Heuristics for clustering risk ranking scores into risk grades: A comparative study

The Capital Requirements Regulation (CRR) permits firms to choose between grade-level and direct estimation of the probability of default (PD) for calculating risk weighted assets. Grade-level estimation is the approach of selecting homogeneous groups of obligors and setting the regulatory PD to the observed average of one-year default rates. A key building block when calibrating grade-level PDs is determining the number of discrete bins. The consequence of too few grades includes loss of risk sensitivity, including under-estimation in higher risk cohorts such as accounts in arrears. Additionally, too many grades could contribute towards volatile PDs that fail to back-test. Under both scenarios, the underlying cyclicality of the risk ranking model may become obfuscated.

To compound issues, binning results are impacted by noise. As well as volatility in observed default rates (typically associated with low volumes), there exists noise in risk ranking scores that is driven primarily by missing value treatments and quantisation of continuous variables. The appropriate number of bins is sensitive to noise in risk ranking scores, and the optimal number of discrete bins may vary across banks depending on data quality and risk-ranking inputs.

While there exist well-established clustering procedures for scorecard building, there is little consensus on clustering approaches or optimisation criteria for calibrated PDs whose output represents an “average of averages” metric. In this presentation, we will explore the performance of various plausible binning heuristics, including quantitative as well as qualitative considerations including execution time and use of human judgement to set or override bin boundaries. We show that in general, simple methods return a greater number of bins, exhibit greater stability in bin boundaries, than complex methods, and with the additional benefit of lower execution times.