Charles University in Prague

Faculty of Social Sciences
Institute of Economic Studies

BACHELOR THESIS

Household Debt in the Czech Republic: Focus on Mortgage Amount Determinants

Author: **Patricie Mittigová**

Supervisor: **doc. Ing. Tomáš Cahlík, CSc.**

Academic Year: **2015/2016**

Declaration of Authorship

The author hereby declares that she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain a different or the same degree.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, May 10, 2016

Signature

Acknowledgments

First and foremost, I would like to thank doc. Ing. Tomáš Cahlík, CSc., my supervisor, for his guidance and invaluable advice during writing of this thesis. I would like to express my gratitude to ČSOB, especially to Ing. Ondřej Brodecký, for enabling me to conduct this study. Finally, I am grateful to my family for their support.

Abstract

The growth of household indebtedness in the last decade led to more attention paid to this issue since it could threaten the stability of financial system in the future. In the Czech Republic, this rise is mainly caused by the increased number of mortgage loans, which are usually the largest financial burden that debtors have to repay. For that reason, I focus on mortgages since their growth has been the most significant in comparison with other loan types. The aim of this thesis is to analyse how personal characteristics affect one's decision on the amount they choose to borrow in order to finance their housing needs. For this purpose, I performed a multiple regression analysis applying several estimation methods. By examining a random sample of ČSOB clients who have taken out a mortgage, I discovered that person's income, age, marital status, education and region a person lives in are significant factors affecting debtor's choice about how much they borrow. Conversely, the number of children affects the amount borrowed only at higher quantiles of the distribution, whereas it is not a significant factor at lower quantiles. This thesis complements rather a limited number of studies about the Czech household debt as it provides new findings about a borrower's decision.

JEL Classification	D12, D14, G21
Keywords	mortgage loans, personal characteristics, household debt, Czech Republic

Author's e-mail	mit.patricie@gmail.com
Supervisor's e-mail	cahlik@fsv.cuni.cz

Abstrakt

Růst zadluženosti domácností v posledním desetiletí vedl k větší pozornosti věnované tomuto problému, protože by v budoucnosti mohl ohrozit stabilitu finančního systému. V České republice byl tento růst způsoben převážně zvýšeným počtem hypotečních úvěrů, které jsou obvykle největší finanční zátěží, kterou dlužníci musí splatit. Z tohoto důvodu se soustředím na hypotéky, protože jejich nárůst byl v porovnání s ostatními typy půjček největší. Cílem této práce je analyzovat, jak osobní charakteristika ovlivňuje, kolik peněz si člověk rozhodne půjčit, aby mohl financovat své bydlení. Za tímto účelem jsem použila mnohonásobnou regresní analýzu a aplikovala několik metod odhadu. Zkoumáním náhodného vzorku klientů ČSOB, kteří si vzali hypotéku, jsem zjistila, že plat, věk, rodinný stav, vzdělání a kraj, ve kterém osoba žije, jsou signifikantními faktory ovlivňující dlužníkovu volbu o výši půjčené částky. Na druhou stranu počet dětí ovlivňuje výši půjčky jen ve vyšších kvantilech distribuce, zatímco v nižších kvantilech to není signifikantní faktor. Tato práce doplňuje poměrně limitované množství studií o dluhu českých domácností, protože poskytuje nové poznatky o tom, jak se lidé v případě půjček rozhodují.

Klasifikace JEL	D12, D14, G21
Klíčová slova	hypoteční úvěry, osobní charakteristika, dluh domácností, Česká republika

E-mail autora	`mit.patricie@gmail.com`
E-mail vedoucího práce	`cahlik@fsv.cuni.cz`

Contents

List of Tables

List of Figures

Acronyms

CZK Czech koruna

ČSOB Československá obchodní banka, a.s.

LAD Least Absolute Deviations

OLS Ordinary Least Squares

Bachelor Thesis Proposal

Author	Patricie Mittigová
Supervisor	doc. Ing. Tomáš Cahlík, CSc.
Proposed topic	Household Debt in the Czech Republic: Focus on Mortgage Amount Determinants

Motivation Household debt has been rising in many countries which led to more attention paid to this issue as its sustainability can be problematic in the future. In the Czech Republic, the increase in household debt is mainly the result of growing number of mortgage loans. Since loans for housing needs are considerably higher compared to consumer credits, they usually account for the highest amount a person has to repay. Therefore, I focus on mortgage loans since their growth has been the most significant relative to other loan types. The aim of this thesis is to analyse how personal characteristics influence the amount people decide to borrow in order to buy or construct a house. For this purpose, a random sample of ČSOB clients who have taken out a mortgage is examined in order to determine which characteristics have a significant effect on the amount borrowed.

Hypotheses

#1: People with higher income borrow higher amounts.

#2: Men borrow higher amounts than women.

#3: A person's age is a significant factor influencing the amount borrowed.

#4: People who have obtained a university degree borrow higher amounts compared to people with primary or secondary education.

#5: People living in Prague borrow higher amounts compared to people living outside Prague.

Methodology In order to examine the effect of particular personal characteristics, a multiple regression analysis is performed. Several regression models are constructed and estimated by OLS or LAD. Furthermore, quantile regression is performed so that it is possible to determine whether the estimated effects vary across the distribution.

Outline

1. Introduction
2. Literature Review
3. Data Description
4. Econometric Analysis
5. Discussion of Results
6. Conclusion

Core Bibliography

1. BIČÁKOVÁ, Alena; PRELCOVÁ, Zuzana; PAŠALIČOVÁ, Renata. Who Borrows and Who May Not Repay?. CERGE-EI Working Paper Series, 2011, 443.
2. CAMERON, A. Colin; TRIVEDI, Pravin K. Microeconometrics: methods and applications. Cambridge university press, 2005.
3. COULIBALY, Brahima; LI, Geng. Choice of mortgage contracts: evidence from the survey of consumer finances. Real Estate Economics, 2009, 37.4: 659-673.
4. FARINHA, Luísa. The effect of demographic and socioeconomic factors on households' indebtedness. Banco de Portugal Economic Bulletin, 2003.
5. HULLGREN, Maria; SÖDERBERG, Inga-Lill. The relationship between consumer characteristics and mortgage preferences: a case study from Sweden. International Journal of Housing Markets and Analysis, 2013, 6.2: 209-230.
6. KEMPSON, Elaine; MCKAY, Stephen; WILLITTS, Maxine. Characteristics of families in debt and the nature of indebtedness. Corporate Document Services, 2004.
7. WOOLDRIDGE, Jeffrey. Introductory econometrics: A modern approach. Nelson Education, 2015.

Chapter 1

Introduction

Over the past decade, many European countries have experienced an increase in household debt. This issue may be problematic as the excessively high indebtedness of households can threaten the stability of financial system. Therefore, household debt has been examined both from the macroeconomic point of view and at the household level. The first approach assesses the causes of the rising indebtedness or its effects on the economy and proposes possible policies in order to sustain or reduce the current level of debt. Chmelar [1] examined the connection between household debt and the last European recession. On the other hand, analysis of the debt at the household level can help to understand which groups are more likely to be indebted (Kempson et al. [2]). In the Czech Republic, household debt has increased from 522,8 billion CZK in December 2005 to 1393,9 billion CZK in February 2016, of which approximately 70% consists of mortgages. They are most likely the main attribute of the increasing household debt as the amount owed for consumer credits or other loans has remained relatively stable. Mortgage loans have increased from 282,6 billion CZK in 2005 to 981,6 billion CZK in 2016 which is more than three times higher than the initial amount [3]. Since loans for house purchases are usually the largest financial burden households have to repay and also the largest part of the total household debt, this thesis focuses particularly on mortgage loans in the Czech Republic.

The main aim of this thesis is to examine personal characteristics which influence the fact how high mortgage amounts people borrow. This approach is different from the previous studies since the variation in amounts borrowed is examined. On the other hand, researchers have examined more frequently the probability of taking out a new loan, having difficulties with repayment or

defaulting a loan based on the household characteristics (such as a number of members, total income, age of household head etc.). Furthermore, the analysis is usually performed on the total household debt without any distinguishing what the money is owed for. As for the household debt in the Czech Republic, Bičáková et al. [4] provides the most detailed analysis of the Czech household debt. In the thesis, I examine a random sample of ČSOB clients who have taken out a mortgage by performing a multiple regression analysis. The individual factors influencing the amount of money a person decides to borrow are examined so that it is possible to assess the effect of specific personal characteristics of each client. I perform OLS estimation, LAD estimation and quantile regression analysis in order to study the sample in detail.

The thesis is structured as follows: in Chapter 2, the summary of relevant literature connected with the household debt is provided with main focus on mortgages as they are the major loan type examined in the thesis. Chapter 3 describes the data that are analysed. Firstly, the general information about the sample is provided and followed by the detailed descriptive statistics of each variable included in the dataset obtained from ČSOB. The summary statistics of each variable are supplemented with general features that can be observed. The econometric analysis is described in Chapter 4. At first, the examined variables are presented. This is followed by the list of hypotheses to be tested. The following section of Chapter 4 discusses applied methods and reasons why a particular method was chosen. In this chapter, 5 models are estimated applying various methods depending on the form of a regression model or a sample size. Since the sample includes a few observations with extreme values, the applied estimation methods have to be chosen so that the results are plausible. In Chapter 5, I comment on the results of all models and provide possible explanation for each estimated effect. Chapter 6 concludes the thesis.

Chapter 2

Literature review

In this chapter, I present the findings of other authors' studies which are related to the examined topic. Since the aim of this thesis is to analyse the effect of personal characteristics on the amount people borrow so that they are able to finance their housing needs, I describe studies related to mortgage loans. Since there has not been, to the author's knowledge, any research that examines how borrower's characteristics influence one's decision on the amount borrowed, I include additional studies which focus on the connection between personal characteristics and household debt in general as they are similar to the analysis presented in this thesis. Finally, I comment on the research examining the situation in the Czech Republic.

2.1 Mortgage Debt Growth

Many countries have experienced the growth of mortgage debt and thus, there are studies aiming at explanation of the increasing indebtedness. Wolswijk [5] analysed 15 European countries by applying pooled regressions and discovered that financial deregulation measures, stock market growth and an increase in house prices positively influenced mortgage lending. On the other hand, after-tax interest rate has a negative effect. Additionally, he considered various fiscal measures which could be used for balancing housing market. The evolution of mortgage debt in Canada is examined by Fortin and Leclerc [6]. They analysed annual time-series data and discovered that the most significant factors explaining the growth of mortgage indebtedness were housing prices and three exogenous variables: nominal interest rate on mortgage, real per-capita income and inflation rate. Crawford and Faruqui [7] points out that not only mortgage

loans, but also consumer credits led to the growth of the Canadian household debt. Georgarakos et al. [8] assessed attitude people living in 12 European countries had towards mortgage debt. They examined which factors led to financial distress among mortgage debtors and concluded that lower education, health problem or unemployment resulted in financial difficulties in most countries.

2.2 Mortgage Default Rate

Another problem that is connected with mortgage loans is default rate. The situation in the United States was discussed by Mayer et al. [9]. They commented on the rise in mortgage defaults and proposed possible explanations. Contrary to previous opinions that the growth of defaults is caused by mortgage features, they concluded that the problem arose when borrowers with the lowest credit scores were provided with mortgage. Furthermore, when house prices declined, debtors had less incentive to pay off the debt. Quercia and Stegman [10] provided a review of the literature analysing mortgage default from the early studies.

2.3 Discrimination in Mortgage Market

Discrimination in mortgage lending market has been examined by many authors. Ladd [11] found the evidence of discrimination against minorities by mortgage lenders. Similarly, Bayer et al. [12] discovered differences in the likelihood of obtaining mortgage depending on the race or ethnicity as the result of analysing 7 large metropolitan markets between 2004 and 2008. Apart from placing a person at a disadvantage due to their race, different types of discrimination can be examined. Dietrich and Johannsson [13] modified the approaches, that were previously used for detection racial discrimination, and applied them for testing the presence of age and gender discrimination. They did not find any evidence of the presence of this effect in mortgage lending.

2.4 Arrears in Repaying Mortgages

Since mortgage loans are usually very high, it is possible that some people might have a problem with repaying the amount borrowed. Magri [14] used

Eu-Silc data for Spain, Finland, France, Ireland, Italy, the Netherlands and Great Britain in order to analyse the frequency of households which are in arrears in repaying their mortgages. Moreover, they examined the effect of various household characteristics on the probability of being in arrears estimating a probit model. They discovered that increasing age of household head resulted in lower probability. The same applies to education, income and working time. A higher number of household members increases the probability as well. The issue of mortgage payment problems is also discussed by May and Tudela [15] who focused only on the case of Great Britain, but they conducted a more detailed analysis compared to Magri [14]. Their results suggest that facing financial difficulties in the past increases the probability of having problem servicing the debt in the present. Unemployment appears to be the most significant factor leading to mortgage payment problems jointly with interest income. The most detailed analysis of British household indebtedness and the risk of being in arrears is provided by Kempson et al. [2], but they do not consider only mortgage debt, but total household debt. They found that younger people were more likely to be in arrears. Having children and higher number of credit commitments also increase the risk of being in arrears. The fact that lower income leads to higher likelihood of having financial difficulties while repaying is in line with the previously mentioned studies. Additionally, Nettleton and Burrows [16] examined the effect of being in mortgage arrears on debtor's health. They discovered that mortgage indebtedness worsened person's well-being. Additionally, men who face difficulties with repayment tend to visit their general practitioners more frequently. The effect of the overall household debt on debtor's health in Germany was examined by Keese and Schmitz [17]. They discovered that household indebtedness strongly affected person's mental health.

2.5 Mortgage Choice Determinants

As the topic of this thesis is to examine decisions people make on the amount they borrow, this section presents studies related to mortgage choice determinants. Coulibaly and Li [18] analysed data from the Survey of Consumer Finances in the United States in order to determine which factors influence a borrower's choice between fixed- and adjustable-rate mortgages in order to revisit previous studies. They estimated a logit model analysing not only borrower's characteristics, but also relative pricing and other terms of a mortgage

contract. They showed that risk-averse people tended to prefer fixed-rate mortgages. Furlong et al. [19] discovered that low-credit borrowers seemed to have higher propensity to choose adjustable-rate mortgage. Furthermore, they examined the effect of house price appreciation on a borrower's choice which mortgage type to take out. Additionally, Vickery [20] found the connection between mortgage choice and movements in retail interest rate. In Australia, determinants of mortgage choice were discussed by Dungey [21] as she examined the effect of global financial crisis on people's decisions which mortgage product to choose analysing mortgage applications that originated between January 2003 and May 2009. The author used a logit model in order to predict probabilities of preferring certain mortgage type for different groups of applicants. She discovered that the recent crisis resulted in a significant change in borrower's preferences towards mortgage type. Moreover, individual characteristics are also an important determinant. In another research paper, Dungey et al. [22] assessed not only the effect of borrower's characteristics on mortgage choice, but also the role of regulatory capital requirements. The situation in Sweden was examined by Hullgren [23] who analysed data collected by questionnaires in Stockholm area. The results show that high income households tend to take out adjustable-rate mortgages. Additionally, people seem to be influenced by their bank advisers while choosing mortgage type. Hullgren and Söderberg [24] also analysed differences between men and women and discovered that men's choices are influenced by age, low level of education and risk averseness, whereas women's decisions are affected by income, low financial literacy and increase in interest rate.

2.6 Personal Characteristics and Household Debt

Bover et al. [25] examined the distribution of household debt across euro area countries and factors that influenced it. As for secured debt, they discovered that age, income and education level of household members had a significant effect, but their importance varied across selected countries. Farinha [26] analysed the situation in Portugal using data from 1994 and 2000 in order to assess the effect of gender, age, education, work status, income, marital status and the number of persons living in the household on the probability of holding debt and also on the outstanding amount of debt. The results suggest that higher income and higher education both lead to higher probability of holding debt. Additionally, these factors also increase the outstanding debt amount.

2.7 Czech Household Debt

Although household indebtedness in the Czech Republic has significantly increased since 2005, there are not many research papers which would focus on this phenomenon and aim at its analysis. Czech Statistical Office [27] issued a short article regarding the growth and development of the debt, but it did not provide any econometric analysis of this problem. On the other hand, Bičáková et al. [4] conducted the most detailed analysis of household debt in the Czech Republic. Firstly, they examined the development of household credit market from 2000 to 2008 and claimed that the share of households having a loan had not significantly changed over this period, whereas the amount of debt outstanding had raised. The amount increased particularly for younger households as a result of increased mortgage lending. Secondly, they used another dataset, the Statistics on Income and Living Conditions for 2005–2008, which included household-level micro data. They estimated probit and tobit models in order to determine the probability of having a loan, the probability of taking out a new loan and the amount of a new loan. They discovered that higher income, higher number of members in a household and presence of children increased the probability of having a loan. On the other hand, higher age and longer duration of marriage have the opposite effect. The probability also decreases with higher education of household members. They discovered that these variables had the same effect on the likelihood of taking out a new loan and the amount borrowed. Finally, they proved that there was considerable variation in borrowing across regions in the Czech Republic which is significantly connected with economic conditions in the given region.

The contribution of this thesis is explaining the variation in mortgage amount by analysing factors that influence the amount borrowed in a way that has not been applied yet. As the availability of micro data containing information about household financial situation is limited, I examine a dataset which has different features compared to data analysed in the previous work and therefore this thesis could possibly discover new phenomena in a borrower's choice.

Chapter 3

Data Description

In this chapter, I comment on the origin of the examined dataset and present a detailed description of its structure in order to determine general features and relations in the sample.

3.1 Origin of Data

Since the main objective of this thesis is to analyse factors affecting the amount of money a person borrows, it was essential to work with a sample of people which contains personal data. For this purpose, ČSOB provided me with a random sample of its clients who have taken out a mortgage. The dataset was created in September 2015.

The anonymised data include information about 2,000 persons who have in total 2,149 mortgage loans. Clients who have more than 1 mortgage loan were also kept in the sample so that the largest possible dataset could be analysed. Furthermore, 70 observations which did not include information about marital status or region a person lives in were deleted. Since only 3.3% observations were removed, this approach should not significantly influence the results. The final examined dataset contains 2,078 entries. Since there are some extremely high values which could distort the results, the size of the sample is not constant for all regression models, as it depends on the robustness of a particular method used for the analysis.

3.2 Descriptive Statistics

In this section, I provide description of all variables which are included in the dataset. Data can be divided into two categories. The first category consists of variables which directly relate to a mortgage contract, whilst variables in the second category provide personal information about a client. Moreover, some variables had to be modified so that they have a proper form which could be examined.

Since data from ČSOB are analysed, I used additional comments in the dataset explaining meaning of each variable which were written by ČSOB employee and information about ČSOB products on its website [28] in order to describe mortgage-related variables that are in the sample.

3.2.1 Mortgage

ČSOB provided two sums of money which stand for mortgage loan. These are initially arranged mortgage amount and actual withdrawn amount. In most cases, the differences between them are negligible or they do not differ at all. Nevertheless, I decided to use the actual withdrawn amount in the models. I removed 1 observation, because the value of actual withdrawn amount was missing. The range of values is relatively wide. The lowest amount borrowed is only 199,599 CZK in comparison with the highest amount which is 20,823,484 CZK. I added the median value into Table 3.1, because it may better express how much people actually borrow as it is not distorted by extreme values. The median value of 1,500,000 CZK suggests that people borrow rather lower amounts. This can also be seen in Figure 3.1 as high amounts are very rare.

Variable	Mean	Median	Std. Dev.	Min	Max
actual	1,860,425	1,500,000	1,559,037	199,599	20,823,484
arranged	1,822,263	1,500,000	1,512,913	200,000	23,255,151

Table 3.1: Actual withdrawn amount and arranged amount (CZK)

3.2.2 Type

People can choose between two mortgage types which are provided by ČSOB and this type depends on what the loan is meant for. The more frequent type is a classic mortgage used for financing housing needs, which is secured by real estate. On the other hand, the second type called American mortgage allows

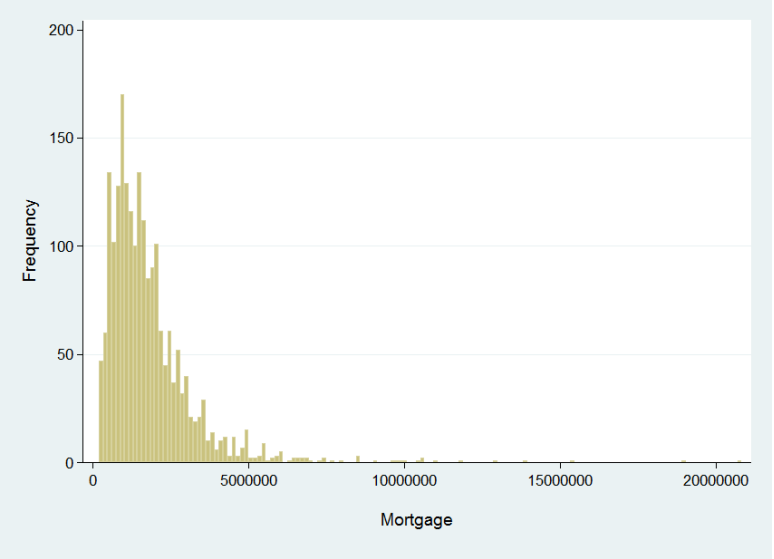


Figure 3.1: Distribution of mortgages

clients to buy anything they want, but the loan has to be secured by real estate as well. It offers lower interest rate than ordinary loan. Only 56 people in the sample have taken out American mortgage, while remaining 2,022 people have a classic mortgage. Since American mortgage is not meant for financing housing, its mean is much lower than the mean of a classic mortgage. This is summarized in Table 3.2.

Another division can be made based on the loan amount to real estate value ratio. The standard ratio is 70%, but in case of a classic mortgage it is possible to have even higher - 85% or 100%. In the sample, there are 1,021 mortgages with standard ratio, 826 clients have ratio equal to 85% and 231 clients borrowed amount up to 100% of the real estate value. I do not differentiate between these types in the analysis and use all observations irrespective of the type.

Variable	Mean	Std. Dev.	Min	Max	N
american	784,797	523,092	200,000	3,200,200	56
classic	1,850,996	1,521,260	199,599	20,823,484	2,022

Table 3.2: American and classic mortgage (CZK)

3.2.3 Interest rate

While considering whether to take out a mortgage or not, people usually regard interest rate as an important determinant. As a result of low interest

rates, mortgages are becoming more affordable to wider range of people [29]. This might be one of the factors responsible for the growth of household indebtedness. In the dataset, current interest rate and average interest rate, which is computed from the beginning of repayment period up to now, can be compared. Table 3.3 shows that there are small differences between these two interest rates.

Variable	Mean	Std. Dev.	Min	Max
average	3.921631	1.166988	1.28	8.98
current	3.410342	1.122128	1.16	9.59

Table 3.3: Average and current interest rate

3.2.4 Repayment

Monthly payments and repayment period are closely related. Thus, these variables are described together. The dataset directly includes information about payments, but the repayment period had to be computed based on the date when a person terminates mortgage withdrawals and the date when the last payment is supposed to be paid.

These variables show how burdensome mortgage can be. Not only does a person have to pay a relatively high sum each month, but also the repayment period can be very long. Thus, debt to income ratio was generated in order to discover percentage of income that is used for repayment. Since some repayment amounts are higher than income, it is probable that these clients are not paying off the loan by themselves, but for instance with a spouse or a partner. Furthermore, the data include information about outstanding debt, which is yet to be repaid. The results are summarized in Table 3.4.

Variable	Mean	Std. Dev.	Min	Max
repayment	9,390	7,818	382	118,180
period	23.280	7.066	5.005	42.312
debt	1,545,176	1,382,225	2490	20,823,484
DTI ratio	0.347	0.277	0.00065	2.47650

Table 3.4: Repayment amount (CZK), repayment period in years, outstanding debt (CZK), and debt to income ratio

3.2.5 Date and Fixation

The dataset includes various important dates related to the loan. The first variable states the date of execution. After signing the contract, clients are allowed to draw the funds. They can decide how they want to draw the mortgages so that it satisfies their needs. The duration of drawing is known, because the date of the first drawing and the date of the last drawing are provided. After terminating mortgage withdrawals, clients commence repayment of the loan.

Another two dates are connected with mortgage repayment. The first one is the date when the next monthly repayment is supposed to be paid. The second one tells the date of the very last repayment. Due to these dates, it is possible to compute the duration of repayment period. Additional variables show information about fixation. During this period, annual interest rate does not change. Both the length of fixation and the date, when annual interest rate will stop to be fixed, are known.

In the following section, I describe variables which characterize each client in the sample.

3.2.6 Gender

One of the most important factors affecting the sum of money borrowed I would like to examine is gender. Figure 3.2 shows that it is much more likely that a person who borrows in order to finance housing needs is male.

The ratio male to female is approximately 7:3. One possible explanation for this could be that in case of married couples husband is usually the one who takes out a mortgage. It is not likely that this disproportion could be the result of gender discrimination in Czech mortgage market. Furthermore, not only do male clients account for 70% of the sample, but they also borrow higher amounts in comparison with female clients. Maybe this is because men earn higher salaries in comparison with women and therefore men can afford to pay higher instalments. This fact applies to the examined sample as well, but specific values about salary are provided in section 3.2.11 Income. Details about mortgage loan amounts depending on gender are summarized in Table 3.5.

Variable	Mean	Std. Dev.	Min	Max
female	1,665,411	1,333,244	199,599	15,379,058
male	1,886,837	1,576,855	200,00	20,823,484

Table 3.5: Mortgage amount by gender (CZK)

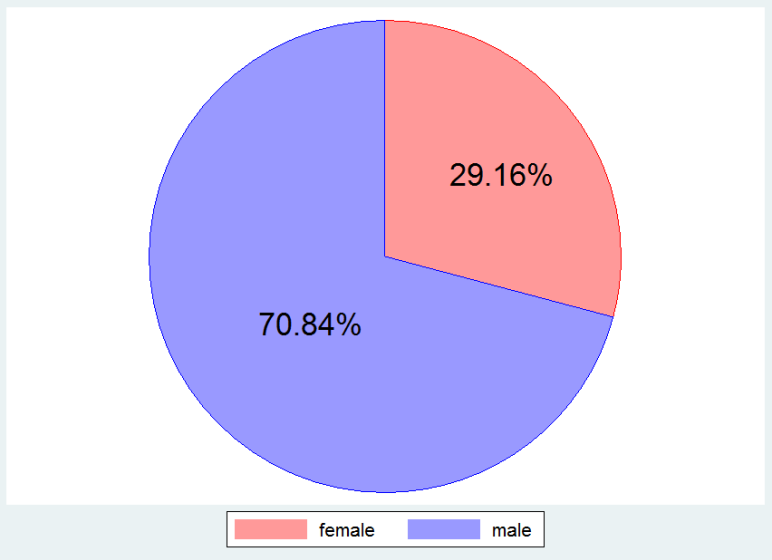


Figure 3.2: Gender distribution

3.2.7 Age

The following variable deals with a person's age. It provides the information about a client's current age. In order to use this variable in the analysis, it had to be modified. Since the dataset includes the date of execution, I was able to compute a person's age at the time of taking out a mortgage, which would help me assess the effect of age on the amount of money borrowed. From now on, I work with the modified variable in the analysis.

I am interested not only in the age effect, but also in the age distribution in the sample, because it may reveal some interesting facts and patterns. The summary statistics are provided in Table 3.6 and the distribution can be seen in Figure 3.3.

The youngest person in the sample is only 18 years old which is the lowest possible age for applying for a loan. The oldest person is 71 years old. These values are very rare as in both cases only 1 person of that extreme age is present in the dataset. The mean age equals 35 years and the median is 34 years. People

aged 30–40 account for 1018 observations in the sample. This suggests that borrowers are probably waiting to take out such a high loan until they have a permanent job and a stable income so that they do not face any financial difficulties while repaying the mortgage. Another possible reason for this could be the connection between house purchases and starting a family, because they usually occur almost simultaneously. As the average age at which couples start a family increased in the past years, this change could possibly influence their decisions on taking out a mortgage and subsequently buying a house. The mean age of women at childbirth in the Czech Republic in 2013 was 29.9 years, which confirms that women have their first child at higher age [30].

Figure 3.6 shows that the distribution is slightly skewed to the right. The coefficient of skewness equals 0.687. This feature indicates that young people borrow more frequently in comparison with elderly people. This might be because the repayment period is usually long and elderly person would not manage to repay the debt.

Variable	Mean	Std. Dev.	Min	Max
age	35.07796	8.22743	18	71

Table 3.6: Age at the time of taking out a mortgage

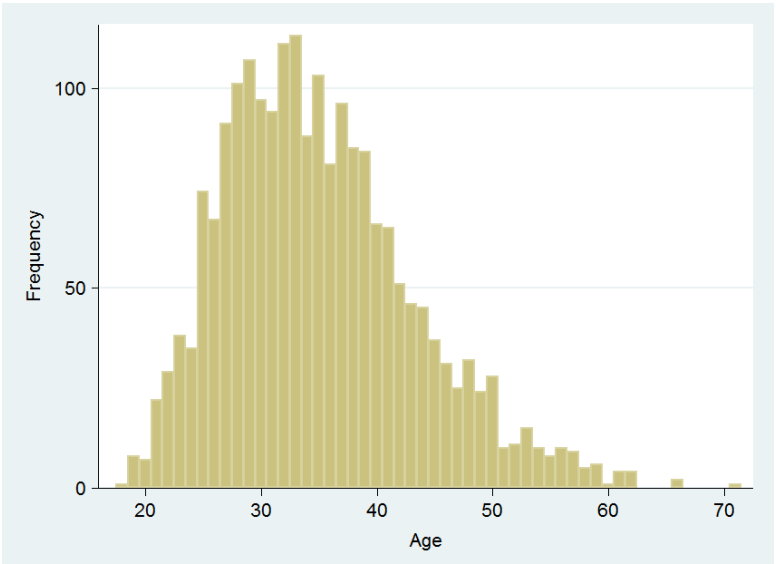


Figure 3.3: Age distribution

3.2.8 Marital Status

This variable tells the information whether a client is married, single or divorced. There are few categories that are very rare. These are widowed, in registered partnership or living with a partner. In order not to have too many variables in the analysis, the least frequently cases are covered in the main categories stated above. I decided to do this so that even more observations are not deleted. In the sample, there are 9 widowed persons who I regarded as single. The same applies for 4 people having a partner, since people who described themselves as single can also have one, but they did not state so. Only 1 person in the sample is in registered partnership who I regarded as married. The description of loan amounts borrowed by main categories is summarized in Table 3.7.

Married people have the highest amount borrowed of all groups on average. This might be because they can use two incomes for repayment. On the other hand, single and divorced people probably have to pay off the mortgage on their own which restricts the loan amount they can afford. Furthermore, divorced clients constitute the smallest proportion and the amount they borrow is the lowest as well. One possible explanation for this could be related to age, since the mean age of divorced people is 40 years and therefore they do not borrow such high amounts.

Variable	Mean	Std. Dev.	Min	Max	N
married	1,952,957	1,566,422	199,599	18,993,469	1108
single	1,729,231	1,514,288	200,000	20,823,484	698
divorced	1,528,616	1,201,393	200,000	9,950,000	272

Table 3.7: Marital status and amount borrowed (CZK)

3.2.9 Number of Children

All observations about the number of children a person has lie between 0 and 9. The mean value equals 0.679 and the median value is even lower - 0 children. Although it is usual to have rather fewer children at this time, the values are lower than expected. The distribution is shown in Figure 3.4. As in the previous cases, the distribution is skewed to the right. The coefficient of skewness equals 10.08.

More than a half of the sample consists of childless persons. One possible reason might be that these people take out a mortgage before starting a family.

The number could possibly be influenced also by a client's age, since elderly people were starting a family at the time when having more than just one child was more common. The effect of number of descendants on loan amount that a parent borrows will be examined in the econometric analysis. The average amount of money borrowed for each group by number of children is summarized in Table 3.8.

Variable	Mean	Std. Dev.	Min	Max	N
childless	1,810,885	1,551,301	200,000	20,823,484	1,266
1 child	1,756,421	1,119,084	199,599	8,500,000	515
2 children	2,151,026	2,161,080	250,00	18,993,469	127
3 children	1,595,273	1,033,404	250,000	6,700,000	98
4 children	2,000,573	1,470,822	480,000	6,700,000	29
5 children	2,033,349	1,354,590	350,000	6,390,000	33
6 children	2,594,741	2,783,737	300,000	7,263,189	7
7 children	10,623,393		10,623,393	10,623,393	1
8 children					0
9 children	3,050,000	3,464,823	600,000	5,500,000	2

Table 3.8: Number of children and amount borrowed (CZK)

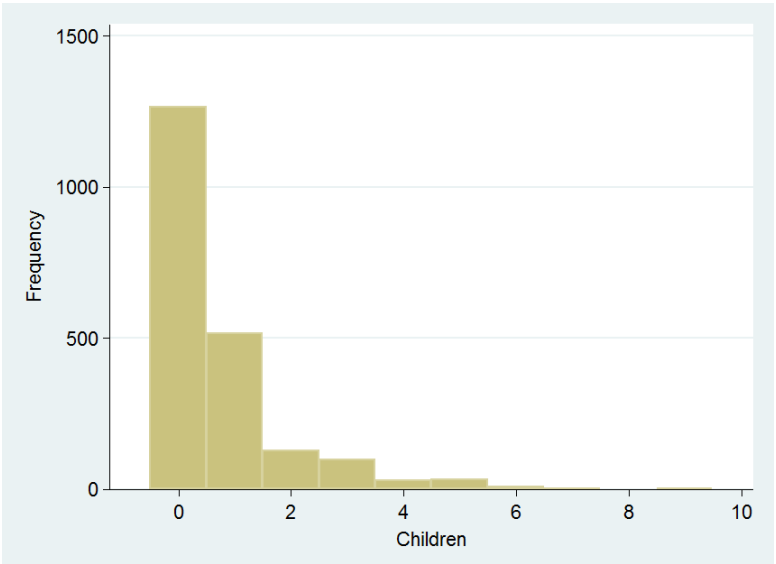


Figure 3.4: Distribution of number of children

3.2.10 University Degree

The data include information whether a client obtained bachelor's degree, master's degree or none. In the analysis, I do not distinguish between degree types. I consider person to have a university degree if they obtained either bachelor's or master's degree, therefore variable where both possibilities are combined without differentiation was generated. In the original dataset, only 36 people have bachelor's degree, while 513 people obtained master's degree. Because of the small number of clients with bachelor's degree, it is better to create only one category in order to avoid having too many variables with few observations. The difference between amounts borrowed by people with university degree and without degree is summarized in Table 3.9. The results show that more than one fourth of the sample has a university degree. Furthermore, they also borrow higher loan amounts.

Variable	Mean	Std. Dev.	Min	Max	N
degree	2,176,388	1,617,533	238,000	18,993,469	549
no degree	1,695,112	1,453,173	199,599	20,823,484	1,529

Table 3.9: University degree and amount borrowed (CZK)

3.2.11 Income

Monthly income is a crucial factor I would like to examine. ČSOB provided me with estimated income of each client. The lowest income is 8,000 CZK, while the highest income equals 1,556,508 CZK. This seems to be extremely high value for someone who needs to borrow money. It might be due to an error. Another explanation could be connected with the method which was used to determine a person's income. For instance, if the client did some profitable business in the past months and the income was estimated as the average value of money earned during this period, it would be feasible to have such a high salary. Nevertheless, it still does not explain why this client borrowed almost 1,800,000 CZK which is relatively close to his monthly income.

The connection between amount borrowed and income will be examined in the next chapter. Therefore, only values connected directly with income are presented. I include not only the mean value, but also the median value as the mean could be affected by extremely high salaries. Additionally, I include these values separately for male and female clients in order to show the difference

between them. This is summarized in Table 3.10. The distribution is shown in Figure 3.5.

The summary table shows that there are quite considerable differences between the mean and the median. The values show that women earn less money in comparison with men. It might be one of the reasons why they are found in the sample less frequently. Based on the income distribution, it is evident that loans are taken out by people with rather lower income as distribution is considerably skewed to the right.

Variable	Mean	Median	Std. Dev.	Min	Max
income	40,992.6	27,650	70,531.69	8,000	1,556,508
female	32,921.29	24,166	33,147.02	8,000	461,500
male	44,315.43	29,287.5	80,835.34	8,000	1,556,508

Table 3.10: Income (CZK)

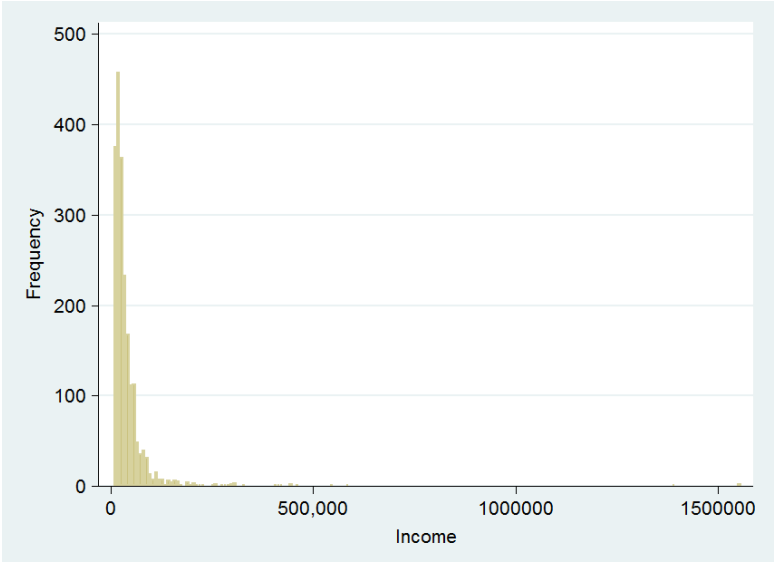


Figure 3.5: Income distribution

3.2.12 Region

Initially, I was interested in the distribution of indebtedness across regions of the Czech Republic. Since such data are not available, I use distribution of people who have taken out a mortgage by region they live in. The examined dataset includes information about postcode for each person. Thus, it was

possible to determine the region where a client lives. Some people do not live in the Czech Republic and therefore they were removed from the sample.

Table 3.11 summarizes the number of people living in each of 14 Czech regions and how much money they borrow. People living in Prague borrow the highest amounts. They also form more than 20% of the sample which is shown in Figure 3.6.

Variable	Mean	Std. Dev.	Min	Max	N
Prague	2,700,197	2,288,715	200,000	20,823,484	445
Central Bohemian Region	1,791,027	1,261,911	250,000	13,924,384	211
South Bohemian Region	1,504,371	976,930.1	250,000	5,000,000	115
Plzeň Region	1,553,488	1,237,972	260,100	9,700,000	112
Karlovy Vary Region	1,308,233	898,900.8	320,000	4,200,000	72
Ústí nad Labem Region	1,465,681	997,450	199,599	6,700,000	123
Liberec Region	1,552,823	1,131,499	200,000	7,000,000	89
Hradec Králové Region	1,707,620	1,463,090	259,000	10,000,000	132
Pardubice Region	1,412,873	732,499.4	200,000	3,600,000	113
Olomouc Region	1,461,447	976,777.7	238,000	6,700,000	139
Moravian-Silesian Region	1,690,093	1,030,006	300,000	5,500,000	157
South Moravian Region	1,712,422	1,029,153	231,172	5,500,000	198
Zlín Region	1,495,288	1,177,448	310,000	7,263,189	112
Vysočina Region	1,457,068	969,126.9	200,000	6,000,000	60

Table 3.11: Distribution of mortgage amount by regions

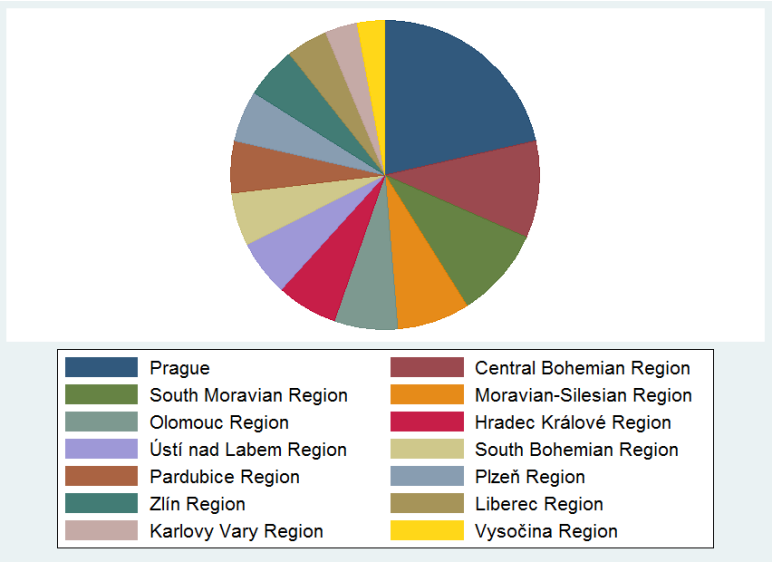


Figure 3.6: Distribution of clients by region

Chapter 4

Econometric Analysis

In this chapter, I provide the analysis about how personal characteristics influence one's decision on the amount they choose to borrow so that they are able to finance their housing needs. In order to determine the effect each dependent variable has on the amount of borrowed money, I use a multiple regression analysis. I apply various estimation methods in order to assess how the amount borrowed changes depending on individual characteristics of each client in the sample. It is very likely that the loan amount is influenced by other factors as well. These might be for instance interest rate or monthly repayment amount. Nevertheless, these variables are not included in the models since I would like to examine only the effect of client's characteristics and not factors that are directly connected with the mortgage. It is possible to leave out mortgage-connected variables as they are probably not correlated with the variables included in the analysis and thus, zero conditional mean assumption should not be violated. Therefore, only variables which are described in the last part of Chapter 3 are used.

As can be seen in the previous chapter, there are few observations that attain extremely high values compared to the rest of the dataset. Even though they are very rare, they could negatively impact the models. Therefore, methods or models which are supposed to prevent this from happening are chosen. Statistical software used for the analysis is Stata 13.0.[1] All tests are performed at 0.05 significance level.

[1] `http://www.stata.com/stata13/`

4.1 Variables

As already mentioned, I analyse only the impact of personal characteristics on the amount borrowed. Some variables expressing this information about each client were modified so that they could be included in the models. The list of analysed variables, their form and necessary modifications is described below.

The main objective is to explain variation in the amount of money borrowed. Thus, the dependent variable in the analysis is the mortgage amount which a client decided to take out. This variable is labelled as *mortgage* and is expressed in CZK.

Independent variables used in the models are each client's personal characteristics so that it is possible to determine how they affect the dependent variable. These are *gender, age, marital status, number of children, university degree, income* and *region*. Variable *gender* was transformed into a dummy which attains only values 0 or 1. Since male persons account for 70% of the sample, male gender is the base group used for comparison. Therefore, the transformed dummy variable is labelled as *female*. The same applies for *region*, but I generated 13 dummy variables for each region except for Prague because it is the base region due to the highest number of people living in the capital city who have taken out a mortgage. They are labelled as follows: *cb, sb, pl, kv, ul, li, hk, pa, ol, ms, sm, zl* and *vy*.

Another variable that needed to be transformed is *marital status*. People were divided into 3 categories according to the division in the previous chapter. Therefore, 2 dummies called *divorced* and *single* are generated, while married people are regarded as the base group the comparison is made with. The last dummy variable that was generated is *university degree*. It attains 1 when a person obtained a university degree and 0 otherwise.

The remaining variables do not need to be modified, because their form has already been suitable for the analysis without any changes. These are *age* expressed in years, *number of children* and *income* in CZK.

4.2 Hypotheses

In this section, I present hypotheses that will be tested based on the results of the regression models. Furthermore, expected outcomes are presented jointly with the possible explanation.

Hypothesis 1: *People with higher income borrow higher amounts.*

I expect that the higher income people have, the higher amount they borrow. Since mortgage loan usually accounts for the largest part of person's debt, its repayment may lead to financial difficulties if a debtor has to use a large part of their income for repayment. Since the amount of monthly instalment is determined by the amount borrowed, it is probable that people with lower income will borrow lower amount so that they are able to repay the debt without difficulties. On the other hand, rich people can afford paying higher monthly payments. Since people with higher salary tend to buy more expensive goods including a house, they might borrow more money in order to buy a more luxurious dwelling. I assume that income is the most influential factor people take into consideration while taking out a mortgage.

Hypothesis 2: *Men borrow higher amounts than women.*

Male persons are expected to borrow higher amounts compared to female persons. Since I assume that income is the main determinant of mortgage amount, men are supposed to borrow higher amounts as they tend to earn more money in comparison with women.

Hypothesis 3: *A person's age is a significant factor influencing the amount borrowed.*

I expect that a person's age is an important factor that influences how much money a person borrows. Younger people do not have such financial means for repayment, therefore they cannot afford to borrow high amount of money. On the other hand, as people get older they are more likely to earn more money and therefore they can borrow more. I expect that this positive effect of age eventually becomes negative since elderly people after retirement usually earn less money compared to their salary while working. Another reason that older people borrow lower amounts could be that they adjust the repayment period so that they manage to pay off the whole debt.

Hypothesis 4: *People who have obtained a university degree borrow higher amounts compared to people with primary or secondary education.*

I assume that people who have obtained either bachelor's or master's degree are supposed to have higher salaries compared to people who have only primary or secondary education. As already mentioned, I expect that higher income leads to higher amount borrowed and therefore people with a university degree are more likely to borrow higher sums for their housing needs.

Hypothesis 5: *People living in Prague borrow higher amounts compared to people living outside Prague.*

Since Prague is the capital city of the Czech Republic, it is supposed to be demanded locality for living. This would result in higher dwelling prices as people are willing to pay more for a house or a flat in Prague. In order to afford a dwelling in the capital city, I expect that people need to borrow more money. Therefore, people living in Prague are expected to owe higher amounts as for mortgage loans.

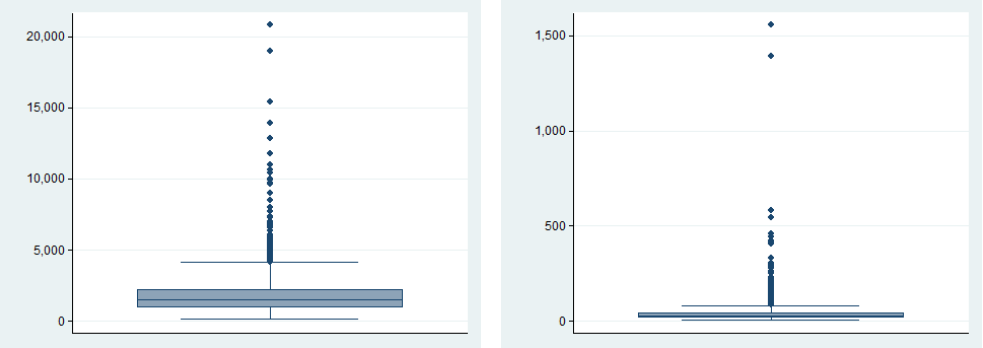
4.3 Methods

In this section, I present analytical methods used for estimating regression models. Each method is described together with the explanation why it was chosen for the analysis. Since the sample evinces feature which needs to be paid attention, I aim to use methods which could possibly correct it. Otherwise, the obtained results would not be plausible.

Firstly, I comment on the problematic values that occur in the sample. Secondly, the description of estimation methods is provided.

4.3.1 Outliers

The sample contains observations which have extremely high values compared to the rest of the dataset. These observations called outliers are problematic due to their unusual values since they can distort the results and therefore special treatment is required. This issue was examined by Rousseeuw and Leroy [31] who summarized robust methods which could be applied so that outliers do not affect estimated coefficients. In the sample, this problem applies to the



(a) Mortgage amount (b) Income

Figure 4.1: Outliers

dependent variable *mortgage* and independent variable *income* which can be seen in Figure 3.1 and Figure 3.5 in the previous chapter where the distributions of both problematic variables are presented.

For displaying outliers Tukey's box plots 4.1 were created where the number of these outlying observations can be visible both for *mortgage* and *income*. All amounts presented in the graphs are expressed in thousands of CZK. In case of *mortgage*, the 95th percentile equals 4,280,000 CZK and the 99th percentile equals 7,363,708 CZK, whereas the highest amount borrowed is approximately 20,000,000 CZK. The range of outlying observations is relatively wide and the values of outliers compared to the rest of the sample are extreme, indeed. In case of *income*, the range of outlying observations is even wider. The 95th percentile equals 95,667 CZK and the 99th percentile is 261,048 CZK. Thus, the highest amount of 1,556,508 CZK has to be treated with caution. I decided to consider observations to be outlying if their value is higher than the 99th percentile. Therefore, incomes higher than 261,048 CZK and mortgage amounts higher than 7,363,708 CZK are regarded as outliers, since lower values still seem feasible. Due to relatively large number of these suspicious observations, it is necessary to adjust chosen estimation method so that the results are trustworthy.

4.3.2 Ordinary Least Squares

As Wooldridge [32] describes this method, OLS is based on the estimation of regression coefficients that minimize the sum of squared residuals and is commonly used for analyses. Due to its computational mechanics, this method is susceptible to extreme values as their residuals can considerably influence

resulting coefficients. Since the sample contains a lot of outliers, it is necessary to either adjust the regression model or the sample size while using OLS.

In the first case, I use logarithmic forms of variables *mortgage* and *income*. This approach suppresses the outlying values as it decreases variation in both variables. Apart from solving outlier problem, Wooldridge [32] also points out, that model with dependent variable in a logarithmic form often more closely satisfies the assumptions behind this method. The disadvantage of this approach is that the estimated coefficients have percentage interpretation.

Using the second approach, the form of problematic variables does not have to be changed, but I adjust the sample size and remove some of the influential observations. In this case, the estimated coefficients show the effect of each variable in CZK. I delete observations with *mortgage* value or *income* value higher than 99th percentile. This has been done in two steps. At first, 20 observations with values of *mortgage* higher than 7,363,708 CZK are removed and after that I remove another 15 observations with values of *income* higher than 261,048 CZK. It means that the adjusted sample has 2,043 observations. Not all extremely high observations are removed, but only those with the highest values since I aim to maintain the size of the sample as large as possible for the analysis. Nevertheless, I believe it is necessary to restrict these suspicious values, even though the number of observations decreases. Since approximately 1.7% values were deleted, this reduction should not influence the results.

4.3.3 Least Absolute Deviations

In case that I want to analyse the dataset without removing any observation and estimate coefficients which provide interpretation in CZK, the use of robust method that is not as susceptible to outliers as OLS is needed. For this purpose, LAD is performed on the original sample of 2,078 observations. Instead of minimizing the sum of squared residuals, LAD is based on minimizing the sum of absolute residuals [33]. By applying both LAD and OLS, I can compare the results and see how influential outliers are.

LAD estimator is also known as median regression estimator as it estimates the parameters of the conditional median of dependent variable given independent variables rather than conditional mean estimated by OLS, since the median is not affected by extreme values as much as the mean. This method is a special case of quantile regression which is described in the following section [32].

4.3.4 Quantile Regression

I perform quantile estimation in order to examine how the effect of each independent variable changes at different parts of the distribution. Compared to OLS method, estimated coefficients are not fixed, but they vary across the distribution. This enables me to study estimated effects in more detail [33, 34].

The conditional quantiles that are examined are $q = 0.1, 0.25, 0.5, 0.75, 0.9$. The median regression is also included so that I can compare the results with other quantiles. I apply this method in order to estimate only the model without logarithmic forms as it evinces more variation in the dependent variable.

4.4 Models

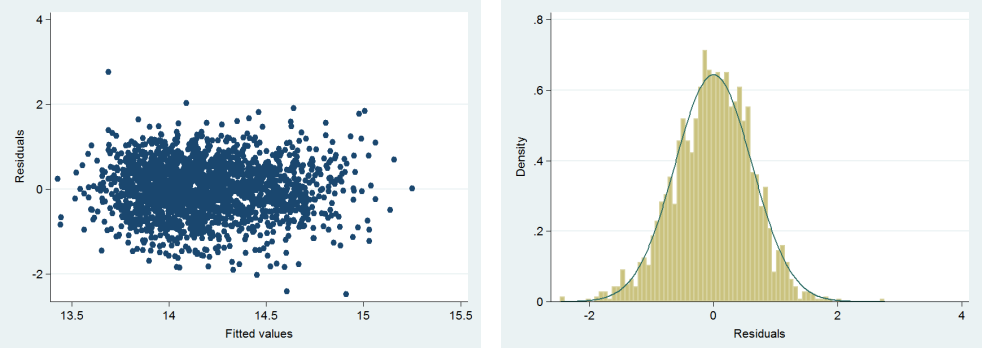
4.4.1 Model I

Firstly, I perform OLS regression with logarithmic forms of variables *mortgage* and *income*. It is possible to use logarithm as both variables attain only positive values. While establishing the regression model, I studied scatter plots showing the relationship between *mortgage* and each independent variable in order to include the most appropriate form of independent variable which best fits the data. I included variable age^2 since it better captures the relationship between *mortgage* and *age*. This variable is labelled as *agesq* in the analysis.

The first regression model denoted as Model I is defined in Equation 4.1.

$$\begin{aligned}\log(mortgage)_i = &\beta_0 + \beta_1 female_i + \beta_2 age_i + \beta_3 age_i^2 + \beta_4 single_i + \\ &\beta_5 divorced_i + \beta_6 children_i + \beta_7 degree_i + \\ &\beta_8 \log(income)_i + \beta_9 cb_i + \beta_{10} sb_i + \beta_{11} pl_i + \beta_{12} kv_i + \\ &\beta_{13} ul_i + \beta_{14} li_i + \beta_{15} hk_i + \beta_{16} pa_i + \beta_{17} ol_i + \beta_{18} ms_i + \\ &\beta_{19} sm_i + \beta_{20} zl_i + \beta_{21} vy_i + u_i\end{aligned} \tag{4.1}$$

I perform post-estimation diagnostic in order to verify whether the model satisfies CLM assumptions. At first, Breusch-Pagan test is performed in order to see whether the model suffers from heteroskedastic errors. Since p-value equals 0.0246, I reject the null hypothesis of homoskedasticity at 0.05 significance level. Therefore, heteroskedasticity-robust standard errors are used so that t statistics are valid. Residual-versus-fitted plot is shown in Figure 4.2a.



(a) Residual-versus-fitted plot (b) Histogram of residuals

Figure 4.2: Model I: homoskedasticity and normality assumptions

Secondly, I plot histogram of residuals and test for normality by Shapiro-Wilk test that yields p-value equal to 0.00030. Thus, I reject the null hypothesis of normally distributed residuals at 0.05 significance level. Since the examined sample is large, this should not be problematic as t statistics and F statistics are asymptotically valid [32]. This graph is shown jointly with normal density function in Figure 4.2b.

In order to exclude the possibility of having a general functional form misspecification, Ramsey's regression specification error test (RESET) is performed. The null hypothesis, that there is no omitted variable, means that the model does not suffer from a functional form misspecification [32]. Since the p-value equals 0.5781, I cannot reject the null hypothesis at 0.05 significance level. In other words, I did not find enough evidence to confirm that the model is misspecified.

The estimated coefficients are summarized together with t statistics in Table 4.1. At first, I test significance of variable *agesq* in order to verify that a quadratic function of *age* more accurately captures relationship between amount borrowed and age of a client than a linear function. The t test confirms that *agesq* is statistically significant at 0.05 significance level. I also test the joint significance of variables *age* and *agesq*. The F test yields p-value equal to 0.0000 and therefore I can conclude that variables *age* and *agesq* are also jointly significant. The quadratic form allows the marginal effect to vary for various *age* values.

The estimated coefficients have percentage interpretation as the dependent variable is in a logarithmic form. Since I have log-level model for the majority of independent variables, the interpretation of β is $\%\Delta y = (100\beta)\Delta x$ where y denotes dependent variable and x denotes independent variable. This does

	OLS
	log(mortgage)
female	-0.0884**
	(-2.94)
age	0.0522**
	(3.22)
agesq	-0.000722**
	(-3.29)
single	-0.0619
	(-1.81)
divorced	-0.171***
	(-3.82)
children	-0.0271
	(-1.87)
degree	0.157***
	(4.74)
log(income)	0.174***
	(6.86)
cb	-0.298***
	(-5.80)
sb	-0.516***
	(-7.65)
pl	-0.471***
	(-7.68)
kv	-0.648***
	(-7.69)
ul	-0.528***
	(-7.87)
li	-0.476***
	(-6.92)
hk	-0.392***
	(-5.72)
pa	-0.470***
	(-7.60)
ol	-0.520***
	(-8.50)
ms	-0.392***
	(-6.65)
sm	-0.346***
	(-7.12)
zl	-0.517***
	(-7.63)
vy	-0.472***
	(-5.64)
_cons	11.89***
	(33.21)
R^2	0.18
N	2,078

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.1: Model I results

not apply to variable log(*income*). The coefficient interpretation of *income* is $\%\Delta y = \beta\%\Delta x$ [32].

Results of Model I

In this part, I comment on the coefficients of each independent variable and provide possible explanation of the estimated effects. The coefficient of *female* suggests that women borrow about 8.8% lower amounts compared to men who are supposed to be the base group. Based on the previous data examination, the ratio men:female in the sample is 7:3. It means that not only do women borrow lower amounts, but also they borrow less frequently. One possible reason for this could be that women earn less money and therefore they cannot allow to repay relatively high monthly repayment sums.

The effect of *age* cannot be seen directly based on the coefficients as a quadratic function is included in the model. The effect is not constant because it depends on the value of the independent variable. As the coefficient of *age* is positive and the coefficient of *agesq* is negative, the regression line has a parabolic shape. It means that there is a maximum point where the independent variable has no effect on the dependent variable. Before this point, *age* has a positive effect on *mortgage* and after this point, the effect becomes negative. The turning point can be computed according to this formula: $age^* = |\hat{\beta}_2/(2\hat{\beta}_3)|$ [32]. After substituting 0.0522 for $\hat{\beta}_2$ and - 0.000722 for $\hat{\beta}_3$, the turning point value is obtained and it equals 36.15 years. It means that for people aged 18–36, one additional year increases the amount of money they decide to borrow. For people who are older than 36 years, the amount of money borrowed decreases as they become older. I compute the effect for the median, lower and upper quantile values so that it is possible to assess how the effect of *age* changes. The median value is 34 years and the corresponding change equals 0.003104. Thus, at the median age one additional year increases the loan amount by 0.31%. The change for the lower quantile value, which is 29 years, equals 1.032% meaning that the effect is more than three times larger than for the median value. Finally, the effect at the upper quantile value of 40 years equals -0.556%. It means that one additional year decreases the amount a person borrows by 0.556%.

One possible explanation for this feature could be that elderly people do not want to borrow high amounts because they might not manage to repay the whole debt. Furthermore, income after retiring is usually lower compared to

money earned while working. On the other hand, young people are more likely to have higher salary as they get older and also gain more working experience. Thus, they can afford to borrow higher amount of money.

The coefficient of *single* is not significant at 0.05 significance level. It means that there is not a significant difference between amounts borrowed by single people and by married people who are the base group. Conversely, the coefficient of *divorced* is significant even at 0.01 level. Divorced people borrow 17.1% lower sums than married people. This could be because of the fact that married people have the use of two salaries, while divorced people probably repay the debt on their own. This might be also related to age as divorced people are usually older and they probably want to avoid being heavily indebted at higher age.

The results suggest that the number of children does not have any effect on the amount a person decides to borrow since the coefficient of *children* is not significant at 0.05 significance level. Based on the distribution of number of children in the sample, it is evident that people who take out a mortgage are mostly childless. This means that people who borrow money in order to finance their housing needs have rather smaller number of children, but it does not influence the amount they borrow.

The coefficient of *degree* is significant at 0.05 significance level. People with a university degree (bachelor's or master's) borrow 15.7% higher amounts compared to people who did not obtain any university degree. A possible explanation might be that people with a university degree are more likely to have higher income. In the sample, the difference between mean values for people with degree and without degree is almost 14,000 CZK. Thus, they can afford to repay the amount borrowed with lower probability of default.

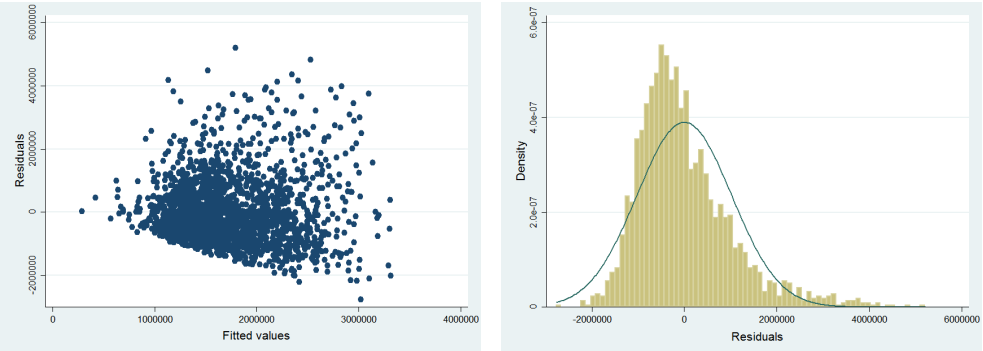
The effect of *income* on *mortgage* is positive. It means that the higher the salary is, the higher amount a person borrows. This might seem a little bit illogical that richer people would be more indebted regarding mortgages. One reason for this could be that richer people are more likely to buy expensive and luxurious goods. This may apply to housing as well. Because it takes some time to earn enough money in order to buy a house, they take out a mortgage so that they can buy it immediately and do not have to wait. As their income is high, it does not seem probable that they would face financial difficulties while repaying the debt.

All coefficients of regions a person lives in are statistically significant at 0.05 significance level. Since I did not include dummy variable for Prague in

order to avoid dummy variable trap, the coefficients show the differences between each region and the capital city. As all values are negative, people living in Prague borrow the highest amounts among all regions of the Czech Republic. People living in Central Bohemian Region borrow 29.8% lower amounts compared to Prague which is the lowest difference. On the other hand, people from Karlovy Vary Region borrow 64.8% lower amounts. This might be connected with dwelling prices. Apartment prices in Prague are higher than in the rest of the country [29]. Since Prague is the capital city, it offers more job vacancies which might be the main incentive why people wish to live in Prague. Because more than one fifth of the sample consists of Prague residents, I decided to examine whether the distribution of people in the sample by region is the same as the distribution of people living in the Czech Republic. For this purpose, I conducted chi-squared test with the null hypothesis of the identical distribution. Based on the test result, the distributions are statistically different at 0.05 significance level. The number of people living in Prague who have taken out a mortgage is 77% higher than the expected frequency based on the distribution of people by regions in the Czech Republic.

4.4.2 Model II

In this part, I estimate the regression model without logarithmic forms of variables *mortgage* and *income* using the trimmed sample of 2,043 observations. Thus, the coefficients do not have percentage interpretation, but express the effect in CZK. Simultaneously, I estimate the identical model using the unmodified dataset of 2,078 observations so that I can compare how much outliers affect the coefficients. I decided to add variable $income^2$ denoted as *incomesq* based on the results of performed Ramsey's RESET test. In case that only variable *income* is included, the test yields p-value equal to 0.0198. Therefore, I reject the null hypothesis of no omitted variable at 0.05 significance level. After adding *incomesq*, the p-value equals 0.0810 and thus, I cannot reject the null hypothesis of no omitted variable at 0.05 level.



(a) Residual-versus-fitted plot (b) Histogram of residuals

Figure 4.3: Model II: homoskedasticity and normality assumptions

Based on the performed RESET tests, the regression model has the following form denoted as Equation 4.2:

$$\begin{aligned} mortgage_i = &\beta_0 + \beta_1 female_i + \beta_2 age_i + \beta_3 age_i^2 + \beta_4 single_i + \\ &\beta_5 divorced_i + \beta_6 children_i + \beta_7 degree_i + \beta_8 income_i + \\ &\beta_9 income^2 + \beta_{10} cb_i + \beta_{11} sb_i + \beta_{12} pl_i + \beta_{13} kv_i + \beta_{14} ul_i + \\ &\beta_{15} li_i + \beta_{16} hk_i + \beta_{17} pa_i + \beta_{18} ol_i + \beta_{19} ms_i + \beta_{20} sm_i + \\ &\beta_{21} zl_i + \beta_{22} vy_i + u_i \end{aligned} \tag{4.2}$$

As in the previous case, Breusch-Pagan test for heteroskedasticity is performed. The p-value of 0.0000 suggests that the model suffers from heteroskedastic errors since I reject the null hypothesis of homoskedasticity at 0.05 significance level. This can be seen in Figure 4.3a. To correct this feature, I estimate the model using heteroskedasticity-robust standard errors. Subsequently, I test normality of residuals using Shapiro-Wilk test. I can conclude that residuals are not normally distributed as the p-value equals 0.0000 and therefore I reject the null hypothesis of normality. Nevertheless, the statistical inference is valid due to large sample size. The histogram of residuals presented jointly with normal density function is shown in Figure 4.3b. Based on the results, Model I satisfies the CLM assumptions better in comparison with Model II. The estimated coefficients are presented in Table 4.2. I include also the estimated coefficients of the identical regression model using 2,078 original observations with all outliers so that the comparison of their effects can be made.

	OLS mortgage	OLS mortgage
female	-154,644.5**	-138,559.3*
	(-3.15)	(-2.25)
age	112,783.2***	68,133.3
	(6.32)	(1.27)
agesq	-1,522.9***	-828.1
	(-6.57)	(-1.11)
single	-82,451.5	-42,102.2
	(-1.40)	(-0.53)
divorced	-259,453.0***	-296,411.5***
	(-3.78)	(-3.56)
children	-41,998.7	-46,097.9
	(-1.56)	(-1.35)
degree	293,257.6***	157,775.7*
	(4.87)	(1.97)
income	10.48***	13.65***
	(4.30)	(5.13)
incomesq	-0.0000287	-0.00000919***
	(-1.69)	(-5.43)
cb	-567,773.8***	-748,919.9***
	(-6.28)	(-6.00)
sb	-782,743.9***	-996,466.0***
	(-7.12)	(-7.71)
pl	-789,514.8***	-997,128.9***
	(-7.60)	(-7.20)
kv	-1,014,686.3***	-1,215,259.6***
	(-8.81)	(-9.11)
ul	-795,782.7***	-1,042,694.8***
	(-7.52)	(-8.36)
li	-753,290.1***	-960,449.6***
	(-5.86)	(-6.73)
hk	-640,927.1***	-757,754.2***
	(-5.47)	(-5.02)
pa	-808,318.1***	-1,023,750.0***
	(-8.47)	(-8.43)
ol	-838,928.4***	-1,034,755.3***
	(-8.74)	(-8.76)
ms	-634,179.4***	-844,014.1***
	(-6.14)	(-7.00)
sm	-627,472.4***	-824,526.4***
	(-6.57)	(-7.39)
zl	-784,903.6***	-1,006,011.4***
	(-6.88)	(-7.18)
vy	-760,295.8***	-950,338.1***
	(-5.59)	(-6.41)
_cons	61,311.6	818,764.5
	(0.18)	(0.90)
R^2	0.17	0.21
N	2,043	2,078

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.2: Model II results using both reduced and original sample

Results of Model II

After estimating model without logarithmic forms, I examine the effect of each independent variable in terms of money. As the coefficients have the same sign as in case of Model I, no possible explanation of observed features is needed since the effect has already been described in the previous part. Therefore, the main focus is put on the magnitude of coefficients and differences using both sample sizes. The coefficient of *female* shows that women borrow on average 154,644 CZK lower amounts compared to men. This is a significant difference as variable *female* is significant even at 0.01 significance level. This does not apply to the second version of Model II. The coefficient is 16,085 CZK lower and significant at 0.05 level. The coefficients of *age* variables evince significant inconsistencies. According to the results of the model using the trimmed sample with the reduced number of outliers, both *age* and *agesq* are highly significant and thus affect the amount borrowed. On the other hand, if I do not reduce the sample and perform the regression analysis without removing some outlying observations, the effect of *age* is not significant. As in the previous model, a quadratic function is used so that the effects vary depending on values of *age*. Based on the signs of coefficients, the quadratic function has a parabolic shape with a maximum value. The turning point when the effect becomes negative is equal to 37.03 years which is 1 year higher compared to Model I result.

The negative coefficient of *single* shows that single people borrow lower amounts compared to married people, but this difference is not different from 0 as *single* is not significant even at 0.1 significance level. On the other, divorced people borrow 259,453 CZK lower amounts and the difference is statistically significant. The number of children does not influence the decision on the amount of money a person borrows as the coefficient of *children* is not significant. Conversely, a university degree has a positive effect and is statistically significant at 0.05 level. People who obtained a university degree are supposed to borrow 293,257 CZK higher amount. As for the model using all observations, the effect is smaller and less significant.

Since I included not only *income*, but also *incomesq*, the effect of money earned is not constant, because it depends on the particular *income* value. Although *incomesq* is significant at 0.1 significance level, F test shows that *income* and *incomesq* are jointly significant at 0.05 level and thus, it is reasonable to include this variable into the model. As already mentioned, the model with both *income* variables is more appropriate based on RESET test results.

The signs of coefficients mean that relationship between *mortgage* and *income* approximates to a parabolic shape. For low income people, higher salary leads to an increase in the amount borrowed. After the turning point, the effect becomes negative resulting in lower loan amount. The turning point equals 182,578 CZK. After such a high income, the effect of money earned decreases. As for the second version of Model II, the turning point equals 742,655 CZK which seems to be extremely high since it is most likely affected by outlying observations. I computed the magnitude of effect for the lower quantile value, the median value and the upper quantile value. The 25th percentile value is 18,426 CZK and the corresponding effect equals 9.42 CZK. It means that an increase in income by 1 CZK results to increase in amount borrowed by 9.42 CZK. For the median value of 27,650 CZK, the effect is 8.89 CZK. The 75th percentile equals 95,667 CZK which is still lower than the turning point and therefore the effect is positive. An increase in salary by 1 CZK means that a person borrows 4.98 CZK higher amount. Only 34 people in the sample earn more than the turning point is. It means that the effect of *income* is positive for the majority of the sample. Nevertheless, including *incomesq* allows the effect to differ for various values.

All dummy variables denoting region a person lives in are significant which means that there are considerable differences between amounts people borrow in Prague and in the rest of the Czech Republic. All estimated coefficients of the second model probably overestimate the differences as all values are higher compared to the model using reduced sample. The smallest difference is in Central Bohemian Region where people borrow on average 567,773 CZK lower amounts compared to the capital city. Conversely, people living in Karlovy Vary Region borrow lowest amounts. This is in line with the results of Model I.

4.4.3 Model III

In this and the following section, I estimate regression models by LAD. This method is more robust to outliers as it does not minimize the sum of squared residuals, but the sum of absolute values of residuals [32]. In the previous sections, I used two approaches in order to decrease the effect of outlying observations on the estimated coefficients by reducing their number and using logarithmic forms of dependent variable *mortgage* and independent variable *income* so that it is possible to estimate the model by OLS and obtain plausible results. Even though including logarithmic forms decreased the variation

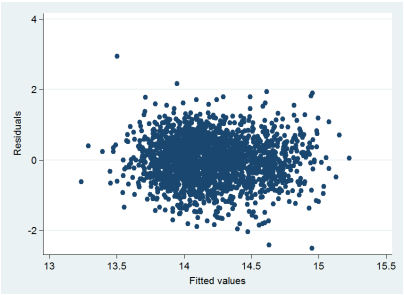


Figure 4.4: Model III: residual-versus-fitted plot

in problematic variables, there still could be rare values compared to the rest of the sample. Thus, both regression models with logarithmic forms and without logarithmic forms are estimated by LAD. The estimated coefficients are presented together with coefficients estimated by OLS in the previous section so that it is possible to assess the effect of outliers.

Firstly, I estimate the model where variables *mortgage* and *income* are in logarithmic forms and therefore the coefficients represent the percentage effect. This model denoted as Model III has the identical form as Model I which is defined in Equation 4.1.

In statistical software Stata, LAD is performed by *qreg* or *qreg2* commands. The 0.5 quantile is set by default and does not have to be changed. The latter option computes robust standard errors in case of heteroskedasticity. Simultaneously, it performs Machado-Santos Silva test for heteroskedasticity in order to determine whether robust standard errors are necessary [35, 36]. Based on Machado-Santos Silva test, I cannot reject the null hypothesis of homoskedasticity at 0.05 significance level as obtained p-value equals 0.139. Thus, the model is estimated using *qreg* command since no robust standard errors are needed because the residuals are homoskedastic. Scatter plot of residuals on the y axis and fitted values on x axis is shown in Figure 4.4. The estimated coefficients are summarized in Table 4.3.

Results of Model III

Almost all coefficients are very similar to those estimated by OLS. Moreover, their statistical significance does not differ at all. Based on the coefficient of variable *female*, women borrow 11.4% lower amounts compared to men. This is approximately 2.5% higher percentage than the coefficient estimated by OLS. The difference between male and female clients in the sample is statistically significant at 0.05 significance level.

	LAD log(mortgage)	OLS log(mortgage)
female	-0.114**	-0.0884**
	(-2.97)	(-2.94)
age	0.0638***	0.0522**
	(4.35)	(3.22)
agesq	-0.000888***	-0.000722**
	(-4.72)	(-3.29)
single	-0.0714	-0.0619
	(-1.63)	(-1.81)
divorced	-0.194***	-0.171***
	(-3.52)	(-3.82)
children	-0.0255	-0.0271
	(-1.55)	(-1.87)
degree	0.185***	0.157***
	(4.54)	(4.74)
log(income)	0.142***	0.174***
	(5.51)	(6.86)
cb	-0.267***	-0.298***
	(-4.09)	(-5.80)
sb	-0.472***	-0.516***
	(-5.76)	(-7.65)
pl	-0.525***	-0.471***
	(-6.36)	(-7.68)
kv	-0.616***	-0.648***
	(-6.22)	(-7.69)
ul	-0.473***	-0.528***
	(-5.95)	(-7.87)
li	-0.515***	-0.476***
	(-5.70)	(-6.92)
hk	-0.465***	-0.392***
	(-6.00)	(-5.72)
pa	-0.465***	-0.470***
	(-5.63)	(-7.60)
ol	-0.534***	-0.520***
	(-7.02)	(-8.50)
ms	-0.322***	-0.392***
	(-4.44)	(-6.65)
sm	-0.416***	-0.346***
	(-6.22)	(-7.12)
zl	-0.650***	-0.517***
	(-7.84)	(-7.63)
vy	-0.524***	-0.472***
	(-4.88)	(-5.64)
_cons	12.06***	11.89***
	(31.62)	(33.21)
R^2		0.18
N	2,078	2,078

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.3: Model III and Model I comparison

Both coefficients of variables *age* are statistically significant and F test confirms that *age* and *agesq* are also jointly significant at 0.05 significance level. The combination of signs of coefficients means that *age* firstly has a positive effect on the amount borrowed and afterwards the effect becomes negative. The turning point, where the effect changes, equals 35.92 years. It is almost identical to the point computed based on OLS coefficients as it equals 36.15 years. Since the magnitude of the effect depends on the value of *age*, I compute it for the median value, lower quantile value and upper quantile value. An increase from the median age of 34 years to 35 years increases the amount of money borrowed by 0.341%. If I substitute the lower quantile value instead of the median, one additional year increases the loan amount by 1.229%. Since the upper quantile value is higher than the turning point, the effect is negative and an increase from 40 years to 41 years decreases the amount by 0.724%. As for Model I which was estimated by OLS, these three effects are 0.31%, 1.032% and -0.556%.

Single persons borrow 7.14% lower amounts than married persons. This differs from the OLS coefficient by 0.95%. Since variable *single* is not statistically significant at 0.05 significance level, the fact whether a person is married or single does not play an important part in deciding how much money they borrow. On the other hand, there is a significant evidence that divorced people borrow lower amounts compared to married people as variable *divorced* is statistically significant at 0.05 significance level. The estimated difference using LAD is 2.3% higher compared to the result estimated by OLS.

The coefficients of *children* are almost identical as their difference is only 0.16%. The magnitude of the estimated effect means that one additional child decreases the loan amount by 2.55%. Nevertheless, the number of children does not have any effect on the amount of money borrowed which is in line with the results of Model I because variable *children* is not statistically significant at 0.05 significance level.

On the other hand, obtaining a university degree increases the amount a person borrows by 18.5%. It means that people with a university degree tend to borrow higher amounts compared to people with primary or secondary education. If I compute the exact number for the median value of 1,500,000 CZK, university degree raise the amount borrowed by 277,500 CZK which is not negligible. The difference is statistically significant at 0.05 significance level.

The effect of *income* can be seen directly. Since both variables *mortgage* and *income* are in logarithmic forms, there is no need to multiply the coefficient by 100 in order to obtain the percentage effect. It means that if income increases

by 1%, the amount borrowed increases by 0.142%. I compute the exact changes for the median values of *income* and *mortgage* in order to assess the effect. If *income* increases by 1% from 27,650 CZK to 27,926.5 CZK, the amount borrowed will increase from 1,500,000 CZK to 1,502,130 CZK.

All coefficients regarding regions people live in are statistically significant at 0.05 significance level. Coefficients of variables *cb, sb, kv, ul, pa* and *ms* are smaller while coefficients of variables *pl, li, hk, ol, zl* and *vy* are larger compared to those estimated by OLS. Since all coefficients are negative, people living in Prague take out the highest mortgage amounts in the Czech Republic. This might suggest that Prague residents are more likely to have higher debt as the costs related to housing in the capital city are probably higher. Central Bohemian Region appears to be the second most expensive region regarding housing which makes people borrow more money. Conversely, people living in Karlovy Vary Region borrow on average 65% lower amounts compared to Prague.

4.4.4 Model IV

In this part, I compare the results of Model II, which were estimated by OLS using reduced sample of 2,043 observations, and a model with identical form estimated by LAD while using the original sample of 2,078 observations. The reduction was made so that the most extreme outliers do not affect the results. Since LAD is a robust method often used in order to obtain more plausible coefficients in presence of outliers, there is no need to reduce the sample size. The regression model denoted as Model IV has the identical form as Model II that is written in Equation 4.2.

As in the previous case, I performed Machado-Santos Silva test for heteroskedasticity detection. Since the obtained p-value equals 0.0000, I reject the null hypothesis of homoskedasticity at 0.05 significance level. Thus, the estimation was performed using Stata command *qreg2* which computes heteroskedasticity robust standard errors by default. Heteroskedasticity can be seen in Figure 4.5 as residuals tend to increase with higher fitted values. The estimated coefficients are presented jointly with Model II coefficients in Table 4.4.

Results of Model IV

Due to the composition of the examined sample which contains a lot of outliers, Model IV is supposed to yield the most plausible results having CZK

	LAD mortgage	OLS mortgage
female	-167,919.5***	-154,644.5**
	(-3.41)	(-3.15)
age	82,523.8***	112,783.2***
	(4.57)	(6.32)
agesq	-1,144.2***	-1,522.9***
	(-4.87)	(-6.57)
single	-87,483.7	-82,451.5
	(-1.58)	(-1.40)
divorced	-220,156.5**	-259,453.0***
	(-3.03)	(-3.78)
children	-44,935.5*	-41,998.7
	(-2.11)	(-1.56)
degree	298,368.8***	293,257.6***
	(4.59)	(4.87)
income	7.456**	10.48***
	(3.27)	(4.30)
incomesq	-0.00000493***	-0.0000287
	(-3.45)	(-1.69)
cb	-473,840.3***	-567,773.8***
	(-4.92)	(-6.28)
sb	-773,331.1***	-782,743.9***
	(-6.67)	(-7.12)
pl	-831,575.3***	-789,514.8***
	(-8.93)	(-7.60)
kv	-898,262.0***	-1,014,686.3***
	(-5.42)	(-8.81)
ul	-738,293.7***	-795,782.7***
	(-6.41)	(-7.52)
li	-823,331.4***	-753,290.1***
	(-8.08)	(-5.86)
hk	-737,316.3***	-640,927.1***
	(-7.59)	(-5.47)
pa	-752,379.2***	-808,318.1***
	(-7.07)	(-8.47)
ol	-847,120.9***	-838,928.4***
	(-7.34)	(-8.74)
ms	-533,571.2***	-634,179.4***
	(-4.15)	(-6.14)
sm	-694,025.8***	-627,472.4***
	(-7.39)	(-6.57)
zl	-965,433.6***	-784,903.6***
	(-10.13)	(-6.88)
vy	-840,863.8***	-760,295.8***
	(-7.90)	(-5.59)
_cons	506,334.4	61,311.6
	(1.50)	(0.18)
R^2		0.17
N	2,078	2,043

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.4: Model IV and Model II comparison

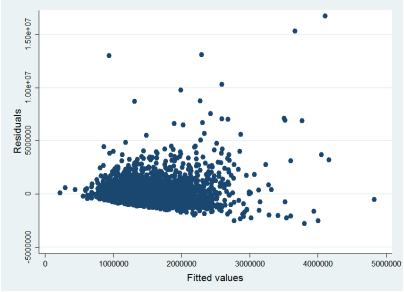


Figure 4.5: Model IV: residual-versus-fitted plot

interpretation. Women are estimated to borrow 167,919 CZK lower amounts than men which is 13,275 CZK higher than OLS coefficient. The difference between female and male persons is statistically significant at 0.05 significance level.

Both *age* variables are individually significant, but also jointly statistically significant at 0.05 significance level. It means that *age* and *agesq* most likely capture the true relationship between mortgage amount and person's age which appears to be changing depending on the value of *age*. The turning point where the effect of age changes from positive to negative is 36.06 years. This is 1 year lower compared to the result of Model II.

As for marital status, married people tend to borrow the highest amounts of all groups. Single people borrow 87,483 CZK lower amounts, but it does not significantly differ from married people as variable *single* is not statistically significant at 0.05 significance level. Divorced people borrow even lower amounts compared to married people. The amount borrowed by a divorced person is on average 220,156 CZK lower and this difference is statistically significant at 0.05 significance level.

The main difference between Model IV and all other models is that variable *children* is statistically significant at 0.05 level. It means that the number of children has an effect on the amount of money a parent borrows. One additional child decreases the amount borrowed by 44,935 CZK. A possible explanation for this effect could be that having children is financially demanding. The more children people have, the more money they need in order to support them. On the other hand, people who have a lot of children might want to buy or construct a larger house which is more expensive, but based on the negative sign of the coefficient this effect does not seem to be probable.

The coefficient of variable *degree* shows the difference between people with a university degree and people with primary or secondary education. Univer-

sity degree increases the amount of borrowed money by 298,368 CZK. This is very significant as variable *degree* is statistically significant at 0.05 significance level. This is almost identical to the coefficient obtained by performing OLS estimation on the reduced sample.

Based on the coefficients of dummy variables denoting region a person lives in, there is a significant difference between Prague and the rest of the Czech Republic. Amounts borrowed in the capital city are considerably higher. People living in Central Bohemian Region borrow 473,840 CZK lower sums which is the smallest difference. All coefficients are lower than those estimated by OLS using the same sample of 2,078 observations. It means that due to outliers the OLS coefficients related to regions were overestimated.

4.4.5 Model V

In the previous sections, I used LAD estimation method which estimated the conditional median. In this section, I extend this concept and estimate other conditional quantiles. This estimation method called quantile regression provides more detailed analysis compared to OLS which computes the mean effect. I can analyse how the effects of particular variables change for various parts of distribution [37]. It means that the estimated coefficients are not fixed, but might differ depending on the chosen quantile. For the estimation of the quantile regression lines the following quantiles are used: 0.1, 0.25, 0.5, 0.75, 0.9. Estimation is performed on the model which provides interpretation of coefficients in CZK, because these coefficients evince larger variation compared to coefficients having percentage interpretation and thus might better express the changes across quantiles. For each quantile estimation, I test whether homoskedasticity assumption is violated using Machado-Santos Silva test and according to the result, corresponding standard errors are computed so that t statistics are valid.

In this part, I estimate the model with variable *mortgage* as the dependent one. Lower index q standing for the given quantile is added so that it is obvious that βs are not fixed. The form of the regression model is shown in Equation 4.3.

$$\begin{aligned} mortgage_i = &\beta_{0,q} + \beta_{1,q} female_i + \beta_{2,q} age_i + \beta_{3,q} age_i^2 + \beta_{4,q} single_i + \\ &\beta_{5,q} divorced_i + \beta_{6,q} children_i + \beta_{7,q} degree_i + \beta_{8,q} income_i + \\ &\beta_{9,q} income^2 + \beta_{10,q} cb_i + \beta_{11,q} sb_i + \beta_{12,q} pl_i + \beta_{13,q} kv_i + \\ &\beta_{14,q} ul_i + \beta_{15,q} li_i + \beta_{16,q} hk_i + \beta_{17,q} pa_i + \beta_{18,q} ol_i + \\ &\beta_{19,q} ms_i + \beta_{20,q} sm_i + \beta_{21,q} zl_i + \beta_{22,q} vy_i + u_i \end{aligned} \tag{4.3}$$

The estimated coefficients corresponding to quantiles 0.1, 0.25, 0.5, 0.75, 0.9 are presented in Table 4.5. All quantile regression models suffer from heteroskedasticity and therefore robust standard errors which are asymptomatically valid in presence of heteroskedasticity are computed using Stata command *qreg2*. Table 4.5 shows not only how the coefficients vary, but also the changes of significance of particular variables. The development of quantile regression coefficients for each variable is plotted in Figure A1 in Appendix so that it is easier to assess the changes across the distribution. The dashed line in each figure represents OLS estimates. The two dotted lines delimit 95th confidence interval for OLS estimates. The continuous line represents the quantile regression estimates. It is complemented with shaded area depicting 95th confidence interval.

Results of Model V

In this section, I comment on the quantile regression results. The interpretation of estimated coefficients is provided together with the overall effects of each variable across the distribution which can be analysed based on the Figure A1. Furthermore, I test whether the estimated coefficients significantly differ for given quantiles.

Variable *female* represents the amount women borrow relative to men. At the 10th percentile of the conditional distribution, the difference is 44,838 CZK, but it is not statistically significant at 0.05 level. As for the 25th percentile, median and 75th percentile, the difference increases reaching the highest value at the 75th percentile. Female persons take out approximately 178,000 CZK lower mortgage amounts compared to male persons. At the 90th percentile the difference shrinks to 152,903 CZK and it is not statistically significant. The results plotted in Figure A1a show that *female* has quite a uniform effect except for the 45th percentile. This deviation can be seen with a few other

	$q = 0.1$ mortgage	$q = 0.25$ mortgage	$q = 0.5$ mortgage	$q = 0.75$ mortgage	$q = 0.9$ mortgage
female	-44,838.1	-99,431.5**	-167,919.5***	-177,860.6*	-152,903.9
	(-1.32)	(-2.69)	(-3.41)	(-2.49)	(-1.22)
age	23,449.9	25,312.8	82,523.8***	153,993.6***	190,407.3***
	(1.66)	(1.72)	(4.57)	(5.92)	(3.69)
agesq	-407.1*	-466.0*	-1144.2***	-2,064.0***	-2,329.6**
	(-2.22)	(-2.47)	(-4.87)	(-6.39)	(-3.21)
single	-92,655.5*	-87,060.6	-87,483.7	-95,931.6	-53,504.6
	(-2.12)	(-1.81)	(-1.58)	(-0.97)	(-0.33)
divorced	-108,280.9*	-117,627.6*	-220,156.5**	-366,830.1***	-548,755.0**
	(-2.32)	(-2.38)	(-3.03)	(-3.90)	(-2.75)
children	-26,983.9	-6,264.5	-44,935.5*	-100,571.3**	-171,462.4**
	(-1.58)	(-0.35)	(-2.11)	(-2.92)	(-2.62)
degree	144,306.2***	211,596.4***	298,368.8***	208,830.5*	262,882.0
	(3.30)	(4.39)	(4.59)	(2.28)	(1.45)
income	1.432*	2.033**	7.456**	16.88***	32.08***
	(2.13)	(2.78)	(3.27)	(9.37)	(8.23)
incomesq	-0.00000106*	-0.00000163**	-0.00000493***	-0.0000111***	-0.0000210***
	(-2.05)	(-2.78)	(-3.45)	(-9.68)	(-8.52)
cb	-239,116.5**	-386,784.7***	-473,840.3***	-761,475.7***	-1,344,335.1***
	(-3.22)	(-5.21)	(-4.92)	(-4.39)	(-4.88)
sb	-422,314.2***	-672,097.8***	-773,331.1***	-994,506.9***	-1,345,294.9***
	(-6.06)	(-7.74)	(-6.67)	(-3.48)	(-4.87)
pl	-346,750.4***	-473,052.7***	-831,575.3***	-1,051,413.7***	-556,729.7
	(-3.97)	(-5.85)	(-8.93)	(-5.92)	(-0.59)
kv	-516,713.9***	-809,811.2***	-898,262.0***	-1,140,968.1***	-1,711,028.9***
	(-6.49)	(-9.69)	(-5.42)	(-4.91)	(-5.67)
ul	-460,784.0***	-621,973.7***	-738,293.7***	-982,524.0***	-1,672,545.3***
	(-5.69)	(-6.60)	(-6.41)	(-5.08)	(-5.98)
li	-417,991.1***	-586,114.1***	-823,331.4***	-1,158,419.1***	-1,501,001.3**
	(-5.45)	(-6.65)	(-8.08)	(-5.31)	(-2.96)
hk	-469,770.3***	-477,970.7***	-737,316.3***	-876,131.5***	-1,032,575.8**
	(-5.58)	(-5.47)	(-7.59)	(-4.24)	(-2.71)
pa	-353,619.1***	-553,001.1***	-752,379.2***	-1,046,218.9***	-1,567,291.6***
	(-4.17)	(-7.06)	(-7.07)	(-5.75)	(-5.22)
ol	-452,821.6***	-634,418.6***	-847,120.9***	-1,174,880.9***	-1,510,356.7***
	(-6.04)	(-7.88)	(-7.34)	(-6.19)	(-4.83)
ms	-389,041.5***	-521,580.9***	-533,571.2***	-883,653.6***	-1,252,985.1***
	(-5.35)	(-6.26)	(-4.15)	(-4.62)	(-4.08)
sm	-301,885.2***	-466,305.7***	-694,025.8***	-812,751.1***	-1,306,050.3***
	(-4.10)	(-5.69)	(-8.39)	(-3.88)	(-4.78)
zl	-430,688.4***	-676,925.8***	-965,433.6***	-1,051,980.9***	-1,280,382.9***
	(-6.02)	(-8.26)	(-10.13)	(-5.22)	(-3.76)
vy	-316,566.3**	-475,833.8***	-840,863.8***	-971,860.4***	-1,718,655.9***
	(-2.69)	(-5.16)	(-7.90)	(-4.53)	(-5.72)
_cons	682,355.7*	1,090,238.1***	506,334.4	-102,498.8	-197,057.3
	(2.46)	(3.89)	(1.50)	(-0.20)	(-0.22)
N	2,078	2,078	2,078	2,078	2,078

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.5: Model VI results for various quantiles

variables as well. A possible reason for that could be that these deviations are caused by outliers as the whole sample of 2,078 observations is used for the analysis. Apart from this deviating difference, all coefficients are within the confidence interval of the OLS estimate. The test for equality of coefficients across quantiles supports the uniform effect as I cannot reject the null hypothesis of identical coefficients.

A person's age is expressed as a quadratic function. The results in Figures A1b and A1c show that variable *age* increases for higher quantiles while variable *agesq* decreases. It suggests that the marginal effect of *age* is not constant for chosen quantiles as the slope of the parabolic curve changes. If I compute the turning point when the effect of age on mortgage amount becomes negative for each quantile, these values are obtained: 28.8 years, 27.16 years, 36.06 years, 37.3 years and 40.86 years. According to the results, *age* has a less significant effect on the amount borrowed at the left tail of the distribution. It means that *age* is more important factor at the higher quantiles.

Variable *single* is significant at 0.05 significance level only at the 0.1 quantile. All coefficients are almost identical except for the 90th percentile and they do not differ from the OLS coefficient. This would suggest that the difference between the loan amount a single person borrows and the loan amount taken out by a married person does not change across the distribution. This is plotted in Figure A1d. On the other hand, divorced people tend to borrow lower amounts relative to married people as we move along the distribution to the right and variable *divorced* is statistically significant at all quantiles. At the 90th percentile the difference is 548,755 CZK. At the left tail of the conditional distribution the effect of being divorced estimated by OLS seems to be underestimated compared to the quantile regression coefficients. Conversely, OLS overestimates the difference at the 60th and higher percentiles. The coefficients are plotted in Figure A1e.

The number of children does not have any effect on the amount of money borrowed at the lower quantiles, but it might be a factor parents take into consideration while taking out a mortgage at the higher quantiles as variable *children* is statistically significant at 0.05 level for the median, 0.75 quantile and 0.9 quantile. The overall effect of *children* is increasing as shown in Figure A1f since one additional child leads to a higher reduction in the amount borrowed as the coefficients decrease across the distribution. As in the previous cases, there is a deviation at approximately 0.45 quantile. At the 90the percentile, one additional child decreases the amount borrowed by 171,462 CZK which is

the largest effect.

People who have obtained a university degree tend to borrow higher amounts compared to people without university degree. The difference is statistically significant at all quantiles apart from the 0.9 quantile. The development of coefficients can be seen in Figure A1g. The highest value of the coefficient is at the median. According to the figure where the quantile regression coefficients are plotted, the estimated effect of having university degree is higher that the OLS estimated almost across the whole distribution.

Income coefficients show the most obvious variation across the distribution compared to the other variables as shown in Figures A1h and A1i. This is confirmed by the test for equality of coefficients, because in case of both *income* and *incomesq* I reject the null hypothesis that coefficients do not significantly differ at 0.05 significance level. As *income* is statistically significant for all examined quantiles, it appears to be the major factor affecting the amount borrowed. Since *income* enters the model as a quadratic function, the differences in coefficients across the distribution result in the shape changes of the parabola. The combination of negative and positive sign means that person's salary has a positive effect at first which at some point becomes negative. Since these turning points for all examined quantiles are extremely high and the examined sample includes less than 0.2% of observations with such a high salary, the effect of *income* in the sample seems to be always positive. It is likely that people with such high incomes did not take out any mortgage and therefore they are not among the debtors. In that case, *income* would indeed have a negative effect as it prevents people from borrowing. Therefore, in the model variable *incomesq* adjusts the curvature of the regression line which seems to fit the data better than a linear line.

As for the variables standing for a Czech region a client lives in, almost all coefficients tend to be decreasing across the distribution. Apart from the coefficient of *pl* at the 90th percentile, all coefficients are statistically significant at 0.05 significance level. This means that the amounts people borrow in the capital city are significantly higher than in other regions of the Czech Republic. As already examined, not only do people living in Prague borrow higher amounts, but they also borrow much frequently as they borrow almost twice as often as they are estimated to given the population distribution.

Chapter 5

Discussion of Results

In this chapter, I firstly summarize the main findings which were discovered based on the econometric analysis. The aim of this study is to explain the variation in mortgage amount by personal characteristics which has not been, to the author's knowledge, examined. Therefore, it is not possible to compare obtained results with other studies. I comment on the influence of each variable on the amount people borrow in order to finance housing needs based on the results of Model I-V together with arguments that most likely explain the estimated effect. The magnitude of the effect is computed as the average of OLS coefficient and LAD coefficient for a particular variable. Furthermore, I test hypotheses which are stated in the previous chapter. The results of Models I-IV estimated by OLS or LAD are summarized in Table A1. Quantile regression coefficients for Model V are presented separately in Table A2. Finally, I comment on the plausibility of estimated results.

5.1 Gender

There is a significant difference between men and women. This does not apply only to the proportion of each gender in the sample, but also the amounts of money borrowed are not identical. Men take out a mortgage more than twice as frequently as women since they account for the 70% of persons who are currently repaying the mortgage. A most likely explanation is that in case of married couple it is usual that husband is the one who arranges and signs the contract. The reason why women borrow lower amounts could be connected with their income. The median income is 5,000 CZK lower compared to the median income earned by men. Therefore, women might not afford to

take out such a high mortgage since they would not be able to repay it. I obtained various results depending on the method, but the coefficients were very similar. Women borrow approximately 160,000 CZK lower amounts. This difference changes only slightly across the distribution. The model using logarithmic form of the dependent variable estimated approximately 10% difference in borrowed amounts between men and women. The hypothesis connected with gender was stated as follows: *Men borrow higher amounts than women.* I test the null hypothesis $H_0 : \beta_1 = 0$ against one-sided alternative $H_1 : \beta_1 < 0$. The result of t test suggests that I reject the null hypothesis that there is no difference between amounts borrowed by men and women at 0.05 significance level. Therefore, gender is an important factor affecting how much a person decides to borrow as men tend to take out higher mortgage amounts.

5.2 Age

A person's age does not have a constant effect on the amount borrowed. People aged 18–36 years borrow higher amounts as they are getting older. Each additional year increases the amount borrowed depending on the current age. One explanation for this could be that as people grow older, they are more likely to have higher salary. This might influence their decision on how much money they want to borrow. Additionally, people usually have a stable job with increasing age and thus, they do not have to be afraid of being without sufficient financial means for repayment. The highest proportion of the sample consists of people aged 30–40 years. This might suggest that people take out a mortgage in order to finance their housing needs after they start a family as people tend to have their first child later compared to the previous generations. At the age of 36, the effect of age becomes negative and each additional year decreases the amount borrowed. A possible reason for this could be that people do not want to be highly indebted at their old age as they might not manage to repay the whole amount. Hypothesis 3 was phrased as follows: *A person's age is a significant factor influencing the amount borrowed.* Since age is expressed as a quadratic function, I test not only individual significance of *age* variables, but also joint significance. After performing t test and F test, I can confirm that age is a statistically significant factor since I reject the null hypothesis that coefficients of *age* variables are equal to 0 at 0.05 significance level.

5.3 Marital Status

Married people account for more than one half of the examined sample. This is very likely connected with the traditional process of getting married, starting a family and constructing or buying a house. Furthermore, married couples can use two salaries for repayment which might encourage them to borrow more. There is not a significant difference between amounts borrowed by married persons and single persons. It is very likely that even though single people are not married, they have a partner who pays off the debt with them. On the other hand, divorced people might take out a mortgage so that they can finance a new house or flat after divorce. Thus, they borrow lower amounts since they might repay the debt on their own. There is a significant difference between divorced people and married people as for the loan amount. A divorced person borrows approximately 240,000 CZK or 18% lower amounts relative to a married person.

5.4 Number of Children

Children might function as possible incentive for taking out a mortgage as parents wish to provide their children with a stable environment. As for the effect of the number of children on the amount borrowed, higher number of children leads to lower mortgage amount. This effect is significant at higher quantiles of the distribution as suggested by the quantile regression results. At the left tail of the distribution, the number of children does not seem to affect the amount parents borrow. One possible explanation for the negative effect might be that as the number of children increases, so does the amount of money spent for their needs. This decreases the sum which can be used for the repayment. On the other hand, it is possible that people with more children wish to have a larger dwelling for them which is more expensive. Nevertheless, this effect is not probably that frequent since the estimated coefficients are always negative. One additional child decreases the mortgage amount by 45,000 CZK at the 0.5 quantile. The percentage effect is about 2.6%.

5.5 University Degree

People with either bachelor's or master's degree account for approximately one quarter of the examined sample. Having a university degree increases

the amount of money borrowed and the difference compared to people without university degree is statistically significant. University graduates borrow approximately 290,000 CZK or 17% higher amounts. A possible explanation could be as in the previous cases connected with income. Since people who have obtained a university degree are more likely to have a well-paid job, they can afford higher monthly repayment sums as it does not burden their income that much. Hypothesis 4 was phrased as follows: *People who have obtained a university degree borrow higher amounts compared to people with primary or secondary education.* In that case, I test the null hypothesis $H_0 : \beta_7 = 0$ against one-sided alternative $H_1 : \beta_7 > 0$. Based on the result of t test, I reject the null hypothesis that education does not have any effect on the amount a person borrows at 0.05 significance level. Thus, people with higher education tend to borrow higher amounts.

5.6 Income

Most of the effects that were examined could be linked with income. It is very likely that when people decide on the amount of money they would like to borrow, income is the main factor. As taking out a mortgage puts a burden on clients, they have to consider how much they borrow so that they do not face financial difficulties while repaying the debt. The effect of income is positive which means that higher income leads to higher amount borrowed. This suggests that richer people tend to borrow more money and therefore are more indebted regarding mortgages. A possible explanation for this could be that they have higher living standard and thus, they wish to buy a luxurious house which is more expensive. Since it takes some time to earn money for that even with high monthly salary, it is more convenient to borrow money and buy a new house immediately. The effect of income is not constant and depends on the level of person's salary. As people earn higher sums, an additional increase in income decreases the effect. Based on the results of model with logarithmic forms, 1% increase in salary leads to 0.15% growth in the amount borrowed. The hypothesis related to salary is formulated as follows: *People with higher income borrow higher amounts.* For testing this hypothesis, I use t statistics obtained by estimating the model with logarithmic form. I test $H_0 : \beta_8 = 0$ against one-sided alternative $H_0 : \beta_8 > 0$. The performed t test suggests that I can reject the null hypothesis at 0.05 significance level which means that higher income indeed leads to higher amount borrowed.

5.7 Region

The results of all models show that there is a disparity between amounts borrowed in Prague and in the rest of the Czech Republic. People living in regions outside the capital city borrow significantly lower amounts for their housing needs. A most likely explanation is that Prague is a desired locality with higher dwelling prices. People have to borrow more money so that they can afford to buy a property there. The average mortgage amount is approximately 500,000 CZK higher than in Central Bohemian Region, where people borrow the second highest amounts. It is possible that people who do not want to live directly in Prague decide to live in Central Bohemian Region as it is close to the capital city. On the other hand, people living in Zlín Region and Karlovy Vary Region borrow the lowest amounts. In both cases, it is approximately 60% lower compared to Prague. Since Prague residents account for more than one fifth of the sample, they seem to be more likely to take out a mortgage. Therefore, chi-squared test was performed and it confirmed that people from Prague borrow almost twice as much as they are supposed to based on the population distribution across regions. The last hypothesis mentioned in the previous chapter was stated as follows: *People living in Prague borrow higher amounts compared to people living outside Prague.* If I test each coefficient of dummy variable expressing the difference between Prague and the region a person lives in, I can reject the null hypothesis $H_0 : \beta_j = 0$ against $H_1 : \beta_j < 0$ at 0.05 significance level in all cases. In other words, I have found enough evidence to claim that people living in Prague tend to borrow higher amounts as the differences between Prague and other Czech regions are statistically significant.

5.8 Unbiasedness

The coefficients obtained from Models I-V were in most cases very similar. Therefore, the analysis provided consistent results which should be plausible. Since I examined only ČSOB clients, it is possible that the results might be biased. Nevertheless, Hypoteční banka, a.s. provided me with the summary statistics of its clients who have taken out a mortgage between 2004 and 2014. These additional data include information about the average age, ratio of men and women, education and average amount of mortgage. The average age is 35 years which is identical with the average age of the analysed sample. The same applies to the proportion of men and women which is 7:3 in both datasets.

The average amount of mortgage taken is 1,628,705 CZK. This is a little bit lower than the average value in the ČSOB which might be affected by outlying observations. Since the values are almost identical, I assume that the analysed sample provided by ČSOB could represent the whole population of people who have taken out a mortgage and therefore the estimated results should not be biased.

Chapter 6

Conclusion

The aim of this thesis was to examine borrower's characteristics and their effect on household debt in the Czech Republic. Since the growth of mortgage loans has greatly contributed to the increase in household indebtedness, I focused on the analysis of factors which influence the mortgage amount that people borrow in order to finance their housing needs as mortgages are usually the largest financial burden that has to be repaid. Most studies analysing household indebtedness considered the effect of personal characteristics on the probability of having a loan or being in arrears while repaying. In this thesis, I examined dataset of ČSOB clients who have taken out a mortgage in order to determine which factors affect how much money they decided to borrow. This is a different approach to household debt compared to the previous studies which could possibly discover new interesting facts about a borrower's choice.

Firstly, I examined the dataset in detail in order to discover patterns which could provide some information about the distribution of people in the analysed sample. Subsequently, I performed a multiple regression analysis so that it is possible to assess not only each variable's effect, but also its magnitude and significance. Since the dataset included a few observations with extremely high values compared to the rest of the sample, estimation methods used for the analysis had to be adjusted so that the obtained results were not affected by unusual values. I estimated 4 models using OLS and LAD. Additionally, the fifth quantile regression model was estimated in order to analyse whether the effects of personal characteristics on the amount borrowed vary across the distribution. I discovered that the mortgage amounts people borrow increased with income. Furthermore, having a university degree also positively affects the amount borrowed. A person's age has firstly a positive effect which becomes

negative at the age of 36. As for gender, men borrow significantly higher amounts in comparison with women. Based on the composition of the examined sample, it appears that people who borrow are very frequently childless suggesting that the number of children affects whether a person takes out a mortgage or not, but the results show that the number of children does not have any effect on the amount borrowed at the left tail of the distribution. On the other hand, the number of children negatively affects parents' decision on the amount they choose to borrow at the higher quantiles. Additionally, there is not any difference between amounts borrowed by married and single people, whereas divorced people take out the lowest mortgages. Finally, I found a disparity among regions in the Czech Republic showing that people living in Prague owe more money for mortgage loans.

This analysis was partly limited by the unavailability of data about total household debt and mortgage debt which would contain personal information about debtors. Therefore, this research could be extended depending on the structure of data. One possible improvement of the performed analysis could be the influence of work status on the amount borrowed, since this variable was not used in the analysis. Another extension which could complement this research question is analysing the probability whether a person takes out a mortgage depending on their personal characteristics. In that case, the examined dependent variable would be a dummy variable expressing whether a person has taken out a mortgage or not. For this approach, the analysis would not be performed on people who have taken out a mortgage, but on a random sample taken from the whole population. Another recommendation for future research could be assessing the effects of borrower's characteristics on the overall level of debt without focus on mortgages, although mortgage loans account for the highest proportion of household debt.

Bibliography

[1] Ales Chmelar. Household debt and the European crisis. *European Credit Research Institute*, Research Paper No. 13, 2013.

[2] Elaine Kempson, Stephen McKay, and Maxine Willitts. *Characteristics of families in debt and the nature of indebtedness.* Number 211. Corporate Document Services, 2004.

[3] Czech National Bank. Time series database - ARAD. URL `http://www.cnb.cz/cnb/STAT.ARADY_PKG.VYSTUP?p_period=1&p_sort=1&p_des=50&p_sestuid=1538&p_uka=1%2C2%2C3&p_strid=AAD&p_od=200512&p_do=201602&p_lang=EN&p_format=0&p_decsep=.` Accessed on April 12, 2016.

[4] Alena Bičáková, Zuzana Prelcová, and Renata Pašaličová. Who borrows and who may not repay? *CERGE-EI Working Paper Series*, (443), 2011.

[5] Guido Wolswijk. On some fiscal effects on mortgage debt growth in the EU. *ECB working paper*, (526), 2005.

[6] Mario Fortin and André Leclerc. Determinants of mortgage indebtedness in Canada. *Cahier de recherche/Working Paper*, 7:11, 2007.

[7] Allan Crawford and Umar Faruqui. What explains trends in household debt in Canada? *Bank of Canada Review*, 2011(Winter):3–15, 2011.

[8] Dimitris Georgarakos, Adriana Lojschova, and Melanie E. Ward-Warmedinger. Mortgage indebtedness and household financial distress. *European Central Bank Working Paper Series*, (1156), 2010.

[9] Christopher Mayer, Karen Pence, and Shane M. Sherlund. The rise in mortgage defaults. *The Journal of Economic Perspectives*, 23(1):27–50, 2009.

[10] Roberto G. Quercia and Michael A. Stegman. Residential mortgage default: a review of the literature. *Journal of Housing Research*, 3(2):341, 1992.

[11] Helen F. Ladd. Evidence on discrimination in mortgage lending. *The Journal of Economic Perspectives*, 12(2):41–62, 1998.

[12] Patrick Bayer, Fernando Ferreira, and Stephen L. Ross. Race, ethnicity and high-cost mortgage lending. Technical report, National Bureau of Economic Research, 2014.

[13] Jason Dietrich and Hannes Johannsson. *Searching for age and gender discrimination in mortgage lending.* Comptroller of the Currency, 2005.

[14] Silvia Magri. Arrears on mortgages: Differences across countries and their effect on the pricing of the loan. In *22nd Australasian Finance and Banking Conference*, 2009.

[15] Orla May and Merxe Tudela. When is mortgage indebtedness a financial burden to British households? A dynamic probit approach. *Bank of England Working Paper Series*, (277), 2005.

[16] Sarah Nettleton and Roger Burrows. Mortgage debt, insecure home ownership and health: an exploratory analysis. *Sociology of health & Illness*, 20(5):731–753, 1998.

[17] Matthias Keese and Hendrik Schmitz. Broke, ill, and obese: The effect of household debt on health. *Ruhr Economic Paper*, (234), 2011.

[18] Brahima Coulibaly and Geng Li. Choice of mortgage contracts: evidence from the survey of consumer finances. *Real Estate Economics*, 37(4):659–673, 2009.

[19] Frederick Furlong, Yelena Takhtamanova, and David Lang. Mortgage choice in the housing boom: impacts of house price appreciation and borrower type. Technical report, Citeseer, 2014.

[20] James Vickery. Interest rates and consumer choice in the residential mortgage market. *Federal Reserve Bank of New York*, 2007.

[21] Mardi Dungey. Mortgage product choice in Australia: The impact of market stress. *JASSA*, (4):44, 2014.

[22] Mardi Dungey, Firmin Doko Tchatoka, Graeme Wells, and María B. Yanotti. Mortgage choice determinants: The role of risk and bank regulation. *Economic Record*, 91(295):417–437, 2015.

[23] Maria Hullgren. Mortgage rate choice: the effect of bank advice. *KTH, Building and Real Estate Economics*, 2013.

[24] Maria Hullgren and Inga-Lill Söderberg. The relationship between consumer characteristics and mortgage preferences: a case study from Sweden. *International Journal of Housing Markets and Analysis*, 6(2):209–230, 2013.

[25] Olympia Bover, José María Casado, Ernesto Villanueva, Sonia Costa, Philip Du Caju, Yvonne McCarthy, Eva Sierminska, Panagiota Tzamourani, and Tibor Zavadil. The distribution of debt across euro area countries: the role of individual characteristics, institutions and credit conditions. *European Central Bank Working Paper*, (1639), 2013.

[26] Luísa Farinha. The effect of demographic and socioeconomic factors on households' indebtedness. *Banco de Portugal Economic Bulletin*, 2003.

[27] Czech Statistical Office. České domácnosti dluží bilion, 2010. URL `https://www.czso.cz/csu/czso/cri/ceske-domacnosti-dluzi-bilion-2009-n7dnrbgk2s`. Accessed on April 16, 2016.

[28] ČSOB. Mortgage, 2015. URL `http://www.csob.cz/portal/people/products/mortgages/mortgage`. Accessed on March 17, 2016.

[29] Czech National Bank. *CNB Financial Stability Report 2014/2015*. 2015.

[30] Czech Statistical Office. Czech demografic handbook 2014, 2015. URL `https://www.czso.cz/documents/10180/20548157/130055151111.pdf/3a4b5fcc-3459-4416-a067-f91bd2863ceb?version=1.0`. Accessed on March 18, 2016.

[31] Peter J. Rousseeuw and Annick M. Leroy. *Robust Regression & Outlier Detection*. JSTOR, 1987.

[32] Jeffrey Wooldridge. *Introductory econometrics: A modern approach*. Nelson Education, 2015.

[33] William H. Greene. *Econometric analysis*. Pearson Education Limited, 2012.

[34] A. Colin Cameron and Pravin K. Trivedi. *Microeconometrics: methods and applications*. Cambridge university press, 2005.

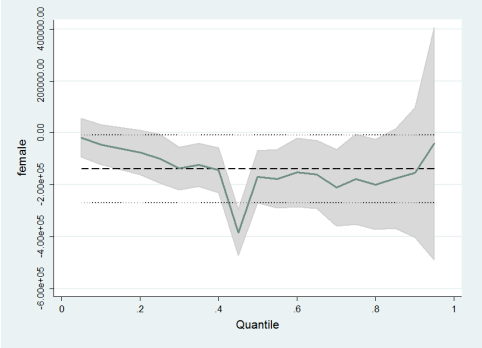
[35] José A.F. Machado and João M.C Santos Silva. Glejser's test revisited. *Journal of Econometrics*, 97(1):189–202, 2000.

[36] José A.F. Machado, Paulo M.D.C. Parente, and João M.C. Santos Silva. qreg2: Stata module to perform quantile regression with robust and clustered standard errors. *Statistical Software Components*, 2015.

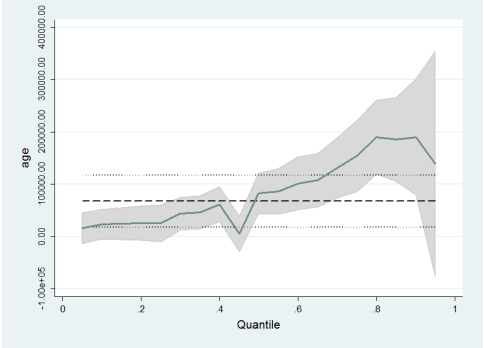
[37] Roger Koenker and Kevin F. Hallock. Quantile regression. *Journal of Economic Perspectives, 15(4):143-156*, 2001.

Appendix

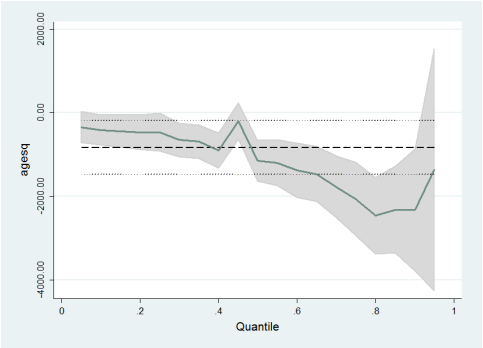
In this part, figures and tables which complement the econometric analysis and summarize the results are presented. Figure A1 shows the quantile regression results so that it is possible to see the variation in coefficients across the distribution. Table A1 provides the summary of all coefficients estimated by OLS and LAD, whereas the results of quantile regression analysis are presented separately in Table A2.



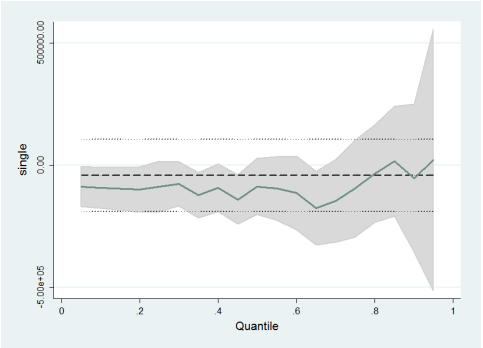
(a) Coefficients of *female*



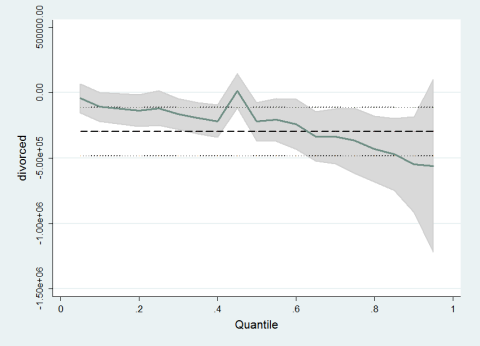
(b) Coefficients of *age*



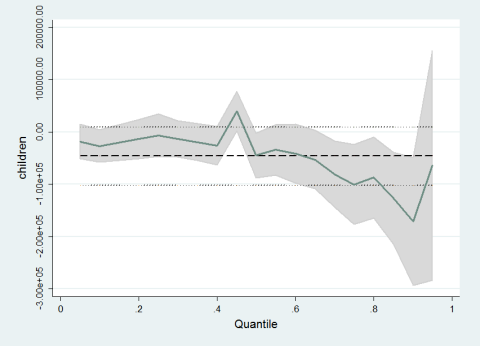
(c) Coefficients of *agesq*



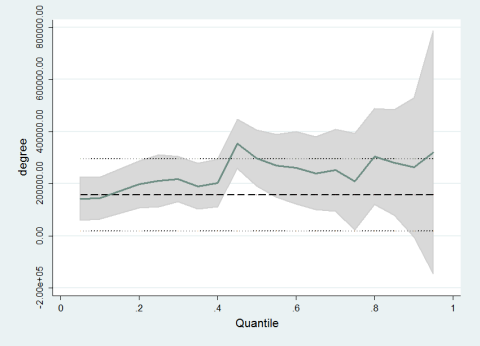
(d) Coefficients of *single*



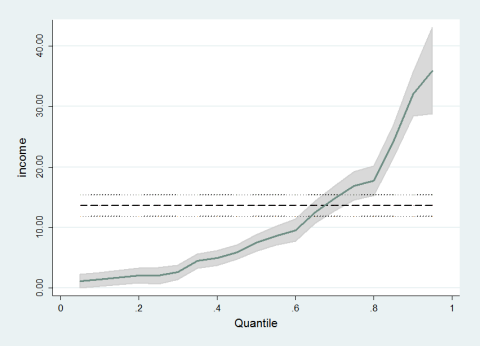
(e) Coefficients of *divorced*



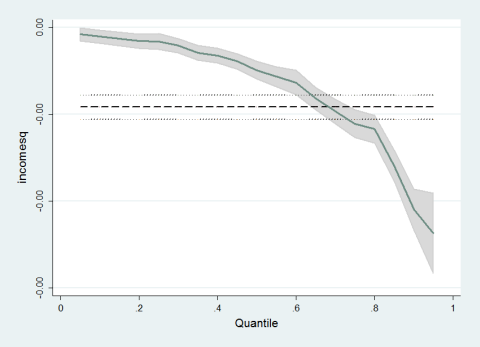
(f) Coefficients of *children*



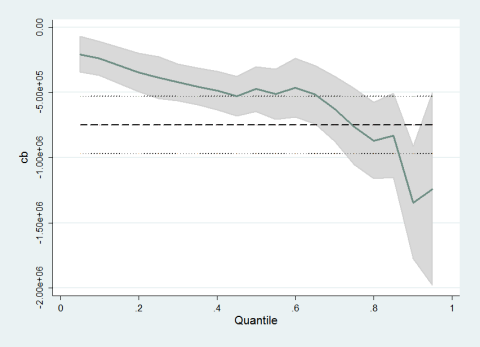
(g) Coefficients of *degree*



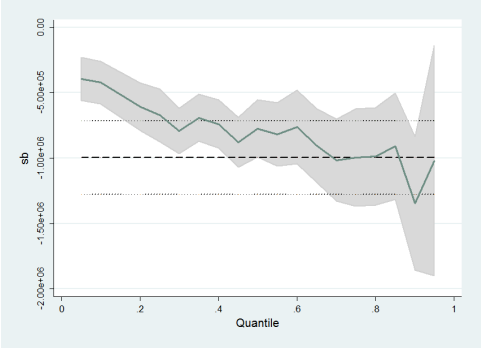
(h) Coefficients of *income*



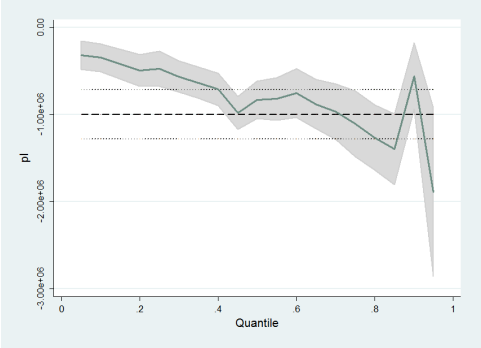
(i) Coefficients of *incomesq*



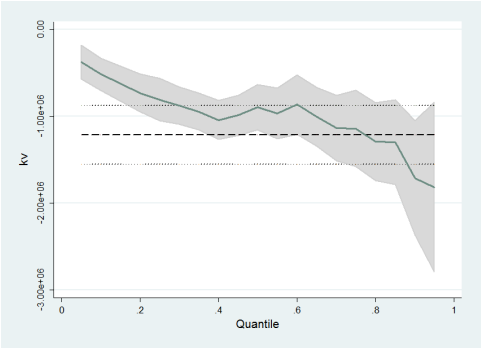
(j) Coefficients of *cb*



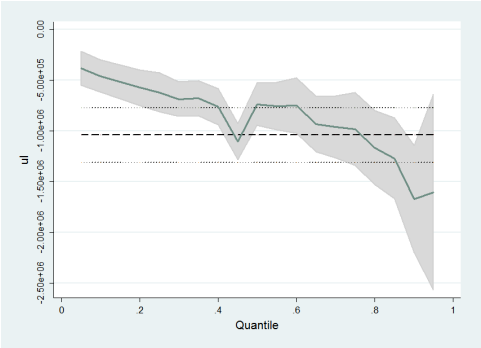
(k) Coefficients of *sb*



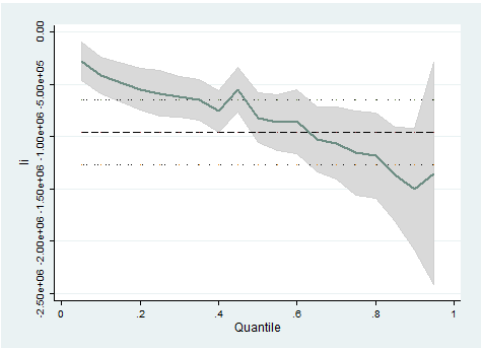
(l) Coefficients of *pl*



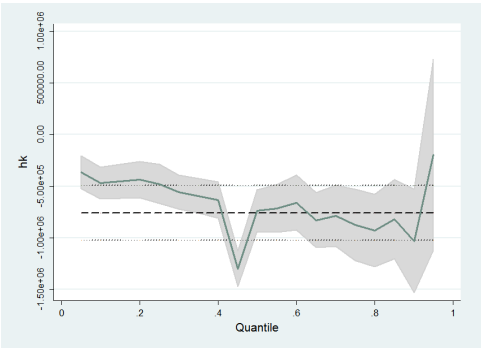
(m) Coefficients of *kv*



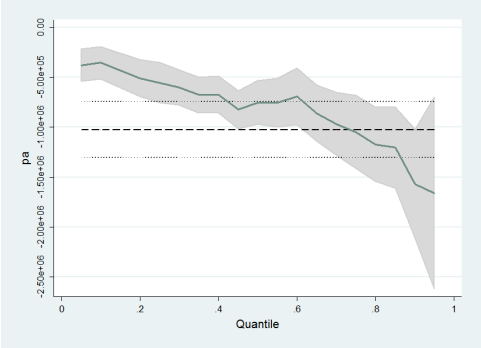
(n) Coefficients of *ul*



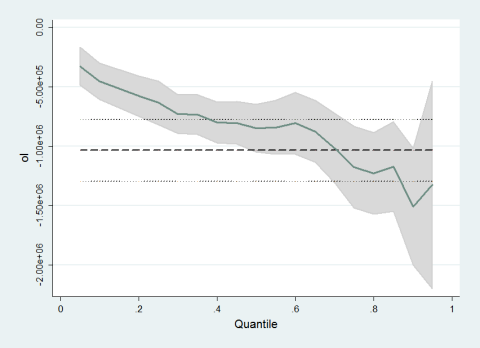
(o) Coefficients of *li*



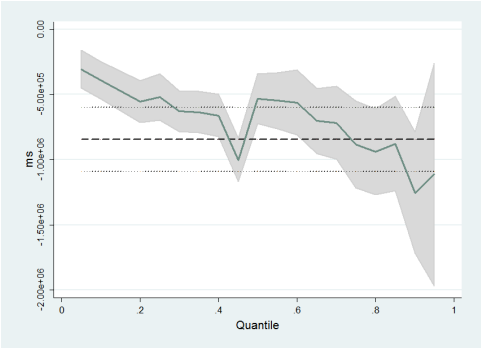
(p) Coefficients of *hk*



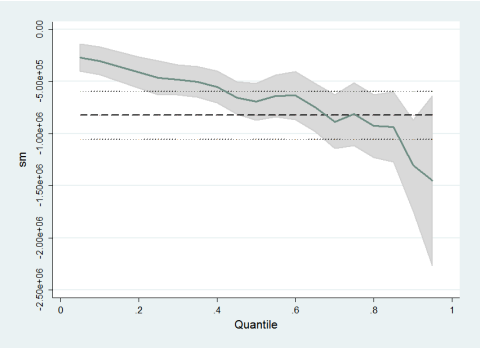
(q) Coefficients of *pa*



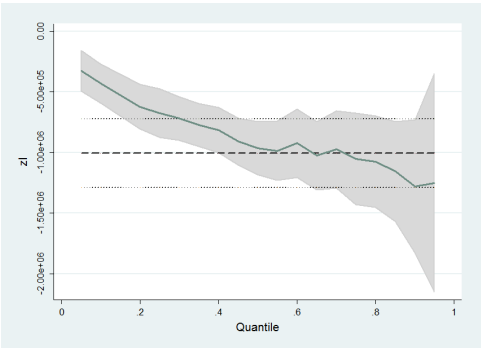
(r) Coefficients of *ol*



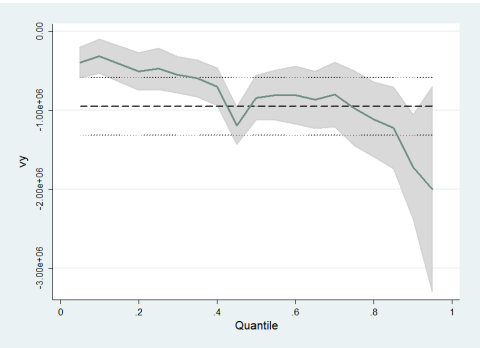
(s) Coefficients of *ms*



(t) Coefficients of *sm*



(u) Coefficients of *zl*



(v) Coefficients of *vy*

Figure A1: Development of coefficients across the distribution

	OLS Model I log(mortgage)	OLS Model II mortgage	LAD Model III log(mortgage)	LAD Model IV mortgage
female	-0.0884**	-154,644.5**	-0.114**	-167,919.5***
	(-2.94)	(-3.15)	(-2.97)	(-3.41)
age	0.0522**	112,783.2***	0.0638***	82,523.8***
	(3.22)	(6.32)	(4.35)	(4.57)
agesq	-0.000722**	-1,522.9***	-0.000888***	-1144.2***
	(-3.29)	(-6.57)	(-4.72)	(-4.87)
single	-0.0619	-82,451.5	-0.0714	-87,483.7
	(-1.81)	(-1.40)	(-1.63)	(-1.58)
divorced	-0.171***	-259,453.0***	-0.194***	-220,156.5**
	(-3.82)	(-3.78)	(-3.52)	(-3.03)
children	-0.0271	-41,998.7	-0.0255	-44,935.5*
	(-1.87)	(-1.56)	(-1.55)	(-2.11)
degree	0.157***	293,257.6***	0.185***	298,368.8***
	(4.74)	(4.87)	(4.54)	(4.59)
log(income)	0.174***		0.142***	
	(6.86)		(5.51)	
income		10.48***		7.456**
		(4.30)		(3.27)
incomesq		-0.0000287		-0.00000493***
		(-1.69)		(-3.45)
cb	-0.298***	-567,773.8***	-0.267***	-473,840.3***
	(-5.80)	(-6.28)	(-4.09)	(-4.92)
sb	-0.516***	-782,743.9***	-0.472***	-773,331.1***
	(-7.65)	(-7.12)	(-5.76)	(-6.67)
pl	-0.471***	-789,514.8***	-0.525***	-831,575.3***
	(-7.68)	(-7.60)	(-6.36)	(-8.93)
kv	-0.648***	-1,014,686.3***	-0.616***	-898,262.0***
	(-7.69)	(-8.81)	(-6.22)	(-5.42)
ul	-0.528***	-795,782.7***	-0.473***	-738,293.7***
	(-7.87)	(-7.52)	(-5.95)	(-6.41)
li	-0.476***	-753,290.1***	-0.515***	-823,331.4***
	(-6.92)	(-5.86)	(-5.70)	(-8.08)
hk	-0.392***	-640,927.1***	-0.465***	-737,316.3***
	(-5.72)	(-5.47)	(-6.00)	(-7.59)
pa	-0.470***	-808,318.1***	-0.465***	-752,379.2***
	(-7.60)	(-8.47)	(-5.63)	(-7.07)
ol	-0.520***	-838,928.4***	-0.534***	-847,120.9***
	(-8.50)	(-8.74)	(-7.02)	(-7.34)
ms	-0.392***	-634,179.4***	-0.322***	-533,571.2***
	(-6.65)	(-6.14)	(-4.44)	(-4.15)
sm	-0.346***	-627,472.4***	-0.416***	-694,025.8***
	(-7.12)	(-6.57)	(-6.22)	(-8.39)
zl	-0.517***	-784,903.6***	-0.650***	-965,433.6***
	(-7.63)	(-6.88)	(-7.84)	(-10.13)
vy	-0.472***	-760,295.8***	-0.524***	-840,863.8***
	(-5.64)	(-5.59)	(-4.88)	(-7.90)
_cons	11.89***	61,311.6	12.06***	506,334.4
	(33.21)	(0.18)	(31.62)	(1.50)
N	2,078	2,043	2,078	2,078

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A1: Estimated coefficients for Models I-IV

	Quantile regression				
	$q = 0.1$	$q = 0.25$	$q = 0.5$	$q = 0.75$	$q = 0.9$
	mortgage	mortgage	mortgage	mortgage	mortgage
female	-44,838.1	-99,431.5**	-167,919.5***	-177,860.6*	-152,903.9
	(-1.32)	(-2.69)	(-3.41)	(-2.49)	(-1.22)
age	23,449.9	25,312.8	82,523.8***	153,993.6***	190,407.3***
	(1.66)	(1.72)	(4.57)	(5.92)	(3.69)
agesq	-407.1*	-466.0*	-1144.2***	-2,064.0***	-2,329.6**
	(-2.22)	(-2.47)	(-4.87)	(-6.39)	(-3.21)
single	-92,655.5*	-87,060.6	-87,483.7	-95,931.6	-53,504.6
	(-2.12)	(-1.81)	(-1.58)	(-0.97)	(-0.33)
divorced	-108,280.9*	-117,627.6*	-220,156.5**	-366,830.1***	-548,755.0**
	(-2.32)	(-2.38)	(-3.03)	(-3.90)	(-2.75)
children	-26,983.9	-6,264.5	-44,935.5*	-100,571.3**	-171,462.4**
	(-1.58)	(-0.35)	(-2.11)	(-2.92)	(-2.62)
degree	144,306.2***	211,596.4***	298,368.8***	208,830.5*	262,882.0
	(3.30)	(4.39)	(4.59)	(2.28)	(1.45)
income	1.432*	2.033**	7.456**	16.88***	32.08***
	(2.13)	(2.78)	(3.27)	(9.37)	(8.23)
incomesq	-0.00000106*	-0.00000163**	-0.00000493***	-0.0000111***	-0.0000210***
	(-2.05)	(-2.78)	(-3.45)	(-9.68)	(-8.52)
cb	-239,116.5**	-386,784.7***	-473,840.3***	-761,475.7***	-1,344,335.1***
	(-3.22)	(-5.21)	(-4.92)	(-4.39)	(-4.88)
sb	-422,314.2***	-672,097.8***	-773,331.1***	-994,506.9***	-1,345,294.9***
	(-6.06)	(-7.74)	(-6.67)	(-3.48)	(-4.87)
pl	-346,750.4***	-473,052.7***	-831,575.3***	-1,051,413.7***	-556,729.7
	(-3.97)	(-5.85)	(-8.93)	(-5.92)	(-0.59)
kv	-516,713.9***	-809,811.2***	-898,262.0***	-1,140,968.1***	-1,711,028.9***
	(-6.49)	(-9.69)	(-5.42)	(-4.91)	(-5.67)
ul	-460,784.0***	-621,973.7***	-738,293.7***	-982,524.0***	-1,672,545.3***
	(-5.69)	(-6.60)	(-6.41)	(-5.08)	(-5.98)
li	-417,991.1***	-586,114.1***	-823,331.4***	-1,158,419.1***	-1,501,001.3**
	(-5.45)	(-6.65)	(-8.08)	(-5.31)	(-2.96)
hk	-469,770.3***	-477,970.7***	-737,316.3***	-876,131.5***	-1,032,575.8**
	(-5.58)	(-5.47)	(-7.59)	(-4.24)	(-2.71)
pa	-353,619.1***	-553,001.1***	-752,379.2***	-1,046,218.9***	-1,567,291.6***
	(-4.17)	(-7.06)	(-7.07)	(-5.75)	(-5.22)
ol	-452,821.6***	-634,418.6***	-847,120.9***	-1,174,880.9***	-1,510,356.7***
	(-6.04)	(-7.88)	(-7.34)	(-6.19)	(-4.83)
ms	-389,041.5***	-521,580.9***	-533,571.2***	-883,653.6***	-1,252,985.1***
	(-5.35)	(-6.26)	(-4.15)	(-4.62)	(-4.08)
sm	-301,885.2***	-466,305.7***	-694,025.8***	-812,751.1***	-1,306,050.3***
	(-4.10)	(-5.69)	(-8.39)	(-3.88)	(-4.78)
zl	-430,688.4***	-676,925.8***	-965,433.6***	-1,051,980.9***	-1,280,382.9***
	(-6.02)	(-8.26)	(-10.13)	(-5.22)	(-3.76)
vy	-316,566.3**	-475,833.8***	-840,863.8***	-971,860.4***	-1,718,655.9***
	(-2.69)	(-5.16)	(-7.90)	(-4.53)	(-5.72)
_cons	682,355.7*	1,090,238.1***	506,334.4	-102,498.8	-197,057.3
	(2.46)	(3.89)	(1.50)	(-0.20)	(-0.22)
N	2,078	2,078	2,078	2,078	2,078

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A2: Estimated coefficients for Model V