Adverse selection, loan choice and default in the Chilean consumer debt market

Carlos Madeira
Central Bank of Chile

Adverse selection, loan choice and default in the Chilean consumer debt market

Carlos Madeira*

Draft

October 2014

Abstract

Using household survey data I estimate a model of consumer loan choice and default behavior. I show that households are sorted into different lenders, with higher income and education being positively associated with choosing Bank loans and negatively associated with other lenders. Debt amounts increase with income, unemployment risk and wage volatility, providing evidence that consumer debt is used to smooth income shocks. Also, debt amounts increase with household size and are quadratic in age, resembling the life-cycle consumption profile. Default behavior decreases with income and increases with higher indebtedness, unemployment and wage risk, confirming the role of adverse selection.

JEL Classification: E21; E24; E32; E51; G01; G21.

Keywords: Consumer credit; Default risk; Unemployment.

*carlosmadeira2009@u.northwestern.edu. All errors are my own.

1 Introduction

Household debt increased consistently in the last decades, both in emerging economies (IMF, 2006) and developed countries (Girouard, Kennedy, André, 2007). This evolution in the quantity of household credit coincided with a period of strong financial innovation, with a great range of loan products being available to consumers. Consumers are able to access credit from a variety of sources, such as credit cards, auto loans, education loans, and for motives as diverse as health, vacations, purchase of durable goods, or a renegotiation of previous debts. Also, the technological evolution has allowed lenders to process larger and better databases on the characteristics of debtors, allowing for an increased use of credit scoring and an heterogeneity of loan terms for each loan applicant (Roszbach, 2004, Edelberg, 2006, Einav, Jenkins and Levin, 2012). Yet despite an increasing availability of consumer credit, several families are still unable to access credit markets or obtain lower loan amounts than desired (Adams, Einav and Levin, 2009). The factors behind consumers' loan choice and their credit constraints have been documented in recent studies for the United States (see Dynan and Kohn, 2007, Attanasio, Goldberg and Kyriazidou, 2008, Adams, Einav and Levin, 2009); however, loan choice in developing countries is still understudied.

This paper studies the consumer credit access, lender choice and repayment behavior of families in Chile. Consumer loans are particularly relevant in Chile, since around 60% of the households have some consumer debt. Using data from the Chilean Household Finance Survey (EFH), I estimate an econometric model in which families choose among a variety of lender types according to their earnings, labor risk, demographics, and unobserved preferences. I find that families are sorted into different lenders according to their labor market risk. Furthermore, household's debt levels, income, and labor market risk have a significant impact on default behavior, which shows evidence of adverse selection and moral hazard (Einav, Jenkins and Levin, 2012) in Chilean loan markets.

In Chile the market for consumer loans has several different providers and their credit offers represent imperfect substitutes for consumers. These loan providers access different customer lists and information, besides being subject to different legal regulations, which affects their loan terms and the ability to target specific markets (see Marinovic, Matus, Flores and Silva, 2011, for a review of the structure and legal framework of different credit providers). The industrial organization literature argues that even small differences across product providers - such as the cost of screening applicants, brand preferences, marketing initiatives, search frictions, travel costs, tied products and asymmetric information - can create substantial frictions for customers' decisions (Nevo, 2011). Therefore it seems adequate to treat loan decisions as a differentiated product model where heterogeneous agents with unobserved preferences choose their preferred loan provider.

Household surveys are an ideal source of information for the study of the market of consumer loans, since the study of problems such as adverse selection requires information on the heterogeneity of agents. Aggregate data can hide the factors affecting the decisions of different consumers. The Chilean Household Finance Survey (EFH) collects detailed information on families' income, assets, loans, debt repayment behavior, and demographic characteristics. From 2007 to 2011 the EFH interviewed 12,264 families, which includes panel data of 2,739 families who were interviewed twice.

The model of loan choice and repayment behavior has three main components: i) a categorical choice between having no debt, wanting debt but being credit constrained, and five different types of lenders, ii) the choice of loan amount, and iii) a categorical outcome of whether the household defaulted or not on at least one payment over the previous year. The five lender types in this categorical model correspond to: Banks, Banks and Retail Stores, Retail Stores, Social Credit (i.e., loans provided by credit and labor unions), and Other Loans (which includes mainly auto loans, educational debt, plus pawn shops and some informal lending). Banks and Retail Stores are the two major lenders in Chile, therefore using both lenders is treated as a separate choice than the option of using just one type of lender. Other types of lenders represent a small proportion of the population and therefore I do not model the interaction of those lenders with other types of debts. Furthermore, there are two options for the families that did not get a consumer loan, which are "No Access to Debt" and "No wish to apply for Consumer Loans". The option of "No Access to Debt" represents families with credit constraints. These are families who applied for consumer loans but were denied credit, plus those who wished to apply for credit but did not do so because they expected to be refused. "No wish for Loans" represents the outside option for all agents, comprising the families who report not having consumer debt and no interest in applying for loans.

All three endogenous variables - the choice of type of lender, loan amount, and repayment outcome - are affected by both observable factors and unobserved preferences. The observable factors include income, education, labor income risk, and demographic characteristics such as the age of the household head and household size. Also, I consider that the choice of lender is affected by the motives behind the indebtedness and by unobservable preferences of the household. The observable motives are broadly classified in three categories: general consumption, the payment of previous loans or debt consolidation, and health needs (Chatterjee et al., 2007, show that a significant part of unsecured debt in the USA is contracted for health reasons). Labor market characteristics are measured in terms of three different variables for the working members of the household: i) the unemployment rate, ii) the wage volatility, and iii) the replacement ratio of income during an unemployment spell (the proportion of the working income that workers still earn after losing a job). Also, as suggested by Shimer (2012), the unemployment rate can be further decomposed in terms of two different mechanisms, the separation rate (the probability of entering unemployment given that one had a job before) and the job finding rate (how quickly workers exit out of unemployment). All of these measures of households' labor market risk provide interesting insights into the nature of their shocks over time.

Unobserved preferences include random effects denoting the taste for each type of lender, a taste for higher or lower loan amounts, and a propensity of each family to repay or not its loans. The random-effects for the taste of each choice are correlated, implying for instance that families with a higher taste for certain lenders may have a higher propensity to default on their loans. Besides the fixed unobserved tastes that are constant for each family, the model also takes into account there are uncorrelated shocks for each time period, implying that agents' decisions may change over time due to some unexplained shock. Assuming a parametric distribution for the unobserved terms, this model can be estimated using Simulated Maximum Likelihood (Train, 2009).

The results show that households with different characteristics tend to sort themselves among different lenders. Households with No desire for Consumer Loans have the lowest wage volatility and the lowest unemployment and job separation rates among all groups. This result seems to confirm that consumer debt is related to smoothing income shocks, therefore households with few income shocks have low demand for consumer debt. Banks are the institution that applies credit scoring and customer specific interest rates on a wider basis, therefore Banks capture the households of highest income and with the lowest unemployment rates among loan applicants. Also, households with loans in Banks suffer the lowest income falls during unemployment. Households with loans in Bank plus Retail and Other debts have both the largest loans in absolute amount and the larger ratios of debt relative to income. Households with No Access to Debt have the lowest income levels and also suffer the strongest income falls during unemployment spells.

Unemployment rates increase the probability of households opting for all loans, but their impact is highest for the clients of Retail Stores, Social Credit and Other Loans. Wage volatility is more strongly associated with households opting for Social Credit, Other Loans and No Access to Debt. Loan amounts increase with income, unemployment risk and wage volatility, therefore it is possible that households use consumer loans to smooth income shocks. The probability of default decreases with income and increases with high levels of debt amount and debt service (debt service includes both monthly amortization and interest payments) relative to income, unemployment risk and wage volatility, which confirms the existence of adverse selection and moral hazard among Chilean debtors. Bank debtors have a significantly lower probability of default even after controlling for observable variables. Since banks resort more to credit scoring and risk-adjusted interest rates, then one should expect banks to capture the customers with lowest risk (Edelberg, 2006), confirming the economic theory of equilibrium in loan markets with adverse selection (Jaffee and Russell, 1976, Jaffee and Stiglitz, 1990). It is also interesting that Health needs are positively associated with default behavior, which confirms the predictions of economic models for health expenses that are unpredictable and uninsurable for households (Chatterjee et al., 2007).

Finally, the probability of getting a loan and the choice of loan amount is increasing in the number of household members and quadratic in age, first increasing with the age of the household head and then falling in its later years. Therefore the demand for consumer debt has an age profile that resembles the findings of life-cycle consumption in the literature (Attanasio and Weber, 2010). In terms of unobservable factors, I find that households with higher income and education are less heterogeneous in their tastes, and that their choice of loan amount is less persistent over time.

This paper is related to a recent and growing literature of empirical models of loan choice and default behavior which measures the impact of observable risk factors and adverse selection (Roszbach, 2004, Adams, Einav and Levin, 2009, Einav, Jenkins and Levin, 2012). It extends that literature in three ways: i) it applies a similar framework for loan choice and default to a developing economy such as Chile, ii) it introduces a wider range of loan options and unobserved preferences by using tools from the applied product choice models in the field of industrial organization (Train, 2009, Nevo, 2000, 2011), and iii) it uses a more diverse characterization of labor income risk by separating overall risk into different variables such as unemployment risk and wage volatility.

This paper is organized as follows. Section 2 describes the consumer credit environment in Chile and the applied model of loan choice and default. Section 3 summarizes the Chilean Household Finance Survey dataset (2007-2011) and the main characteristics of Chilean families. Section 4 describes the sorting of households across different types of lenders in terms of loan amount, income and labor market risk. Section 5 presents the results of the joint model of lender choice, loan amount and default. Finally, section 6 concludes with implications for policy and future research.

2 Credit environment and empirical model of consumer behavior

2.1 The structure of consumer loan providers in Chile

This section starts with a review of the structure of Chilean credit markets and the differences among lenders, whether caused by differentiated product lines or by legal regulations (see Marinovic, Matus, Flores and Silva, 2011, for a review). In Chile all lenders have public access to a commercial registry of debtors who defaulted on payments[1], however this public registry is limited only to agents with negative histories and therefore lenders' information sets on the positive characteristics of loan applicants differ substantially, implying agents' can have different relationships with each lender (Jaffee and Stiglitz, 1990). Banks represent one type of consumer loan provider in Chile, as well as in other countries. Chilean banks have access to a common credit registry with information on all loan amounts and debt default within the banking system[2], but they do not observe loans from non-banking institutions. Banks also make a strong use of credit scoring, changing their loan offers according to agents' credit history, and may even tie their loan offers to other banking products that are signed by their customers, giving for instance preferential treatment to families that have direct deposit of wages, automatic bill payment, mortgages and other financial accounts with them.

Retail stores are another kind of credit provider, with a strong brand image and their own credit cards[3], and which have access to their own private databases on customers' loans and product

[1] See www.dicom.cl/.

[2] See the General Law of Banks of the Chilean Superintendency of Banks and Financial Institutions, www.sbif.cl.

[3] The norms for non-banking credit card providers are detailed in the Chapter III.J.1 of the Compendium of Financial Norms of the Central Bank of Chile.

transactions. Another type of loan providers are credit unions (denoted as Savings and Loans' Cooperatives[4]) and labor unions (denoted as Family Compensation Funds[5]) which are regulated as providers of "social credit". By legislation all Chilean companies must register their workers in one among several Family Compensation Funds, which provide social credit and other services to their affiliates. These labor unions or Family Compensation Funds represent 67.6% of the aggregate "social credit". Family Compensation Funds are chosen by each employer for all its workers and therefore workers do not choose their institution directly. Social credit providers must offer the same conditions to all of their affiliates, therefore they can change interest rates according to loan size and maturity, but are unable to discriminate against characteristics of the debtors such as their income. Also, Family Compensation Funds benefit from being able to deduct loan payments directly from their clients' wage payroll and therefore face little risk of default. Even in the case of a debtor losing its job, its Family Fund is able to deduct a substantial payment from the worker's severance compensation and therefore the risk of default is limited even relative to unexpected unemployment shocks. Finally, there are lenders with more specific goals, such as auto loans at car dealers, education loans, pawn shops[6], and consumer loans provided by insurance companies[7].

In terms of the aggregate amount of consumer credit in Chile, banks represented 71.7% of the total market, while social credit institutions represented 14.8% and retail stores 13.5%, respectively.[8] However, market presence in terms of customers differs from the aggregate loan amounts, since it is estimated that there are around 3.5 million debtors with banking loans, while social institutions and retail stores reach around 2.5 million and 7 million customers, respectively. Therefore retail stores are actually the largest provider of small consumer loans and reach the widest number of customers. Over the last half-decade the market size of each type of lender has differed substantially. Consumer loans in banks at the end of 2013 were 233% as large as their aggregate amount at the beginning of

[4] See the Chilean Government Department of Cooperatives, www.decoop.cl, the General Law of Cooperatives, DFL 5 (2003), www.bcn.cl, and Chapter III.C.2 of the Compendium of Financial Norms of the Central Bank of Chile.

[5] These institutions are regulated by the Chilean Superintendency of Social Security. Each Family Compensation Fund is associated with one of the five labor unions registered at the Confederation of Production and Trade. See the General Statute of Family Compensation Funds, articles 29 to 31 of the Law N°18.833 of 1989.

[6] See www.dicrep.cl.

[7] The regulation of credit by insurance companies is detailed in several norms of the Chilean Superintendency of Assets and Insurance, such as norms NCG 152 of 2002, NCG 208 of 2007 and NCG 247 of 2009.

[8] The aggregate amount of other loans (such as automotive and informal lending) is not entirely known, since credits of smaller and unregulated institutions do not need to be registered for statistical purposes.

2006 (Banco Central de Chile, 2013). Aggregate consumer credit by social institutions was 245% as large in 2013 as in 2006, but credit by retail stores grew only by 57% in the same period.

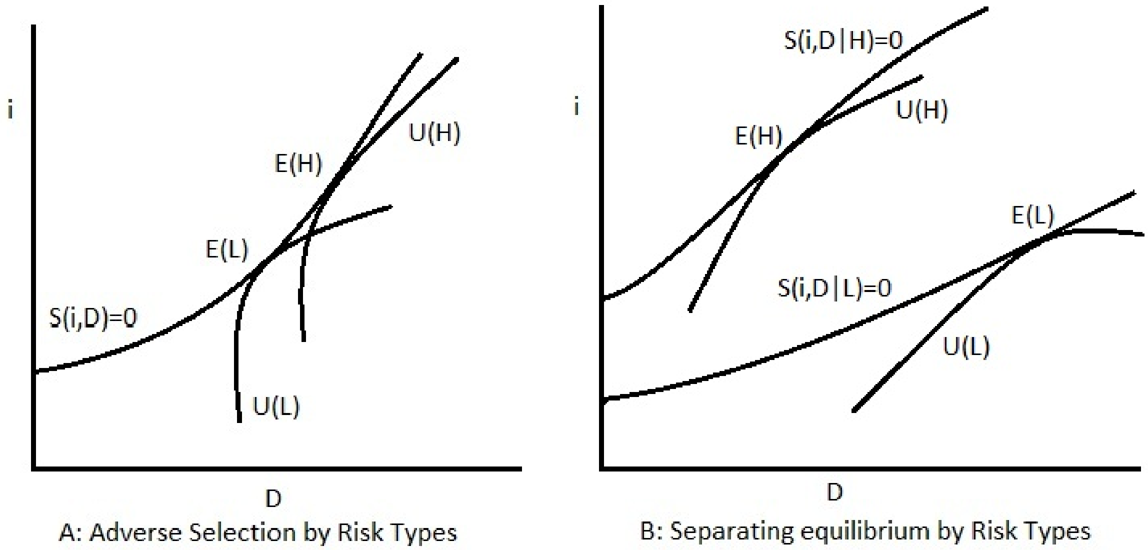
2.2 A review of the economic theory on adverse selection and lending

In this section I develop a simple review of the main results from the theoretical models of loan choice and credit constraints in equilibrium, along lines similar to Jaffee and Russell (1976), Jaffee and Stiglitz (1990), plus some empirical works such as Edelberg (2004), Adams, Einav and Levin (2009), and Einav, Jenkins and Levin (2012). For simplicity, let us think of two types of agent, one of high risk (H) and one of low risk (L). Agents of high risk are willing to pay higher interest rates for each loan amount, because either they are more credit constrained or they actually expect to default in the future (and therefore avoid the payment due to the higher interest rate). If lenders are unable to differentiate the risk type of agents, then the zero profit curve of credit supply will ask higher interest rates (i) from larger loan amounts (D), i.e. $S(i, D) = 0$. This supply strategy may lead different agents to reveal their type by choosing different contracts: i) a contract for low risk agents with low amounts and low interest rates, ii) a contract for high risk agents with a higher interest rate and a higher loan amount. Therefore the theory predicts that if lenders are unable to differentiate agents or apply credit scoring, then loan amounts will be positively correlated with the risk type of agents. Figure 1.A makes a graphical description of this adverse selection case.

The information technology evolution in the last 3 decades has made it easier to differentiate loan offers according to the characteristics and credit scores of applicants. The widespread use of credit scoring makes it more realistic to assume there are different loan supply curves for agents with observable low risk (L) and high risk (H) types. The lenders' credit supply curve will offer lower interest rates (i) at each debt amount (D) for low risk types, $S(i, D \mid L) = 0$, relative to the high risk type isoprofit curve, $S(i, D \mid H) = 0$. Therefore the theory predicts that lenders able to use credit scoring will offer larger loan amounts and lower interest rates to debtors of a low risk type. This case of adverse selection with observable characteristics is illustrated in Figure 1.B.

In general, it is more realistic to assume that there are both observable variables for agents' risk types and unobservable characteristics which credit scoring models are unable to include. Therefore economic theory should predict that better loan conditions (such as larger loan amounts, longer

Figure 1: Theory of self-selection of debtors across different loan contracts



maturities for payment, and lower interest rates) are associated with observable characteristics of lower risk (such as higher income and more secure jobs). However, unobservable risk characteristics (such as a taste for higher loan amounts) may create adverse selection and will be associated with larger loan demand and more frequent default. The standard theory of adverse selection and credit market equilibrium predicts that some types of high risk agents will not be a profitable market segment, due to either legal restrictions (such as usury laws and interest rate ceilings) or fixed costs for loan evaluation. For these high risk agents, lenders will be unable to offer profitable loans, therefore these agents will remain outside the credit market and will be credit constrained.

In summary, according to the economic theory of loan markets we should expect three results: i) lenders will offer better and larger loans to agents with observable characteristics of low risk, ii) unobservable characteristics of high risk will still be associated with both larger loan amounts and default, and iii) agents with very high risk will be credit constrained and without access to loans.

2.3 An empirical model of choice of lender, loan amount and debt default

The consumer choice model considers three endogenous variables: i) a categorical choice between having no debt, wanting debt but being credit constrained, and five different types of loans, ii) the choice of loan amount, and iii) a categorical outcome of whether the household defaulted or not on at least one payment over the previous year. The five lender types in this categorical model correspond to the major loan providers described in the previous section: Banks, Banks and Retail Stores, Retail Stores, Social Credit, and Other Loans (which includes mainly auto loans, educational debt, plus pawn shops and informal lending). It is possible that some consumers have more than one debt type, say debt at Banks and Other Loans (for example, an educational loan), but except for retail store credit (which reaches around 7 million people in Chile) there are few observations with such interactions. For simplicity, I classify the observed lender choice of each household as the one corresponding to the largest loan amount reported by each family. Banks and Retail Stores are the two major lenders in Chile, therefore using both lenders is treated as a separate choice when the household has a positive amount of loans with both lenders.

Families with no consumer loans are classified in two categories: "No Access to Debt" and "No wish to apply for Consumer Debt". "No Access to Debt" represents families with credit constraints, including those who applied for credit but were denied and the ones who did not apply for credit because they expected to be refused. "No wish for Debt" is the outside option for all agents, comprising the families who report no consumer debt and no interest in applying for loans. To be succinct, these options from now on will be denoted simply as "No Access" and "No Debt".

The modelling of a multivariate choice model with several options and many periods incurs into a problem of multidimensionality, since with P possible products there are P^T possible choices in a panel of T periods (Nevo, 2011). Therefore it is useful to apply a parsimonious model that can summarize the choice among the different options in terms of a restricted number of observable and unobservable factors. This is done in terms of a fully specified maximum likelihood model.

Let $U_{i,b,t}$ denote the utility of household i from the option b in period t, with $b \in \{1$ "Bank", 2 "Bank & Retail", 3 "Retail", 4 "Social", 5 "Other Loans", 6 "No Access"$\}$. Furthermore, let us standardize the utility of the outside option, "No wish for Debt", as zero, $U_{i,0,t} = 0$. This standardization is made without any loss of generality, since all that matters for the agents' choice is the difference in utility from each option relatively to the outside option (Nevo, 2000). Consumer chooses the option $Y_{i,t} = b$ of highest utility $(\max(U_{i,0,t}, U_{i,1,t}, .., U_{i,B,t}))$ and then a loan-amount $L_{i,t}$, which are affected by observable characteristics, $x_{i,t}$, plus unobservable preferences for each loan type b, $\varepsilon_{i,b,t}$, and loan-amount, $\zeta_{i,t}$. For simplicity, let us assume the utility of each loan type is both an additive and linear function of the observables and the error term:

$$1)\ U_{i,b,t} = \alpha_{b,t} + \beta_b x_{i,t} + \varepsilon_{i,b,t}.$$

If the consumer decides to have a loan (options 1 to 5) instead of either "No wish for debt" (option 0) or "No Access to Debt" (option 6), then he chooses a log-loan amount which is again an additive and linear function of the observable factors, $x_{i,t}$, plus an unobservable preference $\zeta_{i,t}$:

$$2)\ \ln(L_{i,t}) = \pi_t + \delta x_{i,t} + \zeta_{i,t}.$$

The decision of defaulting at time t, $D_{i,t} \in \{0, 1\}$, is then given by whether a latent propensity to default is positive, $d_{i,t} > 0$. The latent propensity for defaulting on loans is again given by an additive and linear function of the observable characteristics, $z_{i,t}$, plus an unobserved shock $\nu_{i,t}$:

$$3)\ d_{i,t} = \mu_t + \lambda z_{i,t} + \nu_{i,t}.$$

Note that the vector of observable variables that explains default, $z_{i,t}$, differs from the vector of observable variables that explains the choice of the type of loan and the loan-amount, $x_{i,t}$. This is an intentional feature of the model and it is necessary for identification. The reason is because choice models that include an endogenous variable (for example, default in this model) affected by sample selection into different groups (for example, the type of loan chosen by agents in this model) are ill-identified if the same exact vector of variables explains both the endogenous variable choice and the sample selection choice (Vella, 1998). Therefore it is useful if there are at least a few variables that affect sample selection (the loan choice, in this case), but do not affect the default decision directly. In our application there are actually some valid candidates for this role of instruments that affect loan choice, but not default. Note that although for simplicity of exposition all the variables are indexed as being observed at time t, in fact loans have a maturity of several periods (typically, around 1 or 2 years) and therefore the decision of loan choice happened before the repayment period. For this reason it is natural to use the lagged value of some variables as an explanation for loan choice and loan amount (for example, unemployment in the past year), but use the contemporary value of the same variables as an explanation for default. This choice of instrumental variables for loan choice is quite intuitive in economic terms and the validity of this identification approach is often recommended for panel data estimators (Vella, 1998).

To estimate the model it is necessary to specify the distribution of the unobservable random terms, which has some degree of subjectivity since there are several possible distributions that may provide a plausible fit. However, it is desirable that the distribution of the error term satisfies four characteristics: i) it allows for the unobserved preferences of each agent to be correlated over time, with some families being persistent in their behavior; ii) it accounts for some loan types being closer substitutes to each other, therefore the utilities of different options are correlated; iii) the agents' choice of all the distinct outcomes such as lender type, loan amount and default must be correlated, which is predicted by the theory of adverse selection of debtors (Einav, Jenkins and Levin, 2012); and, iv) the distribution should allow for an appropriate degree of heteroscedasticity, since groups are not equally affected by the unobserved shocks. A flexible way for achieving these desired properties is to assume the unobserved tastes for each option $\varepsilon_{i,b,t}$ are given by the sum of an independent extreme valued component plus a normal random-effect that is heteroscedastic and correlated over several choices and time periods (McFadden and Train, 2000, Nevo, 2000, 2011):

$$4.1)\ \varepsilon_{i,b,t} = \bar{\varepsilon}_{i,b} + \tilde{\varepsilon}_{i,b,t},$$

$$4.2)\ \bar{\varepsilon}_{i,b} = 1(1 \leq b \leq 5)\eta_{i,1} + 1(1 \leq b \leq 2)\eta_{i,2} + 1(2 \leq b \leq 3)\eta_{i,3} + \omega_{i,b},$$

with $\tilde{\varepsilon}_{i,b,t} \sim EV(0,1)$, $\eta_{i,a} \sim N(0, \sigma_{\eta_a}(x_i))$ and $\omega_{i,b} \sim N(0, \sigma_{\omega_b}(x_i))$. $1(.)$ is the indicator function, assuming the value 1 if the condition is met and 0 otherwise. $\bar{\varepsilon}_{i,b}$ is the random-effect that represents the time-invariant tastes of the agent for each choice. Equation 4.2) for $\bar{\varepsilon}_{i,b}$ has a simple interpretation in terms of its distinct components, with $\eta_{i,1}$ representing a random factor denoting agent i's taste for any type of loan, $\eta_{i,2}$ being a random factor denoting agent i's taste for both the Bank and Bank plus Retail loan options, and $\eta_{i,3}$ denoting his taste for the options of Bank plus Retail or just Retail. Finally, the random effect $\omega_{i,b}$ is agent i's specific taste for option b. The distribution of all the random-effects is heterocedastic in the vector x_i, which represents the time invariant characteristics of the agent and differs from $x_{i,t}$ which includes time-varying variables.

In a similar way, I assume the unobserved terms for loan amount, $\zeta_{i,t}$, and the propensity to default, $\nu_{i,t}$, are correlated with the unobserved tastes for loan type:

$$\begin{aligned} 4.3)\ \zeta_{i,t} &= \bar{\zeta}_i + \theta\bar{\varepsilon}_{i,b} + \tilde{\zeta}_{i,t}, \\ 4.4)\ \nu_{i,t} &= \bar{\nu}_i + \rho[\bar{\varepsilon}_{i,b}, \bar{\zeta}_i] + \tilde{\nu}_{i,t}, \end{aligned}$$

with $\tilde{\zeta}_{i,t} \sim N(0, \sigma_{\tilde{\zeta}}(x_i))$, $\bar{\zeta}_i \sim N(0, \sigma_{\bar{\zeta}}(x_i))$, $\bar{\nu}_i \sim N(0, \sigma_{\bar{\nu}}(x_i))$, and $\tilde{\nu}_{i,t} \sim EV(0,1)$. The log-loan amount is a continuous variable and for this reason the contemporary shock that each agent faces can be heteroscedastic. Note that the unobserved propensity of default is correlated with both the unobserved tastes for each loan type and the unobserved taste for loan amount $\bar{\zeta}_i$.

The model includes random-effects, which requires panel data to identify the parameters. However, the EFH data contains some purely cross-sectional samples and it is inefficient to ignore such a observations. For this reason the likelihood function includes both the panel and the cross-section samples, which is a specific case of a Full Information Maximum Likelihood (FIML) model. Let $\sum_i \equiv \{Y_{i,t}, L_{i,t}, D_{i,t}, Y_{i,t+s}, L_{i,t+s}, D_{i,t+s} \mid x_i, x_{i,t}, z_{i,t}, x_{i,t+s}, z_{i,t+s}\}$ be the vector containing agent i's choices of type of loan, loan amount and default at both time t and $t+s$, conditional on the observables of both years. Also, let $\varepsilon_i \equiv \left\{\eta_{i,1}, \eta_{i,2}, \eta_{i,3}, \omega_{i,1}, ..., \omega_{i,B}, \bar{\zeta}_i, \bar{\nu}_i\right\}$ be the vector of all the unobservable random-effects. All the random-effects in vector ε_i are independent of each other, therefore the pdf of ε_i is given by $f(\varepsilon_i) = \phi(\frac{\eta_{i,1}}{\sigma_{\eta_1}(x_i)})..\phi(\frac{\eta_{i,3}}{\sigma_{\eta_3}(x_i)})..\phi(\frac{\omega_{i,b}}{\sigma_{\omega_b}(x_i)})..\phi(\frac{\bar{\zeta}_i}{\sigma_{\bar{\zeta}}(x_i)})\phi(\frac{\bar{\nu}_i}{\sigma_{\bar{\nu}}(x_i)})$. This is assumed without any loss of generality, since the same random-effects affect different endogenous variables and therefore the endogenous variables are correlated with each other.

For simplicity of exposition it is easier to write the likelihood of the three endogenous variables given in equations 1), 2) and 3) conditional on the fixed-effects ε_i and then multiply it by the pdf $f(\varepsilon_i)$. Let $\tilde{U}_{i,b,t} = \alpha_{b,t} + \beta_b x_{i,t} + \bar{\varepsilon}_{i,b}$, $\ln(\tilde{L}_{i,t}) = \pi_t + \delta x_{i,t} + \bar{\zeta}_i + \theta\bar{\varepsilon}_{i,b}$, and $\tilde{d}_{i,t} = \mu_t + \lambda z_{i,t} + \bar{\nu}_i + \rho[\bar{\varepsilon}_{i,b}, \bar{\zeta}_i]$, represent the expected means for the latent variables of equations 1), 2) and 3), assuming ε_i is known. The likelihood of observing $\sum_i$ can then be written as a simple product of the multivariate probability of the observed loan option b (given by the traditional multivariate logit ratio), with the probability of loan amount $(L_{i,t})$ and subsequent default $(D_{i,t})$ in both periods:

$$\begin{aligned}
5)\ \Pr(\textstyle\sum_i) \quad = \quad & \int \cdots \int f(\varepsilon_i) \frac{\exp(\tilde{U}_{i,b,t})}{\sum_d \exp(\tilde{U}_{i,d,t})} \phi(\frac{\ln(L_{i,t}) - \ln(\tilde{L}_{i,t})}{\sigma_{\bar{\zeta}}(x_i)})^{1(L_{i,t}>0)} \frac{\exp(\tilde{d}_{i,t})^{D_{i,t}}}{1+\exp(\tilde{d}_{i,t})} \\
& \frac{\exp(\tilde{U}_{i,b',t+s})}{\sum_d \exp(\tilde{U}_{i,d,t+s})} \phi(\frac{\ln(L_{i,t+s}) - \tilde{L}_{i,t+s}}{\sigma_{\bar{\zeta}}(x_{i,t+s})})^{1(L_{i,t+s}>0)} \frac{\exp(\tilde{d}_{i,t+s})^{D_{i,t+s}}}{1+\exp(\tilde{d}_{i,t+s})} \partial\varepsilon_i.
\end{aligned}$$

For the cross-sectional sample, let the vector $\sum_{i,t} \equiv \{Y_{i,t}, L_{i,t}, D_{i,t} \mid x_i, x_{i,t}, z_{i,t}\}$ represent agent i's choices at time t, conditional on the observables $\{x_i, x_{i,t}, z_{i,t}\}$. If one assumes the panel and cross-sectional samples have the same representation in the population, then the likelihood function can be integrated for the same distribution of random-effects as the panel data observations. Note that this does not imply the model is unidentified, since the panel sample allows the model to identify the complete distribution of the unobservables. Therefore this approach is valid as long as the panel data sample is large enough. The likelihood of vector $\sum_{i,t}$ is therefore written as:

$$6)\ \Pr(\textstyle\sum_{i,t}) = \int \cdots \int f(\varepsilon_i)\dfrac{\exp(\tilde{U}_{i,b,t})}{\sum_d \exp(\tilde{U}_{i,d,t})}\phi(\dfrac{\ln(L_{i,t}) - \ln(\tilde{L}_{i,t})}{\sigma_{\tilde{\zeta}}(x_{i,t})})^{1(L_{i,t}>0)}\dfrac{\exp(\tilde{d}_{i,t})^{D_{i,t}}}{1+\exp(\tilde{d}_{i,t})}\partial\varepsilon_i.$$

The log-likelihood of the model is then given by the sum of the log-likelihood of the panel and cross-sectional samples, where $i \in P$ denotes whether the observation is in the panel sample or not:

$$7)\ LL = \sum_{t=1}^{T-1}\sum_{i=1,i\in P}^{N_t}\sum_{b=1}^{B}\ln(\Pr(\textstyle\sum_i)) + \displaystyle\sum_{t=1}^{T}\sum_{i=1,i\notin P}^{N_t}\sum_{b=1}^{B}\ln(\Pr(\textstyle\sum_{i,t})).$$

Besides the time-varying error terms, this model has 11 unobserved random-effects which form the vector ε_i and influence the correlation of different choices and periods. This implies that the likelihood function of equations 5) and 6) is based on a high dimensional integral and it is computationally difficult to calculate precisely. For this reason the choice probabilities are not calculated exactly, but rather based on an approximation which averages a limited number of draws, R, from the distribution of $f(\varepsilon_i)$. This Simulated Maximum Likelihood (SML) method is asymptotically consistent if R increases proportionally with N (Train, 2009). In this application I use 100 draws to simulate the probability of each observation, with the multivariate draws chosen by a Modified Latin Hypercube Sampling (MLHS) method (Hess, Train and Polak, 2006).[9] In general, the MLE asymptotic distribution is also valid for the SML method, but this asymptotic distribution is invalid if the model is not exactly true and if the number of draws R does not converge to infinity (Train, 2009). Therefore the model's standard-errors are estimated from 100 bootstrap replicas, which is asymptotically valid under a general set of conditions (Horowitz, 2001).

[9] I choose the MLHS method, because it chooses pseudo-random draws equally spaced in each dimension of the integral and then randomly paired across dimensions. The reason why MLHS can perform better than standard uniform draws is because uniform draws can have too much randomness and there is a certain probability of obtaining draws too close to each other, while some areas of the integral have few or no draws at all. In this sense MLHS guarantees that all the areas of each dimension are represented with at least one draw and therefore the simulated draws have a wider coverage. Some simulation studies show that 100 MLHS draws can be as efficient as more than 1000 uniform draws (Hess, Train and Polak, 2006). The MLHS method to obtain R multivariate draws basically starts with an equal spaced sequence of values, $\varphi(j) = \dfrac{j-1}{R}$ for $j = 1, ..., R$, in each dimension. Then a scrambled Halton pseudo-uniform number x is added to the draws of each dimension to get $\tilde{\varphi}(j) = \varphi(j) + \dfrac{x}{R}$ for $j = 1, ..., R$. The draws are then transformed using the inverse normal cdf and multiplied by the standard-deviation of the univariate distribution of the integral, $\sigma\Phi^{-1}(\tilde{\varphi}(j))$, to obtain an univariate normal draw. The draws of each dimension are then randomly paired with the R draws from the other dimensions to obtain R multivariate normal draws.

This model of loan decisions has certain implicit assumptions into it, since it assumes choices are well approximated by a function of known characteristics and randomly distributed unobserved preferences. One could assume other models for debt choice, such as an explicit multi-period optimization where agents choose the best option for maximizing expected lifetime utility based on an explicit evaluation of uncertain future paths and punishment costs for defaulting (see for instance, Chatterjee et al., 2007). However, an explicit lifetime optimization framework requires several assumptions about the agents' utility functions, their discount rates relative to future consumption and the knowledge agents have about their uncertain future outcomes. Empirical evidence of agents' cognitive limitations disputes assumptions such as rational expectations, time-consistency and revealed preference (Bertrand and Morse, 2009, Kahneman, 2011). Therefore simple behavioral models are not necessarily less realistic than structural models based on complete optimization. For this reason, the choice model in this paper is more closely related to other works who approximate agents' decisions in a flexible way, such as Edelberg (2006) and Einav, Jenkins and Levin (2012).

3 Data

3.1 The Chilean Household Finance Survey (EFH)

The main source of information for the characterization of the financial behavior of Chilean households is the Chilean Household Finance Survey (in Spanish, Encuesta Financiera de Hogares, hence on EFH). The EFH is a representative survey with detailed information on households' assets, debts, income and financial behavior, and is broadly comparable to similar surveys in the United States and Europe (Eurosystem, 2009). In 2007 and 2011 the EFH interviewed 3828 and 4059 urban families nationwide. In the years 2008 to 2010 the EFH was only implemented in the capital city of Santiago (which represents over 40% of the total national population), therefore the sample size is smaller for those waves. The EFH has a rotating sample, in which part of the sample is re-interviewed. Therefore there are 1792 families which were interviewed both in 2007 and 2011, while 947 families were interviewed both in 2008 and 2009. In total there are 6790 cross-sectional observations (i.e., families interviewed only once) plus 2739 panel observations (Table 1).

Table 1: Panel and cross-sectional sample size of the Household Finance Survey (EFH)

EFH	Panel	Cross-Section	Total
2007	1,792	2,036	3,828
2008	947	207	1,150
2009	947	243	1,190
2010		2,037	2,037
2011	1,792	2,267	4,059
Total	2,739	6,790	12,264

The EFH has a particularly detailed focus of the loans and debt commitments of each household. It asks for the largest 3 debts that each household has for each type of loan, among a total of 13 categories of loans: Banking Credit Card Debt, Banking Line of Credit, Banking or Financial Agency Consumer Credit Loan, Retail Store Credit Card, Retail Store Consumer Loan, Auto Loans, Social Credit, Education Loans, Loans from relatives, Loans from usurers, Pawn shops, Grocery and Shopping on credit (i.e., store tabs), and Other Debts. Therefore the survey may ask up to a total of 39 debts that the household has at the moment, although obviously very few agents will report having debts with all the possible categories of loans.

For two reasons it is easier to work with just 5 types of lenders (or 5 types of loans), therefore my analysis is limited to options that sum all the loans for a given lender type and with each family classified discretely with the lender type representing the largest loan amount: Banks, Banks and Retail Stores (for the families reporting the use of both kinds of loans), Retail Stores, Social Credit, and Other Debts. The first reason is that it is desirable to eliminate the irrelevant alternatives from the choice model (Train, 2009), with a classic example being the inclusion of options such as "red bus" and "blue bus" for agents that do not care about the color of public transport. Several of the 13 types of loans elicited by the survey are similar products and are often offered by lenders to the same customers and for similar purposes (for instance, many customers use Credit Cards and Lines of Credit for similar reasons, although their choices may depend on the specific convenience of the occasion). This is a strong reason for aggregating all the options for credit cards, lines of credit and consumer contracts for each lender, instead of treating them separately. The second reason is related to the curse of multidimensionality, since the number of parameters in the model increases with the number of options and it is difficult to make a reliable analysis of too many options, particularly if some options have few or no observations (for example, loans from usurers are reported by less than 0.07% of the families). For this reason, Other Debts represents the sum of Auto Loans, Education Loans, Loans from relatives, Loans from usurers, Pawn shops, Grocery and Shopping on credit (i.e., store tabs), and Other Debts. Note that this category is largely composed of Auto Loans, Education Loans and Other Debts, with the remaining options representing negligible numbers.

Table 2 shows the proportion of households that chose each of the 5 lender types, plus households with either No Consumer Debt (because the family does not want debt) and No Access to Debt (if the family applied for loans, but was refused). The proportion of households without a wish for consumer debt represents 27% of the Chilean population, while those with No Access to Debt represent close to 13% of the population. Retail Stores are the most popular choice among households, representing more than 40% of the population, with 29% being Retail Store only users and 13% being users of both Bank and Retail Store Loans.

For each debt the EFH survey registers its loan amount (in Chilean pesos), maturity (in months), and other details such as the motivation for contracting the loan (with possible motives including vacations). The survey questionnaire also asks about the loan's interest rates, but less than half the respondents report to remember them.[10] One important aspect of our study is how default of consumer loans is measured. The question used for measuring default is "Approximately, in the last 12 months have you fallen into morosity or late payments for each one of your loans?". I consider that default corresponds to a dummy variable denoting one or more events of morosity.

Table 2 shows the loan amount, maturity and morosity rates of different lender types using the pooled EFH sample, that is all the cross-sectional samples available. I also report the average loan interest rates of different lender types, from statistics of the Chilean Superintendency of Banks and Financial Institutions and the Superintendency of Social Security. There is no information on interest rates from users of several lenders, such as Banks and Retail or Other Debts. Table 2 also reports the share of the consumer loan destined for a given purpose of the household, more specifically "Pay previous debts" and "Health needs". Other motivations such as "general consumption" are not reported, since their classification is too general to be interpreted. The main

[10]This memory problem is explained by debtors tendency to remember the payment amount better than their contract's interest rate. Also, it is difficult to recover an estimate of the implicit interest rate from the financial formula for the present value of the payments of a loan. This is due to ommited variables in the financial formula, such as loan commission fees, and measurement error in the reporting of the loan amounts, payments and maturities.

Table 2: Population*, Maturity (months), Loan amounts (thousands of Chilean pesos), motivation (share of total consumer debt destined for a given purpose), interest rate and morosity rates (EFH)

Type of Debtor	Population	Maturity	Morosity	Loan amount: mean/percentiles				Interest	Pay debts	Health
		Mean	Mean	Mean	p25	p50	p75	Mean	Mean	Mean
Bank	7.8%	25.28	10.2%	2,549	416	1,110	2,649	19%	14.3%	5.0%
Bank+Retail	12.9%	20.45	21.3%	3,015	859	1,703	3,520		16.9%	6.0%
Retail Store	28.9%	12.17	19.0%	492	102	216	486	47%	3.7%	3.3%
Social Debt	5.6%	27.07	11.4%	1,124	307	590	1,131	21%	18.1%	13.8%
Other Debts **	4.6%	32.25	21.5%	4,101	1,207	2,425	4,274		8.0%	3.2%
No Consumer Debt	27.3%									
No Access to Debt	13.0%									

* % of the total Chilean households in urban areas. ** Maturity for Other Debts is for Auto Loans only.

conclusion is that households with Bank, Bank plus Retail, and Social Debt are more likely to have a motivation of paying back previous debts or health needs. Loan consolidation and health needs also motivate a significant part of unsecured debt in the USA (Chatterjee et al., 2007).

Users of Bank credit only have a morosity rate of 10%, which is half the value reported by users of both Bank and Retail credit (Table 2). Also, Bank users have much larger loan amounts and longer maturities than the users of Retail Stores. In Chile neither Retail Stores or institutions of Social Credit are able to offer heterogeneous interest rates to their customers, only Banks offer customer specific interest rates (Marinovic, Matus, Flores and Silva, 2011), so the economic theory predicts that Banks will get the best observable risk types by offering better loan terms such as lower interest rates, larger loan amounts and longer maturities. While Social Debt lenders are unable to risk price their offers, these institutions are able to garnish their clients' wages easily, therefore this high punishment cost should explain their low morosity rates. However, households with both Bank and Retail Store debt have morosity rates as high as the customers of Retail Stores only. Perhaps this can be explained because such debtors have an unobservable taste for high loan amounts. Table 2 shows that households with both Bank and Retail Store debt have much higher loan amounts than the debtors of Bank and Retail Store separately, which could be a sign that these are debtors with particularly high needs for liquidity. The households with Other Debts also have high loan amounts and morosity rates, but perhaps this can be explained by special characteristics of these debtors. For example, education loans are granted to younger agents, who may be more

Table 3: Population of debtors, loan amounts (thousands of pesos) and morosity over time (EFH)

Type of Debtor	Population		Loan amount (median)		Morosity rate	
	2007	**2011**	**2007**	**2011**	**2007**	**2011**
Bank	6.5%	8.2%	968	1,176	8.8%	11.7%
Bank+Retail	13.6%	11.8%	1,435	1,826	18.9%	24.6%
Retail Store	31.9%	25.9%	232	177	21.1%	19.5%
Social Debt	3.8%	7.8%	484	748	12.1%	12.2%
Other Debts	4.6%	4.9%	1,511	2,866	25.2%	20.5%
No Consumer Debt	26.6%	28.7%				
No Access to Debt	13.0%	12.7%				

subject to unemployment risk and unstable income. Also, perhaps education and auto loans have lower punishment costs for morosity, since lenders cannot deduct payments and punishment fees from their clients' bank accounts (as Banks do) or their wages (as Social Credit institutions do).

Table 3 shows the percentage of the population, median loan amounts and morosity rates in the years 2007 and 2011. The biggest changes observed between 2007 and 2011 are that users of only Banks and Social Debt increased respectively to 8.2% and 7.8% of the population. Loan amounts of users of Social and Other Debts increased substantially, while the median loan amount at Banks increased less. It is also noticeable that the morosity rate of Bank users increased somewhat.

Tables 4 and 5 summarize the changes to income and use of consumer loans in Chilean households, using information from the EFH panel sample (2007-2011). In Table 4 I report the transition probabilities from one household income quintile ($Q_{i,t}$) to another between 2007 and 2011, $\Pr(Q_{i,2011} = q \mid Q_{i,2007} = q')$, where 1 denotes the families with the 20% lowest income. The conclusion is that household income has some persistence, but there is substantial income volatility in Chile. The probability that a household of the lowest income (quintile 1) will remain at the bottom of the distribution is 40%, while the probability of a household remaining at the top income level (quintile 5) is 53%. Among the middle income levels (quintiles 2 to 4), mobility is even higher and there is a high chance that households will move into either a higher or a lower income level.

In Table 5 I show the transition probability of a household changing from one lender type to another or towards having either no consumer debt or no access to debt, $\Pr(Y_{i,2011} = b \mid Y_{i,2007} = b')$. The last column in the table replicates the share of the population in each debt status over the whole period of 2007 to 2011. If one compares the diagonal values of the transition matrix, which

Table 4: Transition of families across different income quintiles (EFH Panel, 2007-11)

	Quintile 2011				
Quintile 2007	**1**	**2**	**3**	**4**	**5**
1	**40%**	27%	19%	7%	7%
2	22%	**29%**	27%	15%	7%
3	11%	23%	**26%**	25%	15%
4	10%	14%	22%	**30%**	24%
5	7%	9%	12%	20%	**53%**

Table 5: Transition of households across different debtor types (EFH Panel, 2007-11)

	Debt Status in 2011							**Population**
Debt in 2007	**No Debt**	**Bank**	**Bank+Retail**	**Retail**	**Social**	**Other**	**No Access**	**in 2007-11**
No Debt	**40.9%**	7.0%	6.7%	27.8%	4.2%	2.3%	11.1%	27.3%
Bank	27.8%	**18.1%**	20.5%	19.1%	3.9%	2.1%	8.4%	7.8%
Bank+Retail	16.4%	18.2%	**30.0%**	25.1%	2.2%	1.5%	6.7%	12.9%
Retail Store	20.9%	7.5%	13.6%	**39.1%**	5.2%	0.4%	13.3%	28.9%
Social Debt	36.2%	8.1%	3.5%	34.0%	**14.1%**	0.0%	4.1%	5.6%
Other Debts	34.7%	12.7%	28.4%	12.6%	0.0%	**5.5%**	6.1%	4.6%
No Access	33.4%	4.9%	7.8%	28.9%	6.9%	0.6%	**17.6%**	13.0%

represent the probability of a debtor keeping the same status as previously, with the average debt status of the population, then one gets an idea of how persistent agents are in their choices. It is clear that the probability of an agent keeping the same debt status is above the average rate in the total population and this happens for all categories, therefore choices tend to be persistent. In particular, debtors of Social Debt, Banks or of Bank plus Retail Store are more than twice as likely to keep their choices relative to the average probability in the population. Also, it is striking that debtors of Retail or Bank plus Retail have a probability of only 20% and 16% respectively of moving into a state of No Debt. Therefore these debtors are systematically in need of debt, whether with the same lender or a different one. This confirms the previous results that debtors of Bank plus Retail appear to be agents with higher needs for liquidity relative to other households.

4 The sorting of income risk across different types of loans

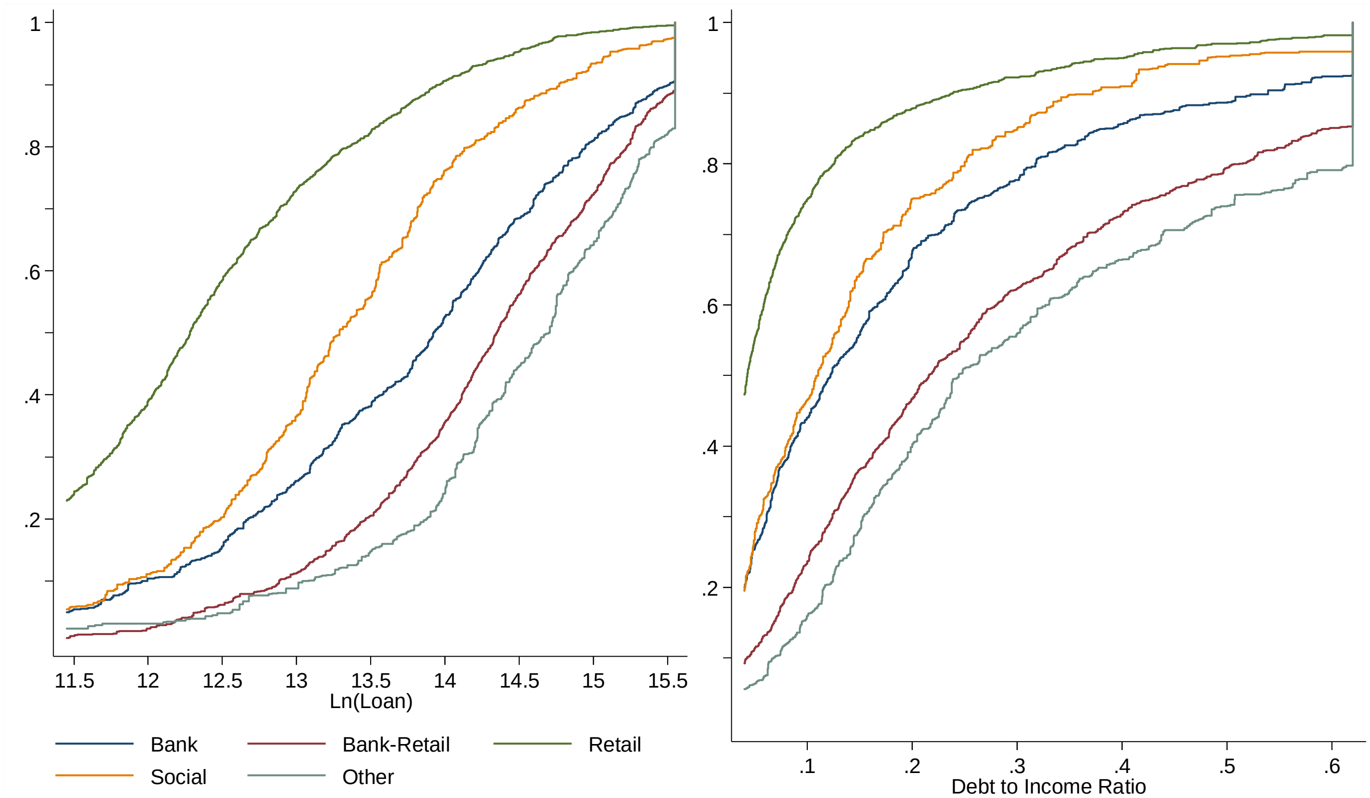
The EFH survey collects detailed information on the income, education, age and other characteristics of each household member, but it has limited data on some aspects, such as their income volatility or stability of employment. For this reason I estimate the income and employment risks of the EFH workers based on the mean statistics for workers with the same characteristics in another dataset.

Based on the quarterly Chilean Employment Survey, ENE, which covers 35,000 households, Madeira (2014) estimated three measures of risk in employment status for the period 1990 to 2012: the unemployment rate ($u_{k,t} = \Pr(U_{k,t} = 1 \mid t, x_k)$), the separation rate ($\lambda_{k,t}^{EU} = \Pr(U_{k,t+1} = 1 \mid t, U_{k,t} = 0, x_k)$) defined as the probability of being unemployed given that one was employed in the previous quarter, and the job finding rate ($\lambda_{k,t}^{UE} = \Pr(U_{k,t+1} = 0 \mid t, U_{k,t} = 1, x_k)$) defined as the probability of being employed after being unemployed in the previous quarter. The vector x_k is composed of 540 mutually exclusive groups, given by $x_k = \{$Santiago Metropolitan city or Outside, Industrial Activity (primary, secondary, terciary sectors), Gender, Age (3 brackets, ≤ 35, $35-54$, ≥ 55), Education (less than secondary schooling, secondary or technical education, college), and Household Income quintile$\}$. Madeira (2014) also computed these groups' labor income volatility even if no job is lost, $\sigma_{\zeta,t}(x_k) = \sqrt{E\left[(Y_{k,t} - E[Y_{k,t} \mid Y_{k,t-1}, x_k])^2 \mid t, U_{k,t} = U_{k,t-1}, Y_{k,t}, x_k\right]}$, and the income loss caused by going into unemployment, $R_{k,t}(x_k) = \dfrac{E\left[Y_{k,t} \mid t, U_{k,t} = 1, x_k\right]}{E\left[Y_{k,t} \mid t, U_{k,t} = 0, x_k\right]}$.

Using these labor risk measures I calculate the expected income $\bar{P}_{i,t}$ of each EFH household i as the sum of their non-labor income, a_i, and its expected labor income, $P_{i,t}$: $\bar{P}_{i,t} = a_i + P_{i,t}$, where $P_{i,t} = \sum_k P_{k,t}$ is the sum of expected labor income of each household member k. $P_{k,t} = W_{k,t}(1 - u_{k,t}) + W_{k,t}R_{k,t}(u_{k,t})$ is each worker k's average labor income during the employed ($W_{k,t} = Y_{k,t}R_{k,t}^{-U_{k,t}}$) and unemployed states. The employment risk of each household is then given by a weighted average of the rates of each member using their labor income relative to the total household labor income: $\bar{u}_{i,t} = \sum_k \frac{P_{k,t}}{P_{i,t}} u_{k,t}$, $\bar{\lambda}_{i,t}^{UE} = \sum_k \frac{P_{k,t}}{P_{i,t}} \lambda_{k,t}^{UE}$ and $\bar{\lambda}_{i,t}^{EU} = \sum_k \frac{P_{k,t}}{P_{i,t}} \lambda_{k,t}^{EU}$. Similarly, the household's weighted labor income volatility (even if no job is lost) and the replacement ratio during unemployment are given by $\bar{\sigma}_{i,t} = \sum_k \frac{P_{k,t}}{P_{i,t}} \sigma_{\zeta,t}(x_k)$ and $\bar{R}_{i,t} = \sum_k \frac{P_{k,t}}{P_{i,t}} RR_{k,t}(x_k)$.

Figure 2 shows the cumulative distribution function of the loan amounts (in logarithm) and the consumer debt to annual income ratio ($\frac{L_{i,t}}{12 \times \bar{P}_{it}}$, where $\bar{P}_{it}$ is the expected monthly income) in the pooled EFH survey (2007-11). Retail only debtors are the ones with the highest probability

Figure 2: The Cdf of the loan amounts chosen by debtors of different loan types



of having low loan amounts, since their cdf is stochastically dominated by either Social and Bank debtors. Bank plus Retail debtors and Other debtors have the greatest probability of having high loan amounts (or the lowest probability of having low loan amounts). One question is if the difference in loan amounts is entirely explained by income, since higher income households may also pay larger loans. The answer is given by the empirical cdf of the consumer debt to annual income ratio. In Figure 2 it is shown that clearly Retail only debtors have lower debt to income values in relation to both Social and Bank debtors. Also, Bank plus Retail and Other Debts users have the highest debt to income ratios. Therefore the differences in the sorting of loan amounts across lender types remains even if we take into account household income.

Table 6 reports the mean values of the household' measures for the unemployment rate ($\bar{u}_{i,t}$), the separation rate ($\bar{\lambda}_{i,t}^{EU}$) and the job finding rate ($\bar{\lambda}_{i,t}^{UE}$) across different loan choices. The groups with No Consumer Debt or only Bank loans are the ones with the lowest unemployment and separation rates. Households with Other Debts are the ones with the highest average unemployment rates,

Table 6: Mean values of labor market risk and household earnings across debtor types (EFH)

Debtor Type	$\bar{u}_{i,t}$	$\bar{\lambda}^{EU}_{i,t}$	$\bar{\lambda}^{UE}_{i,t}$	$\ln(\bar{P}_{i,t})$	$\bar{\sigma}_{i,t}$	$\bar{R}_{i,t}$
Bank	4.8%	2.0%	33.8%	13.56	18.4%	25.8%
Bank+Retail	5.3%	2.3%	35.4%	13.46	18.2%	25.5%
Retail Store	5.5%	2.6%	36.6%	13.01	16.5%	23.5%
Social Debt	5.0%	2.0%	30.9%	13.14	17.4%	22.7%
Other Debts	6.1%	2.2%	34.2%	13.47	20.7%	26.1%
No Consumer Debt	4.2%	1.9%	30.6%	13.13	16.2%	23.0%
No Access to Debt	5.4%	2.2%	31.0%	12.77	17.6%	21.3%

perhaps because of their younger age. The mean job finding rate is between 31% to 37% for all groups. Table 6 also reports the means values for the log household expected income ($\ln(\bar{P}_{i,t})$), the labor income volatility ($\bar{\sigma}_{i,t}$) and its replacement ratio of income during unemployment ($\bar{R}_{i,t}$). Bank only customers are the group of highest income, while those with Retail Store loans or with No Access to Debt have the lowest mean income. Unemployment represents a strong income reduction for Chilean households, since the mean values of $\bar{R}_{i,t}$ imply that agents only keep 21% to 26% of their working income during an unemployment spell. The households with No Consumer Debt appear to be the group least susceptible to shocks, since they are the group with lowest unemployment rate, lowest separation rate and lowest labor income volatility. The permanent income theory of consumption predicts that agents should use debt to smooth temporary income shocks (see Chatterjee et al., 2007, or Dynan and Kohn, 2007), therefore it makes sense that households with the lowest income risk also have the lowest demand for consumer loans.

While Table 6 reports the mean values of households' income, employment risks and income volatility, it is also useful to analyze how heterogeneous households are and how each group deviates from the mean. Figure 3 shows the cdf of the households' expected income ($\ln(\bar{P}_{i,t})$), unemployment rate ($\bar{u}_{i,t}$), labor income volatility ($\bar{\sigma}_{i,t}$, which can also be denoted as the standard deviation of wage shocks) and replacement ratio of income during unemployment ($\bar{R}_{i,t}$) for debtors and non-debtors. For simplicity, I use only 4 groups in the graphical comparison instead of the 7 groups used in Table 6 and the previous tables. Basically, I classify households in the same two options for non-debtors as before (No Consumer Debt, No Access to Debt), but use only two classifications for the groups of debtors: i) users of Retail Store loans only, which represent 29% of the household population

(Table 2) and are the largest group with consumer debt; and, ii) users of Bank, Social Debt and Other Debts, which represent 30.9% of the Chilean population (this figure is obtained by summing the distinct categories of this group in Table 2). Another simplification concerns the problem that often households have a lot of heterogeneity at the extreme margins, but one is mostly concerned with the heterogeneity that affects most of the population and not its extreme points (which could eventually be outliers due to measurement error). Therefore to make the graphs easier to read the cdfs are plotted only in the range of 20% to 90% probability.

Figure 3 shows that in terms of income there is a clear stochastic dominance among the different groups, with households with No Access to Debt having lower income than those with Retail loans and those with Retail loans having lower income than both the households with No Consumer Debt and the households with Bank, Social and Other Debts. Also, it is clear that households with No Consumer Debt have the lowest unemployment rates, which is another confirmation that a partial motivation for consumer loans is to smooth temporary income shocks. Households with Bank, Social and Other Debts also have lower unemployment rates relative to those with Retail loans only or No Access to Debt. Labor income volatility ($\bar{\sigma}_{i,t}$) is highest for the households with Bank, Social and Other Debts, which may imply that consumer debt is used for smoothing income shocks in this group. The replacement ratio of income during unemployment is the lowest for those with No Access to Debt, followed by the users of Retail loans only and those with No Consumer Debt. Users of Banks, Social and Other Debts have the highest replacement ratios during unemployment, therefore this is the group that suffers the lowest loss of income from job loss.

Figure 3 shows that income and labor experiences have a lot of heterogeneity in the population. Unemployment rates can range from as low as 2% to as high as 8%. Labor income volatility has a range between 11% to 27%, while replacement ratios can vary between 18% and 34%.

Besides analyzing unemployment rates, it is also appropriate to look at the employment separation ($\bar{\lambda}^{EU}_{i,t}$) and job finding ($\bar{\lambda}^{UE}_{i,t}$) rates. The reason is because unemployment rates has a different interpretation if it is driven by high separation rates (lots of workers losing their jobs) or by low job finding rates (which implies that unemployed workers have difficulties finding jobs and therefore unemployment spells last a long time). Both of these employment transition rates play a role in explaining labor market shocks in the United States (Shimer, 2012) and in Chile (Madeira, 2014).

Figure 4 shows the cdf of the separation ($\bar{\lambda}^{EU}_{i,t}$) and job finding ($\bar{\lambda}^{UE}_{i,t}$) rates for Chilean households.

Figure 3: The Cdf of labor market characteristics of debtors versus non-debtors

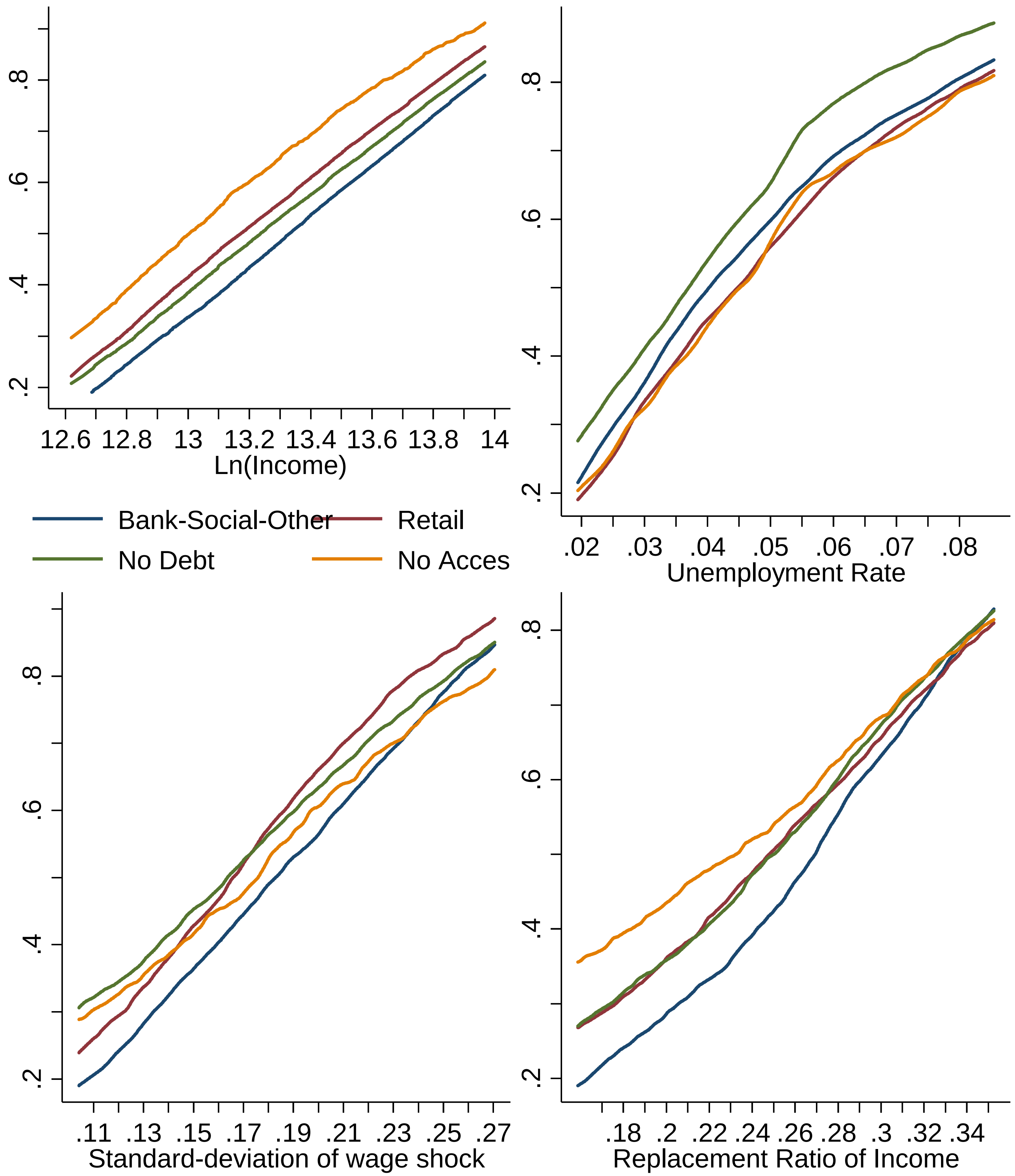
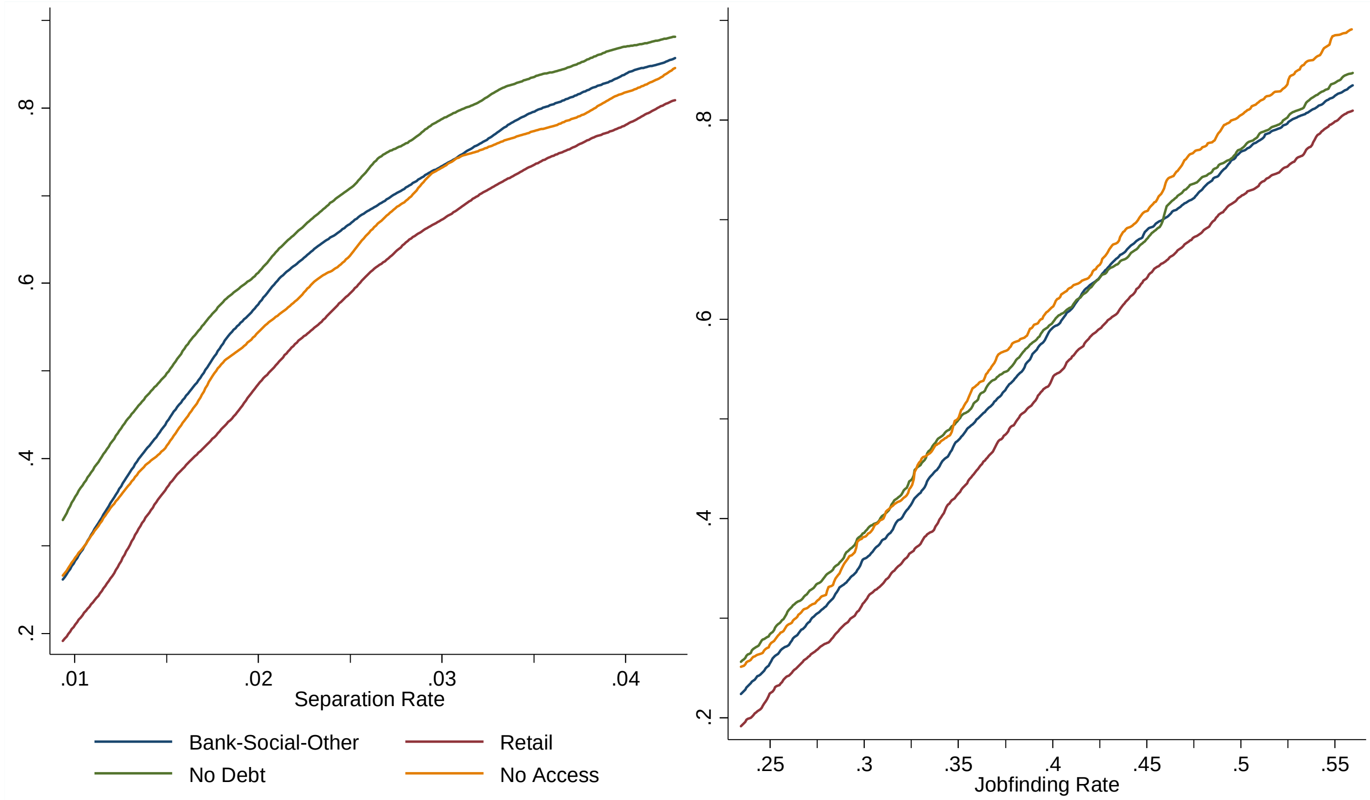


Figure 4: The Cdf of employment transition probabilities for different loan types



There is a lot of heterogeneity in these variables, with the separation rate ranging from as low as 1% to as high as 4% and the job-finding rate varying between 25% and 55%. The separation rate has the most clear differences between debtor and non-debtor groups. Households with No Consumer Debt have lower separation rates than users of Bank, Social and Other Debts, and these last ones have lower separation rates than those with No Access and the users of Retail loans only. The differences in job-finding rates are less clear. Users of Retail loans only have both the highest separation rates and the highest job finding rates, which implies that employment mobility is high in this group. However, the groups with No Consumer Debt, No Access and users of Bank, Social and Other Debts have a similar distribution for the job-finding rate.

Figure 5 shows the differences in income and labor market characteristics of different debtor groups. Users of Retail and Social Debt are the ones with the lowest income, while the users of Bank and Other Debts have the highest income. Also, Bank users have a lower unemployment rate than all the other debts, with Social Debt users being the second group with the lowest unemployment rates and users of Other Debts having the highest unemployment. Retail and Social Debt users, however, have the lowest labor income volatility (or standard-deviation of wage shocks), while users of Bank and Other Debts have the highest wage risk. Users of Retail and Social Debt are the ones with the lowest replacement ratios and therefore suffer the most during a jobless spell.

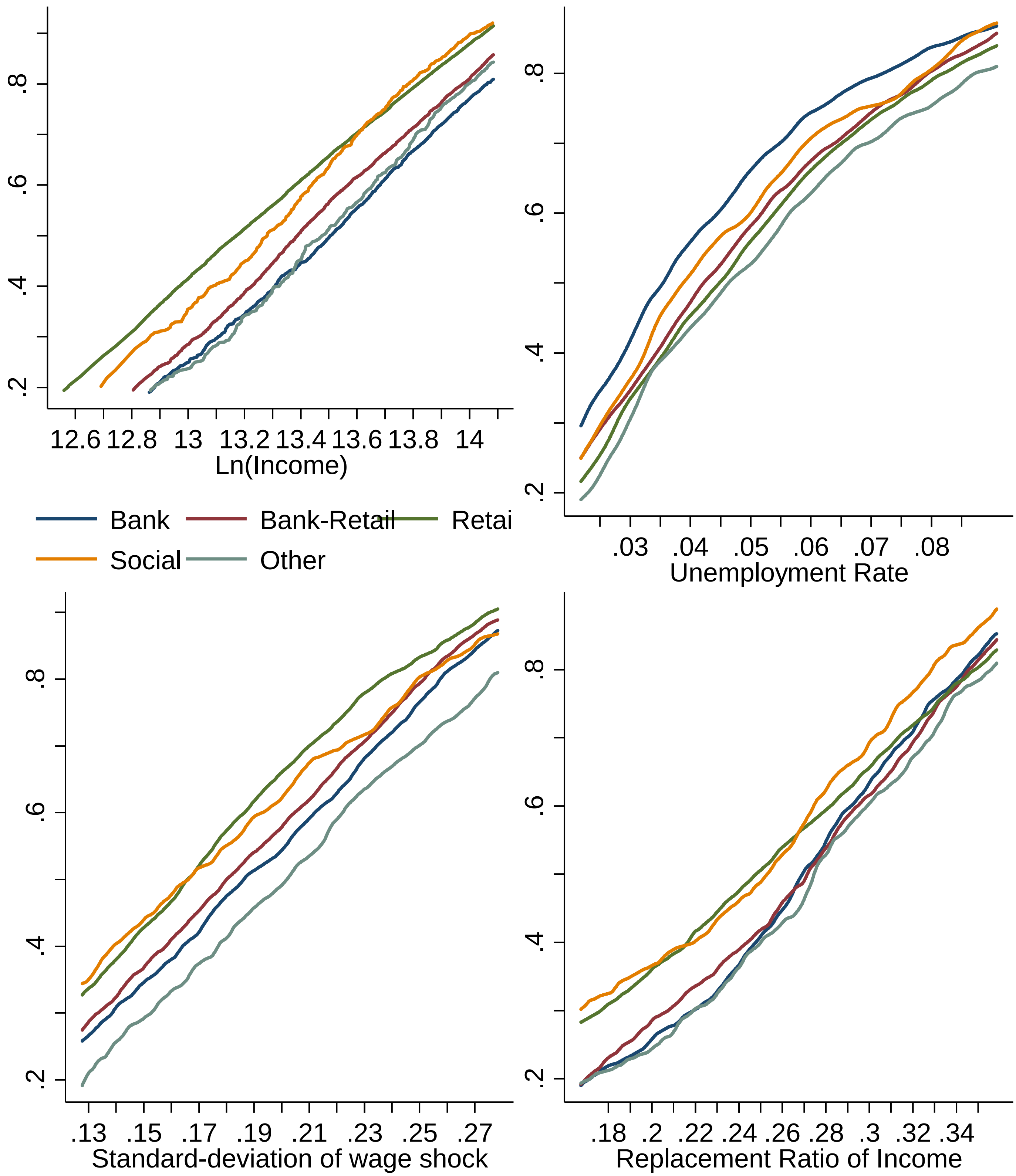
Overall, Figures 2, 3 and 4 portray a clear picture of different income and labor market characteristics across non-debtors and different groups of debtors. Households with No Access to Debt have the lowest income, highest unemployment rates and lowest replacement ratios of income, therefore it is the group most subject to low income and income fluctuations. Households with No Consumer Debt (because of a lack of demand for such loans) have the lowest unemployment rates, separation rates and labor income volatility, therefore it is the group least subject to income shocks. The users of just Bank loans are the ones with the highest income, highest replacement ratio and lowest unemployment rate, but they suffer from substantial wage volatility which may create a demand for smoothing consumption. Users of Other Debts have high income and high replacement ratios in the same way as Bank users, but they are the debtor group most subject to both high unemployment rates and high labor income volatility, therefore it could be seen as a riskier segment relative to Bank users. Finally, users of Retail loans are the ones with the lowest income among debtors (although they have higher income than the group with No Access to Debt), and also have a high unemployment rate and low replacement ratio, which could make them a riskier debt segment. However, Retail users have a low standard-deviation of wage shocks, therefore their income is relatively stable during their employment experience. Users of both Bank plus Retail loans are a segment somewhat in between the exclusive users of either Banks or Retail loans.

5 Results

5.1 The role of demographics, income profile and unobserved preferences

Now I discuss the results from the consumer loan choice and default model exposed in section 2. As explained before, the model requires some variables that affect loan choice, but not default. Since unemployment risk, labor income volatility are measured for several time periods (all the quarters from 1990 to 2012) for each type of worker, then it is possible to create these variables for each

Figure 5: The Cdf of labor market characteristics of debtors of different loan types



EFH household at previous periods than the survey date. It is natural to assume that households were driven by labor market effects that happened at the time of the loan contract, which was a substantial time before the current period t. Consumer loans typically have a maturity of 12 to 24 months, therefore it is reasonable to assume that the labor market conditions that influenced loan choice happened 4 quarters or more before the current period. For this reason the vector affecting loan choice includes expected income ($\ln(\bar{P}_{i,t-4})$,), unemployment risk ($\bar{u}_{i,t-4}$) and labor income volatility ($\bar{\sigma}_{i,t-4}$) with a lag of 4 quarters (although a shorter or longer lag could be used). Note that expected income is a weighted sum of each household member's labor income, its unemployment probability and its replacement ratio, therefore it can be estimated for previous time periods. In addition the vector that affects loan choice includes the education, age and structure of the household (whether it is a couple or a family with many members), and the motivation of the loans (which share of the debt was motivated by needs of "Health" or to "Pay Previous debts"):

$$x_{i,t} = \left\{ \begin{array}{c} \ln(\bar{P}_{i,t-4}), \text{ unemployment risk } \bar{u}_{i,t-4}, \text{ labor income volatility } \bar{\sigma}_{i,t-4}, \\ \text{years of education of household head, age of the household head, age squared,} \\ \text{dummies for each year, dummy for residence being out of the Santiago capital,} \\ \text{dummy for 2 members and dummy for 3 or more members in household,} \\ \text{Share of debt justified by "Pay Previous Debts", Share of debt justified by "Health"} \end{array} \right\}.$$

In a similar way I assume that the vector $z_{i,t}$ that explains loan morosity or default at time t includes some variables that do not necessarily affect loan choice. One variable is the ratio of consumer debt to the annual income ($RDI_{i,t} = \frac{L_{i,t}}{12 \times \bar{P}_{i,t}}$), which can be seen as a measure of long-term solvency of the household. Households with larger loans may feel more stressed about their long term commitments and choose to default on their loans. In the same way, some households may be more worried about this month's specific commitments instead of their long-term expenses. For those households, the current monthly debt service (the debt service of a loan, $DS_{i,t}$, includes both the amortization and the interest payments) relative to this month's income ($Y_{i,t}$) may provide a liquidity motive for defaulting or simply paying a loan with some delay. For this reason I also include the ratio of monthly Debt Service to Income ($RDSI_{i,t}$) as a possible factor affecting household default. The overall vector of observables that explain default includes the financial ratios $RDI_{i,t}$, $RDSI_{i,t}$, plus the current expected income ($\ln(\bar{P}_{i,t})$), unemployment risk ($\bar{u}_{i,t}$) and labor income ($\bar{\sigma}_{i,t}$), and the same demographic variables that affect loan choice:

Table 7.1: Coefficients for the mean value of the Utility of each type of Loan

	1=Bank	**2=Bank+Retail**	**3=Retail Store**	**4=Social Debt**	**5=Other Debts**	**6=No-Access**
2007	-9.706 (0.161)***	-6.357 (0.188)***	4.783 (0.687)***	3.499 (0.959)***	-5.347 (3.809)	9.737 (0.906)***
2008 / 09	-9.912 (0.157)***	-6.924 (0.183)***	4.623 (0.759)***	3.126 (1.003)***	-6.055 (4.068)	9.928 (0.915)***
2010	-9.904 (0.163)***	-7.204 (0.191)***	4.855 (0.676)***	3.904 (0.99)***	-5.225 (3.783)	10.228 (0.924)***
2011	-9.524 (0.165)***	-6.569 (0.19)***	4.868 (0.65)***	4.462 (0.989)***	-4.819 (3.809)	10.251 (0.933)***
Income: $\ln(\bar{P}_{i,t-4})$	0.346 (0.019)***	0.095 (0.022)***	-0.5 (0.055)***	-0.449 (0.056)***	0.086 (0.719)	-0.831 (0.053)***
Education	0.073 (0.01)***	0.046 (0.01)***	-0.109 (0.013)***	-0.077 (0.007)***	0.047 (1.223)	-0.068 (0.009)***
Unemployment $\bar{u}_{i,t-4}$	2.668 (1.052)**	3.139 (0.513)***	4.618 (0.415)***	4.108 (0.292)***	4.344 (1.845)**	2.402 (0.754)***
Wage volatility $\bar{\sigma}_{i,t-4}$	0.152 (0.257)	0.386 (0.232)*	-0.372 (0.38)	0.938 (0.376)**	3.115 (0.599)***	1.502 (0.527)***
Out of Santiago	-0.143 (0.007)***	-0.232 (0.011)***	-0.266 (0.03)***	-0.475 (0.034)***	-0.238 (1.486)	-0.261 (0.014)***
Age of home head	0.115 (0.003)***	0.148 (0.003)***	0.108 (0.006)***	0.001 (0.022)	0.025 (0.424)	0.026 (0.013)**
Age squared	-0.001 (0)***	-0.002 (0)***	-0.001 (0)***	0 (0)	-0.001 (0.002)	0 (0)
2 members in home	0.257 (0.029)***	0.542 (0.02)***	0.74 (0.029)***	-0.01 (0.052)	0.723 (0.498)	0.005 (0.009)
3 or more members	0.567 (0.019)***	1.069 (0.014)***	1.337 (0.071)***	0.29 (0.071)***	1.014 (0.31)***	0.348 (0.018)***
Pay previous debts	5.034 (3.636)	5.286 (3.632)	3.126 (3.622)	4.892 (3.786)	0	
Health needs	5.093 (2.913)*	5.544 (2.917)*	4.398 (2.922)	6.119 (2.838)**	0	

$$z_{i,t} = \left\{ \begin{array}{c} \frac{L_{i,t}}{12 \times \bar{P}_{i,t}},\ \frac{DS_{i,t}}{Y_{i,t}},\ \ln(\bar{P}_{i,t}),\ \text{unemployment risk } \bar{u}_{i,t},\ \text{labor income volatility } \bar{\sigma}_{i,t}, \\ \text{years of education of household head, age of the household head, ...} \end{array} \right\}.$$

Finally, I need to specify the degree of heteroscedasticity in the unobserved tastes for loan choice, loan amount and default, that is the standard-deviation of each element of the normally distributed vector $\left\{\tilde{\zeta}_{i,t}, \varepsilon_i\right\}$, where the vector of random-effects of tastes (i.e., tastes that are constant over time for each agent) is given by $\varepsilon_i \equiv \left\{\eta_{i,1}, \eta_{i,2}, \eta_{i,3}, \omega_{i,1}, ..., \omega_{i,B}, \bar{\zeta}_i, \bar{\nu}_i\right\}$. In this case I assume that all the standard-deviations are exponential functions of a linear-index, $\sigma = \exp(\beta x_i)$, which guarantees that all standard-deviations are positive. The vector x_i that models heteroscedasticity includes a constant, a dummy for the 2008/09 panel, the years of education of the household head, plus the average labor market characteristics of the household (since income and labor market risk is time-varying I apply the mean values over the period 2007 to 2011 for simplicity):

$$x_i = \left\{\text{constant, dummy for 2008/09, years of education of head, } \tfrac{1}{T}\textstyle\sum_{t=2007}^{2011} \left\{\ln(\bar{P}_{i,t}), \bar{u}_{i,t}, \bar{\sigma}_{i,t}\right\}\right\}.$$

Table 7.1 shows the estimates for the coefficients of loan choice, β_b. The coefficients of a multivariate logit model sometimes have a difficult interpretation (Train, 2009), because the agents' choice is made over a multivariate set with $B+1$ choices, $\max(U_{i,0,t}, U_{i,1,t}, .., U_{i,B,t})$, with the first choice being standardized to have value zero. Let us think of a generic variable x and its coefficient on choice b, β_b, which is assumed to be positive. Then $\beta_b > 0$ implies the odds ratio of the probability of option b relative to option 0 is increasing in x, meaning that larger x makes b more likely to be chosen relative to option 0. However, at the same time there could be another option c which has a larger coefficient than b, implying x decreases the chance of b being chosen relative to option c. Therefore in the multivariate case $\beta_b > 0$ does not always increase the probability of b being chosen with larger x. Such is the case only if $\beta_b \geq \max(\beta_1, .., \beta_B)$. This interpretation of the multivariate logit coefficients must be kept in mind while reading Table 7.1.

The coefficient for the lagged household expected income ($\ln(\bar{P}_{i,t-4})$) is the largest (i.e., the most positive) for the option of Bank loans, while it is the lowest (i.e., the most negative) for the No-Access option (Table 7.1). This implies that larger income unambiguously increases the option of a Bank loan and decreases the option of No-Access. The impact of income on the choice of Retail Store loans and Social Debts is negative, therefore larger income increases the likelihood of No Debt in relation to these options. The coefficient of education is largest for the Bank option and lowest for the Retail Store option, which implies that education increases the probability of a Bank loan and decreases the option of Retail Store loans. The coefficient of lagged unemployment ($\bar{u}_{i,t-4}$) increases the probability of all the loan options and the No-Access option in relation to No-Debt. However, unemployment has the effect of increasing more the probability of specific loans, such as the Retail Store, Social Debt and Other Debts options. Being outside of the Santiago capital city has the effect of lowering the probability of all loan options, with its strongest effect on Social Debt. Higher age is positively related to choosing the Bank, Bank plus Retail, and Retail Store options. Households with more members are more likely to choose the option of Retail Store loans. Finally, the motivations for undertaking a loan (Pay previous debts or Health needs) have a special standardization, because families who report positive values for the loan motivation must have chosen one of the loan options 1 to 5 and therefore I standardize the loan motivation coefficients for the last option Other Debts as being 0. Pay Previous Debts does not have an effect on a specific loan type, but Health needs is associated to the choice of Bank, Bank plus Retail and Social Debt.

Table 7.2 shows the heteroscedasticity for the random effects that denote the unobserved tastes for each loan type. The strongest conclusions is that the heterogeneity for the tastes of Bank, Bank plus Retail, and Social Debt, decreases with income and years of education. Similarly, Table 7.3 shows the heteroscedasticity of the random-effects that affect several loan options. The main conclusion is that the heterogeneity for the unobserved taste for all loans ($\eta_{i,1}$) and of the taste for Bank plus Retail and Retail only ($\eta_{i,3}$) is also decreasing in income and education. Therefore

Table 7.2: Coefficients for the standard-deviation (in log) of the random-effect of each type $\omega_{i,b}$

	1=Bank	**2=Bank+Retail**	**3=Retail Store**	**4=Social Debt**	**5=Other Debts**	**6=No-Access**
constant	-0.327 (0.962)	-0.296 (0.092)***	0.519 (0.465)	1.001 (0.394)**	-0.126 (1.797)	0.12 (1.18)
2008/09	0.146 (0.682)	0.278 (0.157)*	0.388 (2.875)	-2.715 (1.683)	-0.213 (2.108)	-0.001 (0.788)
$\frac{1}{T}\sum_{t=2007}^{2011}\ln(\bar{P}_{i,t})$	-2.616 (1.342)*	-2.127 (1.126)*	-0.024 (0.34)	-4.691 (1.604)***	-1.008 (2.387)	-0.625 (1.534)
Education	-2.307 (0.886)***	-1.566 (0.946)*	-0.003 (2.308)	-9.046 (3.245)***	-0.693 (1.595)	-1.246 (1.28)
$\frac{1}{T}\sum_{t=2007}^{2011}\bar{u}_{i,t}$	-0.05 (1.206)	-0.032 (1.002)	-0.658 (0.6)	0.765 (1.179)	-0.009 (1.132)	0.05 (1.127)
$\frac{1}{T}\sum_{t=2007}^{2011}\bar{\sigma}_{i,t}$	-0.082 (0.736)	-0.047 (0.769)	-1.046 (0.706)	-0.007 (1.271)	0.007 (0.422)	0.097 (2.303)

Table 7.3: Coefficients for the standard-deviation (in log) of the factors affecting several choices

	factor 1, $\eta_{i,1}$ **(choices 1 to 5)**	**factor 2,** $\eta_{i,2}$ **(choices 1 to 2)**	**factor 3,** $\eta_{i,3}$ **(choices 2 to 3)**
constant	-0.133 (7.117)	-5.559 (1.522)***	-0.516 (2.898)
2008/09	0.566 (5.986)	0.9 (4.178)	-0.008 (3.415)
$\frac{1}{T}\sum_{t=2007}^{2011}\ln(\bar{P}_{i,t})$	-3.691 (1.921)*	0.481 (1.45)	-6.718 (3.387)**
Education	-1.3 (0.706)*	-0.082 (3.156)	-6.14 (3.041)**
$\frac{1}{T}\sum_{t=2007}^{2011}\bar{u}_{i,t}$	-0.003 (1.942)	3.034 (4.014)	-0.02 (2.27)
$\frac{1}{T}\sum_{t=2007}^{2011}\bar{\sigma}_{i,t}$	0.093 (1.505)	-2.87 (2.58)	-0.088 (0.727)

higher income and more highly educated households have lower heterogeneity of unobserved tastes.

Table 8.1 shows the results for the choice of loan amount ($\ln(L_{i,t})$) and default ($D_{i,t}$). It is worth noting that the expected income ($\ln(\bar{P}_{i,t-4})$,), unemployment risk ($\bar{u}_{i,t-4}$) and wage volatility ($\bar{\sigma}_{i,t-4}$) that affect the loan amount decision have a lag of 4 quarters, while the variables affecting default correspond to the current period t. Basically, income, unemployment rates, wage volatility, households with more members, and debt motivations (especially, the motive of Pay previous debts) are positively related to loan amounts. Education is negatively related to loan amount, which may denote a tendency of highly educated households to better manage their finances over time and resort less to expensive consumer debt (Table 2 showed that interest rates for consumer loans are high in Chile). Also, the estimates show that households with strong unobservable tastes for either Bank plus Retail and Other Debts are more likely to also have a taste for higher loan amounts.

The propensity to default is positively related to high levels of consumer debt relative to annual income ($RDI_{i,t}$), higher debt service ($RDSI_{i,t}$), unemployment risk, wage volatility, households with more members, and to loan motivations (especially, the Health needs). In the year 2010 there was a substantially lower rate of default, even after accounting for the other factors in the model, which could have been due to more selective credit supply policies in the early aftermath of the financial crisis. The choice of loan amount and default behavior are both quadratic in terms of age, first increasing with age and then falling. Default is negatively related to income, but it is not significantly affected by education, which seems to coincide with recent studies for the USA, which show that education, apart from math skills, has no significant impact on debt repayment behavior (Brown, van der Klaauw, Wen and Zafar, 2013). Default is also negatively related to households with a higher taste for Bank loans. This results justifies the behavior of Chilean banks in terms of giving preference to customers with a longer and more exclusive credit history in the banking system, since those are the households of lower risk. Credit history is also related to default in other countries as well (Gross and Souleles, 2002, Roszbach, 2004, Edelberg, 2006).

Both the motives "Pay previous debts" and Health needs have a positive effect on loan amount and the propensity to default. It is also interesting that Health needs has a larger impact on default than the motive "Pay previous debts", since it confirms the predictions of economic models. Economic models of default decision assume that Health expenses are less predictable than other expenses, since tastes for consumption and past loan commitments are already known to the household. Therefore health expenses are an unpredictable shock for households and one that often leads to default even for low amounts of debt (Chatterjee et al., 2007).

Table 8.2 summarizes the estimated heteroscedasticity of the unobserved tastes for loan amount and default behavior, plus a contemporary loan amount shock which is independent over time. Again, we can conclude that households of higher income and education are less heterogeneous in their tastes for loan amount and default behavior. However, the heteroscedasticity of the contemporary unobserved shock for loan amount is increasing with income and education. This shows that higher income and highly educated households are less persistent in their indebtedness, since their loan amounts depend more on contemporary shocks than constant tastes.

6 Conclusions

This paper shows how households' characteristics impact their choice of consumer loans and default behavior. Low labor market risk (as measured by unemployment risk, job separation rates and wage volatility) is correlated with a desire for not having consumer debt, while low income is the strongest

Table 8.1: Coefficients for the mean loan amount (in log) and propensity for morosity

Exogenous variables	**Log-loan amount** ($t' = t - 4$)	**Propensity to morosity** ($t' = t$)
2007	0.074 (3.7)	1.262 (0.17)***
2008 / 09	-0.048 (3.704)	1.326 (0.195)***
2010	0.579 (3.34)	0.568 (0.171)***
2011	0.089 (3.602)	1.369 (0.173)***
Ratio of Debt to Income, $RDI_{i,t}$		1.291 (0.022)***
Ratio of Debt Service to Income, $RDSI_{i,t}$		0.54 (0.052)***
Log-Income: $\ln(\bar{P}_{i,t'})$	6.992 (4.033)*	-0.385 (0.089)***
Years of education of home head	-6.598 (3.32)**	-0.062 (0.126)
Unemployment $\bar{u}_{i,t'}$	0.171 (0.103)*	2.818 (1.585)*
Wage volatility $\bar{\sigma}_{i,t'}$	0.157 (0.092)*	0.854 (0.269)***
Out of Santiago	0.067 (0.369)	-0.028 (1.405)
Age of home head	0.65 (0.309)**	0.109 (0.015)***
Age squared	-0.013 (0.006)**	-0.001 (0.005)
2 members in home	0.291 (0.155)*	0.095 (0.036)***
3 or more members	0.317 (0.154)**	0.407 (0.236)*
Share of loan for "Pay previous debts"	0.588 (0.318)*	0.376 (0.213)*
Share of loan for Health needs	0.184 (0.108)*	0.778 (0.249)***
RE of loan type: $\bar{\varepsilon}_{i,b}$	θ	ρ
1 = Bank	1.245 (0.584)**	-1.674 (0.723)**
2 = Bank+Retail	2.012 (0.78)***	0.626 (0.726)
3 = Retail	1.013 (0.778)	-0.179 (0.728)
4 = Social	1.634 (0.727)**	-0.162 (0.49)
5 = Other	1.808 (0.794)**	-0.342 (1.021)
6 = No-Access	0.002 (0.765)	0.261 (0.66)
RE of log-amount: $\bar{\zeta}_i$		0.341 (1.468)

Table 8.2: Coefficients of the standard-deviation (in log) of the random-effects of loan amount and morosity

Exogenous variables	**Log-loan amount:**	**Log-loan amount:**	**Propensity to morosity:**
	Random Effect $\bar{\zeta}_i$	**Contemporary shock** $\tilde{\zeta}_{i,t}$	**Random Effect** $\bar{\nu}_i$
constant	-0.694 (0.489)	0.137 (1.15)	-0.125 (0.276)
2008/09	-1.026 (1.063)	0.331 (1.021)	0.045 (0.375)
$\frac{1}{T}\sum_{t=2007}^{2011}\ln(\bar{P}_{i,t})$	-10.702 (3.656)***	4.16 (1.251)***	-1.838 (1.065)*
Education	-7.032 (3.551)**	2.552 (0.487)***	-0.885 (0.527)
$\frac{1}{T}\sum_{t=2007}^{2011}\bar{u}_{i,t}$	0.29 (1.479)	0.185 (0.424)	0.015 (0.934)
$\frac{1}{T}\sum_{t=2007}^{2011}\bar{\sigma}_{i,t}$	0.124 (0.054)**	0.235 (0.663)	-0.013 (0.845)

cause of a lack of access to debt and credit constraints. Unemployment rates increase the probability of households opting for all types of consumer loans, but it has a greater impact on lenders who do not apply credit scoring such as Retail Stores and institutions providing Social Credit.

Loan amounts increase with income, unemployment risk and wage volatility, therefore consumer loans may help smooth income shocks. The default probability decreases with income and increases with high levels of indebtedness relative to income, unemployment risk and wage volatility, confirming the existence of adverse selection among Chilean debtors. Bank debtors have the lowest risk levels, which is expected from a lender that applies credit scoring extensively (Edelberg, 2006). Health needs are positively associated with default behavior, which could denote these expenses are difficult to predict and insure (Chatterjee et al., 2007). Finally, the probability of getting a loan and the choice of loan amount is increasing in the number of household members and quadratic in age, resembling the profile of life-cycle consumption (Attanasio and Weber, 2010).

Finally, I show that households are heterogeneous in their loan tastes. This result has broad implications for policy, since economic shocks or new regulations (or deregulation initiatives) affecting a certain lenders would have heterogeneous welfare impact across the population.

References

[1] Adams, William, Liran Einav, and Jonathan Levin (2009), "Liquidity Constraints and Imperfect Information in Subprime Lending," *American Economic Review*, 99(1), 49-84.

[2] Attanasio, Orazio, Pinelopi Koujianou Goldberg, and Ekaterini Kyriazidou (2008), "Credit Constraints In The Market For Consumer Durables: Evidence From Micro Data On Car Loans," *International Economic Review*, 49 (2), 401-436.

[3] Attanasio, Orazio and Guglielmo Weber (2010), "Consumption and Saving: Models of Intertemporal Allocation and Their Implications for Public Policy," *Journal of Economic Literature*, 48(3), 693-751.

[4] Banco Central de Chile (2013), "Financial Stability Report," 2nd Semester.

[5] Bertrand, Marianne and Adair Morse (2011), "Information Disclosure, Cognitive Biases, and Payday Borrowing," *Journal of Finance*, 66, 1865-1893.

[6] Brown, Meta, Wilbert van der Klaauw, Jaya Wen, and Basit Zafar (2013), “Financial Education and the Debt Behavior of the Young”, Federal Reserve Bank of New York Staff Reports 634.

[7] Chatterjee, Satyajit, Dean Corbae, Makoto Nakajima and José-Víctor Ríos-Rull (2007), "A Quantitative Theory of Unsecured Consumer Credit with Risk of Default," *Econometrica*, 75(6), 1525-1589.

[8] Dynan, Karen and Donald Kohn (2007), “The Rise in U.S. Household Indebtedness: Causes and Consequences,” Finance and Economics Discussion Paper 2007-37, Federal Reserve Board.

[9] Edelberg, Wendy (2004), “Testing for Adverse Selection and Moral Hazard in Consumer Loan Markets,” Finance and Economics Discussion Paper 2004-9, Federal Reserve Board.

[10] Edelberg, Wendy (2006), “Risk-based Pricing of Interest Rates for Consumer Loans,” *Journal of Monetary Economics*, 53, 2283-2296.

[11] Einav, Liran, Mark Jenkins, and Jonathan Levin (2012), "Contract Pricing in Consumer Credit Markets," *Econometrica*, forthcoming.

[12] Eurosystem Household Finance and Consumption Network (2009), "Survey Data on Household Finance and Consumption: Research Summary and Policy Use", ECB Occasional Paper 100.

[13] Girouard, N., M. Kennedy, and C. André (2007), "Has the Rise in Debt Made Households More Vulnerable?", OECD Economics Department Working Paper 535.

[14] Gross, David B. and Nicholas S. Souleles (2002), "An Empirical Analysis of Personal Bankruptcy and Delinquency," *Review of Financial Studies*, 15 (1), 319-347.

[15] Hess, Stephane, Kenneth Train and John Polak (2006), "On the use of a Modified Latin Hypercube Sampling (MLHS) method in the estimation of a Mixed Logit Model for vehicle choice," *Transportation Research Part B: Methodological*, 40(2), 147-163.

[16] Horowitz, Joel L. (2001), "The Bootstrap," in Handbook of Econometrics, edition 1, vol. 5, chapter 52, 3159-3228, Elsevier.

[17] IMF (2006), "Household credit growth in emerging market countries", *Global Financial Stability Report*, 46-73.

[18] Jaffee, Dwight M. and Thomas Russell (1976), "Imperfect Information, Uncertainty, and Credit Rationing," *Quarterly Journal of Economics*, 90(4), 651–666.

[19] Jaffee, D. and J. Stiglitz (1990), "Credit rationing," in B. M. Friedman and F. H. Kahn (eds.) *Handbook of Monetary Economics*, vol. 2, Amsterdam: North Holland.

[20] Kahneman, Daniel (2011), "Thinking, Fast and Slow," Farrar, Straus and Giroux Press.

[21] Madeira, Carlos (2014), "Earnings dynamics of Chilean workers: employment transitions, replacement ratio, and income volatility," Central Bank of Chile, mimeo.

[22] Marinovic, Alejandra, José Matus, Karla Flores, Nancy Silva (2011), "Systemic risk of households in Chile", *Revista de Economía Chilena*, 14(3), 5-39.

[23] McFadden, Daniel and Kenneth Train (2000), "Mixed MNL models for discrete response," *Journal of Applied Econometrics*, 15(5), 447–470.

[24] Nevo, Aviv (2000), "A Practitioner's Guide to Estimation of Random Coefficients Logit Models of Demand," *Journal of Economics & Management Strategy*, 9(4), 513-548.

[25] Nevo, Aviv (2011), "Empirical models of consumer behavior," *Annual Review of Economics*, 3, 51-75.

[26] Roszbach, K. F. (2004), "Bank Lending Policy, Credit Scoring and the Survival of Loans," *Review of Economics and Statistics*, 86(4), 946–958.

[27] Shimer, Robert (2012), "Reassessing the Ins and Outs of Unemployment," *Review of Economic Dynamics*, 15(2), 127-148.

[28] Train, Kenneth (2009), "Discrete choice methods with simulation," Cambridge University Press.

[29] Vella, Frank (1998). "Estimating Models with Sample Selection Bias: A Survey", *The Journal of Human Resources*, 33(1), 127-169.