# Validation of Credit Risk Models

## On Favorable P-values in Statistical Tests

Andrija Djurovic

www.linkedin.com/in/andrija-djurovic

# Initial and Periodic Model Validation

- When validating credit risk models, practitioners typically formulate statistical hypotheses to evaluate various aspects of the model.
- The p-value resulting from statistical hypothesis testing is often the sole criterion for reaching a final conclusion.
- When the obtained p-value is unfavorable, practitioners may attempt to explain the test results from a business or practical perspective, seeking reasons why the unfavorable outcome might not be problematic.
- But what about favorable p-values in statistical tests? Should the conclusion be accepted without further investigation?
- The following slides present simulations of p-values from two predictive power tests commonly used to validate the Probability of Default (PD) models.
  The primary goal of these simulations is to show that, in practice, a favorable p-value can be obtained even when the observed default rate (ODR) consistently exceeds the calibrated PDs.
  Practitioners are encouraged to explore these examples further by testing different simulation designs and reconsidering the automatic acceptance of favorable test results based solely on p-values.

# Z-score Test and Simulation Design

### Z-score Test:

One of the most commonly used procedures for testing the predictive power of Probability of Default (PD) models is the z-score test, which is given in the following form:

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

where:

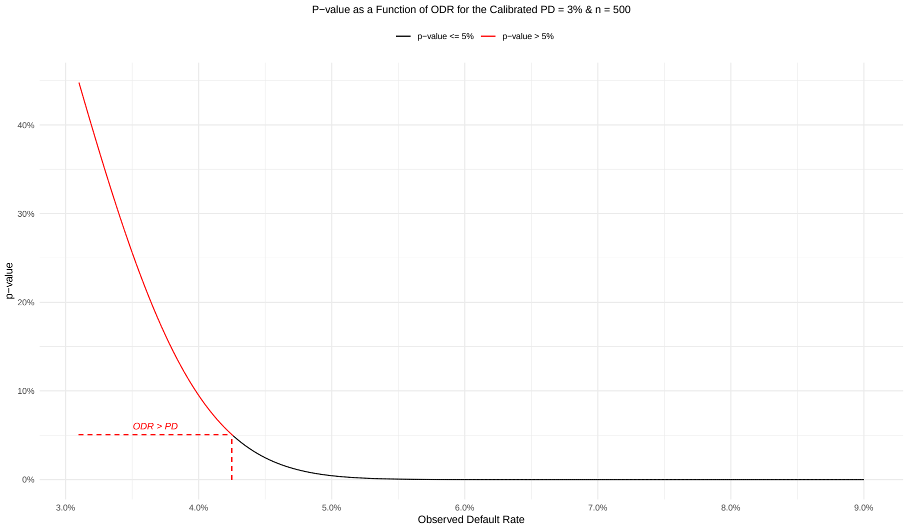
- $ODR$ is the observed default rate;
- $PD$ is the calibrated PD;
- $n$ is the sample size.

Under the assumption that the `z` test statistic follows the standard normal distribution, a `p-value` is calculated accordingly.

### Simulation Design:

The following slide presents the `p-value` as a function of the simulated ODR, with a calibrated PD of 3% and a sample size of 500.

# Z-score Test Simulation Results



# Multi-Period Normal Test and Simulation Design

## Multi-Period Normal Test:

Under the null hypothesis that none of the true probabilities of default in the years `T` exceed their corresponding forecasted `PDs`, the test statistic for the multi-period normal test is calculated as follows:

$$Z_{nt} = \frac{\sum_{i=1}^{T}(\mathrm{ODR}_i - \mathrm{PD}_i)}{\sqrt{T}se}$$

where:

- $PD$ is the calibrated PD;
- $T$ is the number of years;
- $se$ is the standard error defined as

$$\sqrt{\frac{1}{T-1}\left(\sum_{i=1}^{T}(ODR_i - PD_i)^2 - \frac{\left(\sum_{i=1}^{T}(ODR_i - PD_i)\right)^2}{T}\right)}.$$

Assuming that the $Z_{nt}$ test statistic follows the standard normal distribution, a `p-value` is calculated accordingly.

## Simulation Design:

For three years (`T = 3`) ODR ranging from 3.1% to 4.25%, the following slide presents the `p-value` of the normal test for a calibrated PD of 3%. An interactive plot can be found here.

# Multi-Period Normal Test Simulation Results