Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

Credit Burden of Households in Slovakia

Diploma Thesis

Adam Biroš

Bratislava 2011

525ee80a-3c15-4902-b736-9cfdd96015d6

Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

Credit Burden of Households in Slovakia

Diploma Thesis

Adam Biroš

Department of Applied Mathematics and Statistics
9.1.9 Applied Mathematics
Mathematics of Economics and Finance

Supervisor:
Doc. Dr. Jarko Fidrmuc

Bratislava, 2011

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

Úverové zaťaženie domácností na Slovensku

Diplomová práca

Adam Biroš

Katedra aplikovanej matematiky a štatistiky
9.1.9 Aplikovaná matematika
Ekonomická a financná matematika

Vedúci diplomovej práce:
Doc. Dr. Jarko Fidrmuc

Bratislava, 2011

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Adam Bíroš
Študijný program: ekonomická a finančná matematika (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: 9.1.9. aplikovaná matematika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: anglický

Názov : Credit Burden of Households in Slovakia

Cieľ : Estimation of determinants of credit burden in Slovakia using consumption expenditure data of households

Vedúci : doc. Ing. Jarko Fidrmuc, Dr.

Dátum zadania: 11.02.2010

Dátum schválenia: 08.04.2011

....
prof. RNDr. Daniel Ševčovič, CSc.
garant študijného programu

....
študent

....
vedúci práce

Dátum potvrdenia finálnej verzie práce, súhlas s jej odovzdaním (vrátane spôsobu sprístupnenia)

....
vedúci práce

I declare that this thesis was written on my own, with the only help provided by my supervisor and the reffered-to literature.

Adam Biroš

Acknowledgement

I would like to express thanks to my supervisor Doc. Dr. Jarko Fidrmuc, for all of the support and guidance he offered throughout the elaboration of this thesis. I would also like to thank my family for their support and love.

Abstract

We estimate the determinants of the credit burden of households using EU-SILC database from 2005 and 2006. We opt for empirical approach of the Heckman selection model to control for potential selection bias. We find that especially responsibility of the households significantly lowers the probability of high credit burden. Capital adequacy has significant effect on access to credits but its effect on credit burden in negligible. High liquidity and also better earnings ability indicate lower probability of credit burden and are important determinants of households' credit burden. We show that asymmetric information between borrowers and lenders and soft information are important factors that should be used wisely to avoid high credit burden of households. Finally, we find that the credit burden is dependent on trend of economy and it is less severe during boom times.

Keywords: households' loans, Heckman selection model, credit burden, asymmetric information, hard and soft information, informational economics

Abstrakt

Odhadujeme determinanty úverového zaťaženia domácností pomocou databázy EU-SILC z rokov 2005 a 2006. Rozhodli sme sa pre empirický prístup Heckmanovho výberového modelu, aby sme sa vyhli problémom s výberovou odchýlkou. Zistili sme, že hlavne úroveň zodpovednosti finančného správania signifikantne znižuje pravdepodobnosť vysokého úverového zaťaženia. Kapitálová primeranosť má signifikantný vplyv na dostupnosť úverov, ale efekt kapitálovej primeranosti na úverové zaťaženie je zanedbateľný. Vysoká likvidita, ako aj lepšia schopnosť zarábať znamenajú nižšiu pravdepodobnosť úverového zaťaženia a sú dôležité determinanty úverového zaťaženia domácností. Ukázali sme, že asymetrická informácia medzi bankami a ich klientami, ako aj tzv. mäkká informácia sú dôležité faktory, ktoré by banky mali brať do úvahy, aby sa predišlo vysokému úverovému zaťaženiu domácností. Zistili sme tiež, že úverové zaťaženie je závislé na vývoji ekonomiky a je nižšie počas konjunktúry.

Kľúčové slová: úvery domácností, Heckmanov výberový model, úverové zaťaženie, asymetrická informácia, mäkká a tvrdá informácia, ekonómia informácií

Contents

Chapter 1

Introduction

Households' credits were characterized by high growth rates in Slovakia as well as in other new member states of European Union in 2000's. Various authors analyze credits and possible explanations for rapid credit growth. However, this development has been analyzed for aggregated macroeconomic data in previous literature.

In general, there exist two trends to explain this phenomenon: Some economists argue that the rapid credit growth in emerging economies is a natural catching-up process that leads to equilibrium level of credit-to-GDP ratios. Contrary to equilibrium level theory, some authors present an idea that the fast growth of credits is not sustainable in the medium to long run and is better described as a boom-bust feature of growing economies. We present both approaches and try to find out later in empirical analysis which scenario is more likely.

There exist a number of papers that analyze credit growth using macro data some of which we mention throughout this thesis. Similarly, there are some studies analyzing micro data. For instance, paper of Hainz and Nabokin [16] use micro data and similar empirical approach than we do but instead of households they analyze private enterprises. In this paper we analyze The European Union Statistics on Income and Living Conditions (EU-SILC) dataset from 2005 and 2006. It is a unique dataset that contains consumption expenditure data of households. The idea of our paper is to analyze loan determinants and determinants of the credit burden. We want to find out whether banks were acting responsible when granting credits to households or if the rapid credit growth was caused by banks' expansion strategy.

We intend to find out how succesful the banks were in assessing their clients before they granted them credits. We estimate which personal and household characteristics are important loan determinants. We want to find out whether banks can overcome the bank risk connected with asymmetric information between borrower and lender. We test the hypothesis that soft information is important determinant of the credit burden that should be included into banks' scoring models.

Finally, we want to find out what are the effects of upswing on the credit burden. We want to perform this by comparing results from year 2005 with results from 2006, taking into account that GDP growth rate was higher in 2006.

The paper is organized as follows. In chapter 2, we provide literature overview and depict the theoretical background to motivate our analysis. In chapter 3, we describe the empirical strategy. In chapter 4, we describe our dataset. The results from the empirical analysis are presented in chapter 5. Chapter 6 concludes.

Chapter 2

Literature overview

2.1 Credit Boom and Income Convergence

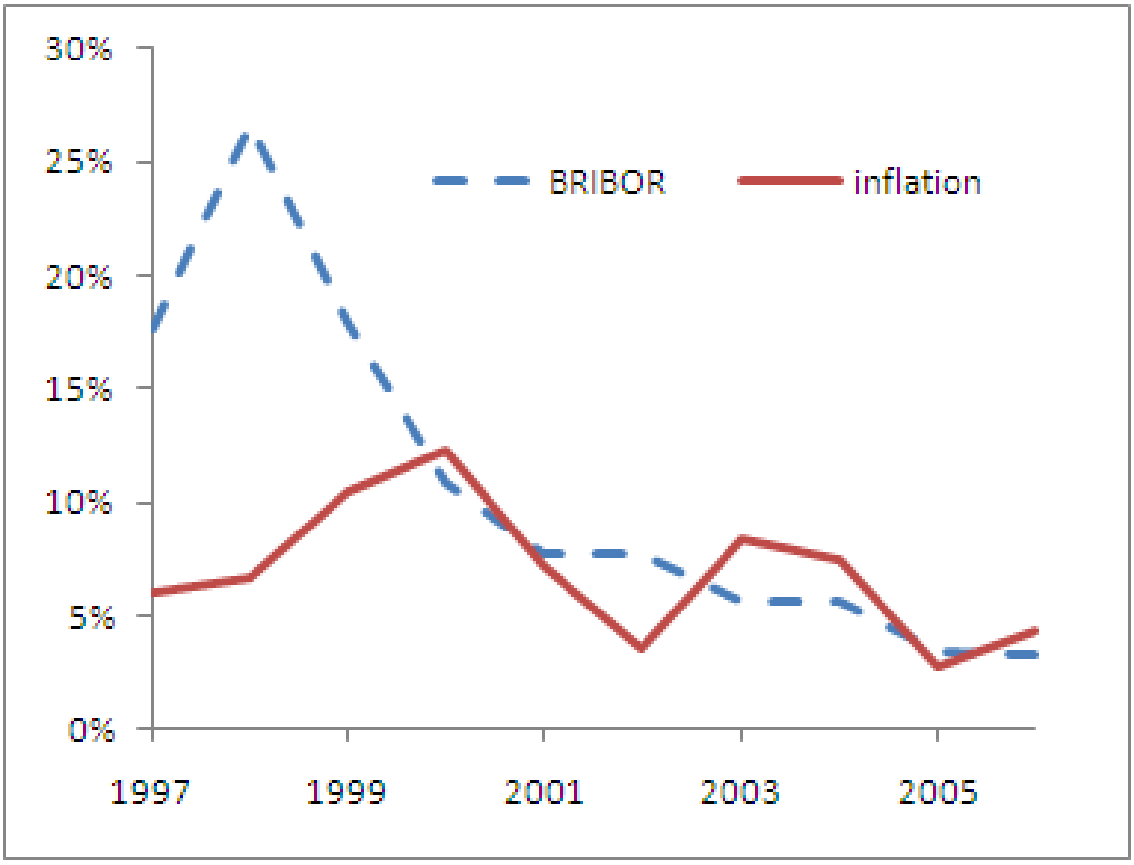
There was a massive growth of credit level in Slovakia since 2000's. There are various factors that help to explain this phenomenon:

- Initial private credit level to gross domestic product ratio in Slovakia as a transition economy was low
- This ratio can be below its equilibrium level
- Credit installment payments decline as interest rate declines. As a consequence, households can borrow more without increasing the repayment burden. This point is of particular importance in contest of our research of credit burden in Slovakia.

The above reasons often appear in an optimistic approach literature on credits stating that private credit-to-GDP ratios in transition economies and emerging economies can still be below their long-run equilibrium level. Égert, Backé and Zumer [2] analyze credit growth in Central and Eastern Europe as a catching-up process that leads to equlibrium level. They define equilibrium level as the level of private credit, which would be justified by the economic fundamentals. Their paper specifically focuses on situation where initial credit-to-GDP level is out of tune with economic fundamentals. They call the situation where the initial credit-to-GDP ratio is higher than what the level of economic development would justify "initial overshooting", if this ratio is lower it is "initial undershooting". Regarding Slovakia, they found out that the initial overshooting might not have been too large and that the CEE countries' credit-to-GDP ratios are still below the equilibrium levels.

Boissay, Calvo-Gonzalez and Koźluk [30] analyze concerns from a financial and macroeconomic stability perspective raised by the fast pace credit growth in Central and South-Eastern European countries. Their paper provides an econometric analysis of the macroeconomic determinants of the growth of credit for 11 transition countries. Authors model credit growth as a function of both macroeconomic fundamentals and the gap between the actual credit-to-GDP ratio and an equilibrium level. They found out that even accounting for a rising trend in the equilibrium credit-to-GDP ratio, a number of countries in the region have experienced "excessive" credit growth in the sense that the observed credit growth is higher than what we would have expected given the evolution of macroeconomic variables.

Figure 1: Evolution of interest rates and inflation in Slovakia



Source: NBS statistics [13], http://www.inflation.eu/ [14]

Figure 1 illustrates the development of interest rates and inflation in Slovak Republic from 1997 to 2006. Note the double-digit rates in the beginning of the period that were gradually declining. Inflation is oscilating but a slight negative trend is visible. From **Figure 1** follows that real interest rates were close to zero. This is one of the reasons why credits were so demanded in 2000s - borrowes took advantage of low real interest rate which made repaying easier.

2.2 Boom and Bust Cycles

Contrary to equilibrium level theory, especially after the financial crisis, some authors present an idea that the fast growth of credits is not sustainable in the medium to long run and is better described as a boom-bust feature of growing economies.

Mendoza and Terrones [12] define a credit boom as an episode in which credit to the private sector grows by more than during a typical business cycle expansion. They propose a method for identifying credit booms, and implement it to study the microeconomic and macroeconomic characteristics of credit booms in industrial and emerging economies. They identify a credit boom as an episode in which credit exceeds its long-run trend by more than a given "boom" threshold, with the duration of the boom set by "starting" and "ending" thresholds. Their results present differences between boom-bust cycles between industrial and emerging economies:

- credit booms and the macro and micro fluctuations associated with them are larger in emerging economies, particularly in the nontradables sector
- not all credit booms end in financial crises, but most emerging markets crises were associated with credit booms
- credit booms in emerging economies are often preceded by large capital inflows but not by financial reforms or productivity gains

Jeanne and Korinek [26] introduce their paper by describing the interaction between debt accumulation and asset prices which contributes to magnify the impact of booms and busts. During booms, increases in borrowing stimulates increasing of collateral prices and vice versa. During busts, credit constrains lead to quick sales of assets and further tightening of credit. They present a model to study the optimal policy responses to booms and busts in credit and asset prices. They found out that agents' borrowing choices in boom times render the economy more vulnerable to credit and asset price busts involving debt deflation in bust times.

2.3 Hard and soft information

Households' credits are indeed of high importance in Slovakia since its financial systems depend maily on banks and less on financing though - so called bank-based (Égert, Backé, Zumer [2]). Commercial banks are the main providers of households' loans. It is essential for the bank to know the claimer well and so it is trying to gather as many information about her as possible.

Our unique dataset contains valuable and detailed information about individual households and members that are living in each household. Besides information that banks collect when assessing credit applicants (income, marital status, education etcetera) we have a variety of information that the banks cannot collect and which they do not employ into scoring schemes. Therefore, in our analysis we are able to investigate further than banks and the results might be new and interesting. We want to test the hypothesis that soft information and asymmetric information is important in analysis of the credit burden.

The terms of hard information and soft information are used in various papers, however, they were not properly and completely defined until 2004 Petersen's paper. In his article, Petersen is naming characteristics which define hard and soft information (see Petersen, [6]):

- Hard information is nearly always represented by numbers while the soft information is often communicated in text. Soft information includes opinions, ideas, rumors etcetera.

- The collection method of hard information need not to be personal. Instead it can be collected by a questionaire using form without the assistance or guidance from a human data collector. Data that we use in this paper were also colected by a questionaire although not every element of the data being collected can be classified as hard information.

- Hard information is more comparable than soft information. This is natural in respect of the representation of hard information (numbers) and soft information (text). Thus the person who collects hard information can be different than the person who evaluates the information and makes a decision.

- Soft information can be assessed by creating a numerical score. For instance, the variable that we use in our analysis *affordability of consumption* is indexed from 1 to 6. However, this does not make this information hard because the interpretation of two different persons as for what is easy and what is difficult varies. As Petersen highlights, with soft information the context under which it is collected and the collector of the information are part of the information. That is why soft information is (or should be) collected in person and the decision maker is usually the same person as information collecter.

- With hard information, the collection and use of information can be separated at different management levels.

Banks that are collecting information about the credit applicant should be familiar with the differences between hard and soft information. Hard information e.g. current account balance, monthly income or family status can be verified, easily recorded as numbers and compared. On the other hand, banks also should collect soft information such as applicants' health. While it may seem that financial intermediaries should not rely on soft information, in our opinion banks should try to develop advanced methods for making use of soft information.

2.4 Asymmetric information

Asymmetric information occurs when one economic agent knows something that another economic agent does not know (Varian [10]).

We present a model example of asymmetric information. The example is designed for simplicity with the purpose to outline the problem although the reality is more complex. Consider a credit market with 2 types of borrowers - solvent and risky. In population there is s of solvent clients and $1-s$ of risky clients. Solvent and risky borrowers are willing to accept interest rates of 6% and 12%, respectively. Banks would charge solvent clients 5% and risky borrowers 10% provided they could differenciate between them. However, if there is information asymmetry and the quality of client is hidden from the bank, clients have to prepare for interest rate of $5\% * s + 10\% * (1-s) = 10\% - 5\% * s$. Solvent clients will borrow from bank only if $10\% - 5\% * s \leq 6\%$ which holds if $s \geq 0.8$. In other words, solvent clients will be interested in banks' credit only in there is less than 20% of risky clients in the population. If there is more than 20% of risky borrowers, the interest rate charged by banks will be too high for solvent clients to accept. As a result, solvent clients are driven out of the credit market. They will not borrow from banks. The only borrowers will be risky clients and thus banks will adjust the interest rate at 10%. From this example it is clear that the problem of information asymmetry is not solely about the bank risk but also about the credit demand and health of banking sector.

What occured in our example is called **adverse selection** which takes place when borrowers and banks have asymmetric information. This situation leads to undesirable results because it is in banks' interest to have solvent clients in their loan portfolios to avoid the bank risk. Moreover, such a situation has negative impact on economy because some people (in our example solvent clients) are excluded from lending process and therefore they cannot afford consumption that they would desire. Demand for money falls, demand for goods falls afterwards and this further causes decrease of production, rise of unemployment and overall economic depression.

Asymmetric information between borrowers and lenders appears in various forms. A debt contract establishes the legal rights and obligations for those who receive financing (borrowers - households) and those who provide it (lenders - financial intermediaries). Essentially, the borrower promises to repay the principal plus required interest in an agreed amount of time (Bebczuk (2003) [11]). The banks wants to assess the riskiness of the client to provide a suitable contract for her.

In the first place, various conditions and household characteristics puts the household's ability to repay in question. This problem can be solved by estimating the probability of full reimbursement and consequently adjusting the interest rate. Still, this approach works best if the borrower provides the bank with accurate information. The credit applicant can try to hide some negative signs of her household to make the bank believe that the repaying of the credit will not be a problem. However, once this household gets credit and spends the funds, often it can occur that the monthly payments associated with repaying the loan are a high burden for the household. Some of those households will still repay the loan, nevertheless, others will stop paying which will result in accumulating of arrears and possibly announcing of default.

Asymmetric information increases the bank risk only if the loan is not secured or if the borrower is a legal body with limited liabilites. Nevertheless, if credit applicant conceals some information to get a credit from the bank, it might have negative impact on his future repayment ability leading to high credit burden. In case of households, banks require that the loans they provide are secured. Banks in Slovakia are securing that the credits will be repayed by either requiring to provide collateral (such as car or a real estate) or guarantor or co-signer who takes responsibility for repaying in case if the primary borrower fails to do so.

Households may have various incentives to borrow from the bank. The household either needs the funds to cover the expenses which are actually higher than available income. In this case, the bank interested in succesful repaying would obviously not grant the loan to the applicant. However, there is still a possibility that on the market there exist some lending institutions that provide credits to risky applicants with intention to get hold of the collateral. One of the results of this thesis we intend to find out is whether this is the case of Slovakia, too.

There may be also systemic bias due to underestimating of default risk by banks. Banks that start to operate in emerging market do not yet have enough information about clients' behavior nor about financial systems in the country. Moreover, the banks may provide credits as a part of expansion strategy (market share competition). Especially in emerging markets where bank systems are not well established, various banks and financial intermediaries are competing by introducing various products to attract new clients, sometimes at a cost of increasing the bank risk. This problem is closely connected with asymmetric information and it is referred to as **moral hazard**. The problem of moral hazard arises for example when bank is secured against risk and thus takes risky actions which it would have avoided without securance.

An interesting study by Myerson [27] connects problem of moral hazard of bankers with boom-bust cycles. He mentions that efficient solution to moral hazard in banking must involve long-term promises of large late-career rewards for individual bankers who maintain good performance records. This requires that bankers must expect long-term relationship with investors. Investors are thus forced to accept limits on the liquidity of their investments, even though their physical investments may be short-term in nature. The idea of long-term career rewards is essential for motivating bankers to identify appropriate investments. Investors thus have to trust their bankers upon expectation about long-term future profits in banking. The value of bankers' positions depends on the recent history of the economy and so it can affect the current investment level. Myerson concludes that by this mechanism, long-term solutions to financial moral hazard can create dynamic forces that drive aggregate economic fluctuations.

The problem of moral hazard also exist with borrowers which may act irresponsible. For instance, holder of a credit card may use it excessively and spend too much money on goods which may lead to her default. Banks are therefore attempting to reduce moral hazard for example by setting a spending limit on credit cards.

Our study of credit burden of Slovak households proposes to add some value to the existing body of literature on informational exonomics by analyzing the effects of asymmetric information and hard and soft information as loan determinants to prevent high credit burden of households.

Chapter 3

Empirical Strategy

Our data came from a survey and thus there may be a problem of selection bias. The reason is that the survey may not be representative of the whole population. This can be described as problem of truncated data. Thus, if we would use OLS, we would get incostistent estimates. Moreover, we analyze determinants of credit burden of households and compare it to loan determinants. We therefore control for potential selection bias in a Heckman selection model. This approach has been used in literature recently.

Hainz and Nabokin [16] use cross-section data on firm-level from the Business Environment and Enterprise Performance Survey to analyze differences between use of credit and acces to credit. Upon the finding that there are significant differences in their determinants, they opt for Heckman selection model. Fidrmuc, Hake and Stix [19] use household data set collected by the Euro Survey project. They analyze determinants of households' plans to take a loan and to take a loan in foreign currency to find out which incentives drive foreign loans demand.

3.1 Truncated Normal Distribution

We say that the data are truncated when sample data are drawn from a subset of a larger population of interest. Truncation is essentially a characteristic of the distribution from which the sample data are drawn. A truncated distribution is the part of an untruncated distribution that is above or below some specific value. (Greene [8]). Consider the situation when the data on dependent variable are available only for values greater than threshold value τ (truncation from below) and denote the observed value of dependent variable by y. y is the incompletely observed value of a latent dependent variable $y^* \sim N(\mu, \sigma^2 I)$ (Golder [7]). Observed value

$y = y^*$ if $y^* > \tau$ and the observations on $y^* \leq \tau$ are lost. In this case the variable $y|y > \tau$ follows a truncated normal distribution. The problem that arises is that we have truncated a part of the original distribution. That means that the distribution has to be re-scaled so that the integral of the distribution function over possible values is equal to one:

$$f(y|y > \tau) = \frac{f(y)}{P(y > \tau)} = \frac{\frac{1}{\sigma}\phi(\frac{y-\mu}{\sigma})}{1 - \Phi(\frac{\tau-\mu}{\sigma})} = \frac{\frac{1}{\sigma}\phi(\frac{y-\mu}{\sigma})}{1 - \Phi(\alpha)}$$

where $\phi(\cdot)$ is the standard normal probability density function, $\Phi(\cdot)$ cumulative distribution function and $\alpha = \frac{\tau-\mu}{\sigma}$.

The truncated normal distribution has the following likelihood function:

$$L = \Pi_{i=1}^{N} \frac{f(y)}{1 - \Phi(\alpha)}$$

$$\ln L = \sum_{i=1}^{N} (ln(f(y)) - ln(1 - \phi(\alpha)))$$

3.2 Inverse Mill's Ratio and Moments of the Truncated Normal Distribution

Inverse Mill's ratio is defined as the ratio of the probability density function to the cumulative distribution function of a distribution:

$$\lambda(\alpha) = \frac{\phi(\alpha)}{1 - \Phi(\alpha)} \text{if the truncation is from above}$$

$$\lambda(\alpha) = -\frac{\phi(\alpha)}{\Phi(\alpha)} \text{if the truncation is from below}$$

Inverse Mill's ratio is also called the hazard function for the standard normal distribution. With the help of inverse Mill's ratio we can express the moments of the truncated normal distribution as follows (Golder [7]):

$$E[y|y > \tau] = \mu + \sigma\lambda(\alpha)$$

$$Var[y|y > \tau] = \sigma^2(1 - \delta(\alpha))$$

where

$$\delta(\alpha) = \lambda(\alpha)(\lambda(\alpha) - \alpha)$$

There are two important results on truncated distribution summarized by Greene [8]:

- If the truncation is from below, then the mean of the truncated variable is greater than the mean of the original one. If the truncation is from above, then the mean of the truncated variable is smaller than the mean of the original one.

- Truncation reduces the variance compared with the variance in the untruncated distribution (because $0 < \delta(\alpha) < 1$ for all values of α).

3.3 Incidental Truncation

The issue of selection bias arises due to an incidental truncation of the sample (Greene (2003) [8]).

As Golder [7] notes, a brief description of incidental truncation will make the Heckman model much easier to understand. Consider two variables y and z with bivariate distribution with correlation ρ. We are interested in the distribution of y given that z exceeds a particular value. Intuitively, if y and z are correlated with the possitive sign, then the truncation of z should push the distribution of y to the right.

The truncated joint density of y and z is

$$f(y, z|z > \tau) = \frac{f(y, z)}{P(z > \tau)}$$

We can express the moments of an incidentally truncated bivariate distribution as follows (Greene [8]):

$$E[y|z > \tau] = \mu_y + \rho\sigma_y\lambda(\alpha_z)$$
$$Var[y|y > \tau] = \sigma_y^2(1 - \rho^2\delta(\alpha_z))$$

where

$$\alpha_z = \frac{\tau - \mu_y}{\sigma_z}$$
$$\lambda(\alpha_z) = \frac{\phi(\alpha_z)}{1 - \Phi(\alpha_z)}$$
$$\delta(\alpha_z) = \lambda(\alpha_z)(\lambda(\alpha_z) - \alpha_z)$$

$\phi(\alpha)$ is the standard normal density, $\lambda(\alpha_z)$ is the inverse Mill's ratio for z (Golder [7]). Note that the expressions involving z are analoguous to the moments of the truncated distribution of x from the previous section. Also, if the truncation is from below, then inverse Mill's ratio changes to $\lambda(\alpha) = -\frac{\phi(\alpha)}{\Phi(\alpha)}$. Like truncation, incidental truncation reduces the variance, because $0 < \delta(\alpha) < 1$ and $0 < \rho^2 < 1$.

3.4 Heckman Selection Model

Heckman selection model basically applies the moments of the incidentally truncated bivariate normal distribution to a data generating process (Golder [7]). To motivate a regression model that makes use of the results of the moments of the truncated normal distribution we present an example that describes the idea of this paper: We analyze effects of various household and individual characteristics on the credit burden. A basic model consists of two equations:

- *Loan equation.* The choice of a household to apply for a loan (demand side) as well as decision of the bank to grant a loan for a household (supply side) is a function of characteristics such as available income, employment status and collateral as well as, for example, age, education and marital status.
- *Burden equation.* The credit burden of a household depends on available income, size of a household and other characteristics.

The problem of truncation surfaces when we consider that the second equation describes credit burden, but an actual figure is observed only in the household has a loan. (From our data, only a participation equation, that is, whether household has a loan, is observable. There is no information on the scale of the selection variable.) We infer from this that supply for loans exceeds demand for loans. Thus, the burden variable in the second equation is incidentally truncated.

To put the preceding example in a general framework, consider the following selection equation

$$z_i^* = w_i\gamma + u_i$$

$$z_i = \begin{cases} 1 & \text{if } z_i^* > 0 \\ 0 & \text{if } z_i^* \leq 0 \end{cases}$$

and the following outcome equation:

$$y_i = \begin{cases} x_i\beta + \epsilon_i & \text{if } z_i^* > 0 \\ - & \text{if } z_i^* \leq 0 \end{cases}$$

We assume that the error terms in the selection and outcome equation have normal distribution and are correlated:

$$u_i \sim N(0,1)$$

$$\epsilon_i \sim N(0,\sigma^2)$$

$$\text{corr}(u_i,\epsilon_i) = \rho$$

3.5 Conditional Means and Marginal Effects in the Heckman Selection Model

To derive the first moment in Heckman Selection Model we have to insert equations from the above section into the relevant equations for the moments of the incidentally truncated bivariate normal distribution (Golder [7]):

$$\begin{aligned} E[y_i|y_i \text{ is observed}] &= E[y_i|z_i^* > 0] \\ &= E[x_i\beta + \epsilon_i|w_i\gamma + u_i > 0] \\ &= x_i\beta + E[\epsilon_i|w_i\gamma + u_i > 0] \\ &= x_i\beta + E[\epsilon_i|u_i > -w_i\gamma] \end{aligned}$$

If the error terms ϵ_i and u_i are independent, then $E[\epsilon_i|u_i > -w_i\gamma] = E[\epsilon_i] = 0$ and we can get consistent estimates of β by OLS. When u_i and ϵ_i are correlated, we obtain (Greene [8]):

$$E[\epsilon_i|u_i > -w_i\gamma] = \rho\sigma_\epsilon\lambda_i(\alpha_u)$$

where $\alpha_u = \frac{-w_i\gamma}{\sigma_u}$ and $\lambda_i(\alpha_u)$ is inverse Mill's ratio:

$$\lambda_i(\alpha_u) = \frac{\phi(\frac{-w_i\gamma}{\sigma_u})}{1 - \Phi(\frac{-w_i\gamma}{\sigma_u})} = \frac{\phi(\frac{w_i\gamma}{\sigma_u})}{\Phi(\frac{w_i\gamma}{\sigma_u})}$$

We can now express the conditional mean in the Heckman selection model as follows:

$$\begin{aligned} E[y_i|y_i \text{ is observed}] &= x_i\beta + \rho\sigma_\epsilon \left[\frac{\phi(\frac{w_i\gamma}{\sigma_u})}{\Phi(\frac{w_i\gamma}{\sigma_u})}\right] \\ &= x_i\beta + \rho\sigma_\epsilon\lambda_i(\alpha_u) \\ &= x_i\beta + \beta_\lambda\lambda_i(\alpha_u) \end{aligned}$$

Using this result we obtain

$$y_i|z_i^* > 0 = E[y_i|z_i^* > 0] + \nu_i = x_i\beta + \beta_\lambda\lambda_i(\alpha_u) + \nu_i$$

Least squares regression using the observed data-for example, OLS regression of credit burden on its determinants, using only data for households that have a loan-produces inconsistent estimates of β. This problem can be describe as omitted variable. Least squares regression of y on x and λ would be a consisten estimator, but if λ is omitted, then the *specification error* of an omitted variable is commited. Even if λ_i were observed, then least squares would be inefficient, because the disturbance ν_i is heteroschedastic (see variance of the incidentally truncated bivariate distribution).

Now we will explore the marginal effects in the Heckman selection model. As Greene [8] points out, the marginal effect of x on y_i consists of two components:

- The direct effect of the independent variable on the mean of y_i which is represented by β.
- The indirect effect of the independent variable that appears in the selection equation is the change of the probability that an observation is in the sample.

The marginal effect of x on y_i in the observed sample is

$$\frac{\partial E[y_i|z_i^* > 0]}{\partial x_{ik}} = \beta_k - \gamma_k \left(\frac{\rho\sigma_\epsilon}{\sigma_u}\right) \delta_i(\alpha_u)$$

where $\delta_i(\alpha_u) = [\lambda_i(\alpha_u)]^2 - \alpha_u \lambda_i(\alpha_u)$. In our analysis of credit burden, the selection variable z^* is not observed. Rather, we observe only its sign, i.e. we observe only whether a household has a loan or not. We can infer the sign of z^*, but not its magnitude, from such information. Since there is no information on the scale of z^*, the disturbance variance in the selection equation cannot be estimated. Thus, we reformulate the model as follows (Greene [8]):
selection mechanism: $z_i^* = w_i\gamma + u_i$, $z_i = 1$ if $z_i^* > 0$ and 0 otherwise;
$P(z_i = 1|w_i) = \Phi(w_i\gamma)$ and
$P(z_i = 0|w_i) = 1 - \Phi(w_i\gamma)$.
regression model: $y_i = x_i\beta + \epsilon_i$ observed only if $z_i = 1$,
$(u_i, \epsilon_i) \sim$ bivariate normal$[0, 0, 1, \sigma_\epsilon, \rho]$. In other words, error termas are normally distributed, $u_i \sim\ N(0, 1)$, $\epsilon_i \sim N(0, \sigma_\epsilon^2)$, but they are correlated, $\text{corr}(u_i, \epsilon_i) = \rho$
We used probit to estimate the selection equation. Thus, σ_u is assumed to be 1. The marginal effect is in this case:

$$\frac{\partial E[y_i|z_i^* > 0]}{\partial x_{ik}} = \beta_k - \gamma_k \left(\rho\sigma_\epsilon\right) \delta_i(\alpha_u)$$

From the marginal effect it is clear that if $\rho \neq 0$ and the independent variable appears both in the selection and outcome equation, then β_k does not indicate the marginal effect of x_k on y_i.

There are two ways of estimating the Heckman model: Heckman's Two-Step Procedure and maximum-likelihood estimation. We will describe those methods in the gollowing sections.

3.6 Estimation by Heckman's Two-Step Procedure

Assumme that u_i and ϵ_i are independent of the explanatory variables with mean zero and $u_i \sim N(0,1)$ (Wooldridge [9]). The model is estimated in two steps (Golder [7]):

- Estimate the selection equation by probit maximum likelihood estimation to obtain estimates of γ. For each observation in the selected sample, compute the inverse Mill's ratio
$$\hat{\lambda}_i = \frac{\phi(w_i\hat{\gamma})}{\Phi(w_i\hat{\gamma})}$$
and $\hat{\delta}_i = \hat{\lambda}_i(\hat{\lambda}_i + w_i\hat{\gamma})$
- Estimate β and $\beta_\lambda = \rho\sigma_\epsilon$ by OLS on x and $\hat{\lambda}$.

The estimators from this two-step procedure are consistent and asymptotically normal. This procedure is often called a "Heckit model".

3.7 MLE Version

The Heckman model can also be estimated by maximum-likelihood estimation (MLE) without using inverse Mill's ratios. Stata uses this approach to estimate Heckman probit model.

MLE requires stronger assumption than the two-step procedure and thus is less general. We assume that both error terms are normally distributed, $u_i \sim N(0,1)$, $\epsilon_i \sim N(0,1)$, but they are correlated, $\text{corr}(u_i, \epsilon_i) = \rho$.

Log-likelihood function is introduced in Stata manual [29]:

$$\begin{aligned}\ln L \;=\; & \sum_{i\in S; y_i\neq 0} \ln\{\Phi_2(x_i\beta + \psi_i^\beta, w_i\gamma + \psi_i^\gamma, \rho)\} + \\ & \sum_{i\in S; y_i=0} \ln\{\Phi_2(-x_i\beta + \psi_i^\beta, w_i\gamma + \psi_i^\gamma, -\rho)\} + \sum_{i\notin S} \ln\{1 - \Phi(w_i\gamma + \psi_i^\gamma)\}\end{aligned}$$

where S is the set of observations for which y_i is observed, $\Phi_2(.)$ is the cumulative bivariate normal distribution function with zero means and correlation ρ, ψ is a scaling variable and $\Phi(.)$ is the standard cumulative normal distribution function.

In the maximum likelihood estimation, ρ is not directly estimated. Directly estimated is atanhρ:

$$\text{atanh}\rho = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)$$

Chapter 4

Data Description and Descriptive Statistics

In this paper we analyze The European Union Statistics on Income and Living Conditions (EU-SILC) dataset. It contains data on individual households such as region or number of members that live in a household as well as data on household members such as their education, economic activity and income. We use household survey data from Slovakia from 2005 and 2006. About 5 thousands households were interviewed using a questionare aiming at collecting comparable cross-sectional multidimensional microdata on income, poverty, social exclusion and living conditions.

The dataset consists of four parts: register of households for cross-sectional survey, register of persons for cross-sectional survey, data on the households for cross-sectional survey and personal data for cross-sectional survey. We merged the whole dataset into one file to be able to analyze loan determinants and determinants of credit burden. We carefully selected essential variables and dropped the rest.

Social exclusion and housing-condition information is collected at household level. Labour, education and health information is obtained for persons aged 16 and over. Income at a very detailed component level is mainly collected at personal level, but some components are included in the 'household' section.

Data that belong to **household level** consist of about 5 thousands observations in both years. Household level answers whether the household has a loan and whether repaying of a loan is a high burden for the household. Having a loan is documented by bank contracts and is therefore a proxy for hard information but credit burden is a proxy for soft information because it is based on subjective statement of a household members. We have also information about ownership status of the dwelling - 79% of households interviewed owns their dwelling (we do not distinguish here whether the dwelling is paid off or not), 21% of households are tenants. Nearly 50% of households have a car. Ownership of dwelling and car is example of hard information because it can be verified e.g. by buying contract. Heads of households have also been asked about how easy is it for their household to manage with their disposable income. The answers have been categorized from 1 to 6, 1 corresponding to the lowest affordability of consumption and 6 to the highest. Financial strength describes whether the household is able to face unexpected financial expenses. Both affordability of consumption and financial strength are proxies for soft information. Moreover, those 2 variables are not used by banks' scoring models and we want to explore whether soft information is important. If those variables will prove to be important determinants of credit burden, banks should consider them when assessing credit applicants to get lower bank risk and to lower the default rate of clients. Another variable that belongs to household level is available income. In original dataset, yearly income was presented in Slovak crowns but we use logarithm of income in our analysis. Therefore, mean income 12.182 stands for approximately EUR 540 per month. Variables low income and high income represent first and third tercils, respectively. 10 percent of households have arrears. Households in our dataset have 1 to 10 members. Note that household size is not measured by number of persons living in a households but modified OECD scale is used in which first adult member has a weight of 1, every other adult member has a weigth of 0.5 and every child under 14 has a weigth of 0.3 (hence the maximum of household size variable is not integer). For more information about structure of Slovak households, see **Figure 3**.

Personal level complements data from household level with information about individual persons that live in the household. Dataset of persons contains about 12 thousands observations. Each person has a unique personal ID and also household ID which assigns to every person household in which she lives in. Personal dataset contains information about each person's marital status, employment status, age, health condition and education. For our analysis we used personal data of the household head, i.e. of the person which is responsible for the dwelling.

Several studies analyze credits and lending but their approaches and datasets are different from ours. Égert, Backé and Zumer [2] use quaterly data obtained from the International Financial Statistics of the IMF. Their data include bank credit to the private sector, credit to the government sector, short-term and long-term interest rate series, the consumer and producer price indices (CPI and PPI), real and nominal GDP, and industrial production. They are thus analyzing macroeconomic time series. Boissay and Calvo-Gonzalez [30] also use macro data in their analysis of credit growth sustainability. Their variables include aggregate real credit supply, demand, and equilibrium levels, real GDP, real interbank rate, real retail lending rate and financial liberalization. Brzoza-Brzezina [31] analyze lending booms in new EU member states using vector error correction model in real loans to the private sector, real GDP and real interest rate. Calza, Manrique and Sousa [32] also use vector error correction to analyze long-run relationship between real loans, real GDP and real composite lending interest rate. The study of Hainz and Nabokin [16] is similar to our paper in their approach but instead of households they analyze private enterprises. Calza, Gartner and Sousa [33] model demand for loans to the private sector in the euro area using real GDP and prices and bank lending rates. Their study is similar to ours because they analyze loan deteminants, but they use aggregate macro data instead of micro data at household level.

In contrast to the existing literature, our paper analyzes micro data of individual households and thus adds new perspective on the credits analysis problem. Descriptive statistics of the variables which we employ are shown in **Table 1**.

Table 1 reveals that in our sample, 5.7% of household heads are unemployed. This is somewhat inconsistent with reality because unemployment rate was about 14% in 2005. Also, 35.3% of people in sample are retired and 17.7% of population is widowed. Real number of retired and widowed people are much smaller among Slovak citizens. This puts the representativeness of EU-SILC survey to question. To solve this issue we can drop observations with retired people to get alternate dataset for our analysis. Unemployment rate of dataset without retired people is 8.7% which is closer to active labour force than the rate from the initial data set.

Variable name	Observations	Mean	Std. Dev.	Min	Max
Household level					
has loan	5147	0.306	0.461	0	1
burden	1576	0.425	0.495	0	1
dwelling	5147	0.790	0.407	0	1
car	5147	0.479	0.500	0	1
affordability of consumption	5147	4.165	1.024	1	6
financial strength	5147	0.374	0.484	0	1
income	5139	12.182	0.653	9.585	14.844
low income	5139	0.333	0.471	0	1
high income	5139	0.333	0.471	0	1
arrears	5147	0.105	0.307	0	1
household size	5147	1.917	0.693	1	5.3
members	5147	2.996	1.535	1	10
Personal level					
married	5142	0.692	0.462	0	1
single	5142	0.052	0.222	0	1
divorced	5142	0.072	0.258	0	1
widowed	5142	0.177	0.381	0	1
unemployed	5142	0.057	0.231	0	1
retired	5142	0.353	0.478	0	1
age	5147	52.734	14.911	16	95
age<=20	5147	0.003	0.052	0	1
age 21-30	5147	0.064	0.244	0	1
age 31-40	5147	0.147	0.354	0	1
age 61+	5147	0.298	0.457	0	1
bad health	5142	0.241	0.428	0	1
good health	5142	0.375	0.484	0	1
low education	5130	0.156	0.363	0	1
high education	5130	0.169	0.374	0	1
business	5147	0.083	0.275	0	1

Table 1: descriptive statistics, 2005

In a few figures we will illustrate some interesting demographic characteristics of Slovak households.

Figure 2: Income distribution of households in Slovak Republic

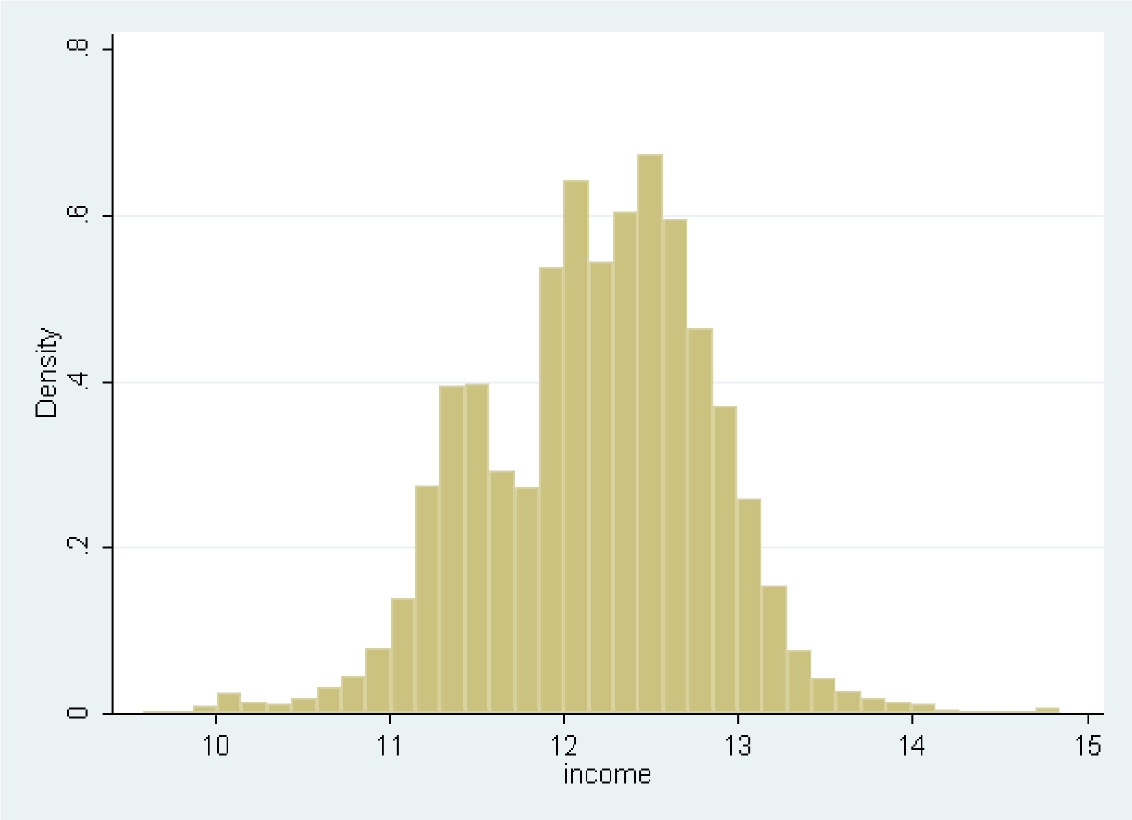


Figure 2 shows distribution of income among Slovak households. Medium income (about EUR 540 per month) is prevalent, while households with very low or very high income make smaller part of the sample. The lowest income that the households have available per year is about 10 in log which corresponds to 22 thousands Slovak crowns per year or 1800 Slovak crowns per month which is about EUR 60 per month. The high-end income is about 14 in log, 1.2 million Slovak crowns per year or 100 thousands Slovak crowns per month which is roughly EUR 3000 per month.

Figure 3: Number of persons living in Slovak households

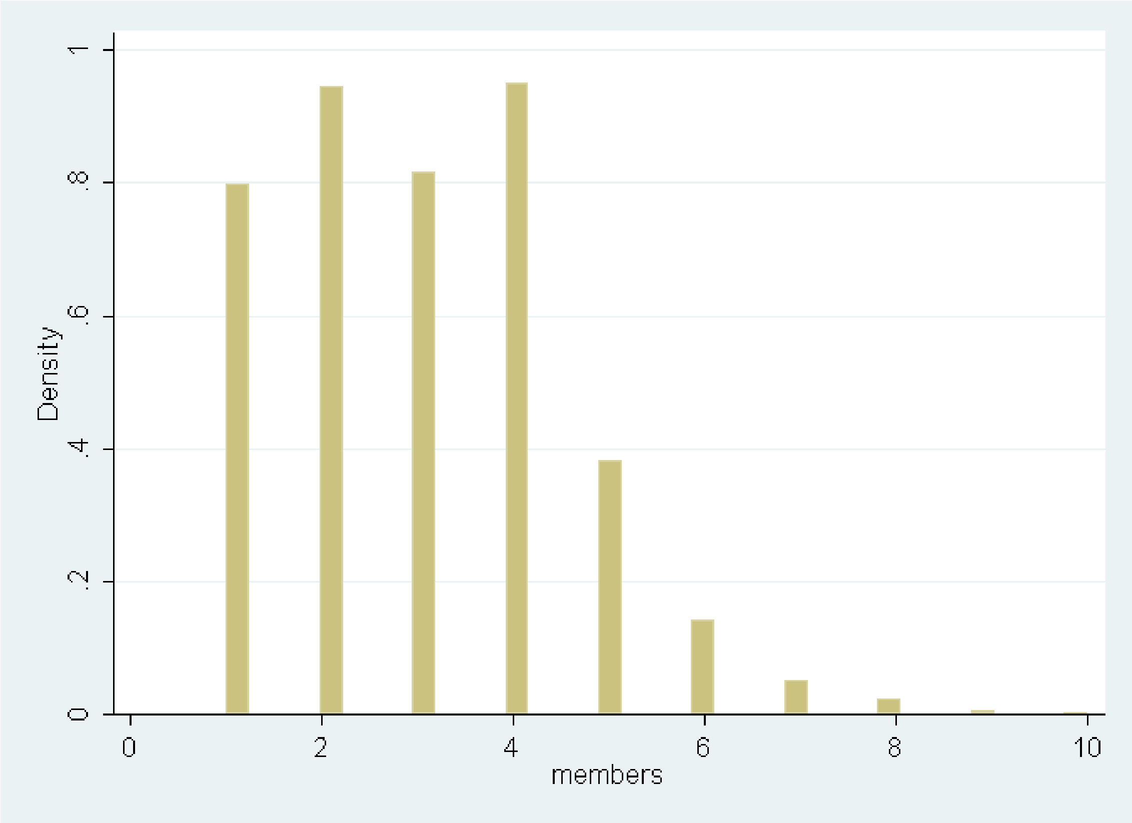


Figure 3 shows the structure of typical Slovak household. The majority of the households have up to 4 members, while number of members are approximately uniformly distributed when looking at households of this size.

4.1 Factors affecting households' default

To assess the effects that individual variables have on credit burden of Slovak households, we will use analogy with CAMEL ratings. CAMEL is a bank monitoring system that was introduced by US Federal Deposite Insurance Corporation. The financial ratios are sorted into five categories: Capital adequacy, Asset quality, Management competence, Earnings ability and Liquidity (Fidrmuc, Süß [15]). In relation to our analysis of households' credit burden, we categorize variables into four groups as follows (we omitted asset quality because it is irrelevant when talking about households and used responsibility as a substitute for management competence):

I. Capital adequacy Into this category we need to put information about possessions that households own. Therefore, *car* and *dwelling* variables fit here. Car as well as real estate can serve as collateral and therefore we expect significant effects of those variables on probability of having a loan. This probability is, however, influenced by two opposite factors: demand for loans and supply of loans. For example, the household that has a car does not have a demand for a car loan. On the other hand, if a household has a car, by securing the credit by the car the bank will me more willing to grant credit. Therefore, it is not an easy task to predict marginal effects of capital adequacy on probability of having a loan. As for the effect of capital adequacy on the credit burden, we expect that it will not be significant. We suppose that other variables (notably *income*) have higher effects on credit burden of households.

II. Responsibility The variables that we can use for measuring responsibility of individual households are *affordability of consumption* and *financial strength*. The first one measures household's competence over its financial affairs, i.e. how easy or how hard it is to manage with disposable income. It is positively correlated with household's *income* (with correlation coefficient 0.25), however, contains somewhat different information because it is measured by personal answer of one of the household's members. That is why this variable if a proxy for soft information. We expect that better responsibility will lower the probability of high credit burden because of better decision making. A responsible person who can manage her finance well will not ask for large credit such that the repayment burden would be too high. Similarly, for the households that are able to face unexpected financial expenses (this is precisely captured in variable *financial strength*), we expect that the credit burden will be less severe.

Another variable that measures responsibility in our analysis is *business*. We expect businessmen to have better managerial skills and also higher income, therefore they likely have less problem with repaying their loans.

III. Earnings ability This category includes disposable *income*. This variable is expected to have significant and negative effect on the credit burden. *Income* is easier to measure than *affordability of consumption* because financial resources such as salaries and allowances are repesented by numbers even though some households can still have problems with calculating total disposable income. We expect strong negative effect of *income* on the credit burden because wealtier households that are able to earn more funds are likely to have less problems with repaying of credit. Banks that are employing scoring models also use income as essencial characteristic to rate credit aplicants.

IIII. Liquidity We chose to measure liquidity of the households by marital status of the head of the household. Most of the persons managing the households are *married*, therefore we chose *married* as the base category. There are various effects of marital status on the financial situation of a household:

- Married people can merge their incomes to share the cost of living.
- Dual income allows the household to pay off debts more quickly and to save for retirement more effectively than in the case of single person.
- Single people are less likely to have children. This fact lowers expenses of singles compared to married couples.
- People that are divorced might have higher expenses than married because they no longer enjoy dual income while they often have to pay alimony payments.

As for the effect of marital status on the credit burden, we expect that divorced people are having significantly higher probability of big repayment burden than married couples due to the facts mentioned above. Assessing the effects of being single on credit burden is more complicated because there are counteracting effects. On the one hand, single people are usually not as responsible as married ones and they can not make use of double income as in the case of married ones. On the other hand, singles have lower family expenses than married people because they usually do not have children.

18% of households have *widowed persons* at the head, therefore it is also important to analyze this category. Financial situation of widowed people is difficult as can be seen from our dataset: 77% of those households where the head is widowed have *low income*, while only 9% of them have *high income*. Therefore, we expect that widowed people will have higher probability of high credit burden than our base category. We could also analyze a special type of households which consist of one-man only. However, we found that 72% of widowed persons are living alone and 65% of households consisting of sole members are widowed persons. Therefore, the correlation between being widowed and living alone is highly positive (0.61) and we do not need to analyze one-man households as a special category.

Another way to measure liquidity is to consider whether the household has *arrears.* If so, it indicates that the household's liquidity is low and therefore we expect positive effect of arrears on the credit burden.

Other variables that we will analyze include *health, age and education level.* We expect that persons with good health have less problems with paying off the loans while people with bad health have higher probability of high credit burden. The reason for this assumption is that overall health condition affects person's work efficiency and therefore it also affects her salary. The disposable income is positively correlated with good health and negatively with bad health and we can think of health condition as a control variable that indirectly affects credit burden through income. Similarly, age does not have direct effect on credit burden but person's salary usually follows evolution over her lifetime: Young people who start working start with lower income. As people work longer and are gaining experience, their income is growing. When people are retired, their income falls down again because pensions are smaller than income from the job (although, the higher the person's income was, the higher her pension is as well). Therefore, age has effect on income and income is affecting the credit burden. The effect of *education* on credit burden is more difficult to predict because person's qualification influences her salary but also her reasoning. Therefore, people with low education might incline to overestimate their repaying capabilty. It is then up to banks' scoring model to refuse to grant too high credits for such applicants. We expect that people with high education have less problems with repaying credits because not only their income but also their responsibility is higher.

Moreover, it is rational to suppose that bigger households face higher credit burden. We have variables *household size* and *members* that we can use to analyze this hypothesis.

Chapter 5

Estimation Results

5.1 Empirical Strategy

Our empirical strategy follows the approach proposed by Heckman. The credit burden of household is observed only if the respondent has a loan. Therefore, the estimation of probability that repaying of credit is difficult for a household would be biased. We estimate probability of having a loan and repaying burden in a two-step setup where selection equation is defined as a probability L that a household has a loan,

$$P(L = 1) = \Phi(W\gamma + u)$$

In the second stage, we estimate a probit equation that repaying burden is high

$$P(B = 1|L = 1) = \Phi(X\beta + \epsilon)$$

where error terms are normally distributed, $u \sim N(0, 1), \epsilon \sim N(0, 1)$, but they are correlated, $\mathrm{corr}(u, \epsilon) = \rho$.

The selection equation includes income, collateral (dwelling, car), employment categories (unemployed, retired), family status (single, divorced), arrears and financial strength in most of the specifications. Variables which are used for the exclusion restriction are employment categories. Those variables are assumed to have impact on access to loans but they are not related to the credit burden. In every specification we use robust standard errors adjusted for clustering at the regional level.

5.2 Loan Determinants

To arrange the presentation of our results of credit burden clear, we will not present the selection equation for all specifications. Therefore, we start the discussion with the loan equation corresponding to selection equation which is estimated by probit. The dependent variable is the probability that a household has a loan. The result of the probit estimation is in **Table 2**.

Table 2 shows that most of identification variables are highly significant. Unemployed and retired people have significantly lower probability of a loan than employed, which are defined as the base category. Similarly, single and divorced individuals (and their households) have lower probability of a loan than married. This result is expected because married people can usually use double income, while divorced people may have responsibilities such as paying alimony, their available income is therefore lower and so they can not afford to have a loan.

Households with higher income have higher probability of a loan. Households which own dwelling has significantly lower probability of having a loan. This is perhaps caused by the fact that we are analyzing loans other than mortgages and therefore people prefer to provide some other assets as collateral. Also, households that are living in their own dwelling have better overall financial situation than tenants and their demand for loans is therefore lower. On the other hand, household which own a car have higher probability of having a loan as well which can be to some extent caused by households' demand for car loans.

Businessmen have higher probability of having a loan, which is probably caused by the fact that they also have higher income and therefore represent a good clients for the bank.

Loans depend significantly on several demographic factors. Young individuals (up to 40 years old) have a loan, while older households (61 years old or more) have a significant and negative effect on marginal probability of a loan. The probability of a loan also increases with the size of household. People with lower education have a loan. Effect of high education on probability of having a loan is not significant but has a negative sign.

We also analyze restricted dataset without retirees. Those specifications have 'R' in their names. As for loan determinants, it turns out that being widowed is not significant when analyzing dataset without retired people. This result is expected because most of the widowed people are among older and retired people.

Specification	2005-1	2005-2	2006-1	2006-2	2005-2R
Dwelling	-0.116***	-0.102***	-0.051**	-0.028	0.010
	(0.018)	(0.018)	(0.022)	(0.021)	(0.023)
Car	0.047***	0.042***	0.082***	0.079***	0.049**
	(0.015)	(0.015)	(0.014)	(0.014)	(0.019)
Income	0.062***	0.040***	0.065***	0.046***	0.053***
	(0.013)	(0.015)	(0.012)	(0.014)	(0.018)
Arrears	0.123***	0.106***	0.157***	0.131***	0.140***
	(0.023)	(0.023)	(0.026)	(0.026)	(0.028)
Household size		0.063***		0.064***	0.056***
		(0.014)		(0.012)	(0.017)
Single	-0.118***	-0.106***	-0.080***	-0.064**	-0.127***
	(0.027)	(0.028)	(0.026)	(0.028)	(0.036)
Divorced	-0.036	-0.007	0.015	0.065**	-0.029
	(0.026)	(0.028)	(0.025)	(0.028)	(0.035)
Widowed	0.079***	0.138***	0.058**	0.117***	0.023
	(0.024)	(0.027)	(0.023)	(0.026)	(0.045)
Unemployed	-0.087***	-0.093***	-0.042	-0.055*	-0.093***
	(0.025)	(0.025)	(0.030)	(0.029)	(0.031)
Retired	-0.210***	-0.071***	-0.183***	-0.077***	
	(0.016)	(0.025)	(0.015)	(0.024)	
Age 21-30		0.131***		0.238***	0.169***
		(0.031)		(0.036)	(0.033)
Age 31-40		0.128***		0.156***	0.136***
		(0.021)		(0.022)	(0.022)
Age 61+		-0.139***		-0.078***	-0.196***
		(0.025)		(0.025)	(0.057)
Low education	0.019	0.045*	0.029	0.047**	-0.006
	(0.023)	(0.024)	(0.023)	(0.024)	(0.042)
High education	-0.036**	-0.016	0.005	0.021	-0.024
	(0.017)	(0.018)	(0.017)	(0.018)	(0.022)
Business	0.073***	0.063**	0.046**	0.044*	0.068**
	(0.026)	(0.026)	(0.024)	(0.024)	(0.028)

Table 2: Loan Determinants Note: The dependent variable is the probability that a household has a loan. Coefficients dF/dx report the average marginal probability effects. Standard errors are adjusted for clustering at the regional level. ***, **, and * denote significance at 1%, 5%, and 10%, respectively.

5.3 Individual specifications

In this section we present various specifications and discuss the explanations. We first present four sets of specifications A, B, C and D showed in **Tables 3-6** that analyze the effects of capital adequacy, responsibility, earnings ability and liquidity on the credit burden. We selected 4 sets of specifications using a simple rule: In A we omited liquidity variables (i.e. arrears, single, divorced, widowed), in B we omited earnings ability variable (i.e. income), in C we omited responsibility variables (i.e. affordability of consumption and financial strength) and lastly in D we omited capital adequacy variables (i.e. dwelling, car). We did not use all of the variables in a single specification in order to avoid multicollinearity problems.

In each specification set, we first started with a simple model with fewer variables and consequently we were adding more variables to select more complex specifications, i.e. we used method Specific-to-General. We present 5 specifications in each table: 3 specifications from 2005 (e.g. A1, A2, A3), one specification from 2006 (e.g. A2+) and one specification from 2005 from a restricted dataset without retirees (e.g. A2R). We name those specifications by letter and number, plus sign indicates year 2006 and 'R' stands for dataset without retirees. We will first discuss specifications from 2005 and will compare the results with that of specifications from 2006 and dataset without retired people afterwards.

Specification	**A1**	**A2**	**A3**	**A2+**	**A2R**
Year	**2005**	**2005**	**2005**	**2006**	**2005**
Dwelling	0.002	0.013	-0.005	0.049	-0.036**
	(0.04)	(0.29)	(-0.15)	(0.81)	(-2.11)
Car	-0.057	-0.001	0.021	-0.065**	0.028
	(-1.50)	(-0.03)	(0.72)	(-2.40)	(1.34)
Affordability			-0.199***		
of consumption			(-11.36)		
Financial		-0.285***	-0.120	-0.294***	-0.186***
strength		(-5.95)	(-3.72)	(-11.82)	(-8.82)
Income	-0.084***	-0.039	-0.019	-0.027	0.023
	(-3.23)	(-1.48)	(-0.93)	(-1.31)	(1.40)
Business	-0.124***	-0.076	-0.025	0.032	0.006
	(-3.16)	(-1.53)	(-0.49)	(0.53)	(0.25)
Log likelihood	-4035.829	-3936.959	-3807.022	-3269.027	-2913.984
Number of obs	5134	5134	5134	4792	3322
Uncensored obs	1568	1568	1568	1225	1258

Table 3: Specifications analyzing the effects of capital adequacy, responsibility and earnings ability on the credit burden Note: The dependent variable is the probability that a household reports high credit burden. Only second stage equation is reported. Coefficients report the average marginal probability effects. z-statistics are adjusted for clustering at the regional level and presented in parentheses. ***, **, and * denote significance at 1 per cent, 5 per cent, and 10 per cent, respectively.

Specification Year	B1 2005	B2 2005	B3 2005	B1+ 2006	B3R 2005
Dwelling	-0.070*	-0.047	-0.038	0.058	-0.058
	(-1.73)	(-1.21)	(-1.07)	(1.04)	(-1.40)
Car	-0.040	0.015	0.025	-0.082**	0.021
	(-0.94)	(0.37)	(0.83)	(-2.25)	(0.69)
Affordability			-0.201***		-0.181***
of consumption			(-16.32)		(-4.97)
Financial		-0.298***	-0.145***		-0.156***
strength		(-12.69)	(-4.73)		(-5.19)
Arrears	0.183***	0.119***	0.008	0.367***	0.031
	(6.42)	(3.93)	(0.28)	(9.84)	(0.52)
Single	-0.014	-0.043	0.006	-0.070	-0.042
	(-0.45)	(-1.44)	(0.24)	(-1.00)	(-0.69)
Divorced	0.084**	0.072**	0.049	-0.011	0.032
	(2.43)	(2.25)	(1.12)	(-0.26)	(0.95)
Widowed	-0.077***	-0.119***	-0.076***	-0.041	-0.055
	(-3.01)	(-5.05)	(-3.27)	(-1.40)	(-1.24)
Business	-0.055**	-0.033	-0.007	-0.015	-0.011
	(-2.07)	(-0.78)	(-0.15)	(-0.28)	(-0.24)
Log likelihood	-4001.154	-3916.493	-3796.823	-3278.766	-2820.278
Number of obs	5142	5142	5142	4794	3329
Uncensored obs	1573	1573	1573	1225	1263

Table 4: Specifications analyzing the effects of capital adequacy, responsibility and liquidity on the credit burden Note: The dependent variable is the probability that a household reports high credit burden. Only second stage equation is reported. Coefficients report the average marginal probability effects. z-statistics are adjusted for clustering at the regional level and presented in parentheses. ***, **, and * denote significance at 1 per cent, 5 per cent, and 10 per cent, respectively.

Specification	C1	C2	C3	C3+	C3R
Year	**2005**	**2005**	**2005**	**2006**	**2005**
Dwelling	-0.052	-0.068*	-0.061	0.071	-0.049*
	(-1.43)	(-1.89)	(-1.60)	(1.23)	(-1.96)
Car	-0.055	-0.046	-0.031	-0.086**	-0.007
	(-1.28)	(-1.13)	(-0.77)	(-2.27)	(-0.34)
Income	-0.077**	-0.116***	-0.098**	-0.118***	-0.042
	(-2.21)	(-2.70)	(-2.29)	(-4.64)	(-1.63)
Arrears			0.162***	0.364***	0.138***
			(5.10)	(10.18)	(6.48)
Household size		0.109***	0.095***	0.081**	0.064***
		(4.13)	(3.76)	(2.41)	(3.74)
Single	-0.052	0.022	0.032	-0.058	-0.047*
	(-1.41)	(0.51)	(0.85)	(-0.67)	(-1.94)
Divorced	0.075**	0.106***	0.092***	-0.015	0.031
	(2.15)	(3.42)	(2.97)	(-0.23)	(1.17)
Widowed	-0.140***	-0.073***	-0.055**	-0.041	-0.021
	(-6.26)	(-2.61)	(-2.18)	(-0.76)	(-0.56)
Business	-0.107***	-0.118***	-0.106***	-0.028	-0.053**
	(-3.25)	(-3.58)	(-3.12)	(-0.47)	(-2.57)
Log likelihood	-4007.536	-3984.467	-3960.633	-3245.746	-2951.426
Number of obs	5134	5134	5134	4792	3322
Uncensored obs	1568	1568	1568	1225	1258

Table 5: Specifications analyzing the effects of capital adequacy, earnings ability and liquidity on the credit burden Note: The dependent variable is the probability that a household reports high credit burden. Only second stage equation is reported. Coefficients report the average marginal probability effects. z-statistics are adjusted for clustering at the regional level and presented in parentheses. ***, **, and * denote significance at 1 per cent, 5 per cent, and 10 per cent, respectively.

Specification Year	D1 2005	D2 2005	D3 2005	D1+ 2006	D3R 2005
Affordability			-0.200***		-0.187***
of consumption			(-15.29)		(-13.30)
Financial		-0.301***	-0.138***		-0.150***
strength		(-13.84)	(-4.91)		(-5.35)
Income	-0.074**	-0.027	-0.015	-0.074**	-0.020
	(-1.97)	(-0.88)	(-0.69)	(-2.47)	(-1.02)
Arrears	0.188***	0.115***	0.004	0.383***	0.023
	(6.81)	(3.75)	(0.15)	(13.89)	(0.43)
Single	-0.019	-0.044	0.003	-0.099	-0.029
	(-0.55)	(-1.30)	(0.13)	(-1.24)	(-0.73)
Divorced	0.073*	0.058	0.033	-0.028	0.026
	(1.71)	(1.45)	(0.74)	(-0.49)	(0.91)
Widowed	-0.087***	-0.127***	-0.079***	-0.079	-0.059
	(-7.14)	(-7.68)	(-5.28)	(-1.63)	(-1.35)
Business	-0.109***	-0.054	-0.014	-0.038	-0.028
	(-2.94)	(-1.10)	(-0.27)	(-0.69)	(-0.60)
Log likelihood	-4007.142	-3919.924	-3802.494	-3286.099	-2813.092
Number of obs	5134	5134	5134	4792	3322
Uncensored obs	1568	1568	1568	1225	1258

Table 6: Specifications analyzing the effects of responsibility, earnings ability and liquidity on the credit burden Note: The dependent variable is the probability that a household reports high credit burden. Only second stage equation is reported. Coefficients report the average marginal probability effects. z-statistics are adjusted for clustering at the regional level and presented in parentheses. ***, **, and * denote significance at 1 per cent, 5 per cent, and 10 per cent, respectively.

We see that in presented specifications, variables *affordability of consumption* and *financial strength* are significant and have negative sign. This means that households that have higher **responsibility** have significantly smaller probability of reporting high credit burden. This is likely caused by better administration of overall household finance. Interesting result is shown in specifications A1, A2, A3: In specification A1, we omited both affordability of consumption and financial strength and income is significant at 1%. In A2 we added financial strength which is highly significant but income loses its significance. Finally, in A3 specification we included affordability of consumption which caused both income and financial strength to lose their significance. This is an example where we can see collinearity between explanatory variables. Throughout **tables 3-6**, variables describing responsibility have the highest absolute magnitude - about 20% for affordability of consumption and nearly 30% for financial strength. That precisely means that household with high affordability of consumption has 20% lower probability of facing a high credit burden than similar household that has low affordability of consumption. The results for responsibility are robust, they keep negative sign also in specifications from 2006 and in specifications without retired people. Variables that describe responsibility (affordability of consumption, financial strength) are examples of soft information. From our results it turns out that those variables are very important determinants of households' credit burden. This suggests that banks really should try to collect soft information and employ it into schemes that assess credit applicants. Interestingly, responsibility is more important determinant of households' credit burden than *income*. This result is logical because even households with high income can suffer credit burden if they can not manage their financial affairs well.

Nevertheless, **earnings ability**, i.e. *income* keeps negative sign in nearly every specifications. In specifications in **Table 5** it is highly significant with important marginal effects of 8 to 12%. The reason why income is more significant here is that those specifications exclude affordability of consumption and financial strength. Income is a continuous variable (unlike for example affordability of consumption, which is a discrete variable) and the marginal effects means that if the income of the household rises by 1%, probability of high credit burden decreases by 0.12%. The above results are in accordance with our expectations formulated before. Income is measured by numbers and therefore is a proxy for hard information. Nevertheless, some households still can have difficulties to evaluate disposable income that can be composed of many elements such as child contributions and family allowances. The banks that collect this information thus have to search further than just looking at payroll slip, for example by monitoring the client's account balance.

Liquidity

First, we use marital status to measure household liquidity. Marital status is hard information which makes it easier to implemant into banks' scoring models than soft information. The effects of household head being *single* on the probability of credit burden are not significant. On the other hand, *divorced* people as household heads indicate significantly higher probability of the credit burden. In C2 specification the magnitude of this variable is almost 11% which unexpectedly makes it as important as income. The reasons why divorced people face higher credit burden are several:

- Before divorce, married couple likely shared rent costs by living together in one dwelling. Divorced individual has to finance dwelling on her own.

- Divorced person often has children which makes it harder to manage with one income only, even with alimony. The other possible situation is that divorced person is not taking care of the children but has to pay alimony which again increases the household's expenses.

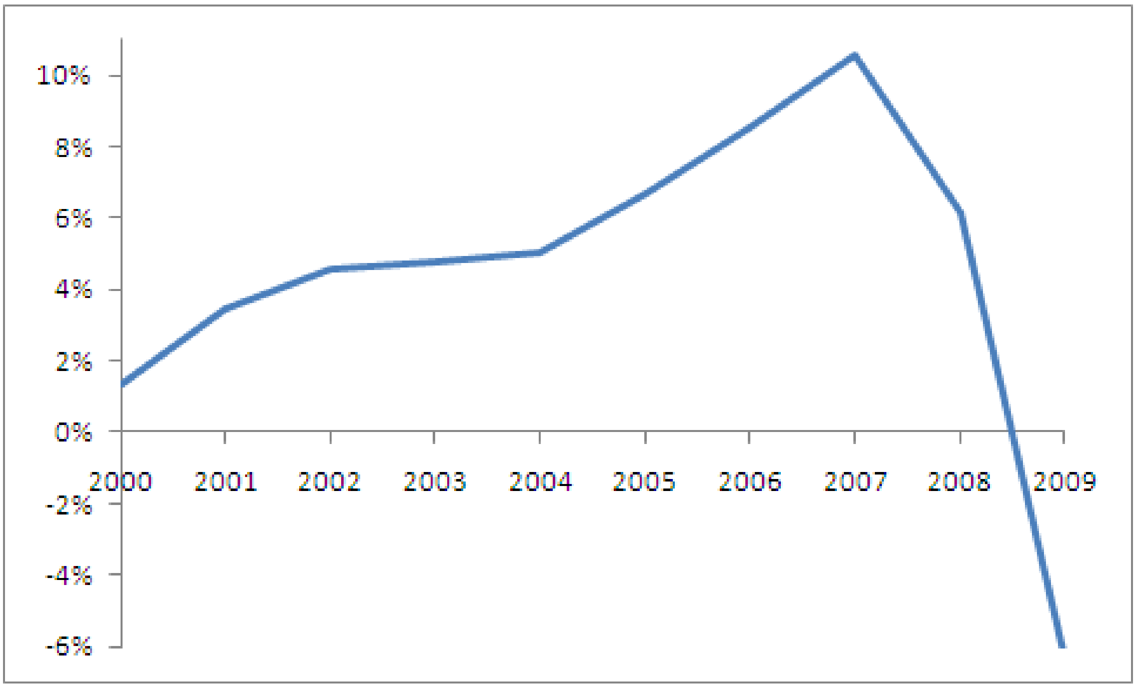
Suprisingly, it follows from our specifications that *widowed* persons have significantly lower probability of the high credit burden. This result is in contrast with our expectation expressed in previous chapter. This results basically implies that banks are careful enough when granting loans to widowed people. More importantly, banks are probably securing risky loans by reqiring life insurance in favor of the bank as a collateral.

Another variable that falls under category **liquidity** is *arrears*. From almost every specification we see that the household that has problem with arrears payments reports high credit burden as well. More precisely, causality runs both directions because household that is reporting high credit burden is also more likely having difficulties with arrears payments. The magnitude of arrears variable is as high as 18% which means that its effect on credit burden is not negligible. The arrears variable is significant in nearly every specification and keeps positive sign which indicates robustness of our models. While amount of arrears can be expresses by numbers, there can be a problem of asymmetric information: The person that asks bank for credit does not have to reveal all relevant documents that show evidence about existence of arrears. Banks thus need to do some research here as well e.g. by cooperating with other banks and financial institutions. Existence of arrears is in essence hard information but it is not easily veriable.

The effect of **capital adequacy**, i.e. **collateral**, is represented by *dwelling* and *car* variables. As expected, those variables do not show significant influence on credit burden in most of the specifications. The result is reasonable because while collateral increses motivation of households to repay, it does not have effect on actual burden that is caused by repaying. Coefficients of *dwelling* and *car* are close to zero and their sign varies through several specifications. However, in specifications where those variables are significant, they always have negative sign. The negative sign of car variable can be interpreted: people that own a car are not bind to work in the region where they live but they can commute and have a better-paid job in different location. Similarly, negative sign of dwelling variable suggests that households that have their own dwelling are slightly better with their overall financial situation than tenants.

In each table, there is one specification analyzing determinants of credit burden using 2006 dataset. Note that both analyzed years were "good" characterized by fast economic growth and low interest rates and in 2006 the growth was higher than in 2005. Years 2008 and 2009 were affected by the global financial crisis and therefore GDP growth rate was falling.

Figure 4: GDP growth (annual %) in Slovakia



Source: World Bank [34]

By comparing the results from 2006 with the ones from 2005 we found out:

- Car is a significant determinant that lowers the probability of high credit burden in 2006 whereas in 2005 it was not significant. In 2006 during economic boom, more people probably found job outside their hometown and car was thus important because they were commuting.

- People were less affected by marital status in 2006 - variables divorced and widowed are no longer singnificant in 2006. Actually, sign of variable divorced changed to negative because people were more succesful in year of higher economic growth. In other words, even people with responsibilities were less burdened by credits repayment in the boom times.

Specifications analyzing dataset without retirees did not reveal very much information, only slightly higher significance of owning a dwelling which lowers the probability of high credit burden.

In **Tables 7-9** we present another 3 sets of specifications E, F and G that show the effects that other variables have on the credit burden. We will discuss the effects of *age*, *health condition*, *education level*, *household size* and finally the effects of household head being a *businessman* on the credit burden.

Specification	E1	E2	E3	E1+	E2R
Year	2005	2005	2005	2006	2005
Age 21-30	0.023	0.036*	0.032*	0.047***	0.055***
	(1.37)	(1.90)	(1.86)	(4.21)	(2.87)
Age 31-40	0.048***	0.049***	0.043***	0.051***	0.059***
	(8.47)	(5.90)	(6.36)	(6.11)	(5.02)
Age 61+	-0.106***	-0.079***	-0.089***	-0.075***	-0.175
	(-7.47)	(-3.69)	(-4.49)	(-15.00)	(-1.40)
Dwelling	-0.056***	-0.058**	-0.051***	-0.005	-0.034
	(-2.88)	(-2.54)	(-2.63)	(-0.30)	(-1.32)
Car	-0.003	-0.009	0.019	0.002	-0.009
	(-0.15)	(-0.43)	(0.95)	(0.34)	(-0.42)
Financial			-0.150***		
strength			(-5.69)		
Income		-0.033	-0.001		-0.036
		(-1.61)	(-0.06)		(-1.61)
Arrears		0.100***	0.07***		0.136***
		(5.80)	(4.82)		(6.70)
Household size		0.058***	0.043***		0.071***
		(4.24)	(3.24)		(4.59)
Business	-0.004	-0.031***	-0.002	0.014	-0.050***
	(-0.32)	(-2.72)	(-0.15)	(0.57)	(-2.86)
Log likelihood	-4021.437	-3950.888	-3869.644	-3312.395	-2924.021
Number of obs	5142	5134	5134	4794	3322
Uncensored obs	1573	1568	1568	1225	1258

Table 7: Specifications analyzing the effects of age on the credit burden

Note: The dependent variable is the probability that a household reports high credit burden. Only second stage equation is reported. Coefficients report the average marginal probability effects. z-statistics are adjusted for clustering at the regional level and presented in parentheses. ***, **, and * denote significance at 1 per cent, 5 per cent, and 10 per cent, respectively.

Specification	F1	F2	F3	F1+	F1R
Year	2005	2005	2005	2006	2005
Bad health	0.110***	0.090***	0.083***	0.136***	0.020
	(3.94)	(2.96)	(2.67)	(5.37)	(0.77)
Good health	-0.092***	-0.086***	-0.053***	-0.011	-0.034***
	(-4.05)	(-5.67)	(-2.87)	(-0.37)	(-4.97)
Dwelling	-0.077*	-0.089**	-0.041	0.035	-0.053***
	(-1.86)	(-1.97)	(-0.95)	(0.61)	(-3.47)
Car	-0.072**	-0.051	0.008	-0.161***	-0.009
	(-1.99)	(-1.25)	(0.22)	(-7.32)	(-0.50)
Financial strength			-0.279***		
			(-7.34)		
Income			-0.060*		
			(-1.88)		
Arrears		0.174***	0.092***		
		(6.49)	(2.79)		
Household size		0.044**	0.044**		
		(2.41)	(2.41)		
Single	0.031	0.068*	0.020	-0.039	-0.114***
	(0.92)	(1.84)	(0.52)	(-0.45)	(-4.03)
Divorced	0.129***	0.132***	0.092***	-0.018	-0.001
	(4.05)	(4.16)	(2.96)	(-0.36)	(-0.04)
Widowed	-0.144***	-0.091***	-0.138***	-0.072**	-0.053*
	(-5.50)	(-2.76)	(-5.22)	(-2.05)	(-1.66)
Business	-0.080***	-0.071**	-0.080*	-0.007	-0.015
	(-2.62)	(-2.32)	(-1.77)	(-0.12)	(-1.11)
Log likelihood	-4016.199	-3967.561	-3873.030	-3334.098	-2991.064
Number of obs	5142	5142	5134	4793	3329
Uncensored obs	1573	1573	1568	1224	1263

Table 8: Specifications analyzing the effects of health condition on the credit burden Note: The dependent variable is the probability that a household reports high credit burden. Only second stage equation is reported. Coefficients report the average marginal probability effects. z-statistics are adjusted for clustering at the regional level and presented in parentheses. ***, **, and * denote significance at 1 per cent, 5 per cent, and 10 per cent, respectively.

Specification	G1	G2	G3	G3+	G2R
Year	**2005**	**2005**	**2005**	**2006**	**2005**
Low education	0.041	0.020	-0.004	-0.029	-0.005
	(1.44)	(0.69)	(-0.13)	(-0.60)	-0.18
High education	-0.121***	-0.096**	-0.045	-0.077***	-0.053**
	(-2.72)	(-2.13)	(-1.10)	(-2.76)	-1.98
Dwelling	-0.072*	-0.077**	-0.050	0.053	-0.044**
	(-1.77)	(-1.97)	(-1.27)	(1.03)	-2.31
Car	-0.050	-0.030	0.020	-0.027	-0.001
	(-1.16)	(-0.78)	(0.53)	(-0.90)	-0.03
Financial			-0.282***	-0.214***	
strength			(-13.80)	(-6.66)	
Income			-0.042	-0.048**	
			(-1.36)	(-2.29)	
Arrears		0.166***	0.105***	0.265***	0.128***
		(5.50)	(3.08)	(6.55)	7.71
Household size		0.052***	0.061**	0.046	0.043***
		(4.54)	(2.45)	(1.49)	5.41
Single	-0.007	0.042	0.001	-0.046	-0.042**
	(-0.19)	(1.33)	(0.03)	(-0.68)	-1.98
Divorced	0.117***	0.115***	0.084***	-0.019	0.031
	(2.97)	(3.50)	(2.81)	(-0.35)	1.16
Widowed	-0.118***	-0.046*	-0.09***	-0.036	-0.023
	(-5.47)	(-1.94)	(-3.47)	(-1.06)	-0.64
Business	-0.065**	-0.051**	-0.057	0.009	-0.021
	(-2.45)	(-2.25)	(-1.45)	(0.16)	-1.51
Log likelihood	-4016.399	-3968.238	-3874.005	-3196.156	-2960.799
Number of obs	5130	5130	5122	4781	3325
Uncensored obs	1572	1572	1567	1224	1263

Table 9: Specifications analyzing the effects of education level on the credit burden Note: The dependent variable is the probability that a household reports high credit burden. Only second stage equation is reported. Coefficients report the average marginal probability effects. z-statistics are adjusted for clustering at the regional level and presented in parentheses. ***, **, and * denote significance at 1 per cent, 5 per cent, and 10 per cent, respectively.

Age has significant effect on credit burden as shown in E specifications. Households with head between 21 and 40 face higher probability of a high credit burden. Age group 21-30 years have a bit smaller difficulty with repaying credits than age group 31-40. This result suggests that banks are not always able to assess the riskiness of the client and they are giving access to credits to some households whose financial situation makes it hard to repay their loans. On the other hand, older households of individuals aged 61 and over have less problems with repaying their credits than our base cathegory. Those households are probably using mainly consumer credits for kitchen accesories or other small loans rather than huge credits such as car loans and therefore they do not report high credit burden. Another reason that can explain this result is that retired people already have certain stable income. Age is a hard information and can be easily implemented into banks' scoring models. However, in analysis of health condition and its effects on repayment ability, the problem of **asymmetric information** emerges.

Specifications F analyze the effects that overall *health condition* of household head have on the credit burden. As expected, bad health indicates higher probability of the credit burden and good health lower probability of the credit burden. Both variables are highly significant at 1%. It follows from those results that health condition is an important factor that should be taken into account by banks that are assessing credit aplicants. However, to truly evaluate health condition of the credit applicant is behind the ability of the bank. The reason is that health condition is a private information and the bank does not have access to applicant's medical record. This information asymmetry makes it hard for the bank to fully explore health condition of the credit applicant and to use this variable in the scoring model. The person who wants to get a loan might conceal information about her health condition in order to get better conditions. Moreover, not even the applicant herself knows her health condition perfectly because of the latent essence of some disesases, e.g. predisposition to heart-attack. The problem of asymmetric information consists in the fact that people with bad health condition are not able to work but they still would like to spend as much as healthy people. However, as bad health will be reflected by lower income, banks can use income as a signaling device that reveals potentional information about clients' health.

Finally, specifications G analyze the effects of *education level* on the probability of high credit burden. Education is an example of hard information because it can be evidenced by degree certificate. The effect of low education on the credit burden is close to zero which suggest that banks' scoring schemes are well-designed and are not underestimating the effects that education have on possibility of repaying. However, the coefficients are not significant so we can not make reliable conclusions here. On the other hand, effect of high education on the credit burden is significant and with magnitude almost 5% which means that banks are possibly underrating repaying abilities of persons with post-secondary or terciary education.

As expected, *household size* has positive effect on the probability of high credit burden and it is significant even though correlation between number of *members* living in a household and *income* is strongly positive (0.61) which means that bigger households usually have higher income. The expenditures rise significantly with the household size because because big households have higher housing expenses and higher consumption which prevails over advantage of higher diposable income.

Businessmen have significantly lower probability of high repayment burden. This is related to businessmen's higher responsibility and higher earnings ability. While business is not always going well and being a businessman thus is not always an advantage, banks can monitor businessmen's activities by looking at various financial ratios of their companies and therefore there is enough information to assess their riskiness.

Interesting results appear when analyzing data from 2006. We keep in mind that 2006 was better year than 2005 because the upswing of the economy was greater.

- Effects of health on the credit burden is less important in a good year - variable good health is not significant in 2006 specifications.

- Significance of household head being a businessman and significance of dwelling ownership on credit burden of households decreases. That means that during boom, even tenants and people that were not businessmen were more succesful than in previous year and their probability of reporting high credit burden was thus lower.

Analysis of households' credit burden on a sample without retired people does not reveal too surprising results. People older than 60 have still lower probability of reporting high credit burden but this effect is not significant when analyzing dataset without retirees. Also, single people are significantly less burdened by credit repaying as shown in F1R and G2R specifications. This can be explained by the fact that singles simply do not need a loan.

Chapter 6

Conclusions

The aim of this thesis was to examine the determinants of credit burden of households in Slovakia using EU-SILC dataset. We analyzed the relationship between household's repayment ability and various household characteristics (ownership of dwelling and car, affordability of consumption and financial strength, income, presence of arrears, household size) and individual characteristics of household head (marital status, being a businessman, age, overall health condition, education level). We classified some of those characteristics into four groups: capital adequacy, responsibility, earnings ability and liquidity. The results suggest that the most important variable is responsibility and that higher responsibility significantly decreases the probability of the high credit burden. Higher earnings ability decreases credit burden as well but this effect is not significant in all of the specifications. Capital adequacy is important in selection equation which suggests that banks require suitable collateral to secure credits. However, capital adequacy does not have significant effect on the credit burden. Higher liquidity indicates lower probability of high credit burden, while lower liquidity represented for example by the presence of arrears implies repayment difficulties.

To summarize, banks in Slovakia are quite succesful in assessing riskiness of their clients and are usually granting adequate credits. We have not found evidence of providing credits to risky applicants with intention to get hold of the collateral. While our results suggest that divorced people face higher difficulties with repaying than married one, there is not much banks can do about it unless they can model probability of divorce. We found that banks are not very succesful in using household responsibility in their scoring models. If banks could find a way to collect information about households' affordability of consumption and financial strength, quality of their scoring models would improve and they could reduce high credit burden of some of the Slovak households. Banks do not have access to claimers' medical record and therefore are in risk of asymmetric information connected with her health. Banks could require life insurance in their favor to reduce bank risk, especially when providing bigger loans. Further, banks should try to get more information about arrears payments associated with the credit applicant to reduce possibility of a high credit burden.

Interesting result that follow from our paper are:

- The banks were acting responsible, they did not grant loans to too risky clients and they required collateral. This means that the rapid lending was not only an implication of boom-bust cycles.
- The credit burden is dependent on trend of economy. During good times, people are less likely to report high credit burden as was shown empiricaly in our specifications.
- Soft information is important determinant of the credit burden and the banks should try to employ it into their scoring models.
- Asymmetric information between borrowers and lenders is important factor that can cause high credit burden of households.

It would also be appropriate to analyze datasets from years of financial crisis to see what are the implications of bad times on the credit burden of households. However, this problem is behind the scope of this thesis and we will leave it for future research.

Chapter 7

Appendix

7.1 Definition of the variables

In this section we will briefly define the variables that we used in individual specifications.

Household level
has loan Dummy variable which takes the value of one if the household has a loan.
burden Dummy variable which takes the value of one if repayments of the credit for hire purchase or loans other than mortgage or loans connected with dwelling are a big financial burden for the household.
dwelling Dummy variable which takes the value of one if the household owns the dwelling (house or appartment). The person is considered owner if he owns title deed whether the house is paid off or not.
car Dummy variable which takes the value of one if the household owns the car.
affordability of consumption Categorical variable measuring the ability of the household to manage with their available income. This variable takes values from 1 to 6, where the higher number indicates higher affordability.
financial strength Dummy variable which takes the value of one if the household can face unexpected financial expenses.
income Total yearly available income of the household in Slovak crowns (log). We have also divided income into 3 groups by tercils to get variables low, medium and high income.
arrears Dummy variable which takes the value of one if the household has arrears of payment.
household size Modified OECD scale is used: weight of 1 for the first adult, weight

of 0.5 for every other adult member and weight of 0.3 for every child of age less than 14 years.

members Number of members in the household.

Personal level

married Dummy variable which takes the value of one if the head of the household is married.

single Dummy variable which takes the value of one if the head of the household is single.

divorced Dummy variable which takes the value of one if the head of the household is divorced.

widowed Dummy variable which takes the value of one if the head of the household is widowed.

unemployed Dummy variable which takes the value of one if the head of the household is unemployed.

retired Dummy variable which takes the value of one if the head of the household is retired.

age Age of the head of the household at the end of the reference period.

bad health Dummy variable which takes the value of one if the head of the household described her overall health condition as "very bad" or "rather bad"

good health Dummy variable which takes the value of one if the head of the household described her overall health condition as "rather good" or "very good"

low education Dummy variable which takes the value of one if the head of the household has either primary or lower secondary education.

high education Dummy variable which takes the value of one if the head of the household has either post-secondary or terciary education.

business Dummy variable which takes the value of one if the head of the household is a businessman.

Bibliography

[1] J. Heckman (1979). *"Sample selection bias as a specification error".* Econometrica 47 (1): 153–61.

[2] B. Égert , P. Backé, T. Zumer (2006). *Private-Sector Credit in Central & Eastern Europe: New (Over) Shooting Stars?* William Davidson Institute Working Paper Number 852.

[3] J. Fidrmuc, R. Horváth, E. Horváthová (2008). *Corporate Interest Rates and the Financial Accelerator in the Czech Republic.*

[4] J. Fidrmuc, C. Hainz (2009). *Default rates in the loan market for SMEs: Evidence from Slovakia.*

[5] J. Fidrmuc, M. Senaj (2010). *Income, Consumption, and Housing Wealth in Slovakia.*

[6] M. A. Petersen (2004). *Information: Hard and Soft.*

[7] M. Golder (2009). *Methods IV: Advanced Quantitative Analysis.* Teaching Material. http://homepages.nyu.edu/ mrg217/homepage.htm

[8] W. Greene (2003). *Econometric Analysis.* New Jersey: Prentice Hall.

[9] J. Wooldridge (2002). *Econometric Analysis of Cross Section and Panel Data.* Cambridge: MIT Press.

[10] H. Varian (1992). *Microeconomic Analysis.* New York: W. W. Norton.

[11] R. N. Bebczuk (2003). *Asymmetric Information in Financial Markets. Introduction and Applications.* Cambridge University Press.

[12] E. G. Mendoza, M. E. Terrones (2008). *An anatomy of credit booms. Evidence from macro aggregates and micro data.* NBER working paper series.

[13] *NBS statistics* NBS statistics, Interest Rates valid until December 31, 2008, BRIBOR / BRIBID

[14] http://www.inflation.eu/inflation-rates/slovakia/historic-inflation/hicp-inflation-slovakia.aspx

[15] J. Fidrmuc, P. J. Süß (2010). *The outbreak of the Russian Banking Crisis.* AUCO Czech Economic Review 5 (2011)

[16] C. Hainz, T. Nabokin (2009). *Access to Versus Use of Loans: What are the True Determinants of Access?*

[17] J. Fidrmuc, M. Senaj (2010). *Income, Consumption, and Housing Wealth in Slovakia: Winners and Losers of Transition.* International conference on Regional Disparities in Smolenice, 17-19 November 2010

[18] M. Brown, S. Ongena, A. Popov, P. Yesin (2010). *Who Needs Credit and Who Gets Credit in Eastern Europe?* 2nd Draft Manuscript for Economic Policy

[19] J. Fidrmuc, M. Hake, H. Stix (2010). *Households' Foreign Currency Borrowing in Central and Eastern Europe.*

[20] C. F. Baum, M. E. Schaffer, S. Stillman (2011). *Using Stata for Applied Research: Reviewing its Capabilities.* Journal of Economic Surveys

[21] T. Beck, M. Brown (2010). *Which households use banks? Evidence from the transition economies.* European Banking Center Discussion Paper No. 2010–25

[22] S. Bridges, R. Disney (2003). *Use of credit and arrears on debt among low income families in the United Kingdom.*

[23] M. Fischer, C. Hainz, J. Rocholl, S. Steffen (2011). *Government Guarantees and Risk Taking Incentives.*

[24] T. Dinha, C. Hainzb, S. Kleimeier (2010). *Collateral and its Determinants: Evidence from Vietnam.*

[25] E. Kočenda and M. Vojtek (2009). *Default Predictors and Credit Scoring Models for Retail Banking.* CESifo Working Paper Series number 2862

[26] O. Jeanne, A. Korinek (2010). *Managing Credit Booms and Busts: A Pigouvian Taxation Approach.*

[27] R. B. Myerson (2011). *A model of moral-hazard credit cycles.*

[28] W. P. M. M. Van de Ven, B. M. S. Van Pragg (1981). *The demand for deductibles in private health insurance: A probit model with sample selection.* Journal of Econometrics 17: 229–252.

[29] *Introduction to Heckman probit with examples (from the Stata manual)* http://www.stata.com/features/heckman-probit/

[30] F. Boissay, O. Calvo-Gonzalez, T. Koźluk *Is Lending in Central and Eastern Europe developing too fast?*

[31] M. Brzoza-Brzezina *Lending Booms in Europe's Periphery: South-Western Lessons for Central-Eastern Members.*

[32] A. Calza, M. Manrique, J. Sousa (2003). *Aggregate loans to the Euro area private sector.* ECB Working Paper No. 202

[33] A. Calza, C. Gartner, J. Sousa (2001). *Modelling the demand for loans to the private sector in the euro area.* ECB Working Paper No. 55

[34] *The World Bank website* http://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG

[35] G. A. Akerlof (1970) *The Market for "Lemons": Quality Uncertainty and the Market Mechanism.* The Quarterly Journal of Economics, Vol. 84, No. 3. (Aug., 1970), pp. 488-500.

[36] Umberto Eco (1977) *Come si fa una tesi di laurea*