JEL Classification: G21, G12, C63, C15
Keywords: credit scoring, retail loans, consumer finance, pricing, simulations

A Simulation Model for Risk and Pricing Competition in the Retail Lending Market

Błażej KOCHAŃSKI - Faculty of Management and Economics, Gdansk University of Technology (blazej.kochanski@pg.edu.pl)

Abstract

We propose a simulation model of the retail lending market with two types of agents: borrowers searching for low interest rates and lenders competing through risk-based pricing. We show that individual banks observe adverse selection, even if every lender applies the same pricing strategy and a credit scoring model of comparable discrimination power. Additionally, the model justifies the reverse-S shape of the response rate curve. According to the model, the benefits of even small increases in the discrimination power of credit scoring are substantial. This effect is more pronounced if the number of offers checked by the applicants before making a decision increases. The simulations illustrate the trade-off between profitability, market share, and credit loss rates. The profit-maximising strategy is to set interest rates slightly lower than the competition; the excessive price reduction turns out to be counterproductive. At the same time, there exists a niche for higher yield players.

1. Introduction

In the modern retail lending market, especially in consumer finance, large quantities of relatively small loans are granted every day. Compared to the other segments of the financial intermediation market, retail lending products are relatively homogeneous. The cost competition has forced automation and uniformisation of lending processes. Currently, virtually all retail lenders use credit scoring to make efficient underwriting decisions quickly. Most of them group loan applicants into risk segments and differentiate the interest rates and other components of the price – this practice is referred to as risk-based pricing (Edelberg, 2006, Staten, 2015). Statistical analysis of large quantities of data and relatively prompt feedback between granting loans and their repayment allow retail banks to modify their pricing and underwriting strategies quite often.

In this paper, we model the retail lending market in a way that combines credit scoring and risk-based pricing. There are surprisingly few researchers who look at this market from such a perspective. Freixas and Rochet (2008), who present a

https://doi.org/10.32065/CJEF.2021.02.01

The author would like to thank Agnieszka Wałachowska, Magdalena Olczyk, two anonymous referees and the editor for their helpful comments and suggestions.

review of the seminal papers in banking microeconomics, mention credit scoring just once and then move to the discussion of the option risk approach to loan pricing. Typical works on credit scoring (Anderson, 2007, Siddiqi, 2017) do not address the details of the microeconomic mechanisms behind credit risk management. Breeden (2010) presents interesting methods to model how credit portfolio losses fluctuate in line with economic cycles and other changes in the environment. His approach helps understand the macroeconomic challenges faced by banks in credit risk management, but the microeconomics and market competition are hardly touched.

The publications explicitly dealing with the microeconomics of credit scoring and pricing include Thomas (2009), Phillips (2013, 2018), and Jankowitsch et al. (2007). Thomas (2009) formulates microeconomic models of credit risk management and pricing mechanisms in consumer finance and retail banking. Jankowitsch et al. (2007) build a model for the economic value of credit scoring. Phillips (2013, 2018) describes best practices in pricing consumer credit products. However, contrary to this paper, all of them view the market-driven mechanisms, like response curves, as external and given.

Adverse selection is observed by credit practitioners as well as celebrated by economists. In the context of credit markets, adverse selection is demonstrated by the fact that higher interest rates attract higher-risk customers and discourage lower risk customers, ceteris paribus. Stiglitz and Weiss (1981) show that adverse selection is behind the otherwise difficult to explain phenomenon of credit rationing. The existence of adverse selection in the consumer lending market has been shown by many researchers. Edelberg (2004) provides evidence for adverse selection in the case of US mortgage loans, Adams et al. (2009) and Phillips et al. (2015) do the same for car loans. Ausubel (1999), Agarwal et al. (2010), and Nelson (2017) show empirical data confirming the existence of adverse selection in the credit card market. Phillips (2018) prefers to refer to the “price-dependent risk” rather than to the “adverse selection”, arguing that this relationship results not only from the asymmetry of information, but also from other factors. Our model shows that negative selection is observed even when the only information asymmetry in the market is that banks do not know the scores and prices set for consumers by other banks. This result is similar to that obtained by Huang and Thomas (2014), who use a different model. Additionally, in our model, banks adjust their pricing to account for the effects of adverse selection.

Blochlinger and Leippold (2006) attempt to answer the question of the economic benefits of credit scoring. They illustrate the microeconomic interplay between banks and borrowers in the corporate lending market, but their model can also be applied to retail finance. It is surprising that out of the hundreds of publications citing Blochlinger and Leippold (2006) we found only three articles where their model is actually used. Agarwal and Taffler (2007) use the model to simulate the economic benefit of using Taffler’s z-score, Hahm and Lee (2011) quantify additional profits resulting from adopting positive bureau information sharing in the Korean market, while Mertens et al. (2018) employ the model to assess the differences in the economic benefit of three popular corporate rating models in the German credit market. In all these cases, the model usage is limited to the economic benefit of credit scoring, and the model is taken without major modifications.

In this paper, we go beyond the original purpose of the Blochlinger and Leippold (2006) model. We expand this model to some extent, to take into account the behaviour of players in the consumer lending market. We introduce a scoring-based segmentation ("score bands") and a pricing strategy based on historical loss rates. The number of banks increases from three in the original paper to ten; in our opinion, it better reflects the number of competitors in a typical retail lending environment. At the same time, it is unrealistic to assume that consumers compare all the loan offers before making a decision. For this reason, we introduce the offer selection mechanism (applicants check c banks, where $c = 3$ in the base scenario).

As it turns out, the model with such modifications remains manageable; at the same time, it can uncover interesting patterns. It explains the adverse selection observed in the retail lending market. It confirms the shape of the response curve, illustrating the relationship between the take rate and the interest rate. It can also shed some light on how minor improvements in credit scoring translate into a significant increase in profits (this is how the model was used by Blochlinger and Leippold (2006) and other authors). Additionally, we show that the scale of improvement depends largely on the loan shopping practices by customers. We also provide a simple example of what pricing strategies might look like in a retail lending environment.

The remainder of the paper is organised as follows. First, we describe the simulation model, its assumptions and parameters. Then the model is used to show that risk-based pricing competition is sufficient to explain the S-shaped response rate curve and adverse selection. The next section builds on the approach proposed in (Blochlinger and Leippold, 2006) to derive the benefits of improving the separation power of a credit scoring model. We show that, to a great extent, the impact of credit scoring depends on the number of loan offers checked by borrowers before the final choice. Finally, we move to analyse the interplay between the market share, the credit losses, and the pricing strategy. The last part discusses possible additional applications of the model and its further development.

All calculations were performed in R. The codes are available from the author on request.

2. Model

2.1 Model Description

Banks and other retail lenders do not act independently of their competitive environment. When they make pricing decisions and build credit scorecards, they need to consider the actions and decisions of their competitors and customers. Similarly to other markets, before taking a loan, the potential customer usually checks several loan offers and chooses the one which is perceived the best for his or her situation. The banks cannot arbitrarily set prices but need to accommodate loan pricing based on applicants' responses, as higher prices result in lower take rates.

The framework proposed by Blochlinger and Leippold (2006) allows for taking into account the credit scorecards used by banks, pricing of loans based on the credit scores, and potential borrowers "shopping" for the best loan offers. The original model was applied to the corporate lending market with individual pricing, where each of the applicants got offer prices based on their risk, and there were only three banks in the economy. It is possible to adapt the model to the retail lending market – for example, short-term consumer finance loans. Compared to the original model, we increase the number of banks in the economy to 10, and we introduce (1) the scoring-based segmentation ("score bands"), (2) the pricing strategy based on the loss rates from the preceding simulation round, and (3) the offer selection mechanism in which potential customers check the price not in all but in some of the banks operating in the market. We use the resulting model to analyse the economics of credit scoring (as Blochlinger and Leippold (2016) do), but also examine the adverse selection effect, the shape of response curves, as well as the interplay between the profitability targets, market share and credit risk.

In this paper, we assume that there are k banks in the economy. It is one of the model parameters: in all simulations, we take $k = 10$, as this value, in our view, best reflects a typical number of competitors in the retail lending market[1]. Each of the k banks has a credit scoring tool and uses it actively to set the prices (interest rates). Credit scores S_1-S_k computed for a sample of potential borrowers by the k banks as well as the values of the latent credit risk factor for these borrowers, Y^*, can then be generated from a multivariate $k + 1$-dimensional normal distribution:

$$(S_1, S_2, S_3, \dots, S_k, Y^*) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

where the means vector consists of zeros:

$$\boldsymbol{\mu} = (0, 0, 0, \dots, 0)^T$$

[1] According to Jurek (2016), 10 biggest banks account for the majority of the total assets in all UE countries, from around 60% in France or Germany, through 70-80% in Central Europe, up to 89% in Greece and 95% in Sweden.

and the covariance matrix has the following form:

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & \rho & \rho & \dots & \rho & \rho_1 \\ \rho & 1 & \rho & \dots & \rho & \rho_2 \\ \rho & \rho & 1 & \dots & \rho & \rho_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho & \rho & \rho & \dots & 1 & \rho_k \\ \rho_1 & \rho_2 & \rho_3 & \dots & \rho_k & 1 \end{pmatrix}$$

Note that all marginal distributions are standard normal; therefore, the covariance matrix is, at the same time, the correlation matrix.

Variables $S_1 - S_k$ represent the credit scores computed for each loan applicant by the k banks in the economy. S_1 is calculated by bank 1, S_2 is calculated by bank 2, and so on: the customer with index p will receive score $s_{1,p}$ from bank 1, $s_{2,p}$ from bank 2, etc. The banks know the scores calculated by themselves but do not know the scores calculated by the competition. The loan applicants do not know their scores, but they get to know the prices offered to them by the banks to which they applied. The last variable, Y^* is the credit risk factor, or "latent risk variable", which itself is not observable by banks or borrowers, but translates into an observable 0/1 default event Y:

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

In Equation 1, d is the default rate in the market (another parameter of the model) and Φ^{-1} is the inverse standard normal cumulative distribution function. In the simulation, we will assume the market default rate at the level of 10% (d = 0.1) – the level typical for consumer finance[2], so $\Phi^{-1}(d) \approx -1.28$. We take the left tail of the distribution rather than the right tail (as is in the original model) because such an approach appears more intuitive to banking practitioners: lower scores usually correspond to higher risk.

The draws from the Monte-Carlo multivariate distribution described above will represent loan applicants. Customer 1 will get a draw of k + 1 variables representing her credit scores and the value of the latent risk variable: $s_{1,1}$, $s_{2,1}$, ..., $s_{k,1}$, y^*_1, customer 2 will be represented by the following vector: $s_{1,2}$, $s_{2,2}$, ..., $s_{k,2}$, $y*_2$, etc.

As can be seen, we assume that the scores are normally distributed, directly or after a monotone transformation. The credit scorecards are not identical, but they are correlated. Following Blochlinger and Leippold (2006), we assume a correlation

[2] For example, 12-month rolling default rate for consumer finance in Czech Republic was about 8% in 2008 and dropped to about 5% 10 years later (Czech National Bank, 2019)

between the bank scorecards at the level ρ = 0.8. In the retail lending context, we may support this level of ρ with the fact that the credit models are based on similar variables (like credit bureau delinquencies, credit history, income, and other financial or demographic data). According to Consumer Financial Protection Bureau (2011), in the US market, the correlations are 0.90-0.92 if the scores are built solely on credit bureau data. Including bank-specific data should result in a slightly lower correlation, therefore 0.8 was adopted. For simplicity, we assume the same level of correlation between scores for each pair of banks.

The scores themselves, or their monotone normally distributed transformations, are correlated with the latent risk variable. Thanks to these correlations (ρ_1-ρ_k in the covariance matrix), the banks gain some insight into the latent risk variable, and in this way, into future default rates, but their knowledge is not perfect. Correlations ρ_1-ρ_k are related to the degree of separation power of scoring models S_1-S_k. In the base scenario, we assume ρ_j = 0.5 for each bank j = 1, ..., k.

It is worth mentioning that the latent risk variable, Y^*, is obviously forward-looking: it determines the future event of default. As such, it cannot be replicated by an ex-ante scorecard computed at the beginning of the observation period. Therefore, Y^* variable could be viewed as implicitly composed of an ideal score (containing all the information available at the beginning of the period) and a residual reflecting future changes and uncertainty, as in the context of Merton's model. In practice, it means that the correlations ρ_j between the scores and the latent variable will be substantially lower than one.

The separation power of credit scorecards is usually measured by what is called by credit risk managers a Gini coefficient (Anderson, 2007, Siddiqi, 2017). This coefficient, another name of which is Somers' D, is zero if credit scoring is just random and one if it is perfect.

The formula for scorecard's Gini in the multivariate normal model we use in this paper is a somewhat complicated integral (Blochlinger and Leippold, 2006, Thomas, 2009), but may be solved numerically:

$$Gini = 2\int F(s|B)dF(s|G) - 1$$

where $F\,(s|B)$ is cumulative frequency of credit scores for bad (defaulted) customers:

$$F(s|B) = \frac{1}{d}\int_{-\infty}^{s} \Phi\left(\frac{\Phi^{-1}(d) - \rho_j x}{\sqrt{1-\rho_j^2}}\right)\phi(x)\,dx$$

and $F\,(s|G)$ is cumulative frequency of credit scores for good customers (those who did not default):

$$F(s|G) = \frac{1}{1-d}\int_{-\infty}^{s} \Phi\left(\frac{-\Phi^{-1}(d) + \rho_j x}{\sqrt{1-\rho_j^2}}\right)\phi(x)\,dx$$

According to the equations, the Gini of the scorecard S_j is a function of the market default rate d and the correlation ρ_j between the scores and the latent variable. As the market default rate is d = 0.1, a correlation ρ_j of 0.5 translates to a Gini coefficient of about 0.542. To take another example, a correlation of 0.6 translates into a Gini coefficient of 0.645. Such Gini coefficients are in line with the experience of consumer credit risk analysts (Beling et al., 2005, Bernardo et al., 2013, Hahm and Lee, 2011, Rezáč and Rezáč, 2011).

The banking market we are modelling is a retail lending market where thousands of loans are granted in a short period by each of the banks. Due to the large number of loans, the usual approach in this market is that lenders group their customers into score bands, that is, segments of applicants who are similar in terms of credit risk. In this paper, we assume that each bank gathers credit applicants into b = 15 groups, based on their percentile rank in the population (according to Witzany (2017), the number of score-based segments is usually between 7 and 25). We use the standard normal distribution's quantiles to get the score bands. For example, score band 1 would contain applicants with scores between $-\infty$ and -1.501, score band 2 between -1.501 and -1.111, et cetera. In general, the interval for score band i would be $\left(\Phi^{-1}\left(\frac{i-1}{15}\right), \Phi^{-1}\left(\frac{i}{15}\right)\right)$.

The banks set the prices (interest rates) for particular score bands. Currently, it is a prevailing approach in consumer finance (Edelberg, 2006, Phillips, 2018, Staten, 2015). As there are k banks in the market and each of them can offer loans to b score bands, there are $k \times b$ prices set in each simulation round. Let the symbol r_{ij} denote the interest rate set by bank j for customers in the score band i. In our model, each loan has the same term and the same ticket (1 unit); thus, the price is the only factor that differentiates the loans. We assume that every bank wants to achieve similar profitability in each score band, so setting the target for the overall return on assets (ROA, i.e., profits as a share of the amount invested in loans) defines the pricing strategy. The banks use the preceding period loss rates in each score band to adjust the pricing. The loss rates are the default rates adjusted by the loss given default (LGD) factor. For example, if the default rate is 10% and we assume that 30% of the loan will be recovered in the case of default, the LGD factor is 0.7 and the final loss rate is 7%. The banks set the interest rates at the level enabling them to cover their credit losses and to get the ROA of approximately $100m_j$ percentage points, where m_j stands for the margin. In the base model, m_j = 0.03 for each of the banks (j = 1, 2, ..., k). Therefore, the price r_{ij} is set based on the ROA target m_j and the loss rate l_{ij}

observed after the preceding period by the bank j in the score band i (see Equation 2).

2.2 Simulation Steps

Before the first round of the simulation, there is an initiation phase. The price matrices (containing interest rates for each bank and score band) are filled with initial values. The initial prices are based on the expected loss rates of all applicants (this means that all banks have the same prices in the analogical score bands at the beginning, in the following rounds the prices will be driven by the loss rates observed by particular banks and their pricing strategy).

One simulation round (one period) consists of the following steps:

(1) At the beginning of each round, n new customers are generated (their scores, the latent risk variable, the score band, the default flag). "Nature" already knows the defaults, but neither the default flag nor the latent risk variable is visible to banks or borrowers before the decision on accepting the loan

(2) Each applicant checks and compares c randomly selected banks before making a decision. She chooses the best price. If two banks have the same price, it is randomly decided which one will be chosen by the customer. If the price offered by each of the banks is higher than the maximum interest rate (μ), the applicant is rejected.

(3) The profits (interest revenue minus credit losses), market shares, default rates, loss rates (default rates adjusted by the loss given default factor), and other summary characteristics are calculated for the banks. This is the output of the model.

(4) Finally, the banks prepare for the next round. They compute the new prices for the next period based on the observed loss rates on loans granted in a given score band. The new prices are set so that the return on assets (return on the aggregated amount of originated loans) in each of the scoring segments is m_j. To achieve this, the new interest rate r_{ij} for the score band i in bank j depends on the observed loss rate l_{ij} and the margin m_j assumed by the bank, according to the following equation:

$$r_{ij} = (l_{ij} + m_j)/(1 - l_{ij}) \quad (2)$$

We divide by $(1 - l_{ij})$ because non-defaulted or recovered customers only pay the interest[3]. If bank j did not grant any loans for the band i in the given period, the price for this bank in this specific score band remains unchanged.

The model has to be run for several periods to enable adjusting of the prices. In particular, banks need to adjust their pricing to accommodate the effects of adverse selection, described in one of the next sections. The tests showed that for the prices to stabilise, at least three periods are needed. This results from the feedback loop in the simulation: prices are driven by the loss rates; loss rates l_{ij} are driven by the characteristics of borrowers and their choices; the prices drive the borrowers' choices.

2.3 Model Parameters

To summarise, the parameters in the model are as follows:

- k – the number of banks (10 banks in this paper),
- the rounding precision of an interest rate, in this paper, the prices are rounded to the nearest quarter of a percentage point,
- b – the number of score bands used by a bank (in the current version of the model, all banks have the same rules for grouping applicants) and use the same b=15,
- d – the market default rate (10% in this paper),
- l – the loss given default (LGD) rate, roughly equal to one minus the recovery rate; set at the level of 0.7, typical for unsecured consumer lending[4],
- n – the number of loan applicants in a given period (1 million assumed in the base scenario to avoid excessive noise in the results),
- $\mathbf{\Sigma}$ – the correlation matrix (ρ_1-ρ_k, the correlations between the scores and the latent variable, set to 0.5 in the base scenario, ρ, the correlations between the scores of particular banks, set to 0.8),
- c – the number of bank offers checked by the applicant before making a choice, c = 3 in the base scenario,
- μ – the maximum interest rate, set to 0.4 (40%) in the simulations to ensure there exists an anti-usury cap in the market (possibly self-imposed),

---

[3] The rationale behind this formula could be illustrated by the following example. We invest 100 000 euro in 100 loans 1000 euro each. Loss rate is 16% (16% of the loans will not be repaid, i.e. the lender will not be able to recover either the principal or the interest). The lender aims at a 5% return on assets (so the target is to have 5 000 euro of profit). The interest rate should then be (0.05 + 0.16)/(1 − 0.16) = 21/84 = 25%. Indeed, if we calculate the revenue and subtract the losses we arrive at the target: (84000) (0.25) − 16000 = 5000.

[4] LGDs in unsecured consumer lending may range from 46% for overdrafts to 77% for credit cards (Konečný et al., 2017)

- m_j – the margin applied by bank j, set to 0.03 in the base scenario, which is a reasonable target margin for a retail lending institution (Vandone, 2009),
- r_{ij} – the interest rate (price) used in the given period by bank j and score band i – this price needs to be set for the first simulation period, and then it is recalculated according to the loss rates and assumed margins (Equation 2)

The first five parameters do not change from period to period in the current version of the model. There is a possibility to alter the rest of the parameters. We will manipulate some of them in the simulations presented in the following sections.

3. Response Rate and Adverse Selection

If a bank accepts the borrower and presents the offer, it is the applicant's turn to make a decision. The applicant may accept the offer and decide to take the loan or reject it and go to the competition. As discussed by Thomas (2009) and confirmed by the banking practice, the take probability may differ: it depends on the credit risk of the customer, measured by the scoring, as well as on the interest rate offered by the bank. When the price increases, the take rate (i.e., the share of the applicants who decide to take a loan at such price, also known as the response rate) drops. The response rate curve, or take rate curve, illustrates the relationship between the interest rate (x-axis) and the take rate (y-axis). It starts with high percentages for the lowest interest rates and ends close to zero for the highest rates. This curve is reverse S-shaped (Agrawal and Ferguson, 2007, Huang and Thomas, 2014, Ma et al., 2010, Thomas, 2009).

Adverse selection is a phenomenon observed by credit practitioners and celebrated by economists (Stiglitz and Weiss, 1981). It is driven by the asymmetry of information between lenders and borrowers. As a result, the probability of default in a given score band is higher for borrowers who took the loan than for the applicants who got the offer but rejected it. It can be illustrated by the following inequality (Thomas, 2009):

$$p(B|r,s,T) > p(B|s) > p(B|r,s,N)$$

where s is a score value, r is the price (interest rate), T is the event that an applicant decides to take the loan, N is the event that the applicant does not take the loan, B is the event that an applicant becomes bad (defaults).

This interplay between risk and price may be driven by several factors (Phillips, 2018, Thomas, 2009):

(1) alternative offers – the higher the price, the fewer applicants finally decide to take the loan as they have better offers from the bank's competitors,

(2) reservation interest rate – the borrowers have some maximum price above which they decide not to take the loan (Edelberg, 2006),

(3) affordability – the default probability increases with the price of the loan, as it is more likely for the borrowers to run into financial difficulties if the loans they take have high interest rates,

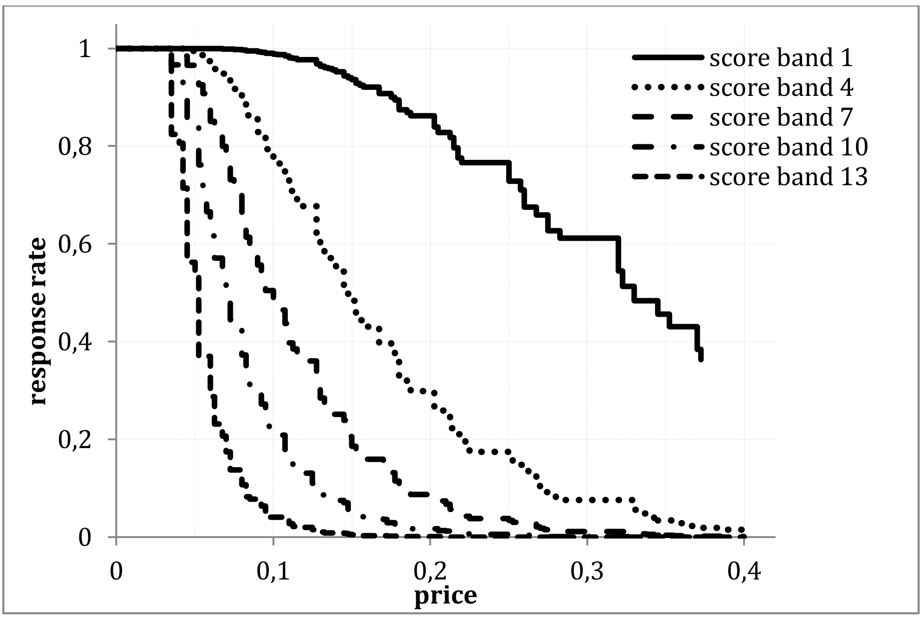
(4) fraud – fraudulent applicants do not plan on repaying, so they are not sensitive to the prices,

(5) behavioural factors – people who value current consumption more than future consumption are the same people who tend to accept higher prices and to be more risky borrowers,

(6) adverse private information (for example, information on coming layoffs in the workplace).

The model presented in this chapter takes into account only the first one of the above factors. There is no reservation interest rate for the customers. The model assumes that they take the loan at any price, provided it is the best offer they found. Additionally, the interest rate does not impact the default probability of a customer – it is the other way round: the default rate expected by the bank in a given score band drives the price. The model does not take into account fraud, behavioural factors or adverse private information.

Figure 1 Response Rate Depending on Price for Bank 1, for Selected Score Bands



The simulation shows that the existence of risk-based pricing and alternative offers is sufficient to explain both the reverse S-shaped response curve and adverse selection.

We run the model with base scenario parameters (k = 10, b = 15, d = 0.1, l = 0.7, n = 1 or 10 million, c = 3, m_j = 0.03 for all js, and μ = 0.4). Figures 1-3 illustrate the outcome of such simulations. The response rate curves in Figure 1, clearly reverse S-shaped, were constructed as follows. We take all applicants who considered bank 1 in their offer selection process and were classified by this bank to one of the score bands, say, band 7. For the borrowers who chose to take the loan from bank 1, the second-best offer is examined. In other words, we check how much bank 1 could increase its price for a particular customer and still remain the preferred lender. For the borrowers who went to the competitor, we take the interest rates of the loans they chose – in this way, we obtain the price bank 1 should have offered to the applicant if it wished to become the number one on her list. The information gathered in this manner enables the computation of the take rate for every possible level of price. Based on the take rates, the response curve is constructed. The same procedure is repeated for other selected bands (1, 4, 10, and 13).

Figure 2 Default Rates in Each Score Band for Applicants Who Took the Loan from the Bank, Did Not Take the Loan and for All Applicants

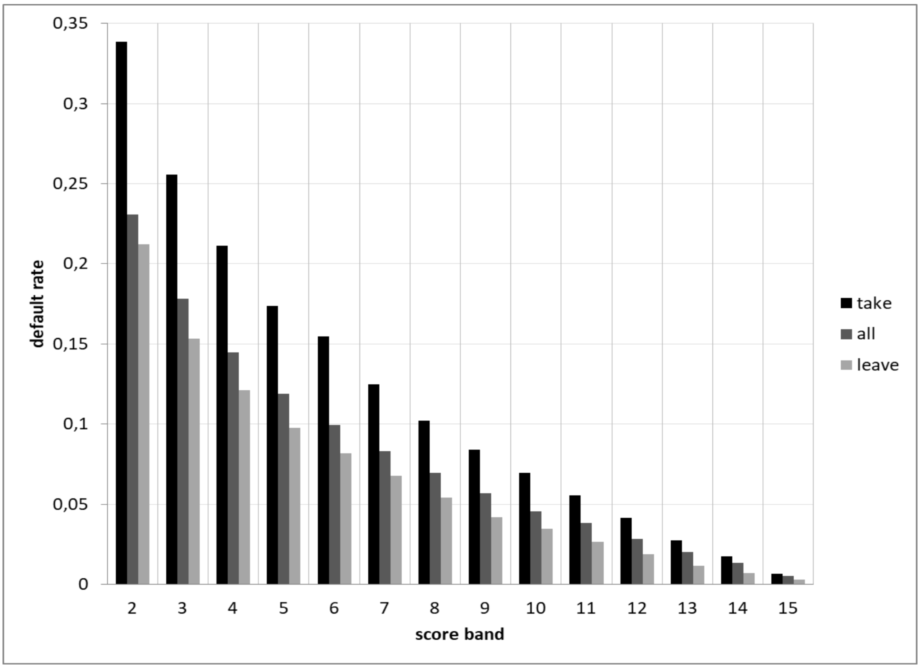
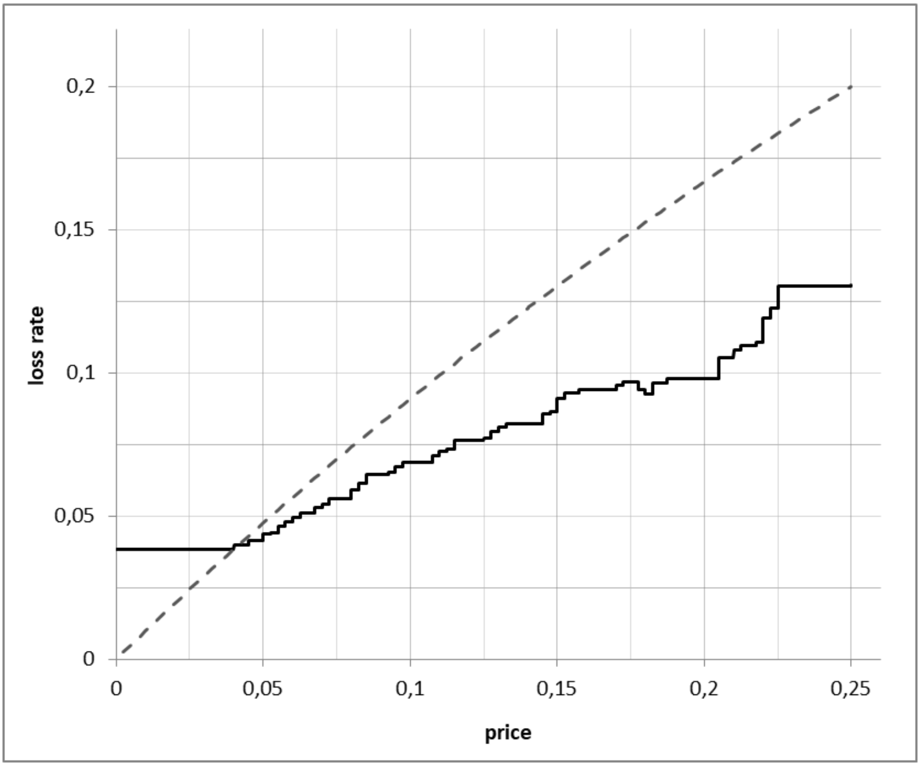


Figure 3 Loss Rate Depending on the Price Offered by Bank 1 for its Score Band 11 Customers (the Solid Line), the Maximum Possible Loss Rate for a Given Price (the Dashed Line)



The simulation also confirms the phenomenon of adverse selection. Interestingly, all banks observe adverse selection in all score bands, even if, and despite the fact, they all have the same pricing strategy and the same discrimination power of their credit scorecards. Figure 2 illustrates this. The observed default rates on the loans granted in each score band, represented by the darkest bars, are the highest for each score band. The default rates of borrowers who decided to take a loan somewhere else (the lightest bars) are the lowest. It is driven by the fact that each customer checks $c = 3$ banks before making decisions, so if she stays with the bank, it means that other banks have classified her into worse score bands and offered higher prices. In fact, this is the only information asymmetry in the model: banks do not know the scores given to the applicants by their competitors.

Figure 3 illustrates another view of the adverse selection phenomenon. If bank 1 decided to set the interest rate for score band 11 at zero per cent, it would gain all applicants, but it would be a loss-making decision. On the other hand, the higher the price, not only the take rates drop (Figure 1), but also the borrowers who remain tend to default more frequently than those who leave (Figure 3). The dashed line represents the maximum loss rate that the bank can accept and still prevent financial losses in this segment. It meets with the solid line at the break-even point. To make money, the bank has to increase the price in this score band even further, despite the falling take rate and rising loss rate.

4. Credit Scoring Economics

The model developed in this paper may be used to assess the impact of improved credit scoring on a bank's profits. The simulation is run with the following parameters: k = 10, b = 15, d = 0.1, l = 0.7, n = 1 million, c = 3, m_j = 0.03 for all js, and μ = 0.4.

Then in one period, the ρ_1 for bank 1 is changed, all other things remaining equal, and we wait three more periods to observe the results. The change in ρ_1:

$$\Delta\rho_1 = \rho_{1(new)} - \rho_{1(old)}$$

can be used as a measure of the improvement of the credit scoring used by bank 1; however, we convert it into the difference in Gini coefficients, $\Delta Gini_1$, so that the results could be transparent to a credit risk practitioner.

Figure 4 Increase in Profits Against Increase in Gini Coefficient

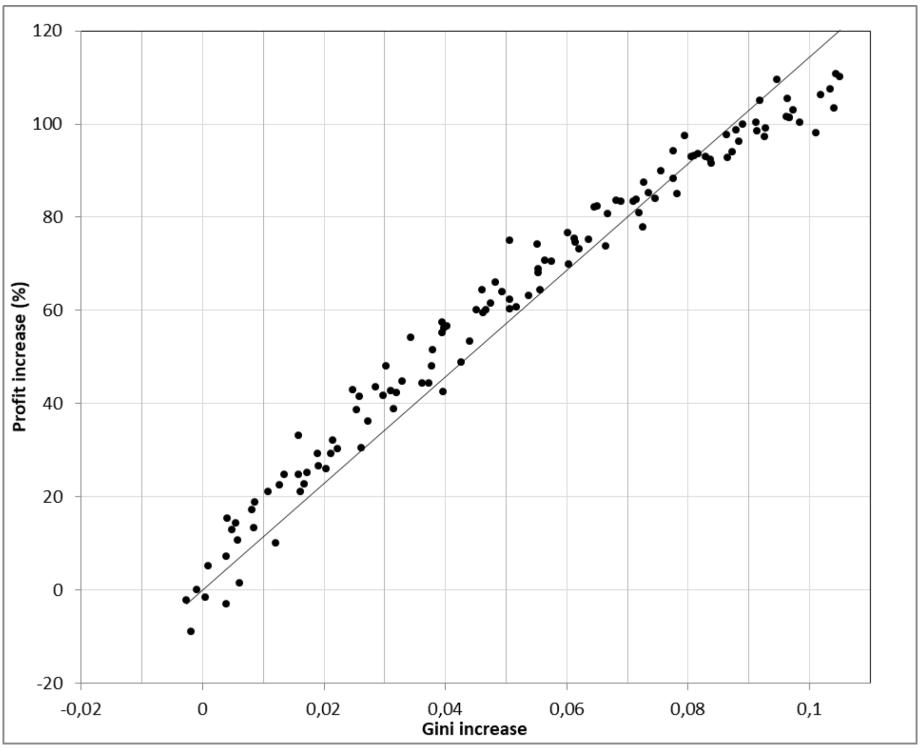
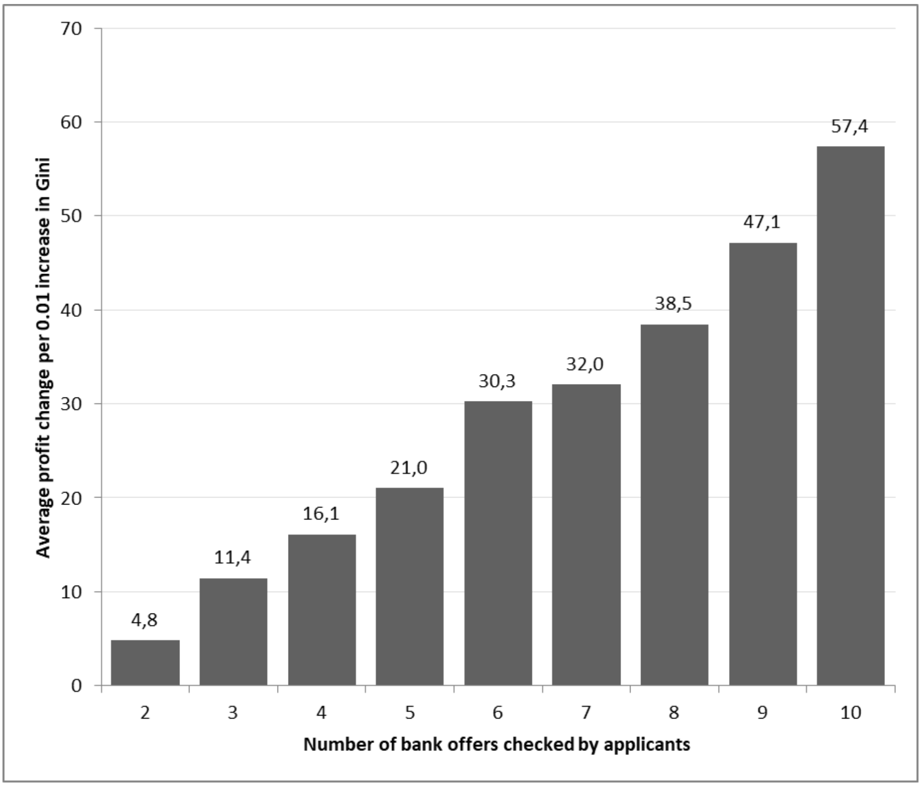


Figure 4 plots the percentage increase in profits against the observed increase in the Gini coefficient of a bank ($\Delta Gini_1$). As the simulations were run for different levels of final ρ_1, we have $\Delta Gini_1$ varying from 0 to roughly 0.10. It appears that the relationship between Gini and profit is close to linear in this range. Consequently, without a big loss of generality, it can be summarised by the regression line. The regression slope (with the intercept forced to 0) is 11.4, which means that $\Delta Gini_1$ of 0.01 is on average associated with 11.4% increase in profits. In our model, the bank with an improved scorecard still uses the pricing strategy based on historical loss rates and three-percentage point margin, so the improvement in the profits comes from the increased market share.

Figure 5 Gini Profit vs Increase Number of Checks (c Parameter)



We can repeat the set of simulations. We do it for various values for c (from 2 to 10). Thus, the loan shopping practice varies from the situation when an applicant checks only two loan offers before the decision to the situation when all possible offers are reviewed.

For each value of c, the regression slope was calculated – the regression slopes are displayed in Figure 5. It seems that the effect of the credit scoring improvement depends to a large extent on c, which is an exogenous parameter showing loan shopping intensity. Nevertheless, even if we assume $c = 2$, the impact of a slight increase in the Gini coefficient is substantial.

The results obtained in this section are similar to those obtained by Blochlinger and Leippold (2006) (in their model, the bank increases profits by 34.3% as a result of a 0.033 increase in Gini) or by Hahm and Lee (2011). At the same time, these results seem to be much higher than the results expected by banking professionals. According to the intuitions of the risk managers we talked with, a 0.01 increase in Gini should translate into 1 or 2% increase in profits, a much lower number than the 5% or 12% obtained in the most conservative simulations.

This difference between the model result and professionals' intuitions can be reconciled in a few ways. First, the model specification or its parameters may be, to some extent, unrealistic or overly simplified. On the other hand, one cannot exclude that bank managers are not aware of the power of the improved scoring models.

In our opinion, both explanations can be valid. It is true that banking managers, even those directly engaged in credit risk management, may frequently be unaware of the financial impact of improvements in credit scoring combined with a proper risk-based pricing strategy. At one of the industry conferences, we heard that a new regulation, which consisted of virtually banning some traditional credit scoring variables, could result in an average drop of Gini coefficients by 0.02-0.04, so, from the bottom-line perspective, "it would not be a big deal".

On the other hand, the model may be overly optimistic in its assumptions. For example, it might be that a substantial share of the applicants does not make financially optimal decisions. Some may prefer to use criteria not directly linked to the loan product, like the cooperation with the bank to date, the reputation and brand, or its marketing campaign. Their choices may be driven by loan tenure, loan amount, instalment amount, or loan covenants, none of which are included in our model. It is also possible that not all banks use the risk-based pricing approach in the way our framework assumes. Another relevant factor may be that the default rates do not translate directly into loss rates – the loan period may be longer than the outcome period used for the development of the scoring model; the LGD may also distort the results.

These considerations should be taken into account if the model proposed by Blochlinger and Leippold (2006) and expanded in this paper is going to be used to guide business decisions. They could also form the basis for the further development of the model.

Leaving aside exact quantification, the results presented in this section show that the effects of improving scoring models in the risk-based pricing regime depend to a large extent on the applicants' loan shopping practices. It can be concluded that an increase in financial awareness and a better understanding of the pricing of credit products by bank customers will result in an increase in the importance of scoring models.

5. Pricing, Profitability and Market Share

Another application of the model is to get insight into the pricing strategy that an individual bank may apply. In this section, we resort to simple modifications of the parameter m_1, representing the margin used by one of the banks.

The graphs in Figures 6-9 are based on the following simulation. We run the model with the standard parameters (k = 10, b = 15, d = 0.1, l = 0.7, n = 1 million, c = 3, m_j = 0.03 for all js, and μ = 0.4) and then change the margin m_1 applied by bank 1 from 0.03 to some other value between 0 and 0.07, wait several periods, and observe the results.

Figure 6 ROA vs. Pricing Strategy

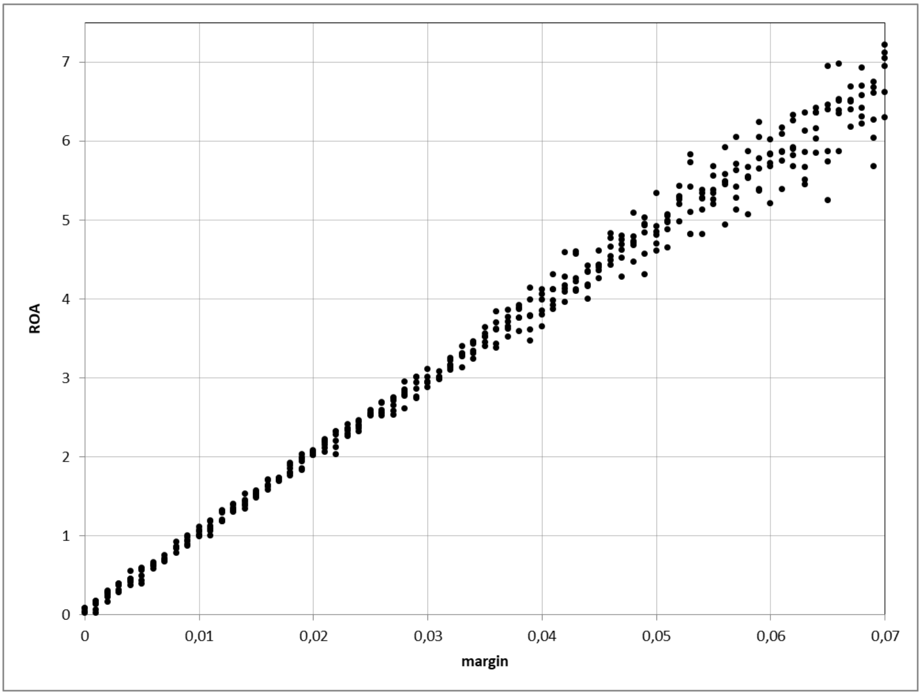
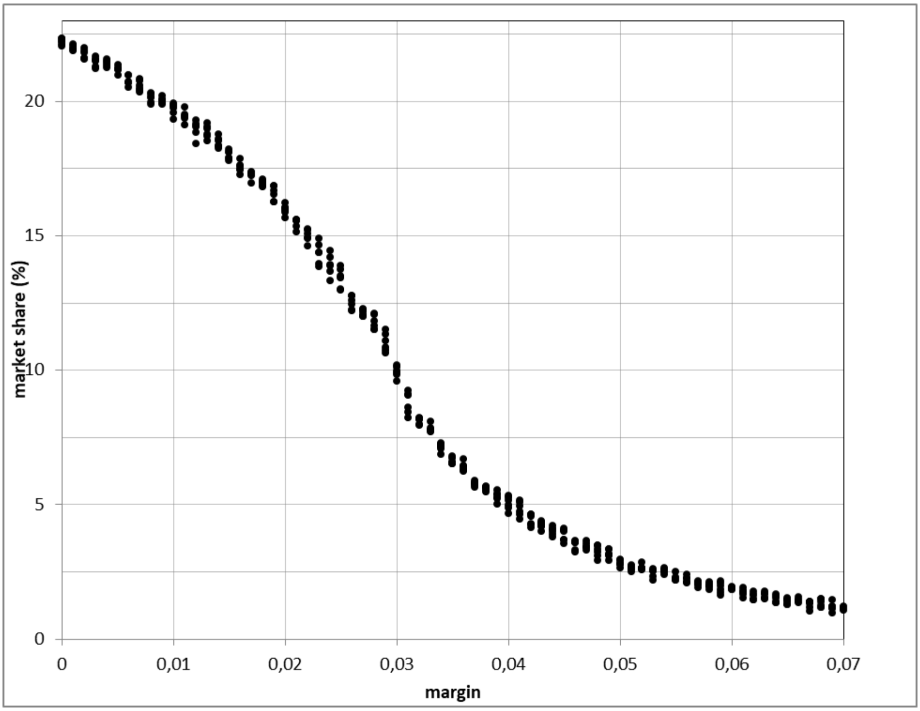


Figure 6 illustrates the fact that a bank in our model can effectively control the return on assets through price changes – the relationship between m_1 and ROA is linear, and the points accumulate along the y = x line. This should not come as a surprise since the m_1 parameter in the simulation is the return on assets expected by the bank. The simulation shows that ROA twice as high as the one of the competition is attainable, although it may be more difficult to achieve: the points are more scattered at higher values of m_1. However, keep in mind that doubling the margins substantially reduces the bank's market share – in this scenario, the bank's assets decline approximately fivefold, to around 2% of the total market. Figure 7 shows this. At m_1 = 0.03 the bank would have a market share at the level of 10% – exactly like its nine competitors who follow the same strategy. When m_1 is lower than that of the competition, the market share increases; when the price is higher, the market share decreases.

Figure 7 Market Share vs. Pricing Strategy



In Figure 8, the profits are plotted against m_1. It seems that if the bank uses slightly lower margins than those of the competition, it can increase its profits. However, a large decrease in margins at one go seems to be a counterproductive pricing strategy. The increased market share does not compensate for the decrease in the return on assets; as a result, the profits drop.

Figure 8 Bank Profits vs. Pricing Strategy

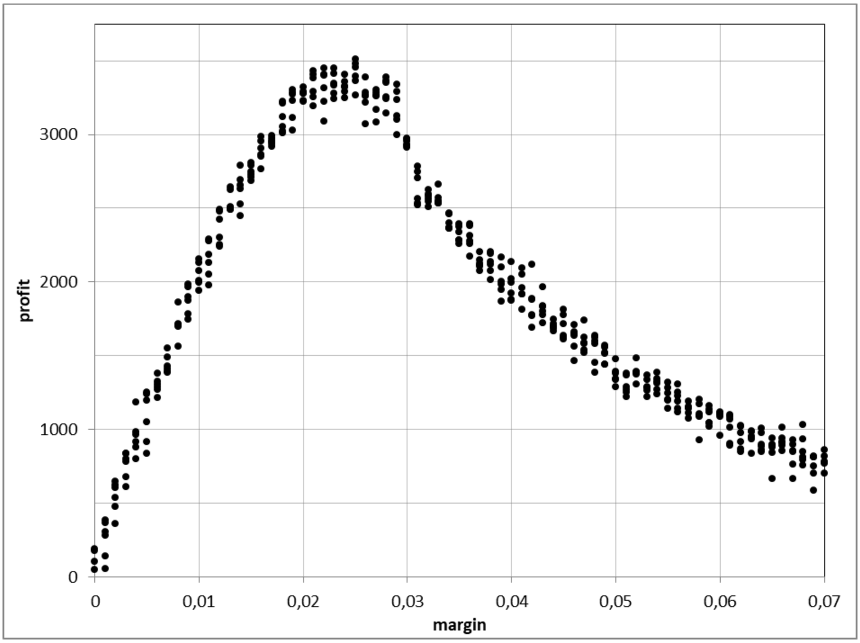
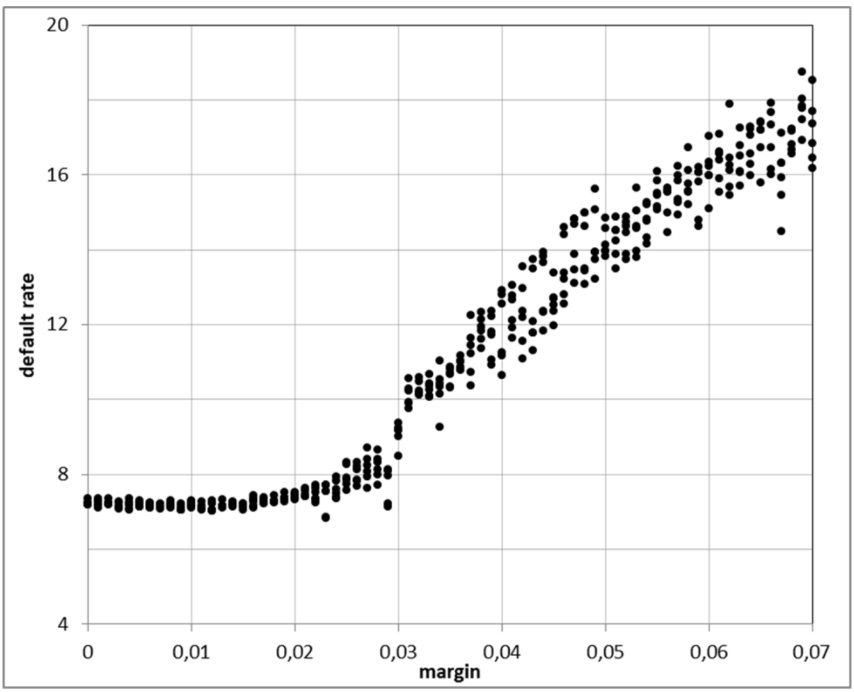


Figure 9 Default Rate vs. Pricing Strategy



Moreover, the model shows that a moderate price increase may be the preferred strategy for a niche lender interested in higher yields. Such an institution should, however, understand that this strategy is associated with a substantial increase in credit loss (note the hockey stick shape of the graph in Figure 9).

We saw that a moderate price cut is a profit maximisation strategy for a single bank. However, if all players adopt this strategy, there will be a gradual erosion of margins across the market.

6. Discussion and Further Research

As described in the preceding sections, the model proposed by Blochlinger and Leippold (2006) can be extended to cover the consumer lending market. In this paper, scoring-based segmentation is added to the model, as well as the pricing strategy whereby the banks set new prices in the next simulation period, based on previously observed loss rates. We introduce the loan shopping parameter (c), and see how the changes in this parameter affect the results. Applying such a model helps understand the economics of credit scoring, which was already shown by Blochlinger and Leippold (2006) and other authors (Agarwal and Taffler, 2007, Hahm and Lee, 2011, Mertens et al., 2018). On top of what other researchers have achieved, we show that the gains from increasing the strength of the credit scoring model depend, in a risk-based pricing regime, on the intensity of loan shopping practices of the customers.

We also go beyond the original application of the model and show that it can explain the reverse S-shaped response curves and the adverse selection phenomenon. The latter result is similar to that provided by Huang and Thomas (2014), where a different modelling approach is used. In addition, we show that banks can subsequently adjust their prices based on the observed loss rates to account for the effects of adverse selection. In the framework outlined in this paper, adverse selection is observed for all lenders even if the Gini coefficients and pricing margins are the same for each bank.

In line with our model, the optimal pricing strategy for a profit-maximising lender is to offer prices slightly, but not much, below the market prices. If all banks follow this strategy, it leads to a gradual erosion of profit margins. At the same time, we show that there exists a niche in the market for banks with a higher appetite for yields, which, however, requires acceptance of elevated credit losses.

We tried to calibrate the parameters of the model to best reflect a typical consumer loan market. Where possible, we used prior research, available data, or credit risk managers' opinions when setting the parameters. In our opinion, the parameters of the model reflect the specifics of the consumer lending market. Note, however, that our tests show that the model will return comparable results for other parameter sets. The exact numerical quantification of the results may be different and clearly depends on the specific calibration, but the overall conclusions presented in the article (reverse S-shaped response curves, adverse selection, increasing importance of credit scoring when loan shopping intensifies, pricing strategies etc.) remain valid even if the values of the parameters, such as default rates, LGD, number of banks, number of score bands, or correlations, change.

Certainly, there are more possible applications of this framework than the ones presented in this paper. For example, we may:

- combine an increase in the discrimination power of a scoring model with the changes in pricing strategy,
- simulate a market in which some banks use risk-based pricing and some do not and observe the results,
- analyse the impact of various levels of the maximum interest rate,
- analyse the impact of granularity (number of banks, number of score bands) on the results,
- simulate a market in which there is a shift in credit losses and the market default rates grow from period to period,
- find the optimal number of offers checked (c) taking into account that there are transaction and other costs associated with obtaining each offer, and the marginal benefits of each check keep decreasing.

There are also possible enhancements to the model, which may help further understand the mechanics of the retail lending market. For example, one could analyse a market with heterogeneous groups of borrowers, where some of the applicants do not pay attention to the prices. Alternatively, customers could select the banks randomly but with probabilities based on banks' market shares. We could also consider a market where particular banks' risk appetites and liquidity constraints vary or where applicants have their reservation rates (the maximum interest rates accepted by borrower segments). Another idea would be to introduce varying loan tickets or repeated loans for the same customers. An analogous modelling framework for the credit card market would also be a challenging task.

The model presented in this paper can be a starting point to analyse the interplay between credit losses, credit risk models, pricing management, market share, and borrowers' decision making. There may be numerous ways of enhancing the model. However, even in the most simple form, the simulations presented in this paper may help, in our humble opinion, better understand the retail lending market.

REFERENCES

Adams W, Einav L, Jonathan L (2009): Liquidity Constraints and Imperfect Information in Subprime Lending. *American Economic Review,* 99(1):9–84.

Agarwal S, Chomsisengphet S, Liu C (2010): The Importance of Adverse Selection in the Credit Card Market: Evidence from Randomized Trials of Credit Card Solicitations. *Journal of Money, Credit and Banking,* 42(4):743–754.

Agarwal V, Taffler RJ (2007): Twenty-Five Years of the Taffler Z-Score Model: Does It Really Have Predictive Ability? *Accounting and Business Research,* 37(4):285–300.

Agrawal V, Ferguson M (2007): Bid-Response Models for Customised Pricing. *Journal of Revenue and Pricing Management,* 6(3): 212–228.

Anderson R (2007): *The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation*. Oxford University Press, Oxford.

Ausubel LM (1999): Adverse Selection in the Credit Card Market. Working paper. https://www.ausubel.com/larry/creditcard-papers/adverse.pdf

Beling PA, Covaliu Z, Oliver RM (2005): Optimal Scoring Cutoff Policies and Efficient Frontiers. *Journal of the Operational Research Society,* 56(9):1016–1029.

Bernardo D, Hagras H, Tsang E (2013): A Genetic Type-2 Fuzzy Logic Based System for Financial Applications Modelling and Prediction. In: *2013 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*:1–8.

Blochlinger A, Leippold M (2006): Economic Benefit of Powerful Credit Scoring. *Journal of Banking & Finance,* 30(3):851–873.

Breeden JL (2010): *Reinventing Retail Lending Analytics: Forecasting, Stress Testing, Capital and Scoring for a World of Crises*. Risk Books, London.

Consumer Financial Protection Bureau (2011): *Impact of Differences Between Consumer and Creditor-purchased Credit Scores. Report to Congress*.

Czech National Bank (2019): *Financial Stability Report 2018/2019*.

Edelberg W (2004): Testing for Adverse Selection and Moral Hazard in Consumer Loan Markets. *Federal Reserve Working Paper.* https://econpapers.repec. org/paper/fipfedgfe/2004-09.htm

Edelberg W (2006): Risk-Based Pricing of Interest Rates for Consumer Loans. *Journal of Monetary Economics,* 53(8):2283–2298.

Freixas X, Rochet J-C (2008): *Microeconomics of Banking*. MIT Press, Cambridge, MA.

Hahm J-H, Lee S (2011): Economic Effects of Positive Credit Information Sharing: the Case of Korea. *Applied Economics,* 43(30):4879–4890.

Huang B, Thomas LC (2014): Credit Card Pricing and Impact of Adverse Selection. *Journal of the Operational Research Society,* 65(8):1193–1201.

Jankowitsch R, Pichler S, and Schwaiger WSA (2007): Modelling the Economic Value of Credit Rating Systems. *Journal of Banking & Finance,* 31(1):181–198.

Jurek M (2016): *Structures of Ownership in the Financial Sector. European Policy Brief.*

Konečný T, Seidler S, Belyaeva A, Belyaev K (2017): The Time Dimension of the Links Between Loss Given Default and the Macroeconomy. *Finance a úvěr-Czech Journal of Economics and Finance*, 67(6):462–491.

Ma P, Crook J, Ansell J (2010): Modelling Take-Up and Profitability. *Journal of the Operational Research Society,* 61(3):430–442.

Mertens RL, Poddig T, Fieberg C (2018): Forecasting Corporate Defaults in the German Stock Market. *Journal of Risk,* 20(6):29–54.
https://www.risk.net/node/5784701

Nelson ST (2017): Private Information and Price Regulation In the US Credit Card Market. *Ph.D. Dissertation. Massachusetts Institute of Technology.*
https://dspace.mit.edu/handle/1721.1/108999

Phillips R (2013): Optimizing Prices for Consumer Credit. *Journal of Revenue and Pricing Management,* 12(4):360–377.

Phillips R (2018): *Pricing Credit Products*. Stanford University Press, Stanford, CA.

Phillips R, Şimşek AS, van Ryzin G (2015): The Effectiveness of Field Price Discretion: Empirical Evidence from Auto Lending. *Management Science,* 61(8):1741–1759.

Rezáč M, Rezáč F (2011): How to Measure the Quality of Credit Scoring Models. *Czech Journal of Economics and Finance (Finance a úvěr),* 61(5):486–507.

Siddiqi N (2017): *Intelligent Credit Scoring: Building and Implementing Better Credit Risk Scorecards*. John Wiley & Sons, Hoboken, NJ.

Staten M (2015): Risk-Based Pricing in Consumer Lending. *Journal of Law, Economics & Policy,* 11(1):33-58.

Stiglitz JE, Weiss A (1981): Credit Rationing in Markets with Imperfect Information. *American Economic Review,* 71(3):393–410.

Thomas LC (2009): *Consumer Credit Models: Pricing, Profit and Portfolios*. Oxford University Press, New York, NY.

Vandone D (2009): *Consumer Credit in Europe: Risks and Opportunities of a Dynamic Industry*. Springer Science & Business Media, Berlin.

Witzany J (2017): *Credit Risk Management: Pricing, Measurement, and Modeling*. Springer International Publishing, Cham, Switzerland.