Beyond the benchmark: Seeking success in non-retail IRB modelling

2025 Non-retail PD model development market survey

Contents

Executive summary

After a decade of 'IRB repair', inconsistencies in firms' approaches remain, and UK Internal Ratings Based (IRB) model approvals have been rare events – limited to mortgages and some niche non-retail portfolios. The risk of inconsistency is greatest in non-retail, where historical data volumes are lowest and subjectivity around design decisions the greatest.

This document is intended to bring fresh perspectives and advance the level of knowledge and understanding of non-retail ratings and Probability of Default (PD) model development. We discuss the principal activities in model development, including regulatory insights, deep-dives into key topics, and benchmarking results from a survey of over 20 banks across the UK and Europe.

How did we get here?
The European Banking Authority (EBA) in 2015 published a study on risk weights, and an associated discussion paper on the future of IRB (EBA/DP/2015/01). Whist supportive of the Internal Ratings Based (IRB) framework for calculating minimum regulatory capital requirements for credit risk, the EBA observed significant variability in both approaches and risk weights, to the extent that trust in the use of IRB models was eroded. To restore confidence in the IRB framework, prudential regulators and market participants embarked on a set of initiatives to address non-risk-based variability in internal models.

This included:

- The issuance of significant volumes of EBA Guidelines and Technical Standards (often referred to as the 'EBA 2020' package).
- The Targeted Review of Internal Models (TRIM) project conducted by the European central Bank.
- Improvements in transparency and availability of information to market participants, including supervisory benchmarking exercises.
- Firms' own redevelopment of internal credit models, to both comply with the 'EBA 2020' requirements and reduce non-risk-based variability in estimates.

Whereas retail data and loss histories typically contain enough observations for robust statistical modelling, the relative paucity of representative default and loss data in non-retail exposure classes necessitates a greater reliance on expert judgement and subjective inference. With subjectivity comes a risk of inconsistency, and that risk increases where model developers are required to make design decisions that prioritise or trade competing requirements.

Whilst credit modelling is generally seen as a well-established science (or indeed art), our own benchmarking indicates that there remain significant areas of divergence between firms' approaches. The focus of this document is therefore non-retail rating and PD model development. We discuss the following topics:

IRB programme success: Beyond modelling (section 3)
Successful IRB programmes are not just about building models; they require a holistic approach that integrates risk management, business strategy, and regulatory compliance.

Segmentation (section 4)
Successful IRB programmes balance the need for homogeneity within segments with regulatory constraints. Successful programmes overcome challenges with standalone versus consolidated turnover and manage boundary effects without resorting to 'grace periods' or blending approaches.

Data challenges: Quality, representativeness, and history (section 5)
Successful IRB programmes deliver sustainable remediation of data quality issues at source, to provide a high-quality risk driver, default and loss dataset that covers a complete economic cycle and adheres to current data management standards. Where use of external data is unavoidable, data quality risks (particularly the consistency of financial statement spreading) are mitigated by extensive controls including sample testing and reconciliations.

Missing data and adjustments: Ensuring unbiased estimates (section 6)
Successful IRB programmes avoid solving data quality problems with layers of imputations and margins of conservatism.
A particular and often overlooked risk arises where missing value imputations introduce 'hidden over-fit' that can inflate model performance estimates. Successful programmes admit only 'neutral' imputations, with strict limits in place to restrict the frequency of use.

PD risk differentiation: Choosing the right model (section 7)
Successful IRB programmes employ a hierarchy of well-established modelling techniques, with an overall preference for default replication. The risk of default prediction models giving reduced discrimination amongst 'good' obligors does not seem to have crystallised. Where firms take the benefit of parental support (implicit or formal), a divergence in approaches remains, in both the grade assignment process and how parental support is reflected in modelling datasets.

Calibration: Aligning with long-run default rates (section 8)
Successful IRB programmes stick to well-established and stable calibration techniques. In practice this means alignment to a 'central tendency' or portfolio-level long-run average default rate. The elimination of direct PD estimation with masterscale mapping under Basel 3.1 has not led to widespread adoption of discretised calibration techniques commonly used in retail portfolios. Most firms seek to incorporate historical overrides into the calibration sample, though considerable divergence exists in the operational implementation.

Margin of conservatism
There exists a general harmonisation of Margin of Conservatism (MoC) frameworks in the market. Firm-specific implementations vary however, especially around the grouping of related uncertainties and approaches to the long tail of immaterial items. MoC is the subject of a separate Deloitte benchmarking exercise, and therefore not discussed in detail.

Moving forward: Continuous improvement in a changing landscape
Since enactment of the first BCBS paper in 1999 which laid the foundation of the regulatory principles for internal models, the Basel IRB Framework has undergone a rich evolution in its 25-year history (and kept many a CRO awake at night). Whilst IRB is reaching maturity amongst most industry participants, firms remain encumbered in their route to approval status, with data limitations, resource constraints and regulatory engagement reflecting the common problems found in this last hurdle. Generally, firms that have continually invested in their data and modelling capabilities have been more successful.

Reaching a final approval status is a sought-after position, for firms as well as regulators, particularly as competent authorities across the world seek to slow-down enactment of regulations in favour of economic growth.

About the survey

Information was sourced from conversations with a representative sample of banks across the UK, SSM and Nordics. The focus of this survey is on their non-retail credit portfolios with an IRB permission. To preserve anonymity, we have not disclosed countries or total assets.

Figure 2.1: Current stage of majority of models

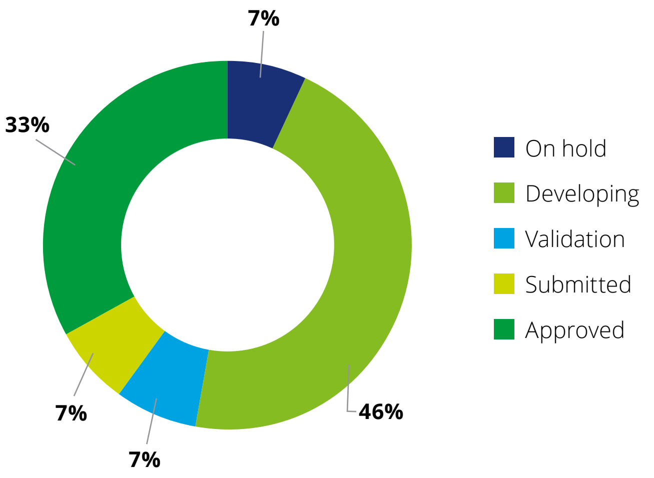
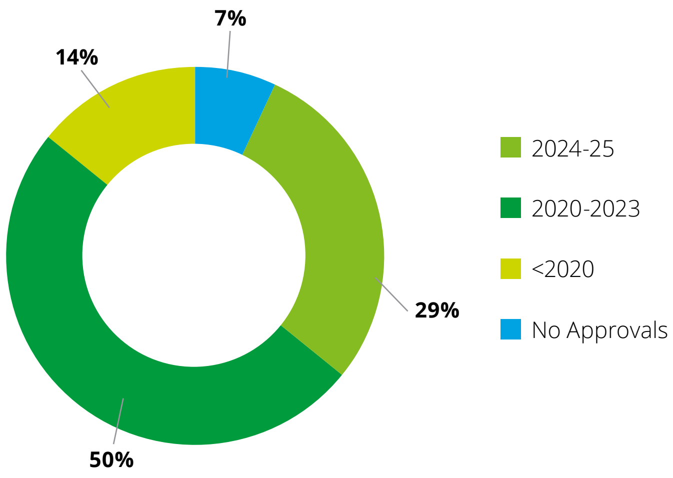


Figure 2.2: Most recent regulatory approval post 2020?



Source: Deloitte 2025 IRB Market Analysis

IRB programme success

This section explores key enablers of IRB programme success that we have observed in the market. Despite IRB being frequently perceived as a pure modelling activity, the credit modelling itself is a reasonably mature field with a degree of consensus around model frameworks, risk drivers, and calibration levels. The modelled assumptions themselves, with perhaps a couple of exceptions, are conspicuously absent from the set of enablers discussed below.

New technical requirements

Cyclicality

The industry has been slow to fully appreciate that cyclicality and model behaviour, especially during a stress, are matters of high importance to the whole bank as well as for macro-stability. Successful submissions have achieved a firm-wide consensus on both the target rating philosophy, and the level that is achieved in practice. In particular, a recognition that historic data may suppress observed cyclicality due to the low interest rate environment and missing risk driver and/or default triggers should carry over into prudence in forward-looking (stress test) assumptions.

Navigating requirements

The 'EBA 2020' package produced a considerable volume of new technical requirements for firms to adopt across their organisations including credit processes, data, IT and Finance processes. Furthermore, instances exist of conflicts between regulatory and business requirements, that firms need to prioritise carefully.

Long term IRB programme success is associated with firms that have been able to make firm-wide consensus trade-offs where technical requirements conflict with business requirements; and achieve a firm-wide embedding of the consensus.

Conversely, firms that regard IRB delivery as a credit modelling project, constraining decision-making to using modelled assumptions in lieu of business model changes, tend to deliver overly complex models that strive to meet all technical requirements but fail to achieve wider business acceptance.

Programme management

Programme design

Successful IRB submissions are typically delivered by a dedicated IRB programme, supported by strong executive sponsorship, to drive day to day activities across risk, finance and IT. Dedicated workstreams typically manage credit model, data, IT delivery; as well as changes to the operating model that may be required to meet regulatory requirements or enhance operating efficiency.

By contrast, the management of IRB changes in BAU (typically modelling) silos is associated with a disconnect from business objectives that manifests itself in prioritising compliance and conservatism over delivering a useful business tool that enables business outcomes; as well as a general lack of firm-wide consensus on key assumptions such as cyclicality.

Senior management awareness & oversight

Senior management need to be able to manage expectations, raise matters with, and respond to, the PRA in close and continuous meetings, from an informed standpoint. Senior management engagement of IRB change programmes has been somewhat variable, with some firms omitting NED engagement altogether.

Project management office

To align with project timelines and objectives, all staff engaged in delivery should share the responsibility for raising risks, stating assumptions, identifying dependencies, and escalating issues. These in turn should be monitored and challenged centrally by the project management office (PMO).

This approach allows the PMO to manage these aspects at a program level, ensuring a coordinated and proactive response to challenges, and in turn, enable wider programme success.

MRM governance weaknesses

There exists a perception that the quality of IRB submissions, and quality of Model Risk Management supporting the submission, has waned since the initial set of IRB waivers some 20 years ago.

Banks have fed back that obstacles include:

- Models that have been over-engineered to accommodate every requirement, resulting in models that are something of a "jack of all trades, master of none".
- Interpretation choices that are perceived as having been made to maintain the capital state.

- Calibrating the appropriate level of business involvement, avoiding both 'analysis paralysis' as well as a perception of ignoring business needs that would jeopardise Use Test compliance.
- The industry experienced a long period where no submissions were made, and experienced resources changed roles or left the organisation.
- Instances where independent validation had failed to adequately challenge material assumptions or decisions.
- Instances where the severity of validation findings were revised down, typically by documenting details of approaches that remained inappropriate for the business requirements or regulatory rules: A mis-specified model that has been fully documented is still a mis-specified model.

Regulatory change fatigue

Banks have fed back that there exists a considerable degree of regulatory change fatigue across the industry. Challenges include:

- Delays in the EBA 2020 package, where the final GLs came with changes from the CPs, requiring rework, as well as perceived ambiguity around whether to align to EBA texts or PRA expectations, in situations where conflicts arose.
- Where new rules were appraised, the focus tended to be on where rules had changed. The need for updates to interpretations and assumptions attached to unchanged line items was sometimes missed, under a culture of aligning to existing (typically pre CRD IV) approaches.
- Beyond model submissions, firms need to implement models and reporting, into IT systems. IRB model outputs stand out as being consumed by (or affecting) a sizeable number of downstream processes including pricing, stress testing, impairment, forecasting, and even remuneration; and therefore present a great many dependencies (up and downstream) for release. Any delays in model approvals or regulatory go-live dates can trigger significant work to untie committed releases.
- Whilst the UK Basel 3.1 near-final text is closely aligned with the global standard, divergence exists in the practical implementation, with respect to the EU. Any decision to diverge between jurisdictions and adopt a 'multi risk rating' approach in non-retail represents a strategic change in direction, which cascades into programme delivery in the form of significant changes in scope, requirements, and approaches.

Resourcing challenges

Banks have suffered from loss of organisational experience where more experienced modellers change roles. More lenders are currently IRB or in the process of applying for IRB, therefore knowledge across the industry is spread thinly.

Successful IRB delivery is associated with firms that have:

- A long-term resourcing strategy to acquire, develop, train, and retain teams with world-class skills and knowledge.
- Made effective use of off-shore or near-shore resources, fully integrating into the onshore teams.
- Recognised that a great deal of the technical work is a mature science and implemented internal standards to foster consistency and make IRB accessible to junior staff.
- Managed close and productive collaboration between business, data, model development, model validation and technology teams.

Data limitations

Data quality

First generation IRB permissions were typically granted to firms despite known data limitations, on the understanding that a step change in data management practices was expected. Modelling teams however continue to be beset by data quality issues, and the industry is perceived as not having progressed sufficiently on data quality, especially at firms that have been asserting BCBS 239 compliance for some time but failed to demonstrate progress in addressing sources of uncertainty that trigger Margins of Conservatism (MoC). A particular challenge, is that the remediation of data quality issues at source can take a considerable amount of time and resource to deliver and can require complex dependency management with other data-consuming processes, including live credit underwriting. Firms may opt to prioritise conservatism and preserve timelines, over remediation. Indeed, cost-benefit analysis of remediation versus MoC usually favours the MoC, albeit without considering the wider firm-wide costs of ongoing unresolved DQ issues.

Default definition

Revisions to the (live, reporting) default definition have brought IRB programmes face-to-face with the real-world operational challenge of sourcing attributes from complex source system landscapes and misaligned expectations around what 'good' and 'material' look like, between Risk, Finance and Technology.

Where firms are making significant changes to parental support frameworks, implementation of the default definition to historical data represents a particular challenge: In particular, firms that have historically applied a 'notch from parent' subsidiary rating need to generate a standalone rating under the appropriate standalone model scope. This potentially adds obligors as well as defaults to the dataset. However, historical default events may be unobservable due to the presence of parental support. Firms typically must choose between biasing assumptions (i.e. rules to infer standalone defaults) and exclusions from the standalone model dataset that can cause representativeness issues and therefore also be biasing.

Margins of conservatism

The requirement to link the margin of conservatism to specific sources of uncertainty has laid bare firms' data quality issues and attached a capital cost to low quality data. As well as the requirement that MoC not be used in lieu of remediation, firms are expected to be able to link MoC to remediation plans, which in turn can be revealing of where firms have regarded IRB as a modelling project.

Segmentation

This section discusses segmentation strategies for modelling, particularly for large corporates. It highlights the tension between aligning model scope with regulatory boundaries (e.g., the Basel 3.1 exposure class boundary of €500m turnover) and ensuring homogeneity within segments for accurate risk assessment.

Regulatory insights

- The CRR aligns 'rating system' boundary with 'type of exposures', which is defined in terms of homogeneous management.
- The CRR also makes general provisions that models should return unbiased estimates, with reasonable and effective input parameters. In practise this typically results in extending the homogeneity requirement to reflect shared risk drivers and relationship with default risk.

Discussion

- Modelling serves to generalise an observed relationship between risk driver inputs and default risk. Segmentation refers to identification of cohorts of obligors that can be grouped together for modelling purposes. Segmentation presents a trade-off, as the introduction of segments reduces the number of free parameters that can be estimated from the available data.
- In low- or lower-default segments, quantitative analysis alone is typically inconclusive and should be augmented with qualitative information including:
- Alignment of model scope with exposure class boundaries can reduce volatility at the boundaries of reporting cohorts but result in non- or less-homogeneous borrower behaviour within the segment.
- Alignment between model scope boundaries to business delineation helps to deliver homogeneity of account management, but again may fail to recognise heterogeneous populations, e.g. by obligor size, industry or region. However, this risks overlapping model scopes between business units, in terms of key risk characteristics such as company size.
- An understanding of obligors' market dynamics can help to inform segmentation. Compared with mid corporates, large corporates in particular are typically better diversified in terms of products, suppliers and customers; have more-sophisticated risk management (and hence ability to respond to competitive threats and exogenous shocks); and exhibit greater financial flexibility (e.g. they can issue or repay debt or equity in the market).

Under Basel 3.1 the large corporate exposure class boundary lies at 500m EUR (EU) or 440m GBP (UK) average consolidated annual revenue over three years. Firms may wish to align rating system scope boundary to the exposure class boundary. Benefits include avoiding volatility or 'flip flopping' of large exposures, between reporting lines. Disadvantages include lack of homogeneity if otherwise 'small' entities enter large corporate model scope on grounds of consolidated turnover alone, as well as representativeness challenges if 'small' entities are mixed with external (e.g. ECAI) data to inform or test modelled assumptions. Additionally, 'small' entities may be managed differently and fail to satisfy the CRR's homogeneity requirements.

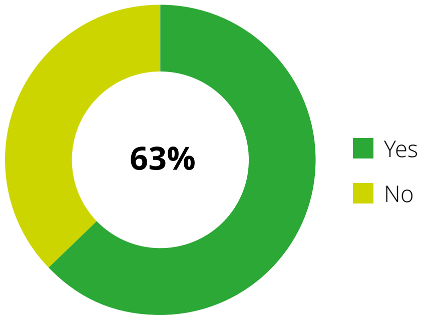
Benchmarks

- Some banks are not harmonising on a mid/large corporate split at 500m consolidated EUR/440m GBP, per the Basel 3.1 exposure classification boundary, and are instead continuing to segment based on standalone turnover.
- A minority of firms would admit obligors with lower turnover but high total assets.
- Borrowers that operate under constraints that would prevent them from behaving like a large corporate in the market typically fall into a scope that is separate from the primary corporate or large corporate model.

Examples include serving a particular industry; have a narrow product set; are highly regulated; are highly concentrated; and/or are special purpose vehicles. This tends to include lending to banks, non-bank financial institutions, funds, real estate development/investment, project finance, sovereigns, sub-sovereigns, regulated utilities, agriculture, and education. In practice, depending on materiality, these exposures would have their own rating models, or calculate RWAs using the standardised approach.

The frequency of flip-flopping or cliff-effects between rating models or segments can be managed by applying a grace period or considering the moving average of turnover. Blending approaches in turnover boundary regions are not observed.

Figure 4.1: Percentage of banks where turnover threshold aligned to CRR requirements of 500M on consolidated group level



Source: Deloitte 2025 IRB Market Analysis

Data sources

This section focuses on data requirements for credit risk models under the CRR, emphasizing the importance of data quality and representativeness. Particular emphasis is given to availability of data, applicability and use of external data sources (for both default identification and risk driver derivation), as well as representativeness challenges found in scope of model application.

Regulatory insights

- CRR Article 179 requires that:
  - PD estimates incorporate all relevant data, information and methods.
  - Firms apply greater conservatism, where there is less data.
  - Data be representative of obligor characteristics, current lending standards, and current or foreseeable economic or market conditions.
- The CRR text on data maintenance (Article 176) emphasises the importance of collecting internal risk driver, default and loss data, as well as the input components for derived quantities such as default and loss.
- CRR Article 174 requires that models are estimated on complete, accurate, representative and appropriate data.
- CRR Article 180 requires that firms use at least five years of data. External or pooled data is permitted.

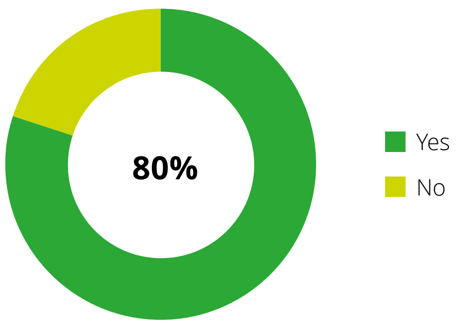
Discussion

Collection and processing of modelling data can be split between objective and subjective measures of data quality. Objective measures such as completeness and accuracy generally align to firm-wide data management policies and ensure that attribute values in datasets are not degraded with respect to the original point of entry or source system feed. Subjectivity is introduced by the need to make the data sample representative of a future obligor profile, set of lending standards and exogenous environment that is by its very nature uncertain.

Representativeness includes external conditions as well as the profile of obligor attributes such as size, country and industry. Whilst corporates' financial ratios exhibit greater overall stability across the cycle than retail model inputs, idiosyncratically large migrations can occur especially where credit quality is deteriorating. It is therefore appropriate to consider a complete economic cycle of data for model estimation, capturing the 2008 stress.

A particular challenge with external data, is robustly demonstrating quality in the absence of inspectable lineage. Common challenges include equivalence of risk driver and default definitions, as well as representativeness of the internal portfolio, especially where external data pertains to ECAI-rated entities or ECAI coverage of the internal portfolio is low. Where external financial information is used, diligence should be performed to ensure that it meets internal standards for data quality, with adjustments being required to align with the internal 'house view' on spreading adjustments, for example to calculate cash flow for debt service or reflect leasing as long-term debt.

Figure 5.1: Proportion of firms surveyed which use external risk driver data

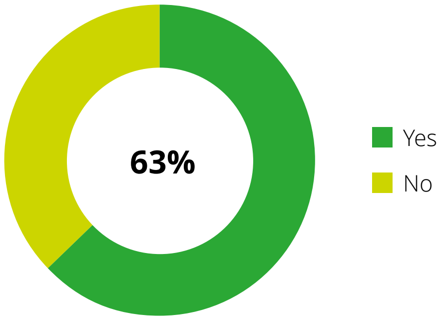


Source: Deloitte 2025 IRB Market Analysis

Benchmarks

- The majority of banks estimate corporate models using a complete economic cycle that starts at or shortly before 2008 and extends until the next peak or present day.
- Representativeness challenges with the 2008 stress – most notably, being a low interest rate stress – are generally accepted as model limitations.
- A minority of firms consider external defaults in addition to internal default, generally in specialised circumstances where it is considered appropriate to model for the 'market portfolio' to achieve an effective range of operation that is most useful to the business.
- Where external data is used for the dependent variable, risk driver information is typically sourced internally due to differences in definition and financial spreading that can lead to significant differences between model estimation and execution data.
- GCD and Credit Benchmark data are generally used for benchmarking purposes only.

Figure 5.2: Proportion of surveyed firms with include 2008 in development period



Source: Deloitte 2025 IRB Market Analysis

Key insights

In sourcing appropriate external data for model development, one firm had to contend with low internal default volumes. This was addressed through supplementing internal default data with external default ratings. Specifically, default volumes were augmented by considering the lowest ranked external grade ratings of exposures in the portfolio. This was blended with the internal DoD rules to create a default flag which incorporated both internal and external information.

The external default data was selected using a combination of expert input (e.g. through identifying jurisdictions comparable to the internal portfolio), along with a bootstrap sampling approach to approximate key risk driver distributions with internal data. To further validate comparability between internal and external default data, distributions were compared through PSI statistics on key risk drivers of both populations.

Later in calibration, the model was calibrated based on the two separate internal and external default populations. Due to perceived differences in the risk profile/level between the two, for conservatism purposes the calibration was targeted on the maximum of the Long Run Average Default rate of the two populations.

Data manipulation

This section delves into the critical aspects of data quality and management in credit risk modelling, particularly highlighting the challenges of missing data and the need for adjustments to ensure representativeness. Firms typically deal with data challenges when legacy bank datasets cannot provide the new level of detail, accuracy and consistency required by regulatory expectations in model development.

Regulatory insights

- CRR Article 179 requires:
  - PD estimates to incorporate all relevant data, information, and methods; and
  - Firms to apply greater conservatism, where there is less data.
- The CRR text on data maintenance (Article 176) emphasises the importance of collecting internal risk driver, default, and loss data, as well as the input components for derived quantities such as default and loss.
- CRR Article 174 requires models to be estimated on complete, accurate, representative and appropriate data.
- CRR Article 178 sets out the default definition, applicable to both standardised and IRB portfolio. PRA expectations are set out in SS3/24, replacing existing rules in SS10/13, SS11/13 and the EBA rules.

Discussion

Missing data

Incomplete attribute values (missing values) present a particular challenge in credit modelling. As well as reducing the effective sample size, missing data is frequently systematic (i.e. not random) and can have a severe biasing effect. Imputation techniques are popular for overcoming missing values, but in-effect introduce additional free parameters into the model, irrespective of whether used for estimation or in the final specification. These hidden assumptions are easily over-fitted (for example, industry-specific imputations can become a proxy for industry) to the extent material distortions are introduced that remain undetected and lead to over-confidence in observed performance metrics. Missing data imputations should only be used where absolutely necessary, and can be shown to be non-biasing.

The biasing effects of missing values can become amplified during model calibration: A common pitfall is to apply prudent imputations during calibration. This becomes biasing in the other direction when the model is applied to a complete set of inputs.

Adjustments

Where historical lending standards and macro conditions are unrepresentative of current or expected future conditions, it is appropriate to make adjustments to observed data. A key example of this is the changes in default definition that have occurred over time.

Historical default definition

In non-retail modelling, default entry criteria are often consistent over time, requiring re-statement only for changes in probation period and materiality of past due amounts. The creation of a historical time series of default events continues to present challenges:

- A common pitfall is to re-state only the default definition, without full consideration of credit policy. For example, if accounts were historically reviewed only semi-annually, historical past dues may be identified that would not occur under today's credit management policies and processes. Such cases are in-effect false positives and can serve to introduce noise into the dataset that biases the model selection. In such a scenario, it is appropriate to exclude from the list of default events (albeit recognising any residual uncertainty and carrying a margin of conservatism, as appropriate).
- Whilst default restatement typically focuses on adding default to the set, erroneous flagging in front-end systems can propagate into time series data. Firms should therefore be mindful that false positive default events, whilst prudent, can distort risk ranking models and calibration levels.
- Some firms' policies require cross-defaulting large cohorts of lending (e.g. all foreign currency lending and/or all bank exposure) in the event of a sovereign default. The resulting spike of obligor defaults can mask the obligor default risk relationship being modelled and can result in an overly backward-looking sovereign risk component to the calibration level.

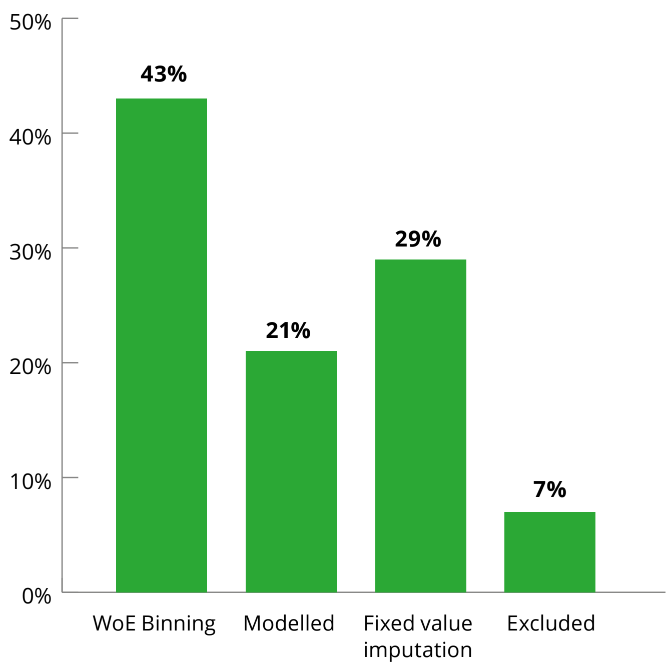
- Where credit experts are engaged to identify default events or verify default events identified using rules, a great deal of human judgement is introduced into the modelling process. Inconsistent levels of understanding of the default definition continue to be reported, especially for niche lending such as financial institutions (bank and non-bank) where even single digit adjustments to historical default volumes can trigger significant movements in calibration level.
- Where obligor exposures exist across systems, business units and legal entities, adjustments are typically required for variations in past due materiality thresholds.

Sample design
Application of textbook modelling techniques would permit stratified sampling to achieve a data sample that is representative of the live portfolio. Sample design also presents an opportunity reduce the risk of over-fitting onto the population where credit was approved, e.g. avoiding repeat or overlapping observations.

Benchmarks

- Tolerance for missing values is generally low, with neutral imputation only applied where the percentage of missing values is low and can be shown to occur randomly. A standback test is to determine if the model specification is robust to missing value treatment. Stricter firms place limits on missing values, in terms of overall percentage (e.g. 5%) as well as the number of missing characteristics for a record to be admitted (one overall, and no key inputs).
- Firms commonly down-sample time series data in order to avoid repeat or overlapping observations and reduce the risk of correlated residuals that would violate regression assumptions.
- We are not aware of firms applying re-sampling techniques such as stratification to enhance representativeness of the application portfolio. Firms instead tend to assume that the historical population is sufficiently representative.
- Modelled approaches to missing value imputations (e.g. conditional on other risk drivers or the dependent variable) are rare, as this would typically be associated with having not captured or remediated sufficient data before starting modelling. Typically, where firms have used a WoE binning approach for risk driver segmentation, a separate bin is designed for missing values, assuming their homogeneity in capturing default risk.

Figure 6.1: Approach for treatment of missing values



Source: Deloitte 2025 IRB Market Analysis

A systematic approach to treatment of missing values

The starting point is a history of risk drivers that is generally well-populated and extends back to the 2008 period.

Acceptance criteria are in place, such that observations may only have a maximum of one missing risk driver, and certain designated key risk must be populated.

Gaps in risk driver information address during the data gathering/processing phase, whereby any missing information is remediated at source.

Residual missing values are subject to neutral imputation, such that the impact of removing the risk driver entirely does not affect the model output derived from the remaining (populated) risk drivers.

PD risk differentiation

This section discusses the challenges that firms face in development of the PD Risk Differentiation approach. This component is a crucial part of the model development, with the overall purpose to derive a rating process which can adequately differentiate low and high-risk accounts on the basis of perceived default risk, using available risk factor information.

Regulatory insights

- The CRR allows firms considerable flexibility in their approaches to risk differentiation. This helps to ensure business model alignment. Indeed, the use of models is not mandated.
- Where firms choose to use rating models, the CRR articulates general principles in favour of applying established science.
- SS4/24 requires that firms document an appropriate policy in relation to novel or narrow rating techniques.
- PS9/24 leaves open the possibility for firms to reflect parental support into rating assignments using either a 'notch from standalone' or 'notch from parent' approach. Under both approaches, CRR article 174 requires firms to rate each separate legal entity, necessitating a standalone obligor grade even where full parental support is achieved.

Discussion

Modelled rating assignments

The art and science of non-retail modelling generally favours three families of models:

- Default Prediction: Regression-family (probit, logit regression) models trained to differentiate observed binary (default, non-default) outcomes. Whilst offering the closest match to internal data, downsides include a relative paucity of data in "good" segments that can serve to reduce risk discrimination as well as representativeness challenges where historical defaults are in segments where banks have reduced exposure.
- Shadow Rating: Regression-family models optimised to maximise alignment with an external benchmark ranking, typically ECAI rating. Whilst the modelling itself is a reasonably mature discipline, challenges include demonstrating equivalence of the benchmark rank order and default definition.
- Expert Lender: Regression-family models optimised to mimic experts' rank order. A key advantage is business buy-in to the ranking, at the risk of biases if expectations are anchored to existing modelled ratings. Other approaches exist beyond regression-family models, most notably the Merton structural model and practical approximations. Additionally, supervised learning techniques beyond regression-family models continue to capture the imagination.

The expectation that firms document an appropriate policy in relation to narrow or novel rating approaches, should generally be interpreted as sticking to well-established and understood techniques supported by standard selection steps (heuristics), avoiding bespoke research and development type projects within IRB delivery.

Parental support

In practice, material amounts of non-retail exposures are assigned grades that are not directly the output of a rating model. As well as the possibility of grade overrides, parental support frameworks continue to receive considerable scrutiny, both internally and from regulators. Under both approaches, firms are expected to provide a standalone obligor rating. A standalone obligor rating is needed for the scenario where parental support is withdrawn or cannot otherwise be assumed, both at reporting date and for scenario analysis or stress testing.

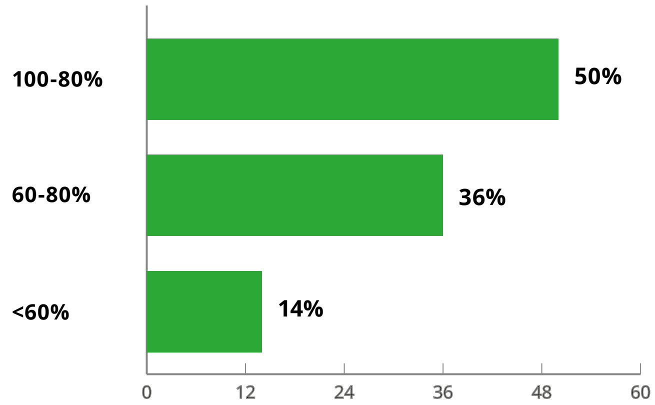
Standalone obligor ratings present modelling challenges, including:

- Unobservable risk driver information, such as subsidiaries that don't produce financial statements.
- Uninformative risk driver information, such as corporate treasury centres that typically show no revenue or assets.
- Distorting effects of adding to standalone model scope, large volumes of supported obligors and unobservable default information (discussed above).

Benchmarks

- The current consensus prefers default prediction models, given the strong link to internal data. The loss of predictive power in 'good' cohorts is not generally seen as a dealbreaker.
- Banks generally operate a hierarchy, with preference for default prediction before reverting to agency replication and internal expert rank order replication.
- The majority of corporate models use default prediction.
- Shadow rating models are used only where internal default volumes are insufficient, and reasonable levels of coverage of the internal portfolio can be achieved.
- Expert lender models are used as a last resort for scenarios without robust default volumes or ECAI coverage of the internal portfolio.
- A consensus is yet to emerge, on how to model supported entities, for standalone obligor ratings.
- Generally, most firms consider a blend of financial and non-financial risk factors in the ranking model, with a slightly higher weighting on financial factors. Other approaches include use of non-financial factors as part of overrides or expert-based notching.

Figure 7.1: Weighting range of financial factors in risk ranking model



Source: Deloitte 2025 IRB Market Analysis

PD Risk quantification/calibration

This section discusses PD Risk Quantification approaches. This is a prescribed part of the regulation that deals with assigning the correct PD estimates to accounts, and feeds directly into the RWA estimates. Navigating the requirements of this step can be challenging, particularly if firms lack sufficient default information to calibrate/benchmark on, which can be a common problem in the Wholesale lending space.

Regulatory insights

- CRR Article 180 specifies a minimum five-year data history. A longer history should be used if available, and calibration should also reflect a representative mix of good and bad years.
- CRR Article 180 also sets the idealised objective for PD calibration, that the PD per grade should be set to the long run average of one year default rates for that grade.
- The CRR previously permitted firms to use direct estimates of risk parameters and assign to grades on a continuous rating scale (e.g. master scale). Under Basel 3.1, the ability to map continuous PD estimates has been withdrawn. However, SS4/24 paragraph 11.23 leaves open the possibility of mapping to grades, i.e. using a master scale, during calibration.
- SS4/24 requires that firms test the homogeneity of grades or pools' default rates.
- SS4/24 requires that firms consider overrides, in the calibration step.

Discussion

Under Basel 3.1, the elimination of direct PD estimates with a masterscale mapping renews the imperative for firms to achieve the idealised objective of aligning grade-level PDs to observed average default rates. However, this is typically not directly achievable in non-retail portfolios, given the paucity of default history. As an example, a minimum threshold of 20 defaults per grade per year (typically required to achieve robust estimates) would require 400 defaults per year in a 20-grade rating scale. In practice firms adopt approximations, such as aligning the calibration sample to a portfolio-level 'central tendency' value.

The calibration from risk ranking score to PD can be regarded as a transformation of the probability mass function. In Retail, the mapping function is typically discretised in order to maximise accuracy, whereas in non-retail the mapping function is typically a polynomial approximation. Under a paucity of data, the polynomial order is typically zero, i.e. an intercept adjustment only. The principal risk with this approach, is that grade-level estimates are unconstrained and may result in PDs that are biased with respect to observed average default rates as well as out-of-line with external benchmarks.

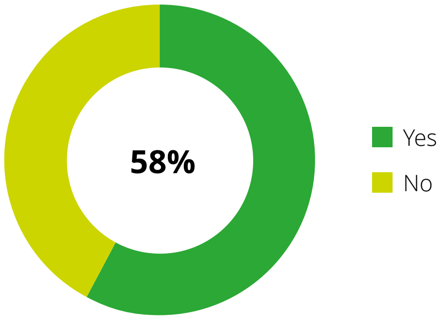
A potential downside of approaches that map to a master scale, is the risk of grade assignments failing homogeneity and heterogeneity tests, including the Games Howell Post Hoc test favoured by the ECB. Such tests are informative within discrete (retail-like) rating approaches, for determining the appropriate trade-off between number of grades and within-grade volatility. But the lookup to masterscale is in-effect a quantisation problem governed by Nyquist-Shannon theory, with the number of grades (and spacing in terms of PD) governed only by the smallest difference that needs to remain discernible after quantisation and should be useful to the business even for a population of one.

The requirement to be calibrated based on post-override grade assignments is not new under Basel 3.1. In the UK market, the requirement was typically satisfied by a general assumption that overrides would be prudent overall, combined with back-testing of final grade-level default rates. However, the consensus has shifted to the point that there is now acceptance that to not reflect override information in the calibration may also be distorting.

Benchmarks

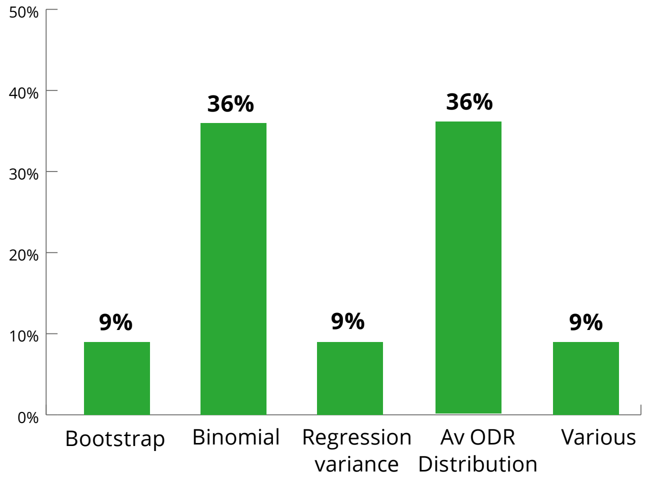
- The current consensus is to perform central tendency alignment using an intercept adjustment.
- The risks associated with an under-constrained transformation function biasing estimates (e.g. using intercept shift only, in situations where a slope adjustment would also be appropriate) are not typically encountered in practice: Use of optimisation techniques, e.g. to fine-tune the calibration in cohorts with a paucity of data, is rare.
- The elimination of direct PD estimates under Basel 3.1 has not resulted in the adoption of Retail (discretised) calibration techniques. Firms with master scales are continuing to use continuous risk ranking estimates, and map to master scale grades. Testing of grades' homogeneity or heterogeneity is less common at firms with a master scale.
- Firms generally apply rules to identify those obligors whose rating would continue to be overridden under the new model, though a wide range of practices exist here in terms of which types of overrides are admitted, and whether to view as absolute or relative with respect to model output. Re-grading exercise (be they large-scale or sample-based) are not generally performed, and most firms surveyed generally do not explicitly apply an adjustment to the calibration sample to incorporate overrides from the new model.
- MoC reflects the statistical uncertainty that arises from PD estimates, as a result of either model/data deficiencies (Category A), changes in lending standards/risk appetite (Category B) and General Estimation Error (Category C). whilst the latter approach is fairly well established, firms do vary in their approaches for quantifying the general estimation uncertainty.

Figure 8.1: Percentage of backs surveyed which do not apply any treatments to calibration sample for overrides



Source: Deloitte 2025 IRB Market Analysis

Figure 8.2: Approach for quantification of MoC C estimation error



Key insights

One bank targets the Central Tendency whilst also preserving discriminatory power at grade level.

A calibration is performed using a polynomial function to continuous scorecard predictions, before application of an inverse logic transform. The polynomial function coefficients were estimated by optimising the Cumulative Accuracy Profile (CAP) curve, under additional constraint that the average LDA PD predictions match the LRA DR (Central Tendency target). Optimisation based on discriminatory power was used, due to the requirement for the model scores to be grouped into pre-defined score bins for the Masterscale, which can impact the overall discriminatory power of the model (depending on the calibration shift).

In subsequent model testing, it was found that heterogeneity was not satisfied for all rating grades. The firm had shown assessments that merging grade results to suffice model testing, would have resulted in a reduction in RWA.

No overrides were applied for the initial model calibration due to expected changes in the override framework in the future. Once finalised, the expected impact was simulated on the calibration sample. Based on the simulation, a temporary expert add-on was quantified which will be in-place until the first recalibration based on production data.

Bringing it all together

Our survey revealed a complex interplay of regulatory demands, data challenges, and modelling intricacies. While the industry has made strides towards harmonising approaches and remediating their non-retail IRB PD models, a closer look reveals areas of inconsistency, persistent hurdles, and emerging role-model practices that are likely to shape the future of these model submissions and approvals. This concluding section distils our key insights, providing a roadmap for banks striving to achieve non-retail IRB excellence.

IRB is more than just modelling
A narrow focus on model mechanics is a recipe for failure. Successful IRB programmes demand a holistic strategy that aligns risk management, business objectives, and regulatory compliance.

Firms must move beyond a siloed approach to IRB. Strong governance, executive buy-in, and cross-functional collaboration are non-negotiable.

Segmentation is a balancing act
Finding the right segmentation granularity is crucial. Firms are grappling with the tension between regulatory boundaries and the need for homogeneous risk segments.

Data: the foundation and the stumbling block
Data quality and representativeness remain paramount yet elusive. Reliance on the 2008 crisis data, while common, presents limitations, and external data sources face scrutiny.

Robust data management practices, including rigorous quality checks, adjustments for historical inconsistencies, and a discerning approach to external data, are essential for model integrity.

The missing data conundrum
Missing data poses a significant challenge, and simplistic imputation methods can introduce bias and overconfidence in model accuracy.

Banks must adopt sophisticated statistically robust strategies for handling missing data, ensuring transparency, and minimising bias in model estimates.

Navigating risk differentiation
Default prediction models reign supreme, but their effectiveness hinges on data availability. Shadow rating and expert lender models offer alternatives, each with its own set of considerations.

Model selection should be driven by data availability and a clear understanding of the strengths and limitations of each approach. Transparency in documenting model limitations is crucial.

Calibration: Aligning with reality
Calibration remains a balancing act between achieving regulatory compliance and reflecting the nuances of non-retail portfolios. The elimination of direct PD estimation under Basel 3.1 adds another layer of complexity.

Banks need to carefully consider the trade-offs between different calibration techniques, ensuring alignment with long-run default rates while acknowledging data limitations and the regulators drive for conservatism.

The path forward: Continuous improvement
The IRB journey is far from over. Continuous investment in data management, model develoment, and regulatory expertise is crucial for long-term success.

Firms must cultivate a culture of continuous improvement, embracing emerging technologies and best practices to navigate the evolving IRB landscape.

Acknowledgements

This report is the result of a team effort that includes contributions by experts of the member firms of Deloitte Touch Tohmatsu Limited not limited to the list below.

Authors

Alexander Marianski
Director
amarianski@deloitte.co.uk

Edward Venter
Director
edventer@deloitte.co.uk

Geyer Bisschoff
Director
geyerbisschoff@deloitte.co.uk

Rohan Gokhale
Senior Manager
rgokhale@deloitte.co.uk

Sachin Galaiya
Senior Manager
sgalaiya@deloitte.co.uk

Subject matter advisors

Ian Wilson
Partner
iwilson@deloitte.co.uk

Thomas Spellman
Partner
thspellman@deloitte.co.uk

Rod Hardcastle
Director
rhardcastle@deloitte.co.uk

Contacts

United Kingdom

Ian Wilson
Partner
iwilson@deloitte.co.uk

Edward Venter
Director
edventer@deloitte.co.uk

Alexander Marianski
Director
amarianski@deloitte.co.uk

Austria

Andreas Hanisch
ahanisch@deloitte.at

Belgium

Roeland Baeten
rbaeten@deloitte.com

France

Herve Phaure
hphaure@deloitte.fr

Germany

Dr. Gerrit Reher
greher@deloitte.de

Italy

Fabio Crepaldi
fcrepaldi@deloitte.it

Ireland

Aoife Bayne
abayne@deloitte.ie

Luxembourg

Jean Philippe Peters
jppeters@deloitte.lu

Netherlands

Marco Folpmers
mfolpmers@deloitte.nl

Sarah Mellet
smellet@deloitte.nl

Nordics

Thomas Clifford
thclifford@deloitte.dk

Brian Skopek Hansen
brianshansen@deloitte.dk

Portugal

Alexandra da Fonseca Ferreira
alexanferreira@deloitte.pt

Spain

Oriol Arnau
jarnau@deloitte.es

Switzerland

Tobias Menz
tmenz@deloitte.ch