HSBC Quants Academy
Project 2 – Corporate Credit Risk

Date: 2025
Prepared by: Philippe De Brouwer, Piotr Kobus, Karol Skrzyszowski
Last update: Philippe De Brouwer

Project 2 – Corporate Credit Risk

Project Description

The purpose of this project is to present basic tools used within credit risk for model building. The core task is linked to development and assessment of probability of default prediction model using tools like logistic/probit regression or linear regression. The project should allow students to analyze data, implement various methodologies, and assess them from both quantitative and qualitative perspective.

Data

The excel spreadsheet provided ("Data Credit IMR.xls") corresponds to a set of qualitative and quantitative information collected from the Credit Risk team in the period 2000-2008 for Middle-Market Wholesale customers.

The fields included in the dataset contain the following information:

- CUSTOMER_ID - internal customer identification number.
- ASSESSMENT_YEAR - year of the credit expert assessment.
- PRODUCT_DEMAND - credit expert's opinion on the competitive environment where the company operates, including its market position and the quality of its portfolio. The variable can show values from 10 to 90, 90 being the best score a company may obtain.
- OWNERS_MANAGEMENT - credit expert's opinion on the quality of the management of the company. The variable can show values from 10 to 90, 90 being the best score a company may obtain.
- ACCESS_CREDIT - credit expert's opinion on the ability of the company to obtain funds from different financial entities. The variable can show values from 10 to 90, 90 being the best score a company may obtain.
- PROFITABILITY - credit expert's opinion on the ability of the company to generate profits based on its current portfolio. The variable can show values from 10 to 90, 90 being the best score a company may obtain.
- SHORT_TERM_LIQUIDITY - credit expert's opinion on the ability of the company to generate cash-flows in the short-term to fulfil short-term financial obligations. The variable can show values from 10 to 90, 90 being the best score a company may obtain.
- MEDIUM_TERM_LIQUIDITY - credit expert's opinion on the ability of the company to generate cash-flows in the medium and long-term. The variable can show values from 10 to 90, 90 being the best score a company may obtain.
- GROUP_FLAG - categorical variable which indicates whether the customer belongs to a Financial Holding. It may have two values: "0" - The counterparty does not belong to a Financial Holding, "1" – The counterparty belongs to a Financial Holding (i.e. the counterparty is a subsidiary of a holding company).
- TURNOVER - it contains the value of the Turnover reported in the financial statements available for the assessment.

- INDUSTRY - it encodes the industry in which the company operates.
- DEFAULT_FLAG - it contains whether the customer has gone into default (i.e. has failed in its financial obligations with the Bank) in a 12-month period after the credit expert's assessment.Questions / Requirements

Questions / Requirements

Student is required to complete the following steps:

- Estimate a statistical model to predict the probability of default (PD) with the information available in the spreadsheet. Explain all the steps followed which should include a data quality assessment, univariate and multivariate analysis, and model validation. Use a logistic regression model for this purpose.
- Estimate the model using a *probit* specification. Which one would be better from a performance perspective? Why?
- The IT department has raised some concerns about the implementation of the model, as a logistic function presents some issues from a systems implementation standpoint. As such, it is now required that you estimate a model using a linear regression approach.
- Do you think this model should be implemented? Please provide your reasons accordingly.
- Several credit experts have proposed a model with the following specification:

$$PD = \frac{1}{1 + e^{-0.1 \times Score}}$$

- Where the Score is the weighted average of the following variables:

	WEIGHT
PRODUCT AND DEMAND	20%
QUALITY OF MANAGEMENT	10%
ACCESS TO CREDIT	10%
PROFITABILITY	15%
ABILITY TO PAY	25%
SOLVENCY	20%

- Demonstrate that the model developed using the statistical techniques that you have considered previously (or any other statistical techniques that you may want to consider) outperforms the model specification proposed by the credit experts. Please provide your justification accordingly by considering business rationale where possible in terms of why you believe your model proposal should be implemented.
- The Head of the Credit team has requested to conduct a segmentation analysis by Financial Holding Type (i.e. holding or subsidiary) and assess whether it is viable to build two bespoke models instead of an overall model for the portfolio. Would you be in favour of this proposal?

(amend as appropriate)