Nova School of Business and Economics

Estimation of Default Probabilities

Author:
Frederike Adams (44091)
Mona Carpio (44170)
Thomas Dornigg (41727)

Instructor:
Prof. João Pedro PEREIRA

An assignment submitted for the course
Credit Risk

Second Semester, T4
2021

April 13, 2021

Ex. 1 — Use the S&P rating transition matrix in Table 2.1 of the handouts. Plot the cumulative default probabilities until 10 years, PD(0, 1), PD(0, 2), ..., PD(0, 9), PD(0, 10) for the ratings classes A and BB. Indicate the precise value you got for 10 years.

Using the S&P rating transition matrix (Table 1), the group first transformed it into a square matrix. This was done by eliminating the column that refers to the future non-rated firms (NR) and adding a row with firms in default, which we assume remain in default in future periods. This transformation yields new probabilities of firms transitioning between the given credit levels at given periods (see Table 2) due to the elimination of the non-rated state. The probabilities of default (PDs) for one year for both classes A and BB, are directly observable in this matrix in column D, rows A and BB.

For all subsequent periods, we used the proposition that the n-period transition matrix is given by

$$P^{(n)} = P^n,$$

where P is the single-period square transition matrix.

From/to	AAA	AA	A	BBB	BB	B	CCC/C	D	NR
AAA	87.03	9.03	0.54	0.05	0.08	0.03	0.05	0.00	3.19
AA	0.54	86.53	8.14	0.54	0.06	0.07	0.02	0.02	4.07
A	0.03	1.83	87.55	5.38	0.35	0.14	0.02	0.07	4.64
BBB	0.01	0.11	3.58	85.44	3.75	0.56	0.13	0.20	6.23
BB	0.01	0.03	0.14	5.16	76.62	6.96	0.66	0.76	9.64
B	0.00	0.03	0.10	0.21	5.40	74.12	4.37	3.88	11.89
CCC/C	0.00	0.00	0.14	0.22	0.65	13.26	43.85	26.38	15.49

Table 1: Rating Transition matrix

From/to	AAA	AA	A	BBB	BB	B	CCC/C	D
AAA	0.8990	0.0933	0.0056	0.0005	0.0008	0.0003	0.0005	0.0000
AA	0.0056	0.9021	0.0849	0.0056	0.0006	0.0007	0.0002	0.0002
A	0.0003	0.0192	0.9180	0.0564	0.0037	0.0015	0.0002	0.0007
BBB	0.0001	0.0012	0.0382	0.9111	0.0400	0.0060	0.0014	0.0021
BB	0.0001	0.0003	0.0015	0.0571	0.8481	0.0770	0.0073	0.0084
B	0.0000	0.0003	0.0011	0.0024	0.0613	0.8412	0.0496	0.0440
CCC/C	0.0000	0.0000	0.0017	0.0026	0.0077	0.1569	0.5189	0.3122
D	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

Table 2: Single-period square transition matrix

For n = 1, 2, 3 . . . 10, one can find the associated cumulative probabilities of default for classes A and BB (see Table 3).

Year	A	BB
1	0.0007	0.0084
2	0.0017	0.0213
3	0.0029	0.0378
4	0.0044	0.0568
5	0.0062	0.0776
6	0.0084	0.0994
7	0.0109	0.1218
8	0.0137	0.1443
9	0.0169	0.1667
10	**0.0204**	**0.1886**

TABLE 3: Cumulative Probability of Default for classes A and BB

Plotting both cumulative probabilities for the firms in class BB and class A (see Figure 1), one can see that both increase over time, however, on a different level. While firms in class A hardly increase, the spike of class BB is much more severe.

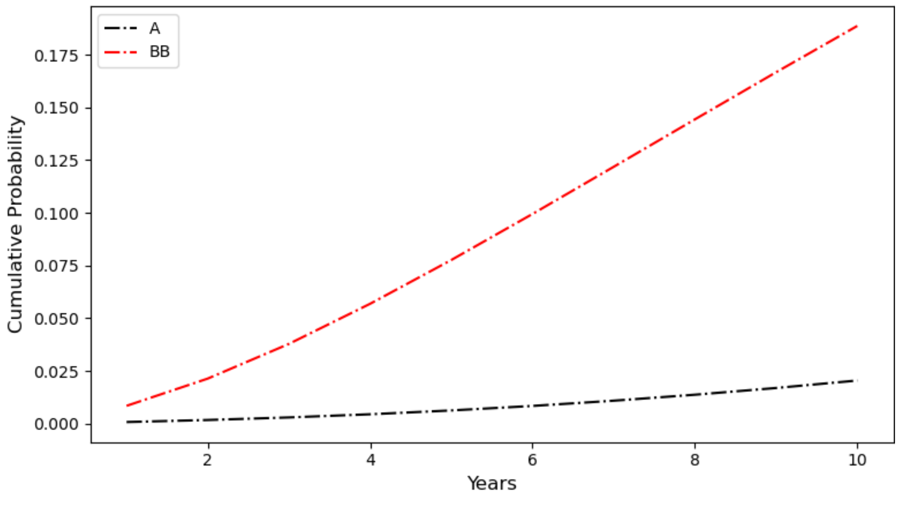

FIGURE 1: Cumulative Probability for A and BB

Ex. 2 — Get the Excel file Logit_data.xlsx posted next to this handout. It represents data on several clients from a given bank. For each firm, we have several financial ratios and an indicator of whether the firm defaulted (y = 1) or not (y = 0) during the year. Estimate a logit model for the probability of default. Do not forget to include a constant, i.e, using the notation in the handouts, set $f = b_0 + b_1 \cdot X_1 + b_2 \cdot X_2$.

In general, a Logistic regression models the probabilities for classification problems with two possible outcomes - in this case, default (1) and no-default (0). It can be regarded as an extension of the linear regression model for classification problems. Instead of fitting a straight line or hyperplane, the logistic regression model uses the logistic function to squeeze the output of a linear equation between 0 and 1 [1]. The logistic sigmoid function produces class membership probabilities for the target class (y=1) given features x:

$$P\left(y = 1|\mathbf{x}^{(i)}\right) = \frac{1}{1 + exp(-z)}$$

The step from linear regression to logistic regression is straightforward. In the linear regression model, one can model the relationship between outcome and features with a linear equation, for instance:

$$z = \beta_0 + \beta_1 x_1^{(i)} + \beta_2 x_2^{(i)} + \ldots + \beta_p x_p^{(i)}$$

For classification however, probabilities between 0 and 1 are preferred. Thus, to achieve that goal, one needs to wrap the right side of the equation into the logistic function [2]. This forces the output to assume only values between 0 and 1, thereby giving the following outcome:

$$P(y^{(i)} = 1) = \frac{1}{1 + exp(-(\beta_0 + \beta_1 x_1^{(i)} + \ldots + \beta_p x_p^{(i)}))}$$

For the estimation of loan defaults with the logit model, the group followed a twofold approach:

- Loan Default prediction with Logistic Regression in Python
- Using Maximum likelihood method to determine the best coefficients in Excel

By using the first approach, modelling defaults in Python via a Logistic Regression, the group made use of the statsmodels library. Fitting the respective function, following output was achieved:

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
Intercept	76.3611	42.820	1.783	0.075	-7.565	160.287
Assets_Liabilities	-0.7461	0.426	-1.753	0.080	-1.581	0.088
EBIT_Assets	-0.2576	0.149	-1.734	0.083	-0.549	0.034

Table 4: Logistic Regression results

By looking at the estimates of Table 4, one can see that all of the variables have significant effects ($p < 0.05$) on the default rate, so this model is both explanatory and predictive to a good degree. Thus, by using these results, we have a logit model that can calculate the probability of default for each one of the 23 firms, p_i.

In the second stage, the group compared the fitted logistic regression model with a logit probabilistic model implemented in Excel. The probability of default is going to be given by $p = \frac{e^f}{(1+e^f)}$. To estimate the value of the parameters of the model, we maximized the Log Likelihood function using the Solver optimization software in Excel. The results for the parameters were the following:

- β_0 = 76.36
- β_1 = -0.75
- β_2 = -0.26

In Table 5, the group summarized the results from both modelling approaches. As expected, the results hardly differ. Thus, for the further modelling in Ex. 3, the group used the estimates which were computed by the Python model.

Model	β_0	β_1	β_2
Python	76.3611	-0.7461	-0.2576
Excel	76.3606	-0.7461	-0.2576

TABLE 5: Model comparison

Ex. 3 — Continuing Ex. 2, suppose a new company comes up to the bank to borrow money. The company has the following ratios:

- Current Assets / Current Liabilities = 105
- EBIT / Assets = 8

1. Using the results in Ex. 2, what is the estimated Probability of Default (PD) for this company?

   To calculate the probability of default for the new company (x), the group used a direct application of the logit model as described in the previous exercise. By plugging in the values for $X_1 = 105$ and $X_2 = 8$, one gets a probability of default of 1.72%.

2. The bank determines the spread it charges on loans through the formula:
   $Spread = ROE \cdot k + PD \cdot LGD$. Assume that ROE = 15%, k = 8%, and LGD = 60%. What is the spread that the bank should charge to this firm?

   The bank "prices" its loans subject to the risk-adjusted return on capital being greater or equal to the return on equity. Hence, for the value of ROE of 15%, capital requirements of 8% and Loss Given Default of 60%, the minimum spread the bank is willing to charge this firm is 2.23%, which satisfies the condition $spread = ROE \cdot k + PD \cdot LGD$. Ideally, the bank should price the loan at a spread higher than 2.23% in order to create value.

Auxiliary

All materials (Python code and pdf) for this assignment are available on the author's Github account.

References

[1] Christoph Molnar. "Interpretable Machine Learning - A Guide for Making Black Box Models Explainable". In: *Leanpub* (Apr. 2021).

[2] Sebastian Raschka. "Interpretable Machine Learning - Book Review and Thoughts about Linear and Logistic Regression as Interpretable Models". In: *Blog* (Aug. 2020).