On Testing the Concentration in the Rating Grades

The Initial and Periodic PD Model Validation

Andrija Djurovic

www.linkedin.com/in/andrija-djurovic

Rating Grade Concentration in PD Scale

- Practitioners usually adhere to certain principles when building the Probability of Default (PD) rating scale. These principles result in specific characteristics of the rating scale.
- One such characteristic, which is assessed during the initial model development and is also regularly monitored over time, is the concentration within the rating grades.
- Ideally, the concentration within each rating grade should remain below a certain threshold and should not increase significantly over time compared to when the model was developed.
- Besides the issue of regulatory non-compliance, some other consequences of the excessive concentration within the rating scale include:
  - loss of risk differentiation and mispricing of risk;
  - inaccurate capital requirements;
  - underestimation or overestimation of PD.
- The following slides present the approach currently mandated by the ECB for IRB validation reporting exercises, which most banks have adopted as the standard for assessing concentration in rating grades. They also explore the challenges associated with this approach and examine potential alternatives to address them.

The Standard Practice for Assessing the Concentration

In the Instructions for reporting the validation results of internal models, the ECB requires banks to report the Herfindahl Index (HI) value for the current application portfolio, along with the p-value of statistical tests used to assess changes in the Herfindahl Index from the model's development phase to its current application.

The formula for the Herfindahl Index is provided as follows:

$$HI = 1 + \frac{\log\left(\frac{CV^2+1}{K}\right)}{\log(K)}$$

with CV being a coefficient of variation:

$$CV = \sqrt{K \sum_{i=1}^{K} \left(R_i - \frac{1}{K}\right)^2}$$

where:

- K is the number of rating grades;
- R_i is the relative frequency of the i-th rating grade.

The Standard Practice for Assessing the Concentration cont.

The change in the HI is assessed indirectly by testing the change in the CV using a normal approximation and the assumption of a deterministic CV from the model's development phase.

The test's null hypothesis is that the HI at the time of the model application is lower than or equal to the HI at the time of development. Based on these assumptions, the p-value is calculated as follows:

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

where:

- Φ is the cumulative distribution function of the standard normal distribution;
- CV_{curr} is the coefficient of variation at the time of the model application;
- CV_{init} is the coefficient of variation at the time of model development (initial validation);
- K is the number of rating grades or bins of the risk factors.

The Challenges Associated with the Standard Practice

While the value of the HI is typically compared directly to specific thresholds - such as 0.20, 0.25, and 0.30 - the change in HI is assessed indirectly through variations in the CV. This raises two key questions: First, does a change in the relative dispersion measure reflect changes in the HI value? Second, how does the assumption of the deterministic CV_{init} influence the test results?

Consider the following data: relative frequencies at the time of the model's development are [0.13, 0.14, 0.23, 0.20, 0.11, 0.10, 0.09], and at the time of the model's application, they are [1, 0, 0, 0, 0, 0, 0]. This simulated data assumes that the HI shifts to a fully concentrated rating scale at the time of the model's application.

Let's examine the HI values during these two periods. The HI value at the time of the model's development is 0.057 while the HI value at the time of the model's application is 1.

Given the assumption of strong concentration, practitioners would expect the p-value for the test of change in the HI to be well below the commonly used significance levels of 0.01 or 0.05. However, the calculated p-value based on the proposed approach is 0.2043.

As observed, the calculated p-value is significantly higher than the commonly used significance levels, leading us to fail to reject the null hypothesis.

The Challenges Associated with the Standard Practice cont.

In addition to the previous example, testing the change in the CV relies on the assumption that the variable is normally distributed. This raises question about the validity of this assumption, particularly when analyzing the relative frequencies of grade in the rating scale at different times.

Although the effects of different underlying distributional assumptions and the relatively small sample size on the assumed test statistics are not explored here, they are important considerations. Investigating these factors is crucial for evaluating the overall validity of the asymptotic normality assumption of the CV, which is given as follows:

$$AN\left(\mu = CV_{sample},\ \sigma = \sqrt{\frac{CV_{sample}^2\left(0.5 + CV_{sample}^2\right)}{K-1}}\right)$$

The Proposed Alternatives

- Instead of indirectly testing changes in the HI based on the CV change, the proposed alternatives employ Monte Carlo simulations to directly assess and test changes in HI values between the two analyzed periods.
- To align with the assumption of deterministic metric value from the time of model development, the first method considers only the variability arising from the data available at the time of model application.
- The second alternative accounts for the variability in both observed relative frequencies while maintaining the portfolio size from the time of model application.
- Details on both alternatives are provided in the following slides.

The Proposed Alternatives - Method 1

Testing process:

1. Calculate the HI metric according to the formula provided on slide 3, using the data from model development. Denote this value as `HI_init`.
2. Based on the application portfolio, for the observed sample size `n` and relative frequencies `f_1` to `f_g` (where `g` is the number of rating grades), simulate `g` observations from a multinomial distribution.
3. Using the simulated observations, calculate the HI value according to the formula provided on slide 3.
4. Repeat steps 2 and 3 `N` times, storing the resulting HI values.
5. Calculate the percentage of simulated HI values greater than `HI_init`, and denote this percentage as `HI_g`.
6. Determine the p-value as `1 - HI_g`.
7. If the p-value is lower than the significance level, conclude that the HI at the time of model application has significantly increased compared to that at the time of model development.

The Proposed Alternatives - Method 2

Testing process:

1. Based on the application portfolio, simulate g observations from a multinomial distribution using the observed sample size `n` and relative frequencies `fa_1` to `fa_g` (where g is the number of rating grades).
2. Using the same sample size `n` from the application portfolio, simulate g observations from a multinomial distribution using the relative frequencies `fd_1` to `fd_g` from the development portfolio.
3. Calculate the HI values using the simulated observations from steps 1 and 2, and denote these results as `HI_curr` and `HI_init`.
4. Repeat steps 1 through 3 `N` times, storing the resulting HI values.
5. Calculate the difference between `HI_curr` and `HI_init` for each simulation.
6. Calculate the percentage of simulated differences greater than 0, and denote this percentage as `D_g`.
7. Determine the p-value as `1 - D_g`.
8. If the p-value is lower than the significance level, conclude that the HI at the time of model application has significantly increased compared to that at the time of model development.

Simulation Study - Method 1

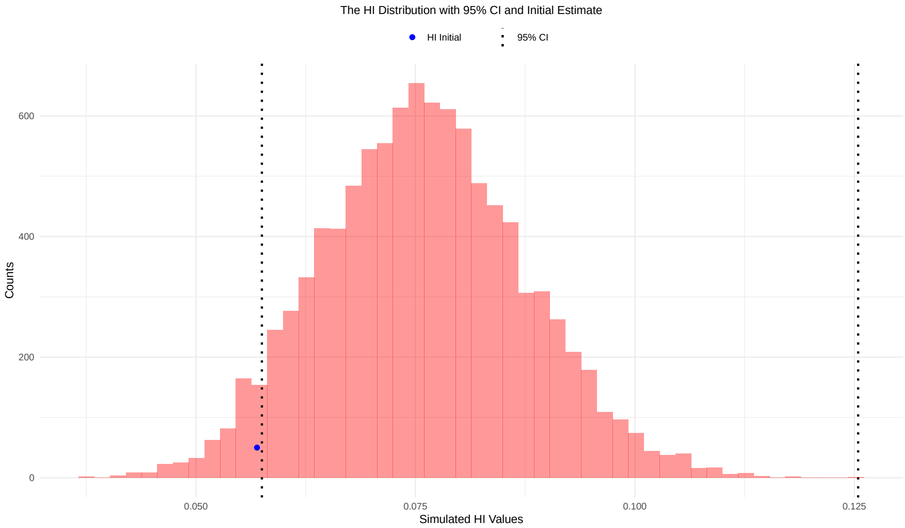
Simulation assumptions:

1. Relative frequencies at the time of model development: 0.13, 0.14, 0.23, 0.2, 0.11, 0.1, 0.09;
2. Relative frequencies at the time of model application: 0.13, 0.14, 0.24, 0.21, 0.11, 0.09, 0.08;
3. Sample size of the application portfolio: `n = 1,000`;
4. Number of simulations: `N = 10,000`.

Simulation results:

1. `HI_init`: 0.057;
2. `HI_curr`: 0.0734;
3. Test `p-value`: 0.047.

Simulation Study - Method 1 cont.



Simulation Study - Method 2

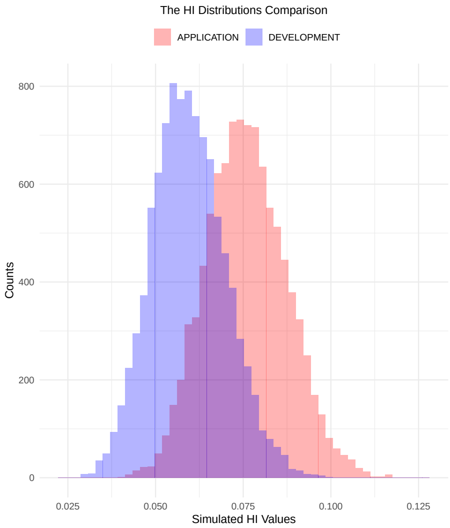
Simulation assumptions:

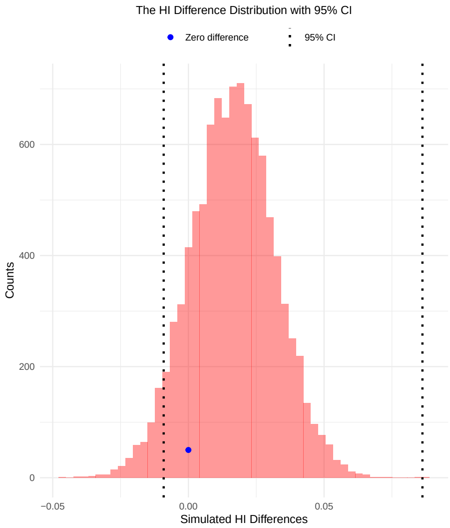
1. Relative frequencies at the time of model development: 0.13, 0.14, 0.23, 0.2, 0.11, 0.1, 0.09;
2. Relative frequencies at the time of model application: 0.13, 0.14, 0.24, 0.21, 0.11, 0.09, 0.08;
3. Sample size of the application portfolio: `n = 1,000`;
4. Number of simulations: `N = 10,000`.

Simulation results:

1. `HI_init`: 0.057;
2. `HI_curr`: 0.0734;
3. Test `p-value`: 0.1464.

Simulation Study - Method 2 cont.





Concluding Remarks

- Besides comparing the HI value to a specific threshold, practitioners often examine whether the HI significantly increased at the time of model application compared to the model development phase.
- The most commonly applied method for reporting on these procedures is the one required by the ECB in the instructions for reporting model validation results.
- In practice, the standard approach has certain challenges. These primarily refer to detecting significant differences in fully concentrated rating scales and the underlying assumptions of the statistical test.
- Alternative methods can be tested and applied under the same or similar assumptions, directly testing changes in HI across two analyzed periods.
- Both proposed methods can identify statistically significant differences even for small changes in HI when applied to larger portfolios. Therefore, it is recommended to combine the test results with the absolute value of the difference (practical and statistical significance) to reach a final conclusion.
- Given the potential weaknesses of alternative methods, it is advisable to properly balance the direct comparison of the HI value against a specific threshold with the observed change in HI over two periods when making a final decision on rating grade concentration.
- Given the above points, practitioners can also extend the use of alternative methods to LGD and EAD models.