WHO IS INTERESTED?

ESTIMATION OF DEMAND ON THE HUNGARIAN MORTGAGE LOAN MARKET IN A DISCRETE CHOICE FRAMEWORK[1]

By

Ákos Aczél[2]

The Central Bank of Hungary

Budapest, Hungary

2016

[1] The paper is based on the thesis that was defended at the Central European University by the author in June 2016.

[2] The views expressed are those of the author and do not necessarily reflect the official view of the Central Bank of Hungary.

ABSTRACT

In this paper I estimate demand on the Hungarian mortgage loan market. I apply a conditional logit model on Hungarian credit registry data. I restrict the choice sets of individuals according to their financial position, banks business models and geographic constraints. The results show that interest rates do matter in the decision of consumers, however taste patterns and restrictions on choice sets lead to constrained decisions.

Table of contents

Chapter 1 - Introduction

Taking a mortgage loan is a transaction that has the highest impact on most people's financials in their whole life. Understanding the motifs behind the decision is important for policy makers and for the providers of the financial services as well. In this paper I analyze the decision making process of consumers. I estimate demand on the mortgage loan market by applying a discrete choice model on Hungarian credit registry data of newly granted loans in 2015.

The analysis focuses on the question that which factors matter in the decision of consumers when they choose among banks. It provides information on price sensitivity and also examines how relevant is that consumers face different opportunities on the mortgage loan market according to their financial background. I test two hypotheses. The first hypothesis states that consumers take into consideration the interest rates when they choose among banks to obtain a mortgage loan, and they prefer the smaller interest rates keeping everything else constant. My second hypothesis is that geographical and financial constraints also affect the decision on mortgage loans.

As aggregated data may mask the factors that I am interested in I estimate a conditional logit model on transaction level data. The dependent variable in this model is the bank that consumers choose. I try to explain the decision by the role of interest rate, demographic factors and some other bank and individual related variables assuming utility maximizing consumers. I apply restrictions on the choice sets of consumers to account for the differences in the business models and target audiences of banks. I argue that a poor customer may have no chance to get a loan from a bank that serves wealthy people, however wealthy people are likely to be eligible for a loan at the majority of the banks. Beside financial constraints, I also apply geographic constraints and assume that consumers consider banks which are present in their neighborhood only. To analyze the impact of the constraints I compare the results of the model that is estimated on restricted choice sets to the results of the model that is unrestricted.

The final results support the idea that the differences in consumers' opportunities matter. Estimating the model without controlling on choice sets results in positive coefficient on interest rate, that can be hardly explained in the context of utility maximization. However, if the model is estimated on the restricted choice sets, the results are plausible and the coefficient on interest rate turns to be negative. Interestingly, introducing demographic variables in the model without restricting choice sets also results in negative sloping demand curve. These results highlight that the microstructure of the market plays an important role and should be taken into account when estimating demand.

The paper is organized as follows. In the next section I briefly show the relevance of the topic by describing some stylized facts of the Hungarian mortgage loan market. I summarize the related literature in section three. As a next step I introduce the dataset that I use. I provide the description of my empirical approach in section five and discuss the results in section six. I apply robustness checks in section seven, then finishing with a conclusion.

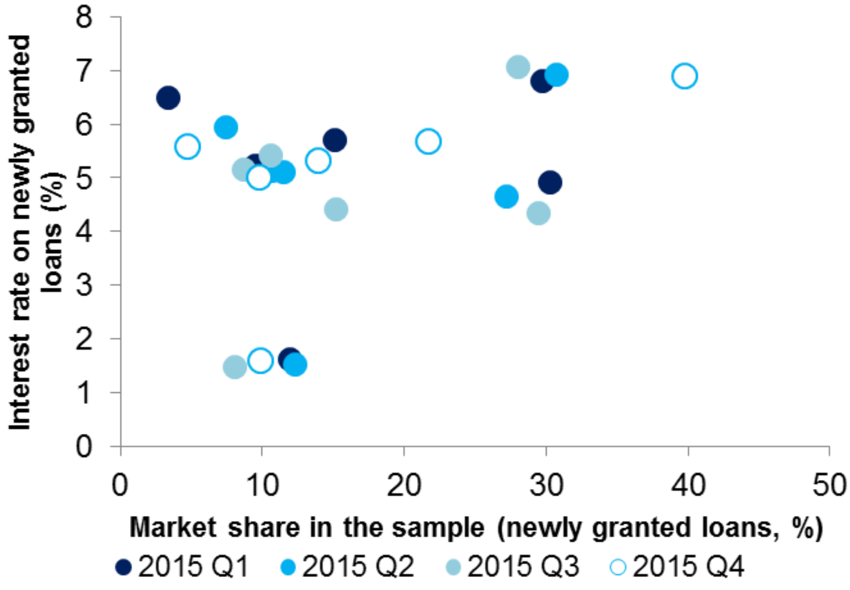
Chapter 2 – Motivation

In this section I show some basic statistics that point to the lack of perfect competition on the Hungarian mortgage loan market. I highlight that geographic constraints may be one of the explaining reasons.

Although it is difficult to formulate well-grounded statements if one looks at only the market shares and price data, this kind of information may help to understand some aspects of the market. On the Hungarian mortgage loan market one can observe that high interest rates do not necessarily result in low market shares of newly granted loans. I calculated market shares and interest rates for key participants of the market for the four quarters of 2015 from the sample that I use for this research. It can be seen that some banks could achieve very high market shares despite of the high interest rates that they set. This pattern could be explained by demand shocks only, however as the market was similar in the last years it is likely to have some structural reasons behind[1].

[1] I do not have the same type of data for previous years, however it is a well known fact that some participants of the market have achieved high market shares in the last years in tandem with setting high interest rates.

Figure 1: Market shares and interest rates of the sampled banks on the Hungarian mortgage loan market[2]

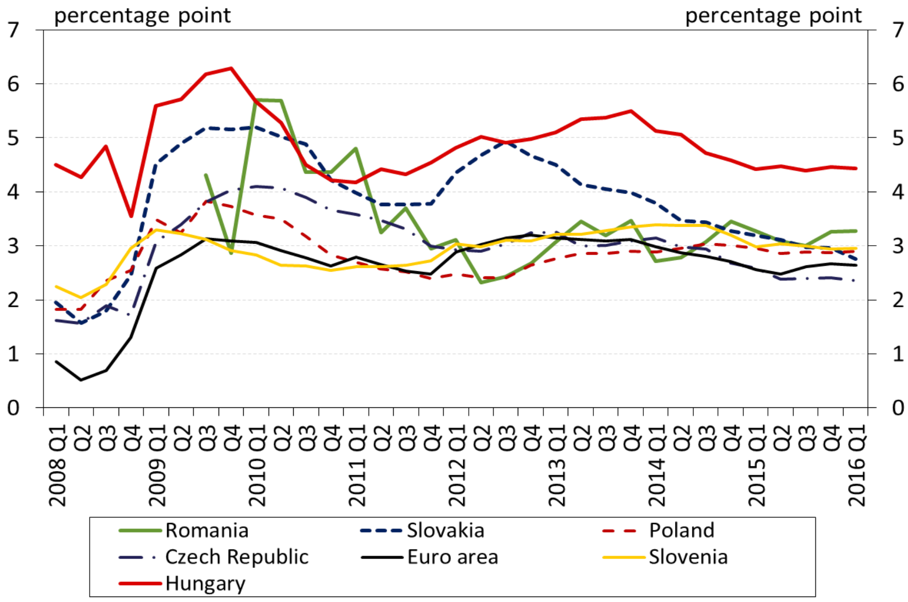


Source: Hungarian Credit Registry, own calculation

Putting this phenomenon into international context also suggests special reasons for the high interest rates. Figure 2 shows the interest rate spread of domestic currency denominated housing loans in six countries in Central and Eastern Europe and the Euro area average between 2008 and 2016 Q1. It can be clearly seen that Hungary has the highest interest rate spread throughout the whole period. Again, there can be several reasons for the high spread including macroeconomic and microeconomic factors. Nevertheless, I argue that these figures together suggest that beside macro and micro related shocks some structural factors can also play an important role in keeping interest rates high.

[2] Due to the sensitive nature of the data I do not disclose the identity of banks throughout the paper. Wherever bank level data is reported I refer to banks as Bank A/Bank B etc. to avoid the possibility of identification of the banks. This restriction does not constrain the explanation of the findings of the analysis.

Figure 2: International comparison of spreads on housing loans extended in domestic currency

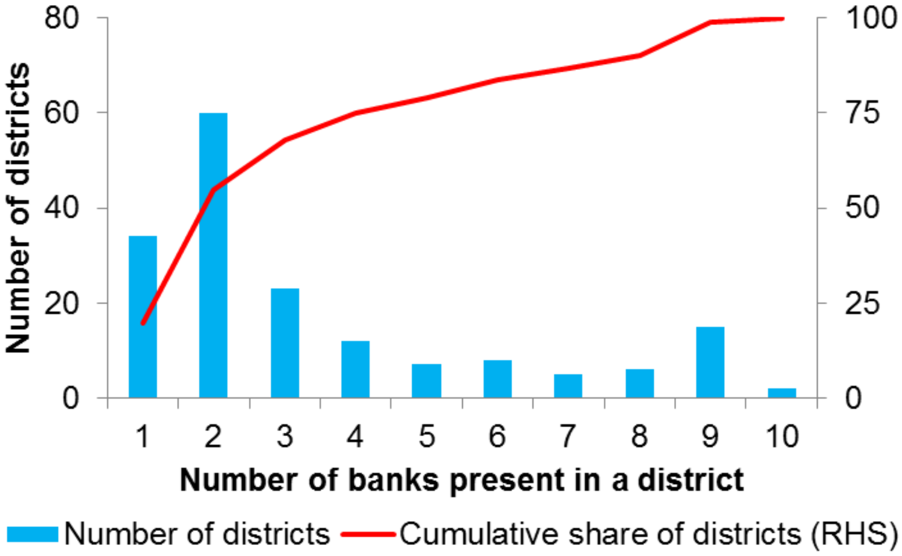


Source: MNB

Although there can be several structural reasons for the high interest rates, at this part of the paper I bring only one thing into relief. Figure 3 shows the presence of the eleven biggest banks of the Hungarian market in districts[3]. These banks accounted for more than 80 per cent of the newly granted housing loans in the previous years. Axis x counts for the number of different banks that are present on the market, and axis y shows how many districts have the given number of different banks. Basically, the picture gives information about the number of banks consumer can choose from. It is a striking fact that in half of the districts there are only one or two banks present and 75 per cent of the districts have no more than four different banks out of the eleven big Hungarian banks. The limited presence of banks in districts highlights that consumers have very limited opportunities when they choose among banks[4].

[3] There are 198 districts in Hungary. There are around 15-20 towns in one district and two cities on average.
[4] It is also interesting to consider the distribution of the population as well. More than one fourth of the population lives in a district where no more than two different banks are available and 40 per cent lives in districts where at most four banks are available out of the eleven big banks. Only the half of population may reach more than five banks out of the elven big banks in their districts.

Figure 3: Distribution of braches among districts



Source: MNB, own calculation

The complete explanation of the high interest rates would require modeling both supply and demand side and would be beyond the scope of this paper. To narrow down the question, I focus on the estimation of demand in the remaining chapters. The fact that banks' presence is limited and varies considerably between districts supports applying a varying choice set model. That model can also account for high market shares if those are related to banks which have broad branch networks. Moreover, consumer patterns based on demographics may also explain a part of the high interest rates. If banks are in the situation of monopolistic competition due to the fact that groups of consumers prefer some banks over other banks, that also points to high interest rates. In a conditional logit framework all of these factors can be modeled.

Chapter 3 – Related literature

There are three streams of papers that are relevant regarding this research. The first one consists researches that develop and apply the discrete choice models of multinomial outcome. The most important paper among these regarding my topic is the work of McFadden (1973). The author describes a conditional logit model that includes alternative specific regressors. Hence, in this model not only the individuals' characteristics but the choice related characteristics matter as well. As McFadden (2001) highlights, one reason that this model became very popular is the fact that it links the choice model to the individual's utility maximization problem. The model showed outstanding performance in applied work which also boosted its popularity. Namely, McFadden (1974) estimated a conditional logit model on travel-mode choices and predicted the market share of a new fixed rail transit system which was about to be introduced. The estimations turned out to be surprisingly precise when they were compared to the actual market shares of the fixed rails system after the launch. There is a series of papers of McFadden that focused on conditional logit models and their further development (McFadden (1978), McFadden et al. (1977), McFadden (1987), McFadden et al. (1978b))

Although the modeling framework differs moderately, there is a broad literature based on the seminal papers of Berry (1994), Berry et al. (1995) (BLP) and Nevo (2001) which is also linked to my research. Most of these researches are based on aggregated data, however during the simulation of individual choices they apply a multinomial logit model. The goal of these models is to find the parameters of the multinomial logit model that make the observed market shares equal to the predicted shares. The difficulty of having no individual level observations is solved by simulation techniques where the parameters are based on the distributions which are observed in census-type data. Berry et al. (2004) also applies this method when both market level and individual level data is available. There are minimum three important contributions of these papers to the literature. Firstly, they develop methods to handle the independence from irrelevant alternatives (IIA) property of the basic conditional logit model that is discussed later. Secondly, they also offer a modeling framework for situations when only aggregate level data is available. Lastly, the simulation techniques makes the application of instrumental variables straightforward, hence handling endogeneity becomes easier. Regarding my research, I follow the estimation technique of McFadden (1973), as I observe individual level data. The difficulty that I face is that handling endogeneity issues in this framework by using the standard instrumental variable technique would be cumbersome. Therefore, I follow an identification strategy that is not based on instrumental variables.

The second set of papers are strictly related to the first in that they also apply the discrete choice framework on aggregated data. The reason for separating them is that they mainly do not contribute to estimation techniques but they apply them on the banking market. The seminal paper of Dick (2002) uses the BLP framework to analyze the deposit market in the US. An important contribution to the literature beside the empirical results is that he shows how the standard discrete choice models can be applied to the banking market. The definition of market, product and price is not obvious in the banking industry and finding valid and strong instruments is also challenging. Dick offers solutions for the arising questions. The offered instrument set was partially utilized in the paper of Molnár et al. (2007) and in Holló (2010) in the analysis of the Hungarian market. Molnár et al. estimated models for the main products that have own markets (three types of deposits and three types of consumer loans), and found that the competition is limited on most of the products’ markets. Holló focused on the cross price elasticities of different products by applying a random coefficient logit framework, and showed that a slight increase in interest rates can motivate consumers to look for another offer. Although the topic of these studies is close to my interest, I depart from this part of the literature as I would like to focus on the microstructure of the market. Namely, I want to examine the relevance of the varying choice sets that consumers face when they choose among different loan offers. This analysis needs consumer level data and the construction of choice sets of all consumers.

The last stream of papers works with consumer level data and models the choice of consumers on the banking market. Although the methodologies applied in these papers are quite close to what I do, most of these papers put emphasis on different questions that I try to answer. Phillips and Yezer (1996) stresses the potential bias concerning discrete choice models applied on the mortgage loan market due to self-selection. Follain (1990) argues that the choice of mortgage loan includes several other decisions from the consumer. Namely, he should decide on the loan-to-value ratio, choose the optimal mortgage instrument and also consider paying or defaulting in the later period of the loan. Rachlis and Yezer (1993) argued that mortgage lending is a set of decisions on: applying for a mortgage loan or not (borrower), accepting the borrower or refusing the application (bank), choosing the mortgage that has optimal conditions (borrower), paying back the whole mortgage or choosing to default (borrower). These studies emphasize the relevance of sample selection bias due to the fact that they mostly focus on discrimination on the mortgage market. I argue that sample selection bias is not a major concern regarding my research as I do not focus on discrimination. Nevertheless, interpretation of my results should be made by care, as the results may not be applicable if the set of potential consumers changes considerably.

Chapter 4 – Dataset

The research is based on the data of the Hungarian Credit Registry. This dataset contains all the credit transactions from 2015 with broad analytics. As I analyze only mortgage loans, I focus on the characteristics of them. Beside the exact type (housing loan/home equity/other mortgage loan), yearly interest rate, size, term of the loan and id of the bank that granted the loan several other characteristics are available. Firstly, there is information regarding the collateral. The value of the property and the total value of the collateral are given. The latter differs from the first as banks apply varying haircuts for real estates. Moreover, additional collateral can be involved in the transaction as well. The frequency of payment and the number of clients involved in the transaction are also available. There are several risk related characteristics in the dataset. The debt service ratio (DSR) and the corresponding wage are given. There is also information on the risk weight that the banks assigned to the transaction. This weight is used for capital requirement calculations, so in several cases it is just calculated according to supervisory rules. Although it reflects the riskiness of the borrower it is not that precise as probability of default and loss given default estimates. These two are only given for a subset of banks, hence they cannot be used for modeling. I also constructed the variable called history that shows whether the individual have had any credit transaction with the given bank in the last eight years. Lastly, the age and place of living of the individual is also available. The location refers to districts only, however this precision is enough for this research.

The second source that I use contains the locations of bank branches in every quarter. I merged this dataset with the credit registry data to obtain a final database with credit characteristics and information on bank presence in the districts.

In order to obtain a coherent database I dropped several observations as their validity was ambiguous. I deleted observations if any of the following conditions applied:

- Loan size lower than 100.000 Forints
- Debt service ratio is not greater than zero or greater than 100[5]
- Wage is equal to zero
- Wage over is 5 million Forints[6] or under 10.000 Forints
- Refinanced loan[7]
- Value of the property is not given
- Term of transaction is under half year
- Interest rate is less than 50bp[8]

Finally, I have 19420 credit transactions in the dataset. I cover the seven biggest participants as I do not have enough observations for other banks. Nevertheless, the market share of these banks on the mortgage loan market exceeds 70 per cent according to the database of the National Bank of Hungary.

Next, I describe the main patterns of the population in the sample and show some basic statistics to give a taste of the data that I use.

The age of the consumers in the sample varies between 18 and 78 years with an average of 39 years. Almost 60 per cent of the consumers have taken credit in the last eight years from their

---

[5] According to the Hungarian legislation the debt service ratio of a newly granted loan can not exceed 60 per cent. It is likely that the loans over this limit were recorded mistakenly, or there are some special technical reasons which may explain the phenomena (http://m.portfolio.hu/finanszirozas/hitel/fekrendszer_blokkolasgatlo_nelkul_az_mnb_szabalyainak_elso_kilometerei.216003.html). In either cases I should exclude these loans from my analysis.

[6] Although there are not many observations with wages higher than 5 million HUF, I decided to delete these observations as extreme wages may be recorded mistakenly

[7] Ideally refinanced loans would be included in the research, however during the observed period the scheme of the MNB of converting households' FX loans into forint loans took place that may distort the data.

[8] I excluded extreme low interest rates as 50 bp (only 32 observations), as such low rates likely to not reflect market mechanisms, but may have some special technical reasons or just not recorded correctly

bank, and 40 per cent had credit transaction with another institution. The distribution of wage is right skewed as the average is over 300.000 Forints, while the median is less than 200.000 Forints. The minimum and maximum wages are according to the data cleaning process that I summarized above. Similarly to wages, the distribution of the value of the properties and the size of the loans are also right skewed. The median property value of 10,8 million Forints is less than the half of the average value of 24,8 million Forints. Loan size varies between 453.900 and 243 million Forints with a median loan of a bit more than 5 million Forints. Interest rates are distributed around 6 per cent, and there are also observations in the sample with interest rates as low as 1 per cent, and as high as 11 per cent.

Table 1: Descriptive statistics

	Age	Wage	Value	Loan size	Interest rate
Mean	39,3	301 007	24 800 000	7 460 000	6,0
Median	38,0	190 900	10 500 000	5 120 000	6,2
Min	18,0	10 000	590 400	453 900	1,0
Max	78,0	4 970 000	1 010 000 000	243 000 000	11,0

Source: Hungarian Credit Registry, own calculation

Looking at these statistics by banks highlights that there is a strong heterogeneity among banks' clientele and business model. The median wage of banks clients varies roughly between 100.000 and 600.000 Forints. Similarly, there is high variation among the median values of the properties and the loan sizes. Variation of average interest rates among banks is not high, except one bank with as low as 1,6 per cent average interest rate in the sample the other banks' average interest rates are within a 2,4 percentage point range (4,7 – 7,1 per cent). The high variation in client characteristics with the low variation in interest rates suggests that the competition is limited on this market and banks may use their monopolistic power in pricing.

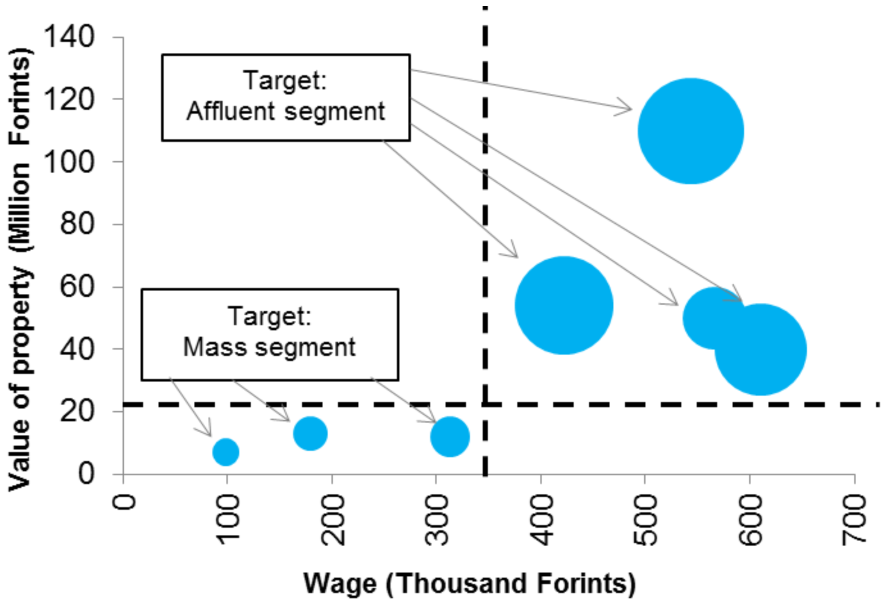
Table 2: Descriptive statistics by banks

	Wage (med)	Interest (mean)	Value (med)	Loan size (med)
Bank A	422 664	5,3	54 000 000	14 000 000
Bank B	543 321	1,6	110 000 000	15 000 000
Bank C	313 536	4,7	12 000 000	5 763 469
Bank D	610 400	5,8	40 000 000	13 000 000
Bank E	179 532	5,5	13 000 000	4 970 124
Bank F	98 302	7,1	7 040 000	3 942 513
Bank G	566 208	6,1	50 000 000	9 972 730

Source: Hungarian Credit Registry, own calculation

Based on these characteristics banks can be separated into two groups. Banks in the first group target clients who have limited financial possibilities, while the other group focuses on wealthy clients. This rough classification highlights the differences among banks target audience and suggests that banks behave differently on the market. Higher end banks are more willing to win affluent clients, while the other banks serve clients with weaker financials.

Figure 4: Differences among banks' clienteles



Note: The size of the circles represent the size of the loan

Source: Hungarian Credit Registry, own calculation

Chapter 5 – Empirical Approach

The modelling framework is based on three pillars. Firstly, the most important part is the discrete choice model that follows the work of McFadden (1973). Secondly, as the main interest is in the factors that affect the choice between banks I also need choice specific information. The theoretical interest rate that banks would have offered to individuals if they had asked for a loan is not observable. Therefore, I mimic the scoring models of banks and estimate the theoretical interest rates based on the information of the transactions that I observe. Thirdly, I restrict the choice set of the individuals based on the distribution of banks among districts and on the observed place of residence of consumers. I also restrict the choice sets according to the business models of banks what I try to pin down by the patterns of lending that I can observe in the data.

Conditional logit model

The discrete choice model that I use is a conditional logit model based on McFadden (1974). That model takes a step forward from the standard multinomial logit framework. As Hoffman and Duncan (1988) underlines, while multinomial logit model focuses on the individuals' characteristics, conditional logit model concentrates on the alternatives by introducing alternative specific variables in addition to the individual specific ones. I describe conditional logit models following Train (2002)[9].

Similarly to the standard discrete choice models conditional logit is based on the utility maximization behavior of consumers. Consumer *i* chooses the alternative that results in the

[9] The detailed derivation of the model can be found in chapter 3 and in the appendix of the book

highest utility. More formally consumer i chooses alternative j if it gives higher utility than all the other alternatives:

$$U_{ij} > U_{ik}, \qquad \forall j \neq k$$

where U_{ij} stands for the utility of consumer i from choosing alternative j. As the researcher cannot observe all the factors that affect the utility of the decision makers the utility can be separated into two parts:

$$U_{ij} = V_{ij}(x_{ij}, s_i) + \varepsilon_{ij},$$

where $V_{ij}()$ is a function of observables such as the vector of alternative specific variables x_{ij} and the vector of demographic variables s_i which is invariant across alternatives. ε_{ij} is an idiosyncratic taste shock which is not observable by the researcher, however it also affects the utility of consumer i from choosing alternative j. The error term is assumed to follow an iid extreme value distribution. This assumption makes the calculations convenient, although it also states that the unobserved factors are not correlated among alternatives.

The probability that consumer i chooses alternative j is the following:

$$P_{ij} = P(V_{ij} + \varepsilon_{ij} > V_{ik} + \varepsilon_{ik} \;\; \forall j \neq k) = P(\varepsilon_{ik} < V_{ij} - V_{ik} + \varepsilon_{ij} \;\; \forall j \neq k)$$

As the idiosyncratic term is considered iid extreme value it's density function can be written as:

$$f(\varepsilon_{ij}) = e^{-\varepsilon_{ij}} e^{-e^{-\varepsilon_{ij}}}.$$

Using the density function one can obtain the probability by solving the following integral:

$$P_{ij} = \int (\prod_{k \neq j} e^{-e^{-(V_{ij} - V_{ik} + \varepsilon_{ij})}}) e^{-\varepsilon_{ij}} e^{-e^{-\varepsilon_{ij}}} d\varepsilon_{ij},$$

which is shown by Train (2002) that can be written in closed form:

$$P_{ij} = \frac{\mathrm{e}^{\mathrm{V_{ij}}}}{\sum_{\mathrm{k}} \mathrm{e}^{\mathrm{V_{ik}}}}.$$

If one assumes that the observed part of the utility function is linear in its' parameters:

$$V_{ij} = \mathrm{x}'_{\mathrm{ij}}\beta + D'_j\gamma s_i,$$

then the probability that consumer *i* chooses alternative *j* becomes:

$$P_{ij} = \frac{\mathrm{e}^{\mathrm{x}'_{\mathrm{ij}}\beta + D'_j\gamma s_i}}{\sum_{\mathrm{k}} \mathrm{e}^{\mathrm{x}'_{\mathrm{ik}}\beta + D'_j\gamma s_i}}.$$

An important characteristic of the quotient above is that it always lies between zero and one, therefore it is convenient to use to model probability. It should be also mentioned that the probability of choosing an alternative depends on both alternative specific factors (x_{ij}) and on the interaction of the individual specific factors (s_i, constant across alternatives,) and the bank dummys (D_j). An example of the alternative specific factors can be the price of different goods, while income or age are instances of individual specific characteristics.

Another important feature of this setting is that it assumes that IIA holds. In the next section I briefly describe IIA and also discuss whether it applies for the model I estimate or not.

Independence from irrelevant alternatives

IIA means that for any two alternatives the ratio of probabilities is fixed regardless of the other alternatives. As Train (2002) highlights, Luce (1959) started the derivation of logit model from stating IIA and treating it as a desired property. Nevertheless, in several situations IIA assumption is unrealistic and makes the results unreliable. A famous example when IIA does not hold considers the probabilities of passengers choosing between two means of transport: red bus or car. If a passenger is indifferent between the two then the ratio of the probabilities is 1.

In this case the probability of taking the red bus is 50% just as the probability of going by car. Now, if we imagine that a new mean of transport, a blue bus is introduced, then the probabilities according to IIA change in an implausible manner. If we assume that the traveler is indifferent between the two buses (as they are similar in all properties except the color), then we can write:

$$\frac{P_{red\ bus}}{P_{blue\ bus}} = \frac{P_{red\ bus}}{P_{car}} = \frac{P_{blue\ bus}}{P_{car}} = 1.$$

The equation holds only if the probability of going by car is 33%, the same as the probability of taking the red bus or taking the blue bus. The decrease in the probability of going by car is unrealistic as we know that the two buses are close substitutes. We can expect that by introducing the blue bus, the probability of going by car remains at 50%, while the probability of choosing either one of the two buses is also 50%.

Although IIA seems to be fairly restrictive in the situation above, I argue that it does not constrain considerably my estimations for three reasons. First of all, the varying choice sets that I apply keep only the alternatives in the model which are relevant for a given consumer. Going forward, the demographic characteristics that I use in the model control for the main substitution patterns, and makes the model less prone to the violation of IIA. Nevertheless, one can still make up a situation in which IIA is violated in this model. However, due to the fact that I do not want to use this model to forecast such market outcomes when a bank enters or leaves a market, I think that the distortion that may be present in the model is limited.

Estimation of interest rates

In this section I focus on one alternative specific characteristic, the interest rate. Due to the fact that there is no data exist about the interest rates that banks would have offered to the consumers if they asked for a loan, I estimated these theoretical interest rates.

I estimated OLS regressions on all banks individually to capture the differences among banks' pricing behavior. The dependent variable was the level of the interest rate, and the list of explanatory variables was the same for all the banks. The models that I estimated can be written in the following form:

$$interest_{ij} = x_i'\beta_j + \epsilon_{ij},$$

where $interest_{ij}$ is the interest rate that consumer i had on the mortgage loan from bank j, x_i is the vector of explanatory variables including 1 for the constant term, and ϵ_{ij} is the shock term.

In the regressions I used three types of variables. The first type consist variables which capture the risk regarding non-performance. These are district dummys and wage. The second type of variables is related to the characteristics of the loan. The type of the transaction is included by a dummy variable. The base category is the housing loan, while the second category is home equity loan. There is also a third category of loans that has mortgage as collateral but they are neither categorized as housing nor as home equity loans. These different types of loans have different risk characteristics that is captured by the regressions. The term of the loan also reflects risk, as the longer the loan is outstanding the higher the risk of repayment keeping everything else constant. The last type of variables is technical. Some banks grant loans through different entities and there is a high difference between the conditions of these loans. This difference is captured by introducing dummys for the entities[10]. A time dummy variable is also introduced.

Figure 5: Results of regressions on interest rate

VARIABLES	Bank A	Bank B	Bank C	Bank D	Bank E	Bank F	Bank G
2.trans_type	0.143***	0.151	0.0940***	-0.0562	0.311***	1.529***	0.128***
	(0.0281)	(0.156)	(0.0282)	(0.0854)	(0.0197)	(0.0152)	(0.0363)
3.trans_type	0.433***						-0.612***
	(0.0532)						(0.0397)
wage	-5.45e-08**	-7.10e-08***	-3.58e-08*	-5.91e-08**	-3.08e-07***	3.64e-09	-7.34e-08**
	(2.57e-08)	(2.54e-08)	(2.11e-08)	(2.96e-08)	(3.83e-08)	(2.28e-08)	(3.32e-08)
2.district	3.551***		4.320***	3.355***	3.066***	4.203***	3.735***
	(0.0789)		(0.0733)	(0.129)	(0.0426)	(0.00616)	(0.0631)
3.district	-2.042***	-1.603***					
	(0.0971)	(0.212)					
4.district	2.223***		2.428***	1.933***	1.966***	2.668***	2.120***
	(0.0372)		(0.0251)	(0.0367)	(0.0174)	(0.00657)	(0.0362)
5.district	0.998***		1.281***	0.595***	0.957***	1.404***	1.089***
	(0.0304)		(0.0113)	(0.0520)	(0.0118)	(0.0238)	(0.0465)
value	-3.01e-10	-4.72e-10***	-1.15e-09	-2.63e-10	-4.93e-09***	-8.88e-10*	-9.34e-11
	(2.65e-10)	(1.40e-10)	(7.10e-10)	(4.05e-10)	(6.40e-10)	(4.99e-10)	(3.72e-10)
term	-0.00206	0.0172***	-0.000122	0.0128***	0.000808	-0.000300	-0.00293
	(0.00171)	(0.00228)	(0.000948)	(0.00436)	(0.000909)	(0.000406)	(0.00229)
Constant	4.363***	3.177***	4.125***	4.368***	4.513***	5.331***	4.491***
	(0.0553)	(0.220)	(0.0223)	(0.0950)	(0.0227)	(0.0223)	(0.0901)
date dummy	✓	✓	✓	✓	✓	✓	✓
entity dummy	✓	✓	✓	✓	✓	✓	✓
Observations	1073	669	4562	457	2330	9248	768
R-squared	0.916	0.317	0.888	0.896	0.898	0.977	0.903
RMSE	0.355	0.326	0.368	0.275	0.277	0.235	0.407

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The signs and magnitudes of the coefficients are intuitive in every case when the variable is significant. In all cases except one, the explanatory power of the models are high (R squared around 90%), while the mean square root of the errors is reasonably low (between 24 and 41 basis points).

[10] I do not disclose directly that which bank grant loans through special entities in order to avoid the possibility of identification of the banks. I only mark that entity dummy is used for all banks, this dummy takes the value of one if loans are granted through special entities as well.

Wage is always negative when significant. Higher wage leads to lower risk of non-performance, therefore the interest rate can be lower. District dummys have always the same sign at every bank and they are always significant. They probably capture the unobserved risk characteristics of the consumers and serves as a proxy for risk. The coefficient on term is positive as interest rate is expected to be positively correlated with the term of the loan due to higher risk of non-performance and higher price of liquidity for longer term loans. The positive coefficient may also reflect the shape of the yield curve.

Despite the fact that the explanatory power of the model is high, and the results are intuitive, some concerns regarding the out of sample performance of the model still remains. To account for these I apply two robustness checks. In one case I make out of sample estimations and compare the results to the observed data. These results are disclosed in the section of Robustness checks. The second check is actually built in the final results. Due to the fact that I should adjust the error terms of the final conditional logit model to account for the estimation error of the interest rates I apply bootstrapping. During this process I reestimate the interest rate models and carry out new predictions on the bootstrap samples and use these estimations for the final model. This method also captures the uncertainty of the interest rate models by estimating them on subsamples.

Finally, self-selection bias can be also a concern. It may be the case that clients self-select into a bank, and they differ considerably from other clients. As I estimate the interest rate model on the sample of the self-selected clients the assumption regarding the randomness of the sample may be violated. However, I argue that this issue is not relevant when I estimate the theoretical interest rate that banks would have set for the following reason. In reality banks use models to set interest rates which are also based on the sample of their clients, consequently they are also exposed to self-selection bias. It is true that these models may be more sophisticated than the model that I use, as banks have more detailed information on the clients. Nevertheless, the high explanatory power of the models that I estimated and the small mean square root of the errors show that my models perform reasonably well. Self-selection is not an issue, as banks would set interest rates in a similar way. Moreover, I also rule out "unlikely clients", who would probably not show up in a given bank, when I apply geographical and financial constraints. The process of ruling out clients is detailed in the next section.

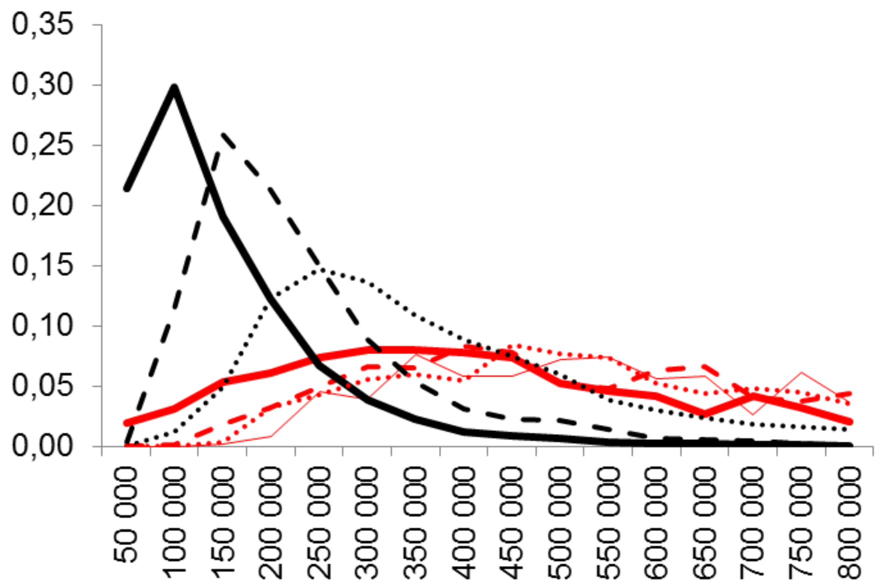
Restrictions on the choice set

I restricted the choice sets of consumers by two methods. Firstly, I assumed that consumers take loans only from banks which are present in their district. This assumption is reasonable for consumers who would like to take a loan in order to buy a property in their district. It is a red herring for banks if a consumer neither lives in the district of the branch where he asks for the loan nor wants to buy the real estate in the branch's district. In this case there is no clear reason of asking for the loan away from the home district. There is a risk that this consumer goes to a branch far away from his town because he would like to hide some information that is known about him in the town where he lives, but not known in a town far away. However, the assumption may be too strong for consumers who migrate from one town to another and asks for a loan in the new place. Banks are happy to serve these clients. Nevertheless, more than 87 per cent of the consumers in the dataset took a loan from a bank which is present in the district where the consumer lives[11]. This high ratio supports the restriction of the choice set based on districts.

[11] Budapest, the capital of Hungary was considered as one district. The clients who took loan from a bank out of their district were not included in the final estimation.

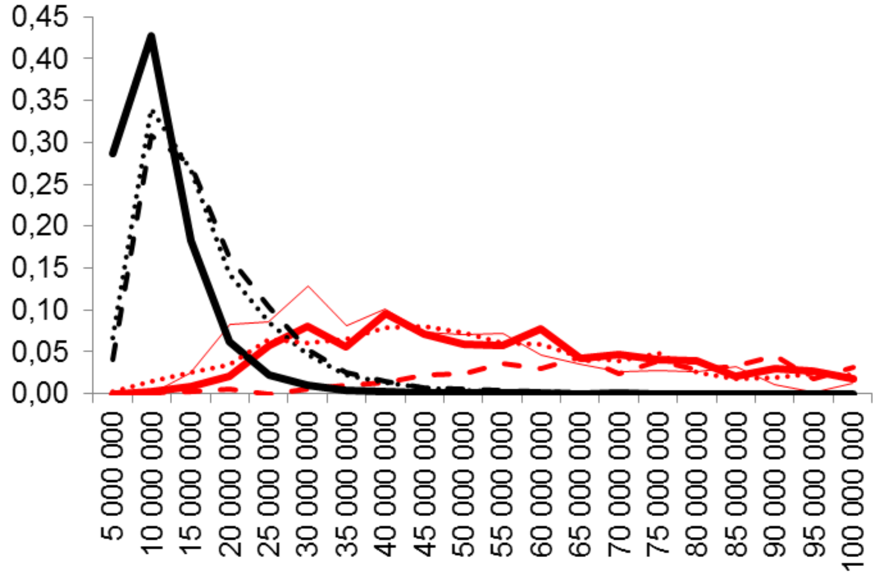
The second group of restrictions is based on the observed patterns in the data. The distribution of clients is fairly heterogeneous among banks concerning the wage, the size of the loans and the value of the property. It is well known that banks target different clientele, consequently they are keen to give credit to a certain group of consumers and does not want to serve other groups of consumers. These strategies can be observed in the data nicely. There is strong heterogeneity among the distributions of banks' customers regarding wage, value of the property and size of the loan. One group of banks serves the clients with limited financial opportunities, while there is another group of banks which focuses on the affluent segment. The figures below show these distributions.

Figure 6: Distribution of customers' wage by banks



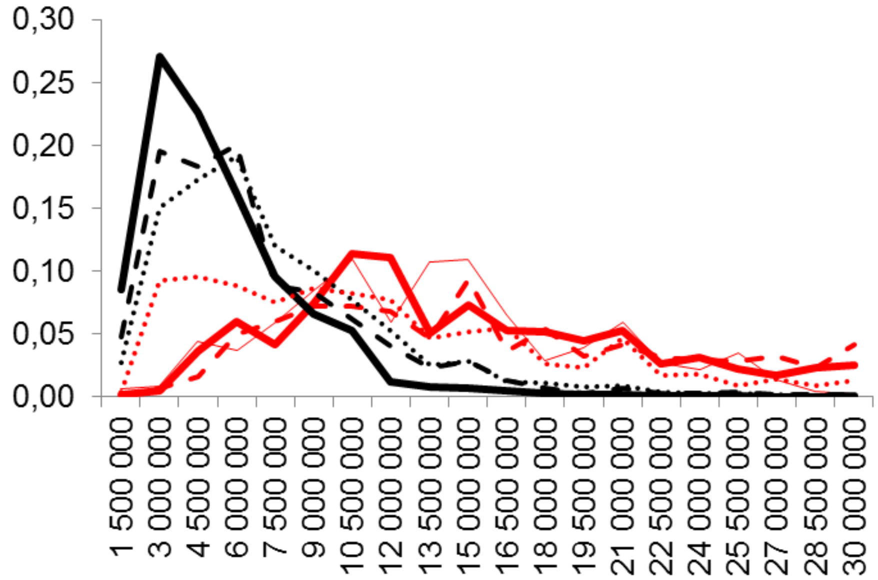
Source: Hungarian Credit Registry, own calculation

Figure 7: Distribution of customers' value of property by banks



Source: Hungarian Credit Registry, own calculation

Figure 8: Distribution of loan size by banks



Source: Hungarian Credit Registry, own calculation

Similarly to Figure 3, the three charts above show strong heterogeneity among banks. According to the charts above three banks focus on the mass segment while the others concentrate on more wealthy customers. On all the three charts the black lines are right skewed, that represents that loans of these banks in 2015 were mainly granted to individuals who have low wage, want to buy a cheap flat and ask for a low amount of credit for the purchase. The opposite is true for the other banks. The strong difference among the characteristics of the clients of different banks supports the idea that banks serves only a well-defined clientele.

I use these distributions to estimate the limits under banks not granting loans. Ideally, I would check the minimum of wage, value and loan size, however the number of observations decreasing considerably at the tails of the distributions. If one considers the minimum of wages as a stochastic variable, then the question arises how to estimate it. Due to the low number of observations at the minimum I decided to use a biased estimator which is the 5^{th} percentile. Although the estimator is clearly biased, it is still more reliable due to the sufficient number of observations. As a robustness check I also estimated the model using the 1^{st} percentile as the limit.

I applied the following rule for inclusion in the choice set. If a consumer's wage is lower than the 5^{th} percentile of the distribution of wages of a given bank, then I excluded this bank from the consumers choice set. Similarly, if the size of the loan or the value of the property is under the limit of the 5^{th} percentile I excluded the bank from the choice set. I argue that the observed loans under these limits were probably granted due to special characteristics of the consumer that I cannot observe. For instance a consumer may not have a legal salary above the limit, however the bank who observes the account of the consumer every day, knows that a stable amount arrives to the account in every month and it brings the consumer over the limit. Special offers tailored to loyal, own customers can also result in that banks accept the credit application, although some parameters are under the limit that are applied to consumers who are not the clients of the given bank.

I also excluded banks which are not present in the district of the consumer. By applying these rules I excluded almost the two thirds of the options. Namely, an average consumer could choose from 7 banks if I did not apply any restrictions, while this number decreased to less than 3 when all the rules were set, and there were several consumers who remained with only one alternative[12].

In the final model I used the estimated interest rates as alternative specific explanatory variables. Hence, the estimated probabilities are:

$$P_{ij} = \frac{e^{\hat{x}'_{ij}\beta + D'_j \gamma s_i}}{\sum_k e^{\hat{x}'_{ik}\beta + D'_j \gamma s_i}},$$

where the variables are the same as in the previous equations except that $\hat{x}'_{ij}$ contains the estimated interest rate instead of the true interest rates.

[12] Consumers who remained with only one alternative were not included in the regressions.

Chapter 6 – Results and policy implications

Models with full choice set

I estimated the conditional logit model by maximum likelihood and used Stata's built-in routine to do it[13]. Due to the uncertainty of the error term of the estimated theoretical interest rates I applied bootstrapping with 100 replications to get the final results. I specified eight different setups to pin down the effect of introducing taste variables and choice set restrictions on the coefficient of interest rate. The results below show both the coefficients on the alternative specific variables (interest rate, bank branch number – the number of branches of a given bank in the district and history) and on the individual specific variables which are interacted with bank dummys. Consequently, the interpretations of the two types of variables differ. Alternative specific variables show the change in the logodds of choosing a given bank if the variable changes by one unit keeping everything else constant. The interacted individual specific variables show the increase in the relative probability of choosing one alternative over another if the variable increases keeping everything else constant. I kept Bank F as the base for comparison in all models.

In the baseline model I included only the interest rate and the number of branches of the selected bank in the district of the consumer (Model: Full choice set, no taste). I did not control on the choice set of individuals. The estimated coefficient of the interest rate is positive in this specification. This result points to endogeneity in the model. The estimated coefficients are unreliable in this case. The main reason for it is that the structural model behind the estimation suggests that the higher the interest rates the higher the utility of the consumer, or put it differently consumers like high interest rates, that is an implausible result.

[13] I applied the asclogit command.

Handling endogeneity

To solve the endogeneity issue first I controlled on consumer characteristics and also included bank fixed effects. Introducing taste variables (age, wage) brought the coefficient on interest rate into the negative territory. The reason for this is likely that consumers have some banks that they like and choose only among them. In this restricted decision interest rate matters, however, taste matters more than interest rate if we consider all the available choices. Introducing bank fixed effects rules out most of the endogeneity concerns as the unobserved bank specific characteristics are captured by the bank dummy. For instance, the concern that some banks may have strong brand and based on that they are able to set higher interest rates should not make the researcher worried. This effect is captured by the bank dummy and pulled out from the error term. Consequently, error term is not correlated with interest rate due to the brand power. The changed sign of the interest rate also supports the idea that most of the endogeneity issues are ruled out. The coefficient on branch number is not significant in this specification.

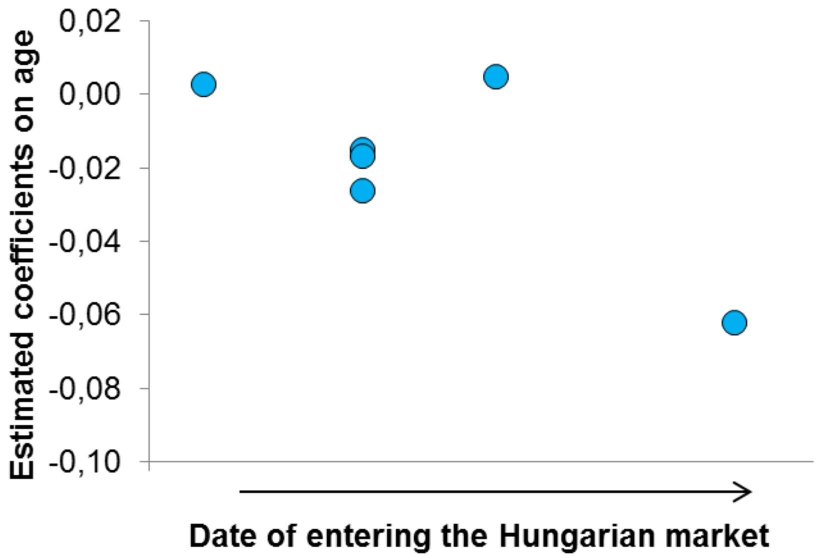
Figure 9: Results of the Conditional logit regressions

BANKS	VARIABLES	Full choice set				Restricted choice set			
		No Taste	Taste	No Taste	Taste	No Taste	Taste	No Taste	Taste
	interest	0.171***	-1.262***	-0.0176	-1.182***	-0.862***	-1.640***	-1.042***	-1.539***
		(0.0222)	(0.124)	(0.0325)	(0.135)	(0.0749)	(0.166)	(0.0987)	(0.185)
	branch number	0.0221***	0.000881	0.0136***	0.00213**	0.0181***	0.00762***	0.00843***	0.00971***
		(0.000617)	(0.000707)	(0.000854)	(0.000959)	(0.00133)	(0.000707)	(0.00156)	(0.00114)
	history			3.037***	2.750***			2.502***	2.750***
				(0.0237)	(0.0240)			(0.0422)	(0.0357)
Bank A	age		0.00311		0.00563		0.00956		0.0167**
			(0.00377)		(0.00405)		(0.00724)		(0.00803)
	wage		1.142***		1.213***		0.401***		0.430***
			(0.0432)		(0.0461)		(0.0373)		(0.0445)
	constant		-6.447***		-6.015***		-2.614***		-1.945***
			(0.182)		(0.183)		(0.340)		(0.374)
Bank B	age		-0.0266***		-0.0115**		-0.0153		-0.00750
			(0.00526)		(0.00499)		(0.0129)		(0.0113)
	wage		1.326***		1.332***		0.428***		0.425***
			(0.0389)		(0.0439)		(0.0372)		(0.0453)
	constant		-10.22***		-9.233***		-3.936***		-2.947***
			(0.374)		(0.385)		(0.648)		(0.619)
Bank C	age		-0.0155***		-0.00626**		-0.0149***		0.00206
			(0.00266)		(0.00289)		(0.00265)		(0.00355)
	wage		0.979***		1.032***		0.439***		0.441***
			(0.0379)		(0.0423)		(0.0289)		(0.0349)
	constant		-3.337***		-2.895***		-1.705***		-1.195***
			(0.117)		(0.134)		(0.159)		(0.184)
Bank D	age		-0.0623***		-0.0511***		-0.0747***		-0.0632***
			(0.00552)		(0.00624)		(0.00920)		(0.0105)
	wage		1.270***		1.289***		0.523***		0.515***
			(0.0408)		(0.0442)		(0.0303)		(0.0375)
	constant		-5.198***		-4.584***		-0.590*		0.283
			(0.214)		(0.237)		(0.356)		(0.428)
Bank E	age		-0.0165***		-0.00558*		-0.0180***		-0.00227
			(0.00286)		(0.00299)		(0.00328)		(0.00405)
	wage		0.480***		0.540***		0.0768**		0.146***
			(0.0321)		(0.0394)		(0.0300)		(0.0344)
	constant		-2.626***		-2.676***		-1.362***		-1.509***
			(0.0856)		(0.111)		(0.123)		(0.143)
Bank G	age		0.00544		0.0136***		0.00448		0.0129*
			(0.00486)		(0.00447)		(0.00628)		(0.00707)
	wage		1.245***		1.275***		0.504***		0.502***
			(0.0411)		(0.0454)		(0.0352)		(0.0412)
	constant		-7.052***		-6.388***		-2.897***		-1.948***
			(0.213)		(0.192)		(0.379)		(0.405)

Detection of taste patterns

It is an interesting result that the age of the consumer is related to the age of the bank in the model. Namely, older people prefer banks that are present on the Hungarian market for a long time[14]. The intuition behind this is that relatively old people know these banks for longer time, consequently there is a higher probability that they have used some services of these. In the same time younger consumers prefer younger banks.

Figure 10: Estimated coefficients on age and presence of banks on the Hungarian market



Source: own calculation

Coefficients on wage are also highly significant. This variable may captures the effect that banks offer different services and target different segments of consumers. The banks that concentrate on the affluent segment are likely to offer services which are tailored to wealthy

[14] Due to the sensitivity of the data I do not write down the age of the banks explicitly as they could be identified after that.

customers, and as a result, wealthy consumers may prefer these banks. Similarly, less wealthy consumers may find suitable services at other banks which focus on the mass segment.

The role of clients' history

I also ran two similar specifications with the distinction that I controlled on previous connection with banks by introducing the variable history. This specification with no taste variables results in not significant coefficient on interest rate, while the history variable is strongly significant. It highlights the fact that consumers value previous connections. This result can have several roots. Just to mention a few, consumers know institutions that they have had connection with better. They may trust more in a bank which they know, and they also save searching costs by not looking for another alternative. The relevance of the variable may also reflect the role of previous decisions. It can also account for a self-encouraging process – looking for another bank would question the optimality of previous decision. This contradiction can be dissolved by not looking around on the market and keep transacting with the original institution. Finding the exact reason is not a goal of this research, here I just stress that previous connections matter. Running a regression that includes history and taste variables as well results in similar coefficients than the specification without history, however branch number now is significant. It means that consumers prefer banks with more branches in their district. The reason of this can be that it is comfortable to get credit from a bank that is easily accessible.

Models with restricted choice set

In the second set of models I introduced restrictions on the choice sets. Interestingly, the only change of narrowing the choice sets leads to negative coefficient on interest rate. This result is consistent with price sensitive consumers whose opportunities of choosing among banks are restricted. To put it differently, consumers choose banks according to their possibilities, and they do care about interest rates at this decision.

If I introduce taste variables to the restricted choice set model, interest rate becomes even more negative. The reason again is that beside restricted possibilities, consumers choose according to their tastes related to banks. They focus on a special set of banks which set is related to demographic characteristics, and take interest rate into consideration when they choose among these banks. The attraction of the offers of other banks is partially offset by the fact that they are not in the preferred set. This mechanism gives monopolistic power to banks which they can use to set higher interest rates than the one related to perfect competition. The results are similar if the variable history is introduced into the models. The fact that history is highly significant suggests that previous connections strongly matter for consumers.

Exogenous factors behind the model

The results above are based on the assumption that endogeneity is ruled out from the model and the remaining variation in interest rates is only due to exogenous factors after controlling for population characteristics. A relevant exogenous factor that is likely to drive differences among interest rates is the changing market structures. Although in a frictionless world banks react perfectly to changing market conditions by opening/closing branches, these actions result in substantial costs. In the observed period in Hungary banks were concentrating on the rationalization of their branch networks, consequently frictions are related to the costs of closure of branches. The main elements of the costs are the penalties that banks should pay due to terminating the contracts of renting the building and the compensation to employees who are layed off from the closing branch. To avoid these costs banks often decide to wait for the termination of contract and may account for fluctuation of the employees. As a result branches are not closed exactly in line with the business cycle but frictions are present. It can be also the case that banks expect improving economic conditions in the future. In this situation it might be rational to keep a loss making branch open if the expected gains in the future exceed the expected losses that the branch is likely to generate in the near future. The too frequent establishment and closure of branches also leads to reputational loss that banks want to avoid.

Conclusion of the results

To sum up the results I found that consumers do care about interest rate, however their own taste and their opportunities also matter. If only interest rate and one or two explanatory variables are included in the models estimates are not plausible as the coefficient on interest is not significantly negative. Controlling on taste variables beside interest rate only, results in downward sloping demand curve. Interestingly, applying restrictions on the choice sets of consumers and not controlling for tastes results in downward sloping demand curve also. This result suggests that limitations of consumer opportunities are important. Finally, previous connection to banks is important in the choice, as it appears strongly significantly in every regression.

Policy implications

I describe in this section three potential policy implications of the model. Firstly, the optimization of the branch network and the altering of the business model potentially affect interest rates, henceforth they are also relevant for the better understanding of the interest rate channel of the transmission mechanism. Secondly, different taste patterns of consumers should be taken into account when one analyzes the impact of changing supply and demand.

Lastly, the model can be used to strategy planning of banks and to supervisory tasks during business model analysis.

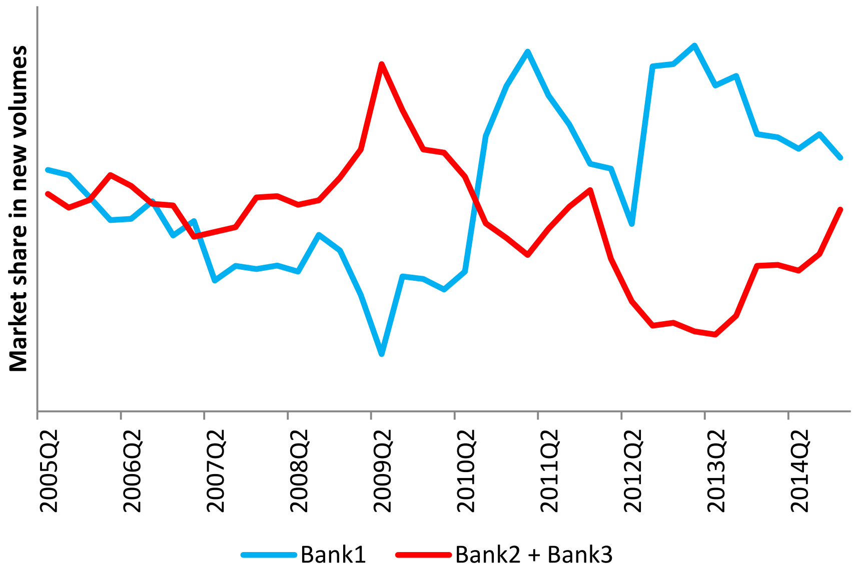
The model helps to understand the impact of branch network rationalization and business model changes of banks on mortgage loan interest rates. The rationalization of branch networks can be a key element of bank’s cost reduction. Lower costs may lead to lower margin that can result in lower interest rates. However, branch closures also reduce the choice sets of clients who may be able to choose from more expensive offers only. As a result the aggregate level of mortgage loans’ interest rate may increase. The change of the market level interest rate depends on the balance of the lower costs of banks and narrowing choice sets of consumers. A consequence is that the change of branch networks may have an effect on the interest rate channel of the transmission mechanism. The model described in this paper helps to estimate the impact of changing choice sets. In addition to the impact of branch network rationalization, choice sets also capture the effect of changing business models (through financial constraints of consumers). Nevertheless, it should be noted that for estimation of the impact of heavy changes, the development of this model to a nested logit one would be more appropriate due to the possible issues that the lack of IIA can cause.

In addition to geographic and financial constraints the model shows that taste patterns are relevant in clients’ decisions. Consequently, the changing supply of banks has different effects on consumers with different tastes. It is also important that the impact of changing demand can be understood better by analyzing the tastes of clients with changing demand. For instance, if the demand for mortgage loans of high-income and middle-aged clients increases, then mainly those banks feel the increasing demand, which are preferred according to the taste of high-income, middle-aged clients. The taste patterns of different groups of clients are captured by the model.

Finally, the model can help supervisors and bank strategists in their work as the structure of the market can be understood more deeply by the results. Own and cross price elasticities can be calculated from the model and these can be used for instance to detect the main groups of competitors or to calculate ceteris paribus effect of the change of different parameters. One application can be seen in Figure 11. The picture shows the market shares of a bank (Bank1) and the sum of market shares of two other banks (Bank2 + Bank3) which were detected by having high cross price elasticities with the first bank[15]. The total market defined here as the sum of the eleven big banks in Hungary. It can be clearly seen that the market share of Bank1 changed just in the opposite way as the market share of Bank2 + Bank3. This phenomenon indicates that Bank2 and Bank3 are the competitors of Bank1, just as the cross price elasticities shows. It is also interesting that the period on which the model was estimated is only the year 2015, while the competing relationship between these banks can be observed through a whole decade before 2015. This finding supports that the results of the model captures the true properties of the market which are stable through time.

[15] Due to the sensitive nature of the data I do not disclose that which banks' market shares are shown on the chart. I removed the numbers from axis y for the same reason.

Figure 11: Market share of three banks among the seven banks included in the model



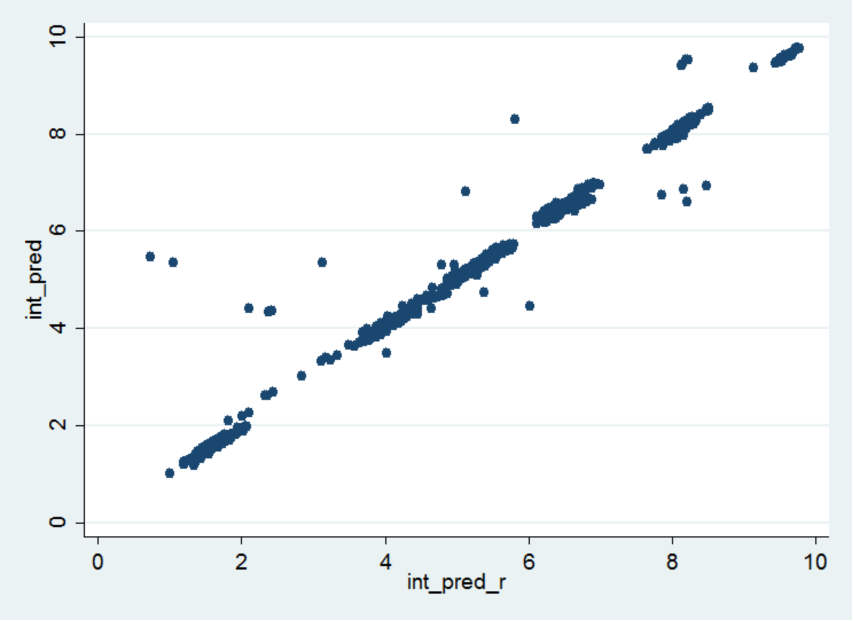
Source: MNB, own calculation

CHAPTER 7 – Robustness checks

Interest rate

Although the interest rates seem to be well explained by the individual bank models, it is not clear whether the out of sample performance of the models are also convincing. As a robustness check, I cut the sample and estimate the models again on one part and run out of sample estimations on the other part of the sample. To keep the random fashion of subsampling I mark one third of the sample randomly by including observations which identifier gives zero if it is divided by three. This subsample serves as the estimation sample. The remaining two-third of the original sample becomes the testing sample. I estimate the OLS regressions on the estimation sample and make predictions on the testing sample. Figure 12 shows the out of sample estimations (x axis, int_pred_r) and the true data (y axis, int_pred). If the dots on the scatterplot lies on the 45 degree line, then the out of sample estimations are the same as the observed data. Although there are a couple of observations away from the 45 degree line, the majority of the dots narrowly scattered around the line. This result suggests that the out of sample performance of the model is good.

Figure 12: Evaluation of out of sample prediciotns of interest rates



Source: own calculation

Choice set restrictions

Here, I examine whether change of the arbitrarily chosen cutoff of 5th percentile of the wage, property value and loan size has material impact on the results. I do this by estimating the final regression (Restricted choice set with taste variables) with the 1st percentile as a cutoff. I apply the same estimation technique as in the original regressions (maximum likelihood, using 100 replications of bootstrap samples to adjust the standard errors). The results of this regression are encouraging, as the coefficients do not change considerably. The coefficients on interest rate, on the number of branches and on the history variable are quite close to each other and lie within the two standard deviation circle of the original regression. The coefficients on the interacted taste variables are also close to their counterparts in the original regression. All in all, based on this exercise I argue that the bias that comes from the fact that I estimate the minimum of the distributions by the 5th percentile does not materially impact my results.

Figure 13: Results of the model applying the 1st percentile as minimum of financials

BANKS	VARIABLES	Restricted choice set Taste
	interest	-1.318*** (0.188)
	branch number	0.00933*** (0.000990)
	history	2.745*** (0.0294)
Bank A	age	0.0126** (0.00622)
	wage	0.648*** (0.0461)
	constant	-3.262*** (0.302)
Bank B	age	-0.0110 (0.00826)
	wage	0.714*** (0.0549)
	constant	-5.496*** (0.600)
Bank C	age	-0.00227 (0.00323)
	wage	0.611*** (0.0386)
	constant	-1.614*** (0.169)
Bank D	age	-0.0557*** (0.00959)
	wage	0.722*** (0.0384)
	constant	-1.309*** (0.360)
Bank E	age	-0.00462 (0.00378)
	wage	0.215*** (0.0369)
	constant	-1.575*** (0.124)
Bank G	age	0.0223*** (0.00540)
	wage	0.740*** (0.0427)
	constant	-4.194*** (0.300)

Conclusion

In this research I estimated demand on the Hungarian mortgage loan market by applying a discrete choice model on Hungarian credit registry data. By this model I could prove the hypothesis that interest rate matters for the Hungarian consumers. I also showed that beside interest rates there are several other factors which are important. One factor is the restricted choice sets of consumers due to their financial position and their location. Another important factor is the taste of the consumers. There are clear patterns on the market of consumer tastes. Beside these, the relevance of previous connection with banks is also detected. The importance of tastes and history in consumers' decisions together with geographic constraints may help banks to act according to monopolistic competition.

Beside the results that are summarized above there are at least three potential policy implications that should be mentioned. The change of branch network and business model affect interest rates, therefore it should be taken into account when one analyzes the interest rate channel of the transmission mechanism. Different taste patterns are detected by the model, and these can alter the effects of the changes of demand and supply. The model can be also used to strategy planning of banks and to supervisory tasks during business model analysis. Further researches may focus on developing this model into a nested logit model in order to relax some assumptions that were applied in this paper.

Reference List

Berry, S. (1994). Estimating discrete choice models of product differentiation. *RAND Journal of Economics*, 242-262.

Daniel McFadden, K. T. (1978b). An application of diagnostic tests for the independence from irrelevant alternatives property of the multinomial logit model. *Transportation Research Record*, 39-46.

Dick, A. (2002). *Demand Estimation and Cosumer Welfare in the Banking Industry.* Finance and Economics Discussion Series Paper 2002-58, Board of Governors of the Federal Reserve System.

Follain, J. R. (1990). Mortgage Choice. *AREUEA Journal*, 125-144.

Holló, D. (2010). Estimating price elasticities on the Hungarian consumer lending and deposit markets: demand effects and their possible consequences. *Oesterreichische Nationalbank Focus on European Economic Integration*, 73-89.

Luce, D. (1959). *Individual Choice Behavior.* New York: John Wiley and Sons.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*.

McFadden, D. (1974). Measurement of Urban Travel Demand. *Journal of Public Economics*, 303-328.

McFadden, D. (1978). Modeling the choice of residential location. In A.Karlqvist, L. Lundqvist, F.Snickars, & J. Weibull, *Spatial Interaction Theory and Planning Models* (pp. 75-96).

McFadden, D. (1987). Regression based specification tests for the multinomial logit model. *Journal of Econometrics*, 63-82.

McFadden, D. (2001). Economic Choices. *America Economic Review*, 351-378.

McFadden, D., Talvitie, A., Cosslett, S., Hasan, I., Johnson, M., Reid, F. A., & Train, K. (1977). *Demand model estimation and validation.* University of California, Berkeley.

Molnár, J., Nagy, M., & Horváth, C. (2007). A structural Empirical Analysis of Retail Banking Competition: the Case of Hungary. *MNB Working Papers*.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica*, 307-342.

Phillips, R., & Yezer, A. (1996). Self-Selection and Tests for Bias and Risk in Mortgage Lending: Can You Price the Mortgage If You Don’t Know the Process? *Journal of Real Estate Research*, 87-102.

Rachlis, M. B., & Yezer, A. (1993). Serious Flaws in Statistical Tests for Discrimination in Mortgage Markets. *Journal of Housing Research*, 315-336.

Steven Berry, J. A., & Pakes, A. (2004). Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market. *Journal of Political Economy*, 68-105.

Steven Berry, J. L. (1995). Automobile Prices in market equilibrium. *Econometrica*, 841-889.

Train, K. (2002). *Discrete Choice Methods with Simulation.* University of California, Berkeley.