Common Inconsistencies in Probability of Default Modeling

Andrija Djurovic

www.linkedin.com/in/andrija-djurovic

PD Modeling

- Most statistical approaches used in Probability of Default (PD) model development, validation, and application rely on specific assumptions.
- In addition to the inherent statistical assumptions of particular procedures, practitioners sometimes introduce additional assumptions they believe are necessary for the overall modeling process.
- When implementing different statistical procedures, it is crucial to understand the appropriateness of these additional assumptions and their potential consequences.
- Practitioners are often inconsistent in applying and relying on various assumptions during model development, validation (both initial and periodic), and application.
- Considering potential inconsistencies, several questions arise:
  - Is it sufficient to argue that certain inconsistencies are widely accepted concepts or standard practices?
  - What are the consequences of these inconsistencies on the final procedure results?
  - How do these inconsistencies influence the discussions between model developers and validators (internal, external, or regulators)?
- The following slides highlight some typical inconsistencies observed in PD modeling.

Example 1: Assumptions on Uncorrelated and Correlated Defaults

Assumptions of Uncorrelated Defaults

When testing the predictive power of the PD model, a commonly used method is the z-score test, defined by the following formula:

$$z = \frac{ODR - PD}{\sqrt{\frac{PD(1-PD)}{n}}}$$

One of the main assumptions of this test is that defaults are uncorrelated. For further details, refer to this document.

Assumptions of Correlated defaults

When practitioners use the IRB PD model to calculate RWA, for instance, in the context of retail residential mortgage exposures, the following formulas are employed:

$$Correlation = R = 0.15$$

$$\text{Capital requirement} = K = LGD \cdot N\left[\frac{G(PD)}{\sqrt{1-R}} + \sqrt{\frac{R}{1-R}} \cdot G(0.999)\right] - PD \cdot \text{LGD}$$

$$RWA = K \cdot 12.5 \cdot EAD$$

In the calculation above for the upper bound of the PD, it is assumed that the PD follows the Vasicek distribution, where one of the parameters is interpreted as the asset correlation.

Example 2: Assumptions of the Deterministic Metric from the Development Sample

Standard Error Calculated from Development Sample

The following formula gives the most common form of the z-score test for assessing predictive power:

$$z = \frac{ODR - PD}{\sqrt{\frac{PD(1-PD)}{n}}}$$

This formula assumes that the standard error of the test statistic is calculated using the calibrated PD obtained from the model development phase.

Deterministic Metric from the Development Sample

When assessing changes in the concentration of a rating scale, a commonly used approach is to examine the change in the coefficient of variation (CV), defined by the following formula:

$$1 - \Phi\left(\frac{\sqrt{K-1}(CV_{curr} - CV_{init})}{\sqrt{CV_{curr}^2\,(0.5 + CV_{curr}^2)}}\right)$$

This formula assumes a normal approximation and that the CV is deterministic based on the model's development phase. For further details, refer to this document.

The same assumption is often applied when testing the discriminatory power of credit risk models.

Example 3: Assumptions of Independent Observations

Autocorrelation in PD FLI IFRS9 Modeling

Ordinary Least Squares (OLS) regression is one of the most commonly used methods for forward-looking Probability of Default (PD) modeling in IFRS9. The standard form of the OLS regression is given by:

$$\hat{y}_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_p x_{pt} + \epsilon_t$$

When using this method, practitioners address the problem of autocorrelation in the residuals (ϵ_t), acknowledging its potential consequences.

Independence in Quantifying the Uncertainty of the Central Tendency

Various methods can be used to manage the variability associated with calculating the Central Tendency (CT) of default rates. The most frequently used method relies on the standard error of the mean, defined as:

$$\frac{\sigma}{\sqrt{n}}$$

When applying this method, practitioners often assume by default that the observed default rates are independent. For further details, refer to this document.