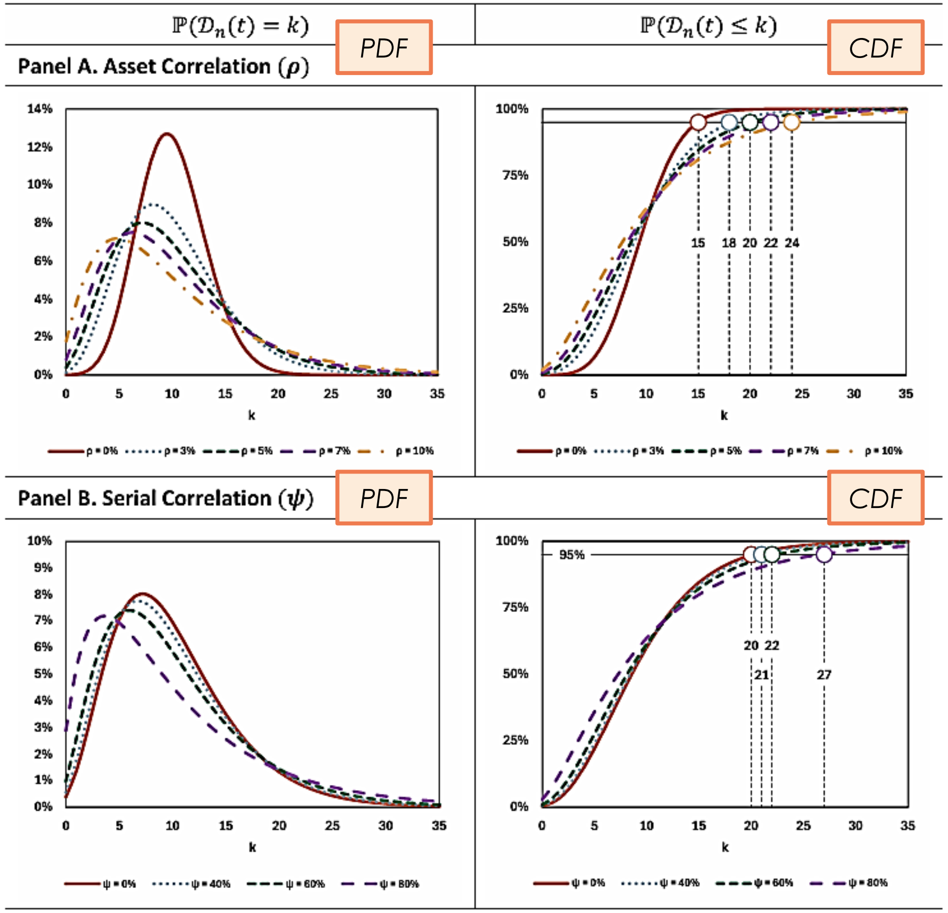
Copyright © 2025 - All rights reserved

Modelling Approach 3/5

Distributional Properties

Ignoring asset correlation (ρ) and serial correlation (ψ) makes the **backtest threshold too strict**, causing **good PD models to be rejected** too often. The figures are essentially showing how uncertainty in defaults increases when defaults become dependent rather than independent.

Panel A	**Effect of Asset Correlation:** asset correlation increases the volatility of portfolio default rates. Instead of observing around 10 defaults most of the time, we see more extreme outcomes, including large default clusters during adverse economic conditions. Asset correlation raises the number of defaults that can be observed before concluding that the PD model is miscalibrated. A threshold of 7.5% is appropriate only under independence; once common systematic risk is introduced, a threshold of about 10% becomes more realistic.
Panel B	**Effect of Serial Correlation:** serial correlation introduces persistence in default behavior. Moderate persistence has only a limited effect, but very high persistence substantially increases the probability of extreme default outcomes. With strong persistence, clusters of defaults become much more likely. Consequently, the maximum acceptable default rate at the 95% confidence level rises from 10% to 13.5%.



Input: *N=200 obligors, PD = 5%, Expected Defaults = 10*

Copyright © 2025 - All rights reserved

Modelling Approach 4/5

Monte Carlo Simulation

Monte Carlo Simulation

The analysis of the PDF and CDF demonstrates that **asset and serial correlation** alter the **distribution of default rates** and, consequently, the appropriate backtesting threshold. To **quantify** this effect, a **Monte Carlo simulation** study evaluates which testing framework achieves the desired **5% Type I error rate under a 95% confidence level**.

			PD = 0.5%			PD = 5.0%		
n	ρ	ψ	BT	BT - C ρ	BT - C ρ & ψ	BT	BT - C ρ	BT - C ρ & ψ
500	0.0	0.0	4.2	-	-	4.1	-	-
500	5.0	0.0	10.3	4.1	-	21.9	4.8	-
500	5.0	40.0	10.9	4.6	4.6	22.7	5.9	4.9
500	5.0	60.0	11.6	5.6	5.6	23.7	7.6	4.8
500	5.0	80.0	12.9	7.6	4.6	25.0	11.2	4.7
1,000	0.0	0.0	6.8	-	-	4.7	-	-
1,000	5.0	0.0	16.4	5.6	-	27.0	5.1	-
1,000	5.0	40.0	16.9	6.4	5.0	27.6	6.2	5.0
1,000	5.0	60.0	17.5	7.6	5.1	28.2	8.0	4.9
1,000	5.0	80.0	17.9	9.7	5.0	28.6	11.7	4.9
5,000	0.0	0.0	4.4	-	-	4.9	-	-
5,000	5.0	0.0	23.7	4.9	-	34.7	5.0	-
5,000	5.0	40.0	23.9	5.8	5.0	34.8	6.2	5.0
5,000	5.0	60.0	24.0	7.1	4.9	34.7	8.0	5.0
5,000	5.0	80.0	23.2	9.7	4.9	33.5	11.9	4.9
10,000	0.0	0.0	5.1	-	-	4.7	-	-
10,000	5.0	0.0	27.5	5.0	-	36.8	5.0	-
10,000	5.0	40.0	27.5	6.0	5.0	36.7	6.2	5.0
10,000	5.0	60.0	27.1	7.4	5.0	36.2	8.1	5.0
10,000	5.0	80.0	25.4	9.9	4.9	34.6	11.9	4.9

Table Legend:
BT = *Standard Binomial Test;* ***BT-C*** = *Corrected Binomial Test;*
$\boldsymbol{\rho}$ = *Asset Correlation;* $\boldsymbol{\psi}$ = *Serial Correlation.*

Evidence

1. The **BT** interprets correlated defaults as evidence of model failure. It works correctly when $\rho = 0$; $\psi = 0$ but over rejects well calibrated models otherwise.
2. The **BT – C** that accounts only for $\boldsymbol{\rho > 0}$ shows that correcting only for asset correlation solves part of the problem.
3. The **BT – C** that accounts for **both** $\boldsymbol{\rho > 0}$; $\boldsymbol{\psi > 0}$ keeps the rejection rate close to the target of **5%.**

Copyright © 2025 - All rights reserved

Modelling Approach 5/5

Multiyear Backtesting

Model validation should **not** be viewed as a **one-off exercise**, but rather as a **continuous and iterative process** (*BCBS, 2005*). Consequently, a key question is whether **a single year of miscalibration** provides **sufficient evidence** of **model deficiencies**, or whether **repeated breaches** over **multiple years** should be required before drawing such conclusions.

To address this issue, ***Blümke (2012)*** proposes a **multi-year backtesting framework** based on **order statistics.** Rather than focusing on a single observed default rate, the approach evaluates the **distribution of the r-th largest default rate observed** over a period of **T years**, allowing persistent miscalibrations to be distinguished from isolated events.

The test derives the **distribution** of the **r-th order statistic** and computes the **probability of observing at least r breaches over T years** under the assumption of a correctly calibrated model:

$$\mathbb{P}\left[\mathcal{D}R_t^{(2)} < x\right] = \sum_{i=k}^{T} \binom{T}{i} \left[\mathbb{P}\left(\frac{\sqrt{1-\rho}\,\phi^{-1}(x) - c}{\sqrt{\rho}}\right)\right]^{k} \left[1 - \mathbb{P}\left(\frac{\sqrt{1-\rho}\,\phi^{-1}(x) - c}{\sqrt{\rho}}\right)\right]^{k-1}$$

$$\mathbb{P}\left[\mathcal{D}R_t^{(2)} < x^*\right] = \alpha$$

x^* is the critical PD threshold corresponding to the chosen confidence level α, obtained from the distribution of the second-highest observed default rate over T years.

PD	T	**ρ = 0%**	**ρ = 5%**	**ρ = 5%**
		ψ = 0%	**ψ = 0%**	**ψ = 40%**
1%	8	0.0%	5.29%	11.22%
	15	0.0%	5.28%	12.01%
5%	8	0.0%	5.10%	11.03%
	15	0.0%	5.11%	11.81%

ψ = 40% represent the empirical value of the average serial correlation

The **results** of the application of the order test to Monte Carlo simulations provide an assessment of its **empirical rejection rate** under **different correlation structures**.

While the test performs as expected when asset correlation is correctly specified and serial correlation is absent, **neglecting serial correlation** results in a **significant increase in false rejections** of otherwise **well-calibrated models**.

Copyright © 2025 - All rights reserved

04

Results

Empirical Results – Baseline Approach vs Innovative Approach

Multiyear Backtesting – General Results

Copyright © 2026 - All rights reserved

Results 1/3

Empirical Results – Baseline Approach vs Innovative Approach 1/2

The **empirical results** obtained with the **standard binomial test** suggest a **non-negligible degree of PD miscalibration** across European IRB portfolios, with **13.4% (301/2,250)** of **rating grades failing the backtesting** exercise and approximately **16.7% of total exposures** classified as **miscalibrated (MCB)**

Furthermore, the distribution of **miscalibrated exposures** appears **highly concentrated**, with a pronounced **right tail** indicating that a limited number of rating grades account for a substantial share of the identified deficiencies **(a)**.

However, these findings prove **highly sensitive** to the assumptions underlying the backtesting framework. Once **asset and serial correlation** are **incorporated** into the analysis, the estimated extent of **miscalibration decreases** markedly. Even under relatively conservative assumptions, the share of **miscalibrated exposures declines** by roughly **one third**, while under more realistic correlation structures it falls to **between 2.9% and 3.5%** of total exposures **(b)**.

Impact of Asset and Serial Correlation on Miscalibration Estimates

*These results suggest that a large fraction of the **miscalibrations** identified by the standard binomial test are **attributable** to its **unrealistic independence assumption** rather than to genuine deficiencies in model calibration. Accounting for the **dependence structure of defaults** therefore leads to a substantially **more accurate** assessment of PD model performance, **reducing identified miscalibrations by approximately 80%** relative to the conventional approach.*

A. Binomial Test

	Sample	MCB	%	Mean	SD	P5	P25	P50	P75	P95	RG
(a)	Total	967,444	16.7	3,214	4,716	31	372	1,394	3,895	12,071	301

B. Binomial Test Corrected (ρ & ψ)

	ρ	ψ	MCB	%	Mean	SD	P5	P25	P50	P75	P95	RG
	1	40	647,671	11.1	2,699	4,023	20	271	1,018	3,388	10,984	240
	3	40	398,218	6.9	2,505	3,739	18	263	952	3,396	10,561	159
	5	40	317,281	5.5	2,498	3,957	17	184	809	3,205	10,561	127
	7	40	251,612	4.3	2,396	3,851	17	186	806	3,203	10,242	105
	10	40	205,501	3.5	2,418	4,065	15	154	772	3,205	10,242	85
	1	60	613,927	10.6	2,729	4,112	22	279	983	3,396	11,111	225
	3	60	383,287	6.6	2,625	3,870	19	186	998	3,471	10,561	146
	5	60	256,890	4.4	2,357	3,807	15	174	806	3,203	10,242	109
	7	60	214,320	3.7	2,381	3,966	15	154	789	3,205	10,242	90
	10	60	174,777	3.0	2,330	4,220	13	118	597	2,375	10,756	75
	1	80	443,935	7.6	2,466	3,619	20	235	966	3,300	9,986	180
	3	80	256,681	4.4	2,399	3,830	15	184	806	3,205	10,242	107
	5	80	204,006	3.5	2,458	4,104	15	154	772	3,396	10,242	83
	7	80	171,566	3.0	2,486	4,362	15	184	772	2,859	10,756	69
(b)	10	80	167,916	2.9	2,624	4,498	17	169	789	3,300	10,756	64

Copyright © 2026 - All rights reserved

Results 2/3

Empirical Results – Baseline Approach vs Innovative Approach 2/2

To gain further insight into **the distribution** of model deficiencies, the identified **miscalibrations (MCB) are decomposed** into **Investment Grade (IG)** and **Non-Investment Grade (Non-IG)** portfolios, using a **PD threshold of 1% (Section B)**. This decomposition allows the analysis to assess whether calibration issues are concentrated in higher or lower risk segments of the rating spectrum.

Real effects of miscalibrations (binomial test)

Subset	MCB	% MCB	Tier 1	Δ
A. Overall Sample				
Total	967,444	16.7	15.7	-10.3
B. Across Rating Categories				
IG	573,966	20.3	15.6	-13.8
Non-IG	393,478	13.2	15.8	-3.0
C. Over Time				
2017	130,085	23.2	14.8	-6.2
2018	165,164	26.0	15.2	-9.2
2019	156,476	20.9	15.7	-7.0
2020	130,552	17.3	16.5	-8.4
2021	77,119	10.0	15.4	-17.9
2022	80,157	10.2	15.5	-17.8
2023	134,970	17.2	16.3	-18.2
2024	91,921	12.1	16.3	-4.6

The estimated reduction (Δ) in the Tier 1 capital ratio amounts to approximately 10.3 basis points at the aggregate European banking system level.

The analysis also reveals a gradual improvement in model calibration over time, with the share of miscalibrated exposures declining between 2017 and 2024.
In the last four years of the sample the number of miscalibrated exposures is substantially lower than in the pre-COREP sample.

Real effects of miscalibrations (binomial test corrected with $\rho = 5\%, \phi = 80\%$)

Subset	MCB	% MCB	Tier 1	Δ
A. Overall Sample				
Total	204,006	3.5	15.7	-3.8
B. Across Rating Categories				
IG	192,706	6.8	15.6	-6.1
Non-IG	11,300	0.4	15.8	-0.2
C. Over Time				
2017	22,132	3.9	14.8	-1.7
2018	42,289	6.7	15.2	-2.5
2019	22,195	3	15.7	-1.3
2020	37,189	4.9	16.5	-2.1
2021	42,715	5.5	15.4	-9.5
2022	18,455	2.3	15.5	-8.9
2023	19,033	2.4	16.3	-8.9
2024	0	0.0	16.3	0.0

Where MCB and % indicates the miscalibrated exposures in EUR millions and percentages, respectively. Tier 1 is expressed in percentages and the corresponding economic impact Tier 1 Δ is expressed in bp

Copyright © 2026 - All rights reserved

Results 3/3

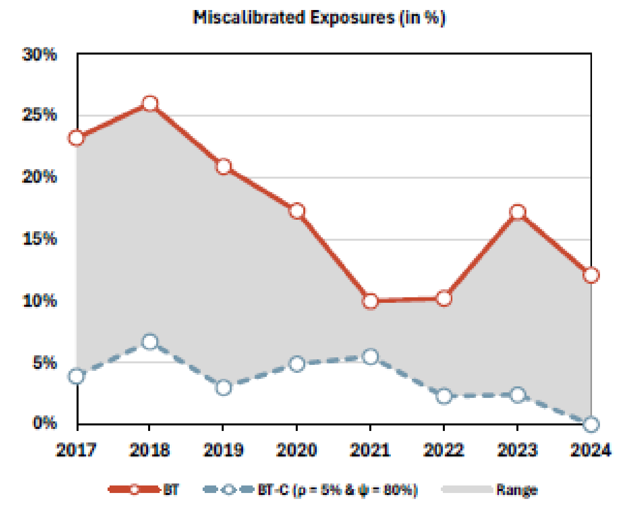
Multiyear Backtesting – General Results

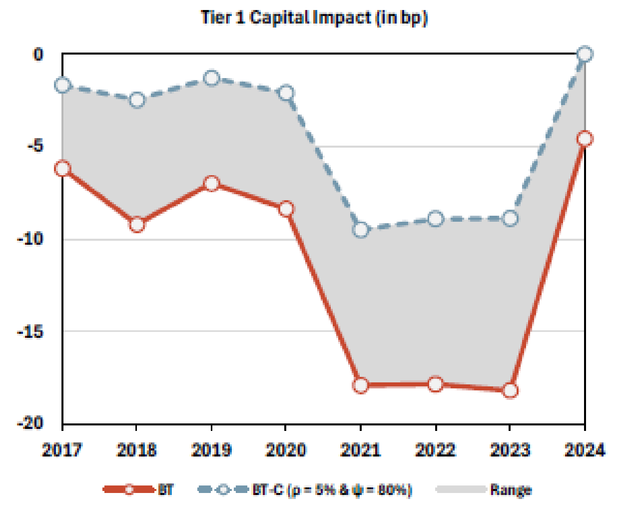
Applying the **order test** over the entire observation period provides a **complementary perspective** on model calibration. Unlike the standard binomial test, which evaluates each year separately, the **multi-year framework** focuses on the **persistence of miscalibrations** over time. As a result, the **share of EAD** associated with **miscalibrated rating grades** reaches **16.1%**, compared with 12.1% under the single-year binomial test.

Interestingly, the **rating grades identified** by the **order test** exhibit **lower average PDs** than those flagged by the **standard binomial test**. This suggests that the **multi-year approach** is able to **detect calibration deficiencies** in relatively **low-risk portfolios** that may remain undetected when backtesting is performed on a year-by-year basis.

Multiyear Test (ρ=5% & ψ=0%)	
% EAD MCB	16.1
Weighted Avg. Critical PD MCB	1.41
Weighted Avg. 2nd Highest DR MCB	1.84

Final results





The figure combines the results of the **different backtesting approaches** to derive a plausible range for both the share of **miscalibrated exposures** (on the left) and the associated **Tier 1 capital impact** (on the right).

The **standard binomial test** provides a conservative **upper bound**, whereas the **corrected framework** incorporating **asset** and **serial correlation** yields a more **realistic estimate**.

Together, these results suggest that **the true** extent of **miscalibration** and its capital impact are likely to **lie between these two extremes**.

Copyright © 2026 - All rights reserved

05

Conclusions and Key Takeaways

Copyright © 2026 - All rights reserved

Conclusions and Key Takeaways

Just in Time

Systematic, data-driven backtesting of PD models can **complement** traditional supervisory practices, **improving transparency**, **scalability**, and **confidence** in EU banks' internal models.

01

New methodology: generalised binomial test correction

The paper **proposed** a **theoretically grounded correction** to the canonical **binomial test** that simultaneously accounts for **asset correlation and serial correlation**. Simulation evidence confirms it keeps the **type I error rate** at the **nominal level**, **outperforming** both the **standard test** and **asset-only corrections**.

02

Miscalibrations are present but declining

Under the **conservative binomial test**, up to **16.7%** of **SME corporate exposures** are **miscalibrated**. **Under realistic correlation assumptions** (ρ = 5–10%, ψ = 80%) this falls to **~3%**. A clear **downward trend** in **recent years** suggests banks are increasing conservatism in their PD estimates.

03

Capital impact is meaningful but manageable

Prudent recalibration of failing models would **reduce** system-wide **Tier 1 capital ratios by 4–10 basis points** under the **binomial test**, and by **~4 basis points** under the **corrected test**. Investment-grade exposures drive the bulk of the impact, highlighting their systemic relevance.

04

Multiyear order statistics detect persistent miscalibration

Using the **second-highest observed default rate** across years as a test statistic captures rating grades with **persistent undercalibration** that single-year tests may miss, reinforcing that model validation must be a continuous, iterative process rather than an annual snapshot.

05

A public dashboard can restore market confidence

Regularly **publishing aggregated backtesting results,** combining the standard binomial test as an upper bound with correlation-adjusted estimates, would **enhance supervisory transparency**, support **macroprudential oversight**, and help rebuild trust in IRB internal models across the EU banking sector.

Copyright © 2026 - All rights reserved

ESSENTIAL SERVICES FOR FINANCIAL INSTITUTIONS

iason is an international consulting firm that has been supporting both financial institutions and regulators in topics related to Risk Management, Finance and ICT since 2008

Strategy

Strategic advisory on the **design** of **advanced frameworks** and **solutions** to fulfil both **business** and **regulatory needs** in Risk Management and IT departments

Methodology & Governance

Implementation of the designed **solutions** in bank departments **Methodological support** to both **systemically important financial institutions** and **supervisory entities**

Solution

Advanced **software solutions** for **modelling, forecasting, calculating** metrics and **integrating** risks, all on cloud and distributed in Software-as-a-Service (**SaaS**)

Company Profile

iason is an international firm that consults Financial Institutions on Risk Management. **iason** integrates deep industry knowledge with specialised expertise in Market, Liquidity, Funding, Credit and Counterparty Risk, in Organisational Set-Up and in Strategic Planning.

Marco Musto

Nicola Bertoli

Valeria Zancan

Matteo Vantaggio

This is an **iason creation**.
The ideas and the model frameworks described in this presentation are the fruit of the intellectual efforts and of the skills of the people working in iason. You may not reproduce or transmit any part of this document in any form or by any means, electronic or mechanical, including photocopying and recording, for any purpose without the express written permission of iason.

www.iasonltd.com

Copyright © 2026 - All rights reserved