Validation of Credit Risk Models

Does the P-Value Provide Sufficient Insight for Model Validation?

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 used in reaching a final conclusion.
- Relying solely on the p-value raises several questions, such as:
  - Should practitioners adopt a unified approach based on the p-value for all portfolio types?
  - Should practitioners adopt a unified approach based on the p-value for all test types?
  - Is the p-value a sufficient criterion for making validation decisions?
- Practitioners rarely supplement the p-value from statistical tests with additional criteria, such as practical significance measures, during validation.
- The following slides present some of the most commonly used statistical procedures for testing predictive power as a function of different sample sizes.
  The main objective of the simulation is to demonstrate that, in practice, a p-value can show statistically significant results even when the change in the tested metric is not meaningful from a business perspective.
  Practitioners are encouraged to extend these examples to different simulation designs.

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:

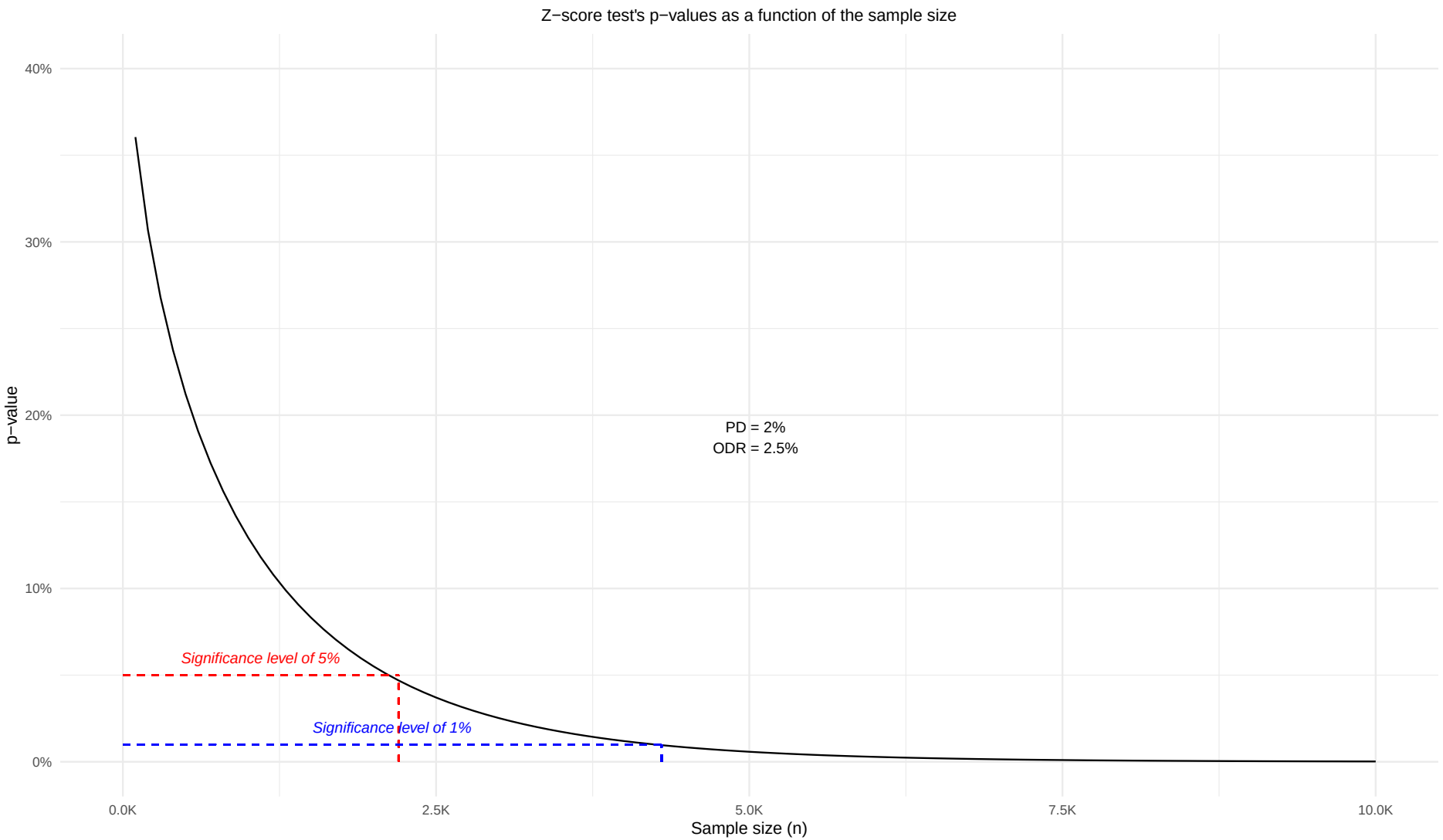
- ODR is the observed default rate;
- PD is the calibrated PD;
- n is the sample size.

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 sample size for a calibrated PD of 2% and an ODR of 2.5%.

Z-score Test Simulation Results



Paired T-test and Simulation Design

Paired T-test:

A typical procedure for testing the predictive power of the Loss Given Default (LGD) and Exposure at Default (EAD) models is the paired t-test. The test statistic is calculated as follows:

$$t = \frac{\bar{x}}{\frac{s}{\sqrt{n}}}$$

where:

- $\bar{x}$ is the mean of the difference between model estimations and observed LGD/CCF values;
- s is the sample standard deviation of the difference between model estimations and observed LGD/CCF values;
- n denotes the sample size.

Assuming that the `t` test statistic follows a t-distribution with `n-1` degrees of freedom, a `p-value` is calculated accordingly.

Simulation Design:

The following slide presents the p-value as a function of the sample size for an average LGD estimate of 36.86%, an average LGD pair difference of -1%, and a standard deviation of the LGD pair difference of 0.32.

Paired T-test Simulation Results