Heterogeneity and Homogeneity Testing in IRB LGD/EAD Models

Is the Mann–Whitney U Test Compliant with Regulatory Requirements?

Andrija Djurovic

www.linkedin.com/in/andrija-djurovic

Regulatory Requirements

- Heterogeneity and homogeneity are key aspects in analyzing the distribution and allocation of obligors and facilities in IRB models.
- In this context, heterogeneity refers to adequate differentiation in risk profiles across ratings, pools, or buckets.
- Homogeneity, however, means that all exposures within a grade or pool exhibit similar risk characteristics.
- Paragraph 175, Credit risk, of the ECB Guide to Internal Models outlines regulatory expectations regarding meaningful risk differentiation and the quantification of LGD models.
- This paragraph explicitly addresses the requirements for heterogeneity and homogeneity in LGD models, in addition to expectations about the adequate distribution of facilities across pools.
- While it does not prescribe specific statistical methods, points (b) and (c) of Paragraph 175, Credit risk set precise requirements for demonstrating sufficient levels of heterogeneity and homogeneity between and within pools based on average realized values.
- Although not explicitly stated, these exact requirements are, in practice, also applied to EAD models.
- The explicit requirement to provide empirical evidence based on the average realized values raises the question of which statistical tests are appropriate and compliant with regulatory expectations.

Heterogeneity and Homogeneity Testing in Practise

- Heterogeneity and homogeneity testing are typically conducted based on realized values within and between pools.
- The most commonly used statistical tests are the Welch t-test and the Mann–Whitney U test.
- Practitioners often assess whether the values follow a normal distribution to determine which test is more appropriate for the distribution of LGD or CCF realized values. If normality is confirmed, the Welch t-test is used; otherwise, the Mann–Whitney U test is applied.
- Given the explicit regulatory requirement to base conclusions on average realized values when assessing heterogeneity and homogeneity, it is questionable whether the commonly used approach is appropriate. Specifically, the Mann–Whitney U test does not test for differences in means, which may lead to incorrect conclusions and non-compliance with regulatory expectations.
- The following slides will present the test statistics for both methods, demonstrate how they can yield conflicting results, and show how applying the Mann–Whitney U test under the wrong null hypothesis may lead to an inflated Type I error for typical distributions of realized LGD or CCF values.

Welch T-Test

The test statistic for the Welch t-test, which is used to compare the means of two samples with potentially unequal variances, is defined as:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where:

- $\bar{x}_1$ and $\bar{x}_2$ are the sample means;
- s_1^2 and s_2^2 are the sample variances;
- n_1 and n_2 are the sample sizes.

This test statistic follows a t-distribution with degrees of freedom given by the Welch–Satterthwaite equation:

$$\mathrm{df} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

To assess whether the means differ significantly, a p-value is computed from the t-distribution using the calculated test statistic and degrees of freedom.

Mann–Whitney U Test

The Mann-Whitney U test statistic is defined as:

$$U = \sum_{i=1}^{n_x} R(x_i) - \frac{n_x(n_x+1)}{2}$$

where $R(x_i)$ is the rank of the i-th observation from group x in the combined sample.

The mean and standard deviation of U under the null hypothesis are:

$$\mu_U = \frac{n_x n_y}{2}, \quad \sigma_U = \sqrt{\frac{n_x n_y(n_x+n_y+1)}{12} - \sum_t \frac{t^3-t}{(n_x+n_y)(n_x+n_y-1)}}$$

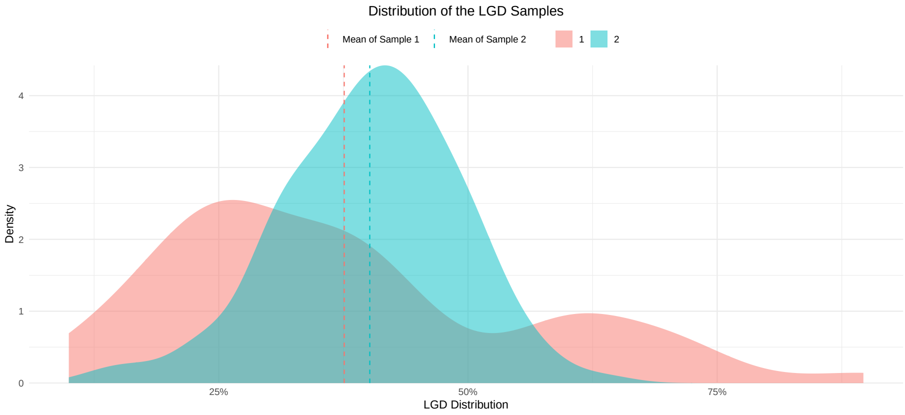
where n_x and n_y are the sample sizes and t represents the number of tied ranks in the data. The standardized test statistic is:

$$Z = \frac{U-\mu_U}{\sigma_U}$$

is approximately normally distributed under the null hypothesis, and the p-value is calculated accordingly. Unlike the Welch t-test, which tests for a difference in means, the Mann–Whitney U test assesses stochastic superiority.

Practical Challenges of Conflicting Results

It is relatively easy in practice to observe or simulate situations where the Welch t-test and Mann–Whitney U test lead to different conclusions. This is especially true for the distribution of realized LGD and EAD models, where the variances can differ across pools. One such example is available here. The observed p-values are 34.54% for the two-sided Welch t-test and 1.37% for the Mann–Whitney U test. The graph below shows the distribution of simulated values per sample.



Simulation Study

If practitioners perform the Mann–Whitney U test under the incorrect assumption that it tests for equality of means, it is easy to show that the Type I error (the percentage of times the null hypothesis is incorrectly rejected) is significantly inflated compared to the Welch t-test. The steps below outline one such simulation design and reveal the simulation results. Simulation assumes the realized values come from a beta distribution with specific parameters, featuring equal means (0.40) but different variances (0.20 and 0.1069).

Simulation Design

1. Simulate values for the first sample by randomly drawing 50 values from a beta distribution with shape parameters $\alpha = 2$ and $\beta = 3$.
2. Simulate values for the second sample by randomly drawing 150 values from a beta distribution with shape parameters $\alpha = 8$ and $\beta = 12$.
3. Calculate p-values for the two-sided Welch t-test and the Mann–Whitney U test.
4. Repeat steps 1 to 3 10,000 times and store the resulting p-values.
5. Calculate the Type I error as the percentage of simulations that reject the null hypothesis at a significance level of 10%.

Simulation Results

Running the above simulation results in Type I errors of 9.69% for the Welch t-test and 18.46% for the Mann–Whitney U test, respectively.