

Mortgage IRB PD Model

Model Development Report

Prepared by:

Risk Management Department
First European Bank

December 28, 2024

Contents

1	Introduction	1
1.1	Purpose	1
1.2	Range of Application	1
1.3	Portfolio overview	1
1.4	Overview of the process and governance	1
1.5	Role of expert judgement	1
1.6	Segmentation and representativeness considerations	1
2	Reference Dataset	2
2.1	Data sources and data collection	2
2.2	Analysis of data quality and data cleansing	2
2.3	Reference Dataset development	2
3	Univariate and bivariate analysis	3
3.1	Segmentation and representativeness analysis	3
3.2	Differentiation dataset creation	3
3.3	Handling of missing values	3
3.4	Creation of the Long List	3
3.5	Selection of risk drivers	3
3.6	Overview of the Short List	3
4	Risk differentiation model development	4
4.1	Review of methodology	4
4.2	Developing model candidates	4
4.3	Review of model candidates	4
4.4	Overview of the final differentiation model	4
4.5	Model testing	4
5	Model calibration	5
5.1	Segmentation and representativeness analysis	5
5.2	Calibration dataset creation	5
5.3	Rating philosophy	5
5.4	Calculation of Long Run Average default rates	5
5.5	Pooling	5
5.6	Model testing	5
6	Deficiencies and Margin of Conservatism	6
6.1	Identification of deficiencies	6
6.2	Appropriate Adjustments	6
6.3	Margin of Conservatism	6
7	Implementation and use	7
7.1	Implementation	7
7.2	Integration into everyday processes	7
7.3	User training and documentation	7

1 Introduction

1.1 Purpose

Defines the main objectives of the PD model development, ensuring alignment with regulatory guidelines and internal risk management goals. Establishes how the results will be used for decision-making and compliance. Summarizes model scope (mortgage portfolio) and overarching business rationale.

1.2 Range of Application

Specifies the segments of the mortgage portfolio to which the model will apply. Clarifies any inclusions/exclusions based on product types or risk profiles. Ensures consistency with overarching rating system requirements.

1.3 Portfolio overview

Outlines the key attributes of the mortgage portfolio: size, geography, and collateral types. Highlights any known concentration risks or unique characteristics. Establishes the context for later segmentation and data representativeness checks.

1.4 Overview of the process and governance

Provides a high-level description of the entire PD model lifecycle, from data preparation to calibration. Sets out responsibilities for model approval, validation, and periodic reviews, in line with regulatory expectations. Emphasizes the need for robust governance to ensure model integrity.

1.5 Role of expert judgement

Explains how and when human insight can override or complement statistical methods. Describes policies for applying overrides, including documentation and validation of assumptions. Stresses the importance of balancing objectivity with practical domain knowledge.

1.6 Segmentation and representativeness considerations

Summarizes why segmentation is necessary, referencing the need for homogeneous risk pools. Describes the checks to ensure representative datasets across different segments. Points out potential challenges in covering changes in default definitions, market conditions, or lending standards.

2 Reference Dataset

2.1 Data sources and data collection

Describes all internal and external data sources used to build the PD model. Explains the process for gathering relevant risk drivers, default events, and economic indicators. Highlights any reliance on third-party data.

2.2 Analysis of data quality and data cleansing

Outlines eight dimensions of data quality and how they are evaluated. Covers how missing or erroneous records are identified, scored, and either corrected or excluded. Mentions the scoring framework to decide which data elements can be used in subsequent steps.

2.3 Reference Dataset development

Details the creation of the final “clean” dataset, incorporating segmentation, data validation, and alignment with the model’s scope. Specifies how historical default events are combined with key risk drivers. Ensures consistency and traceability back to original sources.

3 Univariate and bivariate analysis

3.1 Segmentation and representativeness analysis

Validates whether each segment remains homogeneous and truly represents real-world exposure. Uses statistical tests to confirm distribution similarities, default definition alignment, and stable lending standards.

3.2 Differentiation dataset creation

Explains the subset of data specifically used to rank or differentiate risk (i.e., “development sample”). Ensures the sample adequately reflects defaulted and non-defaulted observations for robust statistical analysis.

3.3 Handling of missing values

Describes methods (e.g., MCAR tests, imputation, or separate “missing” categories) to address data gaps without biasing the model. Notes any business rules where missing information is itself informative.

3.4 Creation of the Long List

Lists all initial candidate variables from univariate exploration. Notes preliminary results: each variable’s correlation with default, stability, and coverage. Sets the stage for further reduction.

3.5 Selection of risk drivers

Discusses how variables transition from the Long List to the Short List based on quantitative (performance, stability) and qualitative (expert judgement) criteria. Emphasizes consistent application of predefined tests.

3.6 Overview of the Short List

Introduces the final set of key drivers likely to have predictive power. Summarizes each driver’s rationale for inclusion, highlighting potential synergy among variables.

4 Risk differentiation model development

4.1 Review of methodology

Provides an overview of the chosen modeling approach (e.g., logistic regression, staged blocks). Justifies why this methodology is appropriate, referencing regulatory expectations and internal use cases.

4.2 Developing model candidates

Describes the iterative process of model building using different subsets or transformations of the Short List. Uses both statistical criteria and business reasoning to refine model structure.

4.3 Review of model candidates

Highlights the performance, stability, and interpretability checks for each candidate. Ensures no overfitting or underfitting, with an eye on future calibration.

4.4 Overview of the final differentiation model

Outlines the final selected model architecture and variables, including scoring/weighting details. References relevant metrics (e.g., ROC/AUC) for model validation.

4.5 Model testing

Explains the preliminary back-testing done to confirm the model's discriminatory power and stability. Demonstrates alignment with the chosen rating philosophy and business expectations.

5 Model calibration

5.1 Segmentation and representativeness analysis

Ensures that each calibration pool aligns with previous segmentation assumptions. Confirms no major shifts in distribution that would bias calibration.

5.2 Calibration dataset creation

Describes the final dataset used for calibrating PD estimates, including how overrides or special cases are handled. Mentions any pooling or smoothing techniques.

5.3 Rating philosophy

Establishes the chosen approach (Point-in-Time, Through-the-Cycle, or Hybrid) for calibrating PDs. Explains how this aligns with internal uses and regulatory requirements.

5.4 Calculation of Long Run Average default rates

Details how historical default rates are aggregated and averaged over a representative period. Explains how “bad years” and “good years” are accounted for to ensure accurate reflectiveness.

5.5 Pooling

Describes grouping of exposures with similar risk profiles to ensure stable estimates. Mentions how the final PD is derived for each pool, referencing key drivers.

5.6 Model testing

Reassesses the model’s predictive accuracy post-calibration. Cross-checks default rate alignment with the long-run average.

6 Deficiencies and Margin of Conservatism

6.1 Identification of deficiencies

Explains how the model's data or methodology might fall short (e.g., biases, missing triggers). Includes referencing possible categories (A, B, C) of deficiencies.

6.2 Appropriate Adjustments

Covers how identified weaknesses are corrected (e.g., adjusting data, refining model assumptions). Ensures these adjustments are consistent with best practices and regulatory guidance.

6.3 Margin of Conservatism

Describes the application of the framework for quantifying and adding conservatism to PD estimates (e.g., categories A, B, C). Emphasizes that MoC reflects the uncertainty due to data or methodological limitations.

7 Implementation and use

7.1 Implementation

Describes how the developed model is implemented in the IT systems.

7.2 Integration into everyday processes

Discusses how the new PD model is embedded into credit risk management, underwriting, and capital calculation. Mentions IT systems updates, reporting, and automation of rating assignments.

7.3 User training and documentation

Highlights the need for thorough user training on the final PD model and processes. Ensures that all procedures, limitations, and override options are clearly documented.

References

- [1] EBA. *Guidelines on PD estimation, LGD estimation and the treatment of defaulted exposures*. European Banking Authority, 2018.
- [2] Internal document. *PD Modelling Guidelines*. Credit Risk Modelling, 2024.
- [3] Internal document. *PD Modelling Procedures*. Credit Risk Modelling, 2024.