IRB LGD Modeling: Technical Documentation and Model Report

Melina Hafelt

July 9, 2025

Contents

1 Abstract

This document presents a comprehensive modeling framework for **Loss Given Default (LGD)** estimation within the IRB (Internal Ratings-Based) approach. The work is conducted on a synthetic dataset with variables structured to resemble realistic credit portfolio data.

Key features:

- Exploratory data analysis and LGD target construction
- Feature engineering, winsorization, and scaling
- Multiple model development: OLS, Random Forest, XGBoost
- Evaluation using R^2, RMSE, MAE, and residual plots
- Calibration analysis and downturn stress testing
- Regulatory considerations and validation recommendations

Disclaimer: All data and variables are simulated and do not represent any real institution or customer.

2 Theoretical Background

2.1 Definition of LGD

Loss Given Default (LGD) represents the portion of a credit exposure a lender expects to lose if a borrower defaults. Formally:

$$\text{LGD} = 1 - \frac{\text{Recovered Amount}}{\text{Exposure at Default (EAD)}} \tag{1}$$

LGD is expressed as a percentage or decimal between 0 and 1. A fully recovered exposure yields LGD = 0, while complete loss implies LGD = 1.

2.2 IRB Context and Downturn LGD

According to Basel II/III:

- LGD estimates must be **conservative** and reflect conditions during an economic downturn.
- Estimates should be based on historical recoveries and take into account *seniority*, *collateral*, and *macro conditions*.

2.3 Modeling LGD: Regression Perspective

In contrast to PD (a classification task), LGD modeling is formulated as a **regression problem**. The objective is to predict the continuous variable LGD using borrower and loan characteristics.

Assumptions:

- LGD is bounded: $0 \leq \text{LGD} \leq 1$
- Only defaulted observations (default_flag $= 1$) are used

3 Data Preparation

3.1 LGD Target Construction

$$\text{LGD}_i = 1 - \frac{\text{RecoveredAmount}_i}{\text{ExposureAtDefault}_i} \tag{2}$$

Observations with ExposureAtDefault $= 0$ or LGD < 0 are removed. The target is validated to ensure it lies within $[0, 1]$.

3.2 Outlier Treatment

To prevent distortion in model fitting, we apply **winsorization** at 1st and 99th percentiles:

$$X_i^{\text{winsorized}} = \begin{cases} P_1(X) & \text{if } X_i < P_1(X) \\ X_i & \text{if } P_1(X) \leq X_i \leq P_{99}(X) \\ P_{99}(X) & \text{if } X_i > P_{99}(X) \end{cases}$$

3.3 Feature Scaling and Transformation

Numerical features are standardized via:

$$z_i = \frac{x_i - \mu}{\sigma} \tag{3}$$

Where μ and σ are the mean and standard deviation from the training set.

4 Modeling and Evaluation

4.1 Models Compared

Three models were trained using the same features:

- Linear Regression (OLS)
- Random Forest Regressor
- XGBoost Regressor

4.2 Evaluation Metrics

- R^2: Proportion of variance explained
- **RMSE**: Root Mean Squared Error
- **MAE**: Mean Absolute Error

4.3 Results

Model	R^2	**MAE**	**RMSE**
Linear Regression	0.388	0.617	0.835
Random Forest	0.378	0.594	0.842
XGBoost	**0.437**	**0.575**	**0.801**

Table 1: LGD model performance on test data

4.4 Residual Analysis

Residuals ($e_i = y_i - \hat{y}_i$) were plotted to assess bias and heteroskedasticity. The XGBoost model exhibited:

- Residuals centered around zero
- No clear heteroskedastic pattern
- Stable spread across predicted range

4.5 Calibration Assessment

We group predictions into deciles and compare:

$$\text{Calibration Error}_j = \overline{y}_j - \overline{\hat{y}}_j, \quad j = 1, \ldots, 10$$

A diagonal pattern in the calibration plot confirms that predicted LGD aligns with realized losses in each bin.

5 Stress Testing: Downturn LGD

A downturn scenario is simulated by increasing predicted LGD by 20%:

$$\text{LGD}_{\text{stressed}} = \min(1.0, 1.2 \cdot \text{LGD}_{\text{predicted}})$$

Result:

- Baseline mean LGD: 0.51
- Stressed mean LGD: 0.61
- Relative uplift: +19.6%

This simulates deterioration in recovery rates and aligns with Basel expectations on conservatism.

6 Regulatory Considerations

6.1 IRB Guidelines

According to CRR Article 181(1)(b):

> LGD estimates must be appropriate for an economic downturn and reflect expected recoveries under adverse conditions.

6.2 Governance and Validation

- **Monitoring**: Track R^2, RMSE, feature drift quarterly
- **Recalibration**: Annually or after portfolio shifts
- **Audit Trail**: All modeling steps documented and reproducible

7 Conclusions

- XGBoost model outperforms others with $R^2 \approx 0.44$
- Model shows good calibration and stability across test segments
- Stress testing confirms regulatory conservatism
- Documentation and preprocessing follow IRB-compliant standards

Recommendation: Use XGBoost for production, with linear regression retained for transparency in regulatory disclosures.