Enhancement of Heterogeneity Testing for IRB Models

Analysis of the Disruption of Monotonicity in the Rating Scale

Andrija Djurovic

www.linkedin.com/in/andrija-djurovic

Model Heterogeneity

- A typical step in building credit risk models is discretizing the model output into ratings, pools, or buckets.
- Practitioners generally follow established principles for this discretization.
- These principles result in specific characteristics, some of which are mandatory, while others vary by model and are desirable but not essential.
- Monotonicity and heterogeneity are typically regarded as mandatory characteristics.
- In this context, heterogeneity refers to adequate differentiation in risk profiles across ratings, pools, or buckets. It is commonly tested in Probability of Default (PD), Loss Given Default (LGD), and Exposure at Default (EAD) models, often using tests like the two-proportion test and t-test.
- Heterogeneity is assessed during the initial model validation and monitored over time.

Practical Challenges in Heterogeneity Testing

- A common approach to heterogeneity testing relies on observed (realized) average values between adjacent grades, pools, or buckets.
- Heterogeneity is typically tested on the full modeling dataset during initial model validation and at specific reference dates or across a few consolidated reference dates during periodic model validation.
- Given these different levels of testing, practitioners often face challenges in assessing overall heterogeneity.
- Monotonicity is a minimum requirement in heterogeneity testing, as the test checks whether a “better” rating has a lower average risk parameter than a “worse” one using a one-sided test.
- Statistical significance depends on both monotonicity and effect size - a sufficiently significant difference between averages is needed to confirm heterogeneity.
- Test outcomes vary significantly depending on the number of observations per rating grade and whether the full dataset or specific reference date(s) are tested.
- Due to these challenges, practitioners may supplement statistical tests with observed monotonicity and use simulations to estimate the likelihood of monotonicity disruptions.
- The following slides present two methods for estimating this likelihood for the PD model. Applying these methods can help practitioners answer the following question: if the model is well-calibrated, what is the likelihood of having more than `m` rating pairs that fail the monotonicity condition? Practitioners may combine statistical test results with observed monotonicity, using the latter as a softer criterion for evaluating this aspect of model performance.
- The presented methods can be extended to other risk parameters, LGD and EAD, by adjusting the framework to accommodate continuous target variables.

Monotonicity Disruption - Hypothesis Testing Framework

Test Statistics

Under the assumption of a discretized rating scale and a well-calibrated model, the probability of monotonicity disruption between two adjacent grades can be calculated by modifying the standard hypothesis testing framework.
The following formula shows how to calculate the test statistic for this purpose:

$$Z = \frac{p_2 - p_1}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

where:

- p_1 is the calibrated PD of the "better" rating grade;
- p_2 is the calibrated PD of the "worse" rating grade;
- n_1 and n_2 denote the sizes of the "better" and "worse" ratings.

Probability of Monotonicity Disruption

The two-proportion test assumes that the test statistic follows a standard normal distribution. Consequently, the probability of monotonicity disruption is calculated as $1 - \Phi(Z)$, where Φ is the cumulative distribution function of the standard normal distribution.

Use in Periodic Model Validation

After obtaining the probability of monotonicity disruption for each adjacent grade pair, it is straightforward to derive the probability that a given number `m` of pairs will or will not fail the monotonicity condition. This can serve as a threshold for the periodic monitoring of model heterogeneity.

Monotonicity Disruption - Simulation Framework

Simulation Design

An alternative method for calculating the probability of monotonicity disruption is based on Monte Carlo simulations. The following steps outline the simulation process:

1. Define the true default rates - p_1 and p_2 (calibrated PDs) - and the sample sizes, n_1 and n_2.
2. For rating grade 1, simulate the default rate as the average of default indicator values of size n_1 drawn from a binomial distribution with the probability of success equal to p_1.
3. For rating grade 2, simulate the default rate as the average of default indicator values of size n_2 drawn from a binomial distribution with the probability of success equal to p_2.
4. Repeat steps 2 and 3, N times, and collect the simulated default rates.
5. Estimate the probability of monotonicity disruption as the proportion of simulations where the simulated default rate for rating grade 1 exceeds that of rating grade 2.

Use in Periodic Model Validation

After obtaining the probability of monotonicity disruption for each adjacent grade pair, it is straightforward to derive the probability that a given number `m` of pairs will or will not fail the monotonicity condition. This can serve as a threshold for the periodic monitoring of model heterogeneity.

Simulation Study

Dataset

The dataset used for the following simulation is shown below:

```
##   rating   no     pd
## 1    RG1  500  0.57%
## 2    RG2  640  1.05%
## 3    RG3  975  1.69%
## 4    RG4 1505  3.10%
## 5    RG5  845  5.30%
## 6    RG6  455  7.93%
## 7    RG7   91 14.51%
## 8    RG8   48 25.90%
```

Simulation Design

For the above rating scale, the first task is to calculate the probability of monotonicity disruption for each pair of adjacent ratings using both the hypothesis testing and simulation frameworks.
After obtaining these probabilities, the next task is to graphically present the likelihood that more than `m` out of a total of `n` adjacent pairs will not fail the heterogeneity condition. These probabilities can further serve to define a threshold, the exceedance of which may trigger additional analysis of model heterogeneity during periodic model validation.

Simulation Results

Probability Monotonicity Disruption - Hypothesis Testing Framework

```
##   rating   no     pd P(monotonicity disruption)
## 1    RG1  500  0.57%
## 2    RG2  640  1.05%                     18.03%
## 3    RG3  975  1.69%                     13.36%
## 4    RG4 1505  3.10%                      1.02%
## 5    RG5  845  5.30%                      0.68%
## 6    RG6  455  7.93%                      3.81%
## 7    RG7   91 14.51%                      4.59%
## 8    RG8   48 25.90%                      5.99%
```

Probability Monotonicity Disruption - Simulation Framework

```
##   rating   no     pd P(monotonicity disruption)
## 1    RG1  500  0.57%
## 2    RG2  640  1.05%                     17.80%
## 3    RG3  975  1.69%                     13.43%
## 4    RG4 1505  3.10%                      1.13%
## 5    RG5  845  5.30%                      0.57%
## 6    RG6  455  7.93%                      3.47%
## 7    RG7   91 14.51%                      3.73%
## 8    RG8   48 25.90%                      5.46%
```

Simulation Results cont.