ECONOMICS OF CREDIT SCORING MANAGEMENT

Błażej Kochański

Abstract

Credit scoring models constitute an inevitable element of modern risk and profitability management in retail financial lending institutions. Quality, or separation power of a credit scoring model is usually assessed with the Gini coefficient. Generally, the higher Gini coefficient the better, as in this way a bank can increase number of good customers and/or reject more bad applicants. In the paper a simple simulation framework for analysis of microeconomics of credit scoring management is presented. The model takes into account competition among banks (there are 10 competing banks in the model), risk-based pricing (the banks differentiate prices based on their credit scoring models), "loan-shopping" practices by credit applicants (each applicant checks the price offered by three randomly selected banks). Such a setup enables us to perform a simulation where one of the banks improves the credit scoring model used and benefits from it. As the simulation shows, even small changes in Gini coefficient may lead to substantial improvement of bank's standing measured by its profitability and market share.

Key words: credit scoring, lending profitability, economics of banking

JEL Code: G21, C15, C38

Introduction

Credit scoring models constitute an inevitable element of modern risk and profitability management in retail financial lending institutions (Siddiqi, 2009, Thomas, 2017). Quality, or separation power, of a credit scoring model is usually assessed with the area under receiver operating characteristic curve, AUROC (Fawcett, 2006, Hanley & McNeil, 1982) or its function, the Gini coefficient (where $Gini = 2 \cdot AUROC - 1$). Generally, the higher Gini coefficient the better, as in this way a bank can increase number of good customers and/or reject more bad applicants (Řezáč & Řezáč, 2011, Finlay, 2014). On the other hand, in order to get a better credit scoring model, one has to allocate analysts' time and/or invest in new tools and data sources. All of this increases costs.

For example, a credit analyst tells the bank management that she can increase the Gini of the credit scoring model by 5 percentage points. This increase is obviously not free of charge: in order to achieve better separation power, access to the new databases/software/modelling tools is needed. Also, she needs to spend considerable amount of her, and her subordinates', time (which could be used for example to prepare impressive PowerPoint reports for the supervisory board instead).

So, how big an increase of bank's profits can be brought by the 5 percentage points increase in Gini coefficient? In the article below a simple framework is presented which could help assess economic benefits of a better scoring model.

1 The framework

Intuitions tell us that the model where a bank using a credit scorecard is independent of its competitive environment is usually overly simplistic. Before taking a loan, the customer usually checks several possibilities and chooses the one which seems to be the best for his or her situation. We should therefore assume that there is not only one bank in the economy. In the model we assume competitive situation where there are 10 banks. The approach to modelling the lending market with 10 banks (and 10 scoring models) can be based on the approaches used in similar contexts by Blochlinger & Leippold (2006) or Einav, Jenkins & Levin (2013), modified accordingly.

Let us assume that each of 10 banks has a credit scoring tool of similar separation power. The economy can be then modelled with a multivariate 11-dimensional normal distribution with means vector:

$$\boldsymbol{\mu} = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)^{\mathrm{T}} \qquad (1)$$

and covariance matrix:

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & \rho & \rho & & \rho_1 \\ \rho & 1 & \rho & \cdots & \rho_2 \\ \rho & \rho & 1 & & \rho_3 \\ & \vdots & & \ddots & \vdots \\ \rho_1 & \rho_2 & \rho_3 & \cdots & 1 \end{bmatrix}. \qquad (2)$$

So:

$$(S_1, S_2, S_3, \dots, S_{10}, Y^*)^T \sim N\,(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \qquad (3)$$

The first 10 out of 11 variables (S_1-S_{10}) represent credit scores allocated to customers by each of the ten banks in the economy. The last variable, Y^* is the credit risk factor, which itself is not observable, but translates itself into observable 0/1 default event Y:

$$Y = \begin{cases} 0 & if \quad Y^* < \Phi^{-1}(d) \\ 1 & if \quad Y^* \geq \Phi^{-1}(d) \end{cases}, \tag{4}$$

where d is a given default rate in the economy and Φ^{-1} is the inverse standard normal cumulative distribution function. In the simulation we will assume default rate d to be at the level of 10%.

Correlations ρ between S_i and S_j (where $i,j = 1, 2, \ldots, 10$ and $i \neq j$) are all set to $\rho = 0.75$, as credit models are based on similar variables (like, in the case of consumer finance, credit bureau delinquencies, credit history, income and marital status and so on, or in the case of corporate loans financial statement based variables and credit history), but are not the same.

Correlations ρ_1, ρ_2, ρ_3, etc. represent degree of separation power of scoring models S_1, S_2, S_3, etc. In the base scenario: $\rho_1 = \rho_2 = \rho_3 = \cdots = \rho_{10} = 0.5$, which (with the default rate of d=10%) translates to a Gini coefficient of about 0.541.

2 First simulations and pricing assumptions

Now let us simulate[1] 1 million loan applications coming from customers. We get 1 million vectors, each of them contains 10 scores from 10 banks and a value of unobservable credit risk, which can be translated into observed 0/1 default indicator.

First several vectors in such a simulation may for example look like those in Table 1. Please note, that Y column is added, which is 1 if Y^*<-1.282, zero otherwise, which corresponds to 10% default rate.

Tab. 1: Example of simulation results for 5 customers

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	Y^*	Y
1.	-1.899	-1.418	-1.138	-2.277	-0.608	-1.280	-1.788	-2.183	-1.416	-1.259	-0.948	0
2.	-0.424	0.471	-0.927	-0.290	0.629	-0.927	-0.066	-1.518	0.453	0.461	-0.866	0
3.	-0.410	-0.883	-0.252	-0.360	-1.472	-0.558	-0.493	-0.657	-0.815	-0.779	-0.401	0
4.	-0.890	-1.675	-0.802	-0.677	-1.227	-0.198	-0.652	-0.420	-0.529	-0.395	-1.882	1
5.	1.938	1.104	1.586	1.123	1.371	1.190	0.801	1.835	0.854	1.978	2.081	0

Source: simulation performed by the author.

If we assume that customers randomly choose one of 10 banks with equal probability, each bank will get approximately 100 thousand applications. So, each of the banks should

[1] The simulations were performed in R, and the codes are available from the author on request.

observe similar Gini coefficient and similar default rate. As a test run shows, this is really the case (Table 2).

Tab. 2: Simulation results summary for 1 million customers.

Bank	Number of loans	Default rate	Gini coeff. on all loans	Gini coefficient on granted loans
1	99117	0.100	0.541	0.544
2	98497	0.097	0.541	0.548
3	97731	0.098	0.542	0.544
4	100794	0.099	0.540	0.536
5	99065	0.098	0.540	0.549
6	101279	0.101	0.540	0.536
7	102145	0.111	0.541	0.549
8	100322	0.098	0.541	0.533
9	101529	0.099	0.542	0.539
10	99521	0.099	0.540	0.537

Source: simulation performed by the author.

Next, we can assume that each bank groups customers with similar score to one of 20 groups ("score bands"), according to the percentile rank. Banks do it in order to set interest rates adequately to perceived credit risk – such practise is called "risk-based pricing" (Edelberg, 2006, Walke, Fullerton, & Tokle, 2018). For simplicity, we will use normal distribution's quantiles to get the score bands, for example score band 1 would contain applicants with scores between $-\infty$ and -1.645, score band 2: (-1.645, -1.282), etc. Generally, interval corresponding to score band *i* would be $\left(\Phi^{-1}\left(\frac{i-1}{20}\right), \Phi^{-1}\left(\frac{i}{20}\right)\right)$, where Φ^{-1} is the inverse of the CDF of the standard normal distribution.

In order not to be a charitable institution, the bank should make profit in each score band. Let us assume that based on the previous historical records each bank sets the interest rate at the level 3 percentage points higher than the default rate (and rounds it to the nearest 0.5 pp). For simplicity, we assume that there are no costs (or the interest rate is net of additional costs) and the difference between the interest rate and the default rate can be called the profit of the bank. Table 3 presents result of price setting in one of the banks after the simulation. Other banks set quite similar prices. Please note that no interest rate cap nor similar anti-usury regulation is assumed, which means that any level of interest rate is possible. As a result, each bank should show a profit of about 3 thousand units (provided that amount of each of ~100 thousand loans is 1 unit). The simulations performed by the author show that this is exactly the case.

Tab. 3: Simulation of interest rate setting process in one of the banks.

Score band	Default rate	Interest rate
1	0.3844	0.415
2	0.2605	0.290
3	0.2086	0.240
4	0.1729	0.205
5	0.1457	0.175
6	0.1288	0.160
7	0.1101	0.140
8	0.0966	0.125
9	0.0849	0.115
10	0.0753	0.105
11	0.0636	0.095
12	0.0568	0.085
13	0.0478	0.080
14	0.0386	0.070
15	0.0343	0.065
16	0.0271	0.055
17	0.0221	0.050
18	0.0159	0.045
19	0.0103	0.040
20	0.0042	0.035

Source: simulation performed by the author.

3 „Loan-shopping“ assumption

Now let us consider a slight change in the model: an applicant, before taking a loan, verifies the offer in 3 randomly selected banks (out of total 10 banks) and then chooses the bank with the lowest interest rate (if more banks have the same lowest rate, the selection is random from within these banks). We may call this approach “loan-shopping”. If we run simulation with this slight change, the results are quite different (see Table 4).

As can be seen in Table 4, it turns out that the banks’ financial profits are negative. The second thing that should be noted is that the Gini coefficient of the bank’s scoring model is higher if the information on default is collected only on the loans granted by the bank than if it would be calculated on all loans granted in the market. This initially surprising result may be easily explained by the fact that the bank gains additional knowledge on underlying credit risk of the applicant if, after checking two other banks, the applicant accepts the offer. The default rates by score band and the structure of score bands change, which is illustrated in Table 5.

Tab. 4: Simulation results summary for 1 million customers with loan-shopping and wrong prices.

Bank	Profit	Number of loans	Default rate	Gini coeff. on all loans	Gini coefficient on granted loans
1	-346.420	100721	0.097	0.540	0.565
2	-827.880	102087	0.111	0.539	0.606
3	-489.085	97531	0.101	0.540	0.569
4	-330.900	100628	0.098	0.540	0.563
5	-434.200	99870	0.098	0.541	0.574
6	-548.380	99615	0.100	0.538	0.564
7	-442.955	100193	0.100	0.540	0.563
8	-334.095	100192	0.098	0.541	0.567
9	-306.420	98207	0.098	0.542	0.571
10	-344.985	100956	0.097	0.540	0.563

Source: simulation performed by the author.

Tab. 5: Simulation of interest rate setting process in one of the banks, after introduction of the „loan-shopping" assumption.

Score band	old price	observed default rate	new price
1	0.420	0.5356	0.565
2	0.290	0.3966	0.425
3	0.240	0.3201	0.350
4	0.205	0.2800	0.310
5	0.175	0.2192	0.250
6	0.160	0.2022	0.230
7	0.145	0.1856	0.215
8	0.130	0.1515	0.180
9	0.115	0.1265	0.155
10	0.105	0.1148	0.145
11	0.095	0.1005	0.130
12	0.085	0.0847	0.115
13	0.080	0.0794	0.110
14	0.070	0.0589	0.090
15	0.065	0.0534	0.085
16	0.055	0.0386	0.070
17	0.050	0.0306	0.060
18	0.045	0.0262	0.055
19	0.040	0.0139	0.045
20	0.035	0.0059	0.035

Source: simulation performed by the author.

With the new prices, the financial results of the banks are positive again (Table 6). This initial simulation brings however quite an important lesson: if the competitive environment changes, the interconnected system has to accommodate it with price changes in order to maintain profitability.

Tab. 6: Simulation results summary for 1 million customers with loan-shopping and correct prices.

Bank	Profit	Number of loans	Default rate	Gini coeff. on all loans	Gini coefficient on granted loans
1	3149.73	102180	0.094	0.538	0.549
2	2996.04	100292	0.112	0.538	0.603
3	3054.84	102410	0.092	0.537	0.546
4	2871.21	101682	0.109	0.536	0.597
5	2965.55	99942	0.103	0.539	0.589
6	3070.37	101414	0.100	0.537	0.575
7	2974.13	98489	0.099	0.539	0.548
8	2915.89	97642	0.096	0.538	0.547
9	2985.95	99882	0.101	0.538	0.576
10	2837.22	96067	0.094	0.538	0.555

Source: simulation performed by the author.

As we have the base scenario with loan shopping ready, we may then simulate what happens if the power of the credit scoring in one of the banks improves. We may simulate it with one following change in the $\mathbf{\Sigma}$ matrix: $\rho_1 = 0.6$, while the rest of the correlations remain at the previous level $\rho_2 = \rho_3 = \cdots = \rho_{10} = 0.5$.

In the first run of the simulation we get the results which are presented in Table 7.

Tab. 7: Simulation results summary for 1 million customers after bank 1 improves its credit scoring.

Bank	Profit	Number of loans	Default rate	Gini coeff. on all loans	Gini coefficient on granted loans
1	***5628.85***	***102440***	***0.076***	***0.643***	***0.677***
2	2590.60	96283	0.097	0.540	0.558
3	2610.42	96382	0.095	0.540	0.552
4	2692.77	99984	0.112	0.540	0.602
5	2656.23	101045	0.101	0.540	0.566
6	2829.85	99504	0.099	0.542	0.560
7	2662.14	98185	0.105	0.539	0.561
8	2918.14	106577	0.107	0.541	0.592
9	2856.82	102904	0.111	0.540	0.593
10	2760.41	96696	0.095	0.541	0.560

Source: simulation performed by the author.

As the ρ_1 parameter went up from 0.5 to 0.6, the underlying Gini went up from around 0.54 to 0.64, and observed Gini went up from around 0.57 to 0.68. This has helped to better separate good and bad customers which resulted in reducing default rate from around 10% to 7.6% and increasing the profits significantly (from around 3 thousand to more than 5.5 thousand units). In the next round, the bank can accommodate its prices to the new situation. which can help increase the market share (see Table 8). Other banks also adjust the prices. The profit advantage is comparable to the previous round, but this time it is achieved through increasing the market share.

Tab. 8: Simulation results summary for 1 million customers after bank 1 improves its credit scoring and all the banks adjust their prices accordingly.

Bank	Profit	Number of loans	Default rate	Gini coeff. on all loans	Gini coefficient on granted loans
1	***5689.485***	***169227***	***0.0623***	***0.640***	***0.625***
2	2642.325	92362	0.1054	0.541	0.567
3	2398.930	93070	0.0977	0.540	0.544
4	2494.870	90624	0.1012	0.539	0.553
5	2514.055	94207	0.1054	0.540	0.558
6	2700.175	91719	0.1066	0.540	0.572
7	2529.910	94849	0.1144	0.540	0.590
8	2627.110	92812	0.1166	0.539	0.598
9	2525.375	92009	0.1154	0.539	0.593
10	2555.580	89121	0.1029	0.540	0.576

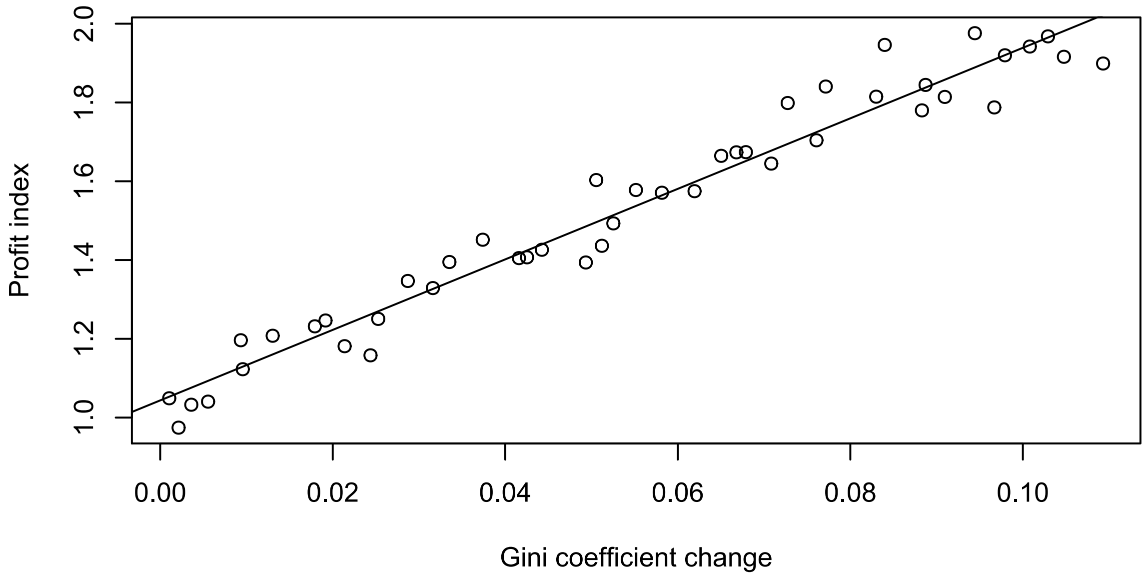
Source: simulation performed by the author.

Finally, it turns out that increasing the ρ_1 correlation from 0.5 to 0.6 (which corresponds to increase in the objective Gini from about 0.54 to 0.64, and observable Gini from about 0.57 to 0.63), almost doubles the profits of the bank in this simulation (the ratio of the new profit to the old profit is ~1.84).

4 Simulation results for various Gini coefficient increases

Similar simulation can be run for various levels of ρ_1 parameter (various levels of separation power increases). Figure 1 illustrates results of 45 simulations for ρ_1 parameter values from the interval [0.5, 0.6]. Profit index shows how much the profit increased after increase in Gini and two rounds of simulations compared to the base scenario (profit index of 2.0 means that the profit doubled, 1.5 that it increased by 50%). The linear regression of the profit index on Gini coefficient change has also been run and the line has been added to the graph. The slope coefficient of the regression line is around 8.95. This means that in the simulation settings 1 percentage point increase in the Gini translates into ~9% increase in the profit of the bank, which shows importance of even slight changes in the credit scoring models.

Fig. 1: Simulation results – profit increase vs Gini coefficient increase.



Source: simulation performed by the author.

Conclusion

The framework presented in the article constitutes a simple starting point to model the lending market from the perspective of credit scoring management. The framework model takes into account competition among banks, "loan-shopping" practices by credit applicants, risk-based pricing and differences in scoring models between the competing banks.

The framework model may help understand microeconomics of credit scoring management. As the example shows, with quite reasonable assumptions, the profits of the bank can be increased by 9% if the Gini coefficient rises 1 percentage point over the market average. Further work using this framework may go in different directions. It may be used to answer other questions, for example what if the loan shopping practices intensify (the applicants checks more then 3 banks before making decision), what if the probability of checking a loan offer depends on the market share of the bank, what happens in the scenario when pricing strategies differ between banks, there is an interest rate cap, there are funding constraints etc.

References

Blochlinger, A., & Leippold, M. (2006). Economic benefit of powerful credit scoring. *Journal of Banking & Finance 30*(3) 851–873.

Einav, L., Jenkins, M., & Levin, J. (2013). The impact of credit scoring on consumer lending. *The RAND Journal of Economics, 44*(2), 249-274.

Edelberg, W. (2006). Risk-based pricing of interest rates for consumer loans. *Journal of Monetary Economics*, 53(8), 2283–2298.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861–874.

Finlay, S. (2014). *Credit scoring, response modelling and insurance rating: A practical guide to forecasting consumer behaviour*. Palgrave Macmillan.

Hanley, J. A., & McNeil, B. J. (1982), The meaning and use of the area under a receiver operating characteristic (ROC) curve, *Radiology*, 143, 29-36.

Řezáč, M., & Řezáč, F. (2011). How to Measure the Quality of Credit Scoring Models. *Czech Journal of Economics and Finance (Finance a Uver), 61*(5), 486-507.

Siddiqi, N. (2006). *Credit risk scorecards: Developing and implementing intelligent credit scoring*. Hoboken (N.J.): J. Wiley.

Thomas, L. C. (2009). Introduction to consumer credit and credit scoring. *Consumer Credit Models*, 1-99.

Walke, A. G., Fullerton, T. M., & Tokle, R. J. (2018). Risk-based loan pricing consequences for credit unions. Journal of Empirical Finance, 47, 105–119.

Contact

Błażej Kochański
Faculty of Management and Economics
Gdańsk University of Technology (Politechnika Gdańska)
ul. Gabriela Narutowicza 11/12
80-233 Gdańsk
bkochanski@zie.pg.gda.pl