Emerging Markets Review 43 (2020) 100677

Contents lists available at ScienceDirect

Emerging Markets Review

journal homepage: www.elsevier.com/locate/emr

Measuring households' financial vulnerabilities from consumer debt: Evidence from Chile

Rodrigo Cifuentes[a], Paula Margaretic[b,*], Trinidad Saavedra[a]

[a] Central Bank of Chile, Agustinas 1180, Santiago, Chile
[b] University of San Andrés, Vito Dumas 284, San Fernando, Buenos Aires, Argentina

ARTICLE INFO

Keywords:
Consumer debt
Financially vulnerable households
Emerging markets
Financial stability
Cluster analysis

JEL classifications:
G83
D12
D14
G20

ABSTRACT

This paper is concerned with measuring the financial vulnerability (FV) of households arising from consumer debt. Our case of application is Chile. Our main finding is that by applying a methodology that allows for households' heterogeneities and that accounts for contextual factors (like motives for asking for debt, exposure to shocks, family structure, holdings of assets and perspectives of future income), we better quantify the risks that financially vulnerable households may entail to the financial system.

1. Introduction

This paper is concerned with measuring the financial vulnerability (FV) of households arising from consumer (unsecured) debt. We consider a household as financially vulnerable when her financial commitments are perceived as being a too heavy burden for her. In the literature, there are several approaches to identify financially vulnerable households (hereafter, FVH); in this paper, we make use of households' self-assessed debt burden.

FV depends on several factors, which are beyond the standard metrics such as the stock of debt and the amount of its service. This is because similar amounts of debt and its service in relation to income may imply various levels of vulnerability to different households, depending on other contextual factors, like motives for asking for debt, exposure to shocks, family structure, holdings of assets and perspectives of future income, among others. The contribution of this paper is that we propose a framework that incorporates these contextual factors to identify FVH and that accounts for households' heterogeneities. Thanks to the richness of our dataset, we can then examine the actual financial situation (as measured, for instance, by the prevalence of credit arrears or default, and debt service to income ratio) of the identified FVH. Our case of application is Chile, an emerging, medium-income economy.

The importance of considering contextual information in the analysis of FV arises from several elements specific to the provision of consumer debt. The first element is that in many countries, there is a wide supply of consumer credit in low amounts with low

* Corresponding author.
E-mail addresses: rcifuent@bcentral.cl (R. Cifuentes), pmargaretic@udesa.edu.ar (P. Margaretic).

https://doi.org/10.1016/j.ememar.2020.100677
Received 6 June 2019; Received in revised form 4 November 2019; Accepted 20 January 2020
Available online 23 January 2020
1566-0141/

requirements (typically, pre-approved or unconditional credit lines). The industry calls this practice solicitation. The reason for it is that, through solicitation, the industry invests in reducing the asymmetries of information and generating credit histories.[1] Financial intermediaries then use these credit histories to offer loans more closely tailored to the risk characteristics of the households (Durkin et al., 2014).

The problem with this mechanism is the potential aggregate consequences it may have: If information available to credit providers is not complete in relation to how much credit individuals are getting,[2] and if several providers are using this mechanism to identify good payers, some customers could profit from the easy access to credit from multiple sources and end up taking more debt than what lenders would have been willing to lend to them individually if applying their credit assessment models. From a systemic risk perspective, the latter is not only an important issue for financial institutions providing these credits, but also for policy makers and authorities with a mandate on financial stability.

A second element specific to the provision of consumer debt is that the industry often uses low rates, even as low as zero, for very short term credits (lower than 6 months) in order to, hopefully, form a habit in consumers. In turn, this practice induces some users to take credit even if they do not really need it: Because they value the liquidity they obtain when they can access free credit, these users keep the funds they would have otherwise spent as precautionary liquidity; later on, they use the liquidity to pay the credit.[3] We call this motive "precautionary" or "transactional". The implicit risk of it is low, since debtors have the resources to cover it. However, careless use of these offers may lead to the accumulation of large payments later on, which could potentially become a burden for them.

The arguments above suggest that, beyond information on debt levels and income, one would need information about the circumstances in which households acquire debt, in order to identify the underlying risks that different types of households face and assess their financial vulnerabilities. As an illustration, the underlying risks of a household holding multiple loans and asking for consumer credit to pay previous loans should substantially differ from the ones of a household taking unsecured debt to benefit from low rates. Moreover, given the business model described (decisions based on history and expecting substantial non-payments, in a context of imperfect and asymmetric information), the industry incorporates this contextual information to a minimum only. Fortunately, the Chilean Survey of Household Finances or SHF provides us with rich contextual information, together with information on households' balance sheets and financial commitments.

Chile offers an interesting case of study, since it is an emerging economy, which has experienced a recent rapid growth in its income and access to credit. As a matter of fact, the importance of consumer debt in Chile is comparable to the median share that it represents within emerging economies. To provide some figures, while consumer debt in Chile accounts for 47% of households' loans (Survey of Household Finances, 2015a), Madeira (2019) shows that unsecured debt represents 56% of total households' loans in a sample of 67 emerging economies, in 2017, with both fractions being well above the 28% that it represents in a sample of 23 advanced economies (Madeira, 2019).

To assess households' FV, our approach consists of three steps. In the first step, we look at households having at least one consumer loan and examine whether accounting for contextual factors (motives for asking for consumer debt, exposure to a negative shock, unfulfilled demands for credit), and socio-economic factors (holdings of liquid assets, income strata, household size and age of the household head) allows us to classify households and identify meaningfully different types of consumer credit users. For the classification, we make use of cluster analysis techniques.

In the second step, we rely on the information on households' self-assessed debt burden, available in the SHF, and study, by credit user type, what drives the probability that households perceive themselves as highly or excessively indebted. To choose the socio-economic, demographic and regional determinants to include in the probit models, we follow the literature and consider the same factors than researchers use (in particular, Ruiz-Tagle and Vella, 2016; Cifuentes and Martínez, 2019). The intuition for following the second step is that if there are different types of consumer credit users, we want to allow for the possibility of heterogeneous parameter estimates. If heterogeneity matters, we should then observe that the value and/or significance of the estimated parameters vary across types.[4]

In the third step, we set the credit user-specific thresholds of FV at the type-specific mean proportion of households reporting themselves as highly or excessively indebted. We then define that a household is financially vulnerable when she has a predicted probability of being highly or excessively indebted (obtained in the second step) above the corresponding threshold in her type. After identifying the FVH, we conduct several analyses: i) We quantify the importance of financially vulnerable households' indebtedness on the population of households with consumer debt and on the total consumer debt; ii) we compare the findings in i) (assuming type-specific thresholds of FV) to the ones obtained supposing a homogenous threshold; iii) we compare our measure to identify FVH based on the self-assessed debt burden, with other financial metrics used in literature (relying on the debt service to income ratio or DSR,

[1] It has also been possible thanks to the significant reductions in the costs of screening, soliciting and processing unsecured loan applications, over the last 20 years (Drozd and Nosal, 2011; Livshits et al., 2010, 2016; Sanchez, 2018).

[2] Supervisors and researchers can deal with the lack of coverage of credit registries with surveys of household finances.

[3] Durkin et al. (2014) show that this argument can explain the co-existence of debt and financial assets.

[4] One could question the appropriateness of the previously described two steps (consisting of using clustering algorithms to pre-process the data and then regression analyses by identified group), and instead, attempt to account for households' heterogeneities by incorporating contextual factors and interaction terms into a single model specification, with self-assessed debt burden as the dependent variable. However, we believe that our methodology is superior and has several advantages relative to the latter approach, namely, it is more parsimonious; it allows to better study effects that vary by groups; and it provides more efficient inference for regression parameters. On top of that, Trivedi et al. (2015) shows that clustering analysis can improve prediction accuracy.

payment arrears or default and the multiplicity of loans).

We highlight the following findings. First, the cluster analysis applied to our dataset delivers six distinct types of users of consumer debt: Those who have assets but value keeping precautionary liquid balances (Durkin et al., 2014),[5] those who finance the purchase of durable goods with consumer loans, those who use consumer debt to finance business activities or make financial investments, those who suffer a negative shock (for instance, those who take a consumer loan to pay medical treatments) and, finally, those who are re-financing previous debts. It is worth adding that the motives for asking for consumer debt, exposure to a negative shock, unfulfilled demands for credit and liquid asset holdings, have been the most informative factors to separate between groups.

The second main result of the paper is that the determinants of households' probability of perceiving themselves as highly or excessively indebted considerably vary across consumer credit user types, both in the level and significance of their estimates. To cross-check this finding, we conduct several statistical tests (mean comparison tests, for the continuous variables, and non-parametric Pearson Chi-square and Kendall's tau tests, for the discrete variables), which also confirm it. To give some illustrations of this second result, on the one extreme, among the group of households who are re-financing previous debts, we find that the age, being married and the number of persons in the household have statistically significant effects on the probability of being highly or excessively indebted. On the other extreme, for those who have assets but value keeping precautionary liquid balance, their probability of perceiving themselves as highly or excessively indebted decreases with years of education, a proxy for current and future income, and increases if the household is paying a mortgage or a rent.

Third, allowing for heterogeneous thresholds of FV, we find that the identified financially vulnerable households represent 22% of the total population of households with consumer debt. Oppositely, when assuming a single threshold of FV, the share of financially vulnerable households on the total population with consumer debt equals 30%. Therefore, in the latter case, we would be considering too many households as financially vulnerable and this way, we would be overestimating the importance of FVH and the risk they could entail on the financial system.

Finally, the comparison between our self-assessed measure with other financial metrics used in literature reveals, on the one hand, that our measure is not only consistent with these alternative criteria but, on top of that, it encompasses much of the information captured in the other indicators. On the other hand, we show that FVH according to the self-assessed debt burden measure, are considerably more likely to exhibit larger DSR, to hold a larger number of loans and to experience payment arrears or default, relative to those who are not financially vulnerable. Furthermore, the evidence shows a considerable disparity across credit user types. The latter thus reinforces the relevance of having a methodology that allows for heterogeneity.

Our paper has two main advantages over the existing literature. First, thanks to the methodology we use and the detailed nature of the statistical source we employ, proving complete information about Chilean households' balance sheets, we are able to identify and distinguish between different households' behaviors. As a case of application, policy makers or regulators with a mandate on financial stability, could apply this approach to identify FVH, monitor their evolution through time and eventually, tailor specific policies to them.

Second, we propose a methodology to identify and measure the importance of financially vulnerable households, which could be applied in other cases, for example, to assess the effect of a policy change, a negative shock to the economy or the impact of the financial deregulation process in emerging economies being at earlier stages than Chile.

The remainder of the paper is organized as follows. Section two describes the methodology, both for the cluster analysis, as well as for the study of the determinants of households' probability of being highly or excessively indebted. Section three presents the data. In turn, section four first displays the results of the cluster analysis, it then examines whether there are significantly different factors driving the probability of households perceiving themselves as highly or excessively indebted; finally, it quantifies the importance of FVH and compares alternative measures to identify FVH. Section five concludes. The appendix contains a more comprehensive description of the SHF methodology, details on some robustness checks, and additional descriptive statistics, absent in the main text.

1.1. Related literature

This paper relates to two strands of literature. On the one hand, there is the literature that assesses financial stability risks arising from indebted households. Within this strand, there are some papers that identify the FVH, by simulating the impact of economic shocks on households' balance sheets (Herrala and Kauko, 2007; Karasulu, 2008; Dey et al., 2008; Albacete and Fessler, 2010; Madeira, 2018); while there are some other authors that identify them, by establishing critical thresholds of FV, based on financial ratios or subjective probability measures (Betti et al., 2007; Del Rio and Young, 2008; Martínez et al., 2013; D'Alessio and Iezzi, 2013). We contribute to this literature, by proposing a way to better account for households' heterogeneity when assessing the risks that financially vulnerable households may entail on the financial system.

On the other hand, our work relates to the literature that investigates the determinants of households' indebtedness (Keese, 2012; Jappelli et al., 2013; D'Alessio and Iezzi, 2013; Ruiz-Tagle and Vella, 2016; Cifuentes and Martínez, 2019). We contribute to this literature, because acknowledging the existence of different types of consumer credit users, we show that this heterogeneity results in significantly different determinants driving the probability that households perceive themselves as highly or excessively indebted.

[5] We relate this type with the borrower-saver household behavior in the so-called Credit Card Puzzle (Gross and Souleles, 2002; Bertaut and Haliassos, 2006; Ponce et al., 2017), which states that by simultaneously having expensive credit card debt and low yielding savings in deposit accounts, households forgo profitable investment.

2. Methodology

This section describes the methodology we rely on to classify households and to assess households' financial vulnerabilities.

2.1. Cluster analysis methodology

Cluster analysis is the analytical technique for developing meaningful subgroups of individuals or objects. The aim of a cluster analysis is to classify a sample of entities into a small number of mutually exclusive groups based on similarities among the entities (Fraley and Raftery, 1998). In any cluster analysis, there are a number of decisions to take, namely, the method of clustering; the notion of distance (which quantifies the similarity between observations); the variables to include in the analysis; the number of clusters to use and the validity criteria to assess how well the clustering scheme fits the dataset. We now describe the choices we have made.

To begin with, we apply the k-means method to classify consumer credit users. The k-means method looks for a partition of the data into k clusters that minimizes the within-cluster sum of squares (sum of distance functions of each point in the cluster to the k center) (Chiang and Mirkin, 2010). Also, the k-means is one of the simplest and fastest nonhierarchical methods to partition the data into disjoint smaller sub-groups and it performs well with large datasets. Second, regarding the notion of distance, we choose the Euclidean distance.

Third, in relation to the variables to include, in addition to being in accordance with the objective of the cluster analysis, we require that the final set of variables satisfy the following two standard criteria: That the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy be above 60% and that we reject the null of the Bartlett's test of sphericity. We then standardize the variables and apply the method of principal components (PCs) to obtain a (sub)set of orthogonal composite variables, which are linear combinations of the original ones and which account for the variance in the original data.[6,7] This practice also corrects for possible outliers (See Anderberg, 1973, for details).

Finally, there is the question about the optimal number of clusters to use, k, and the validity of the results of the clustering algorithm. In fact, both issues continue to be subject of several research efforts. Regarding the first dimension, although there is no absolute method to determine the correct number of clusters, we do a visual inspection of a commonly used metric, that is, the within cluster sum of squares (SS), against the number of clusters. Intuitively, because increasing the number of clusters reduces the dissimilarity within each cluster (and therefore, the within cluster SS), but at the expense of a description of the data which has more degrees of freedom and is therefore less parsimonious (Gough and Sozou, 2005), we look for the minimum number of clusters, such that the reduction in the within cluster SS is substantial.

As a robustness check, we also apply Frey and Van Groenewoud (1972)'s index. Two interesting features of this index are, on one hand, that it relies on a hierarchical method of clustering, thus being complementary to the within cluster SS criteria. On the other hand, to determine the optimal k, the index examines the ratio between the between-cluster and the within-cluster distances from each two possible levels in a hierarchy. More specifically, the optimal k is the very last k at which the ratio between the mean between-cluster and the average within-cluster distances from two levels in the hierarchy becomes above one.

Concerning the second dimension, that is, the validity of the clustering results, we characterize the identified groups of households, using both internal and external pieces of information (that is, variables which we consider and do not consider, respectively, in identifying the clusters). On top of that, we apply a non-parametric chi-square test to examine whether the way and the motive for getting access to the financial system vary between groups.[8]

Wrapping up, the cluster analysis allows us to evaluate our hypothesis that there are different types of consumer credit users, which differ in the way or the motive they get access to the financial system. If our hypothesis is correct, the clustering results should deliver meaningful subgroups of households, with similarities within the subgroups in observable dimensions.

2.2. Assessing households' financial vulnerabilities

To study the factors driving the probability that different types of consumer credit users perceive themselves as highly or

[6] Given a data set of n variables, which may be correlated, the method of PCs allows to construct a set of n or fewer orthogonal composite variables. The PCs are of great interest, not only because their coefficients define a set of orthogonal vectors, but also because of their maximum variance properties: The first PC has the largest variance of any linear combination of the variables represented in the data matrix; the second PC has the largest variance of any linear combination orthogonal to the first PC, and so forth.

[7] The KMO test measures how suited the data is for PC analysis. The test statistic is a measure of the proportion of variance among variables that might be common variance. As a rule of thumb, KMO values above 0.6 indicate that the sample is adequate for PC analysis. The Bartlett's test of sphericity, in turn, tests if the observed correlation matrix diverges significantly from the identity matrix, which is the theoretical matrix under the null that the variables are orthogonal. We can perform a PC analysis, only if we reject the null hypothesis.

[8] Theodoridis and Koutroumbas (2009) describe the existing approaches to investigate cluster validity. Among them, there is an internal criterion, which uses the information for the clustering analysis to evaluate how well the results fit the data, without reference to external information. Our use of internal pieces of information to characterize the identified groups of households, as well as the non-parametric chi-square test belong to the internal criterion. In addition, there is the external criterion, which consists of comparing the results of the cluster analysis to externally known results, such as externally provided class labels. Since we do not have external class labels, we rely on external pieces of information to characterize and compare the identified groups.

excessively indebted, the model specification we consider is,

$$P_{i,k}(S_{i,k} = 1 \mid X_i) = \Phi(X_i \times \beta_k), \tag{1}$$

with $P_{i,\ k}(\cdot)$ denoting the probability; $S_{i,\ k}$ taking the value of one, if household i belonging to group k responds that her level of indebtedness is high or excessive, or zero, if household i responds that her level of indebtedness is moderate or low; X_i, the household i's socio-economic and financial variables and β_k, the to-be estimated parameter vector.

The reason for estimating separate models for each group of households is to allow for the possibility of heterogeneous parameter estimates, consistent with our hypothesis that there are different types of consumer credit users that differ in the way and/or the motive they enter into the consumer credit contract.

Because we work with micro data, which contain missing values and incorporate population weights, we rely on a bootstrap procedure for estimation and computation of standard errors that use both the population weights and Rubin's rules for the variance of missing data (Shao and Sitter, 1996).

When investigating the determinants of households being highly or excessively indebted, an econometric obstacle may arise: Households report their self-assessed debt burden, conditional on having a positive demand of credit and having found a willing lender. In other words, a sample selection problem may emerge. To assess whether this potential selection bias plays any role, as a robustness check, we fit maximum-likelihood probit models with sample selection.

More specifically, for each credit user type, we model the selection mechanism as a binary probit model for the consumer debt choice, which we link to a binary choice model for the probability that a household belonging to a given group perceives herself as being highly or excessively indebted. Without loss of generality, the two possible outcomes in the selection equation of each consumer credit user type are consumer debt type k and no consumer debt.

Regarding identification, we follow closely the identification strategy of Ruiz-Tagle and Vella (2016). Precisely, in addition to imposing some exclusion restrictions, we exploit the geographical variation in households' access to credit, through the availability of bank and retail stores within the Chilean municipalities. The objective is to obtain variables that affect households' access to credit, but not their perceived level of indebtedness.

Finally, to quantify the importance of FVH, we set the credit user-specific thresholds of FV at the type-specific mean proportion of households reporting themselves as being highly or excessively indebted. We then define that a household is financially vulnerable when she has a predicted probability of being highly or excessively indebted above the corresponding threshold in her type.[9]

3. Data

Section three begins by describing the SHF. It then presents the data we use for the cluster analysis and for the study of the determinants of households' probability of perceiving themselves as highly or excessively indebted. Table A1, in the appendix, exhibits the response rates of the variables we use in this paper.

We consider the 2014 version of the SHF. In Chile, the SHF is the most comprehensive statistical source that provides complete information about the balance sheet of households and their ability to service financial commitments. In 2014, the SHF interviewed a representative sample of 4502 urban households, eliciting detailed information on their income, labor status, real state ownership and financial assets, debts, access to credit, pensions, insurances, savings, perceptions about debt service and default behavior. The SHF is answered by one respondent, who responds questions about the whole household or when corresponding, about all members of the household. The respondent is the household head, or the person who has the greatest knowledge of the financial situation of the household. Appendix A1 provides details on the methodology and the questionnaire of the SHF.[10]

In particular, we focus on households holding consumer debt, with consumer loans including bank and retail credit card loans; bank credit lines; installment loans of retail stores; bank and union loans. In Chile, consumer debt has distinct features, relative to other debt types like mortgage, which justify studying it separately. For example, there is widespread access to consumer debt among Chilean households, with 56% of households holding consumer debt, well above the 19% having mortgage debt (Survey of Household Finances, 2015a). Furthermore, consumer debt holding is relatively invariant to household head's education level, age or income. In contrast, mortgage debt holding is more prevalent among middle to high-income households, it is increasing in household head's education level and it exhibits the life cycle inverted U-shaped profile (Survey of Household Finances, 2015a).

To identify the consumer credit user types through the cluster analysis, we consider the following information. First, we rely on the motives that households report for holding consumer debt, namely, purchase of household articles; purchase of cars; vacations; medical treatment; pay other debts; finance business activities; purchase of financial assets; house-renovations and other.

Second, we incorporate the information on whether the household has been exposed to a negative shock (unexpected income decrease or spending rise), together with access to credit, *i.e.*, whether the household has asked for credit, the number of times credit

[9] One could question the appropriateness of relying on the predicted probability of being highly or excessively indebted for the identification of the FVH. Alternatively, we could have directly used households' self-assessed debt burdens and say, for instance, that all households reporting that their debt burden is excessive are financially vulnerable. The reason for not doing it is that by controlling for the socio-economic, demographic and regional heterogeneities across households, and estimating separate models, we allow for non-linearities, we gain granularity and we better identify the FVH.

[10] For details on the sample design of the SHF, refer to Survey of Household Finances, 2015b. For details on the imputation procedure at the SHF, see Martínez et al. (2019).

has been denied and the number of times credit has been partially granted; all of them, over the previous 12 months, prior to the interview date. Finally, as socio-economic and demographic covariates, we include the number of persons in the household, the age of the respondent, the household's income strata and whether the household holds liquid assets (including fixed income, mutual funds, voluntary pension savings and saving accounts).[11]

It is worth making three additional observations about the variables we use for the cluster analysis. First, concerning the criteria to select them, we do not include information on financial performance (like the debt service to income ratio or DSR, the proportion of households with payment arrears or households' perception of their indebtedness level), because the aim is to separate the reasons and/or conditions for a household to drive into consumer debt, from how it is doing in managing its debt. That is why we consider the motives, information on unexpected shocks and access to credit. On top of that, we incorporate some socio-economic and demographic factors, in order to enrich the analysis and, potentially, allow for the groups to better distinguish between themselves.

The second observation is that the final set of variables satisfy the standard criteria described in Section 2.1. Lastly, regarding the missing values in the variables used for the clustering analysis, we replace those values with the average of the 30 multiply imputed data sets provided by the SHF to researchers. Still, because the proportion of missing observations in these variables is low, as reported in Table A1, in the appendix, we believe that this does not bias the clustering results.

Concerning the study of the determinants of households' probability of being highly or excessively indebted, we consider, as dependent variable, the information on self-assessed debt burden. Specifically, the question the SHF asks reads as follows: “Taking into account all the household's debts, how do you describe your household's level of indebtedness?” The answer choices are “excessive”, “high”, “moderate” or “low”. The dependent variable we construct takes the value of 1 if the household answers “excessive” or “high” and zero otherwise.

The reason for using households' self-assessed debt burden is because these measures have two main advantages. On the one hand, since they are self-evaluations, they consider the overall household situation, including her budget situation and expectations of future income; on the other hand, they allow not only to identify households who are already experiencing debt payment problems, but also those that are in a complex financial situation and thus are likely to experience these problems in the future (Cifuentes and Martínez, 2019). Interestingly, Betti et al., 2007 show that households are generally honest in reporting their debt situations in household surveys.

As socio-economic, demographic and regional variables, we consider age and years of education of the respondent; total number of persons in the household; total number of employed individuals in household; dummy variables for gender (1 if respondent is male), marital status (1 if the respondent is married), self-employment (1 if the respondent self employs) and geographical controls (North, Center, South and Metropolitan area). These variables are standard in the literature to capture the social, economic and regional heterogeneity present in the SHF data (Ruiz-Tagle and Vella, 2016; Cifuentes and Martínez, 2019).

In addition, we include home ownership (distinguishing between home paid, owner with a mortgage, rental home and other housing regimes), as a proxy for real wealth; the proportion of households with financial assets (including liquid assets, as well as equities, shares on other companies and insurance savings) and the DSR. In particular, the inclusion of the DSR is to measure to what extent the fraction of households' income spent to pay total financial obligations implies a debt burden. In order to reduce the effect of possible outliers in the DSR, we winsorize 1% in each tail.

To investigate whether there is a sample selection problem, we follow closely the model specification of Ruiz-Tagle and Vella (2016) and employ, as primary exclusion restrictions, variables related to credit accessibility and financial depth; more explicitly, the number of banks per inhabitant and the number of retail stores per inhabitant, both at the municipality level, to measure the facility to access the financial system in each municipality.

As additional exclusion restrictions for identification, we incorporate, as they do, dummy variables for spouse (1 if spouse is present), signed contract (1 if the respondent has signed a job contract), current account (1 if at least one member of the household owns a current account), internet banking (1 if the household uses internet banking), pension system (1 if respondent is affiliated to the pension system) and insurance (1 if at least one member of the household has one voluntary insurance); pension savings; households' real estate assets (including principal residence and other investments in real estate) and the total regional income.

Finally, for a separate robustness check, we consider dummy variables for bank debt (1 if the respondent holds a bank credit card, a credit line or a bank loan) and retail debt (1 if the respondent has a retail credit card or an installment loan of retail stores).

4. Results

Section four starts by presenting the clustering analysis results. It then examines the determinants of households' self-assessed debt burden. Finally, it presents the (heterogeneous) thresholds of financial vulnerability and quantifies the importance of financially vulnerable households on the population of households with consumer debt and on total consumer debt.

4.1. Identifying types of consumer credit users

We start with the determination of the optimal number of clusters to use, *k*. To do so, we first do a visual inspection of the within cluster SS, against the number of clusters, as depicted in Fig. 1. Second, we report the optimal *k* that Frey and Van Groenewoud (1972)'s index selects.

[11] As a robustness check, instead of the household's income strata and the proportion of households with liquid assets, we consider the household's income and the amount of liquid assets, respectively. Results are robust to these alternative variables.

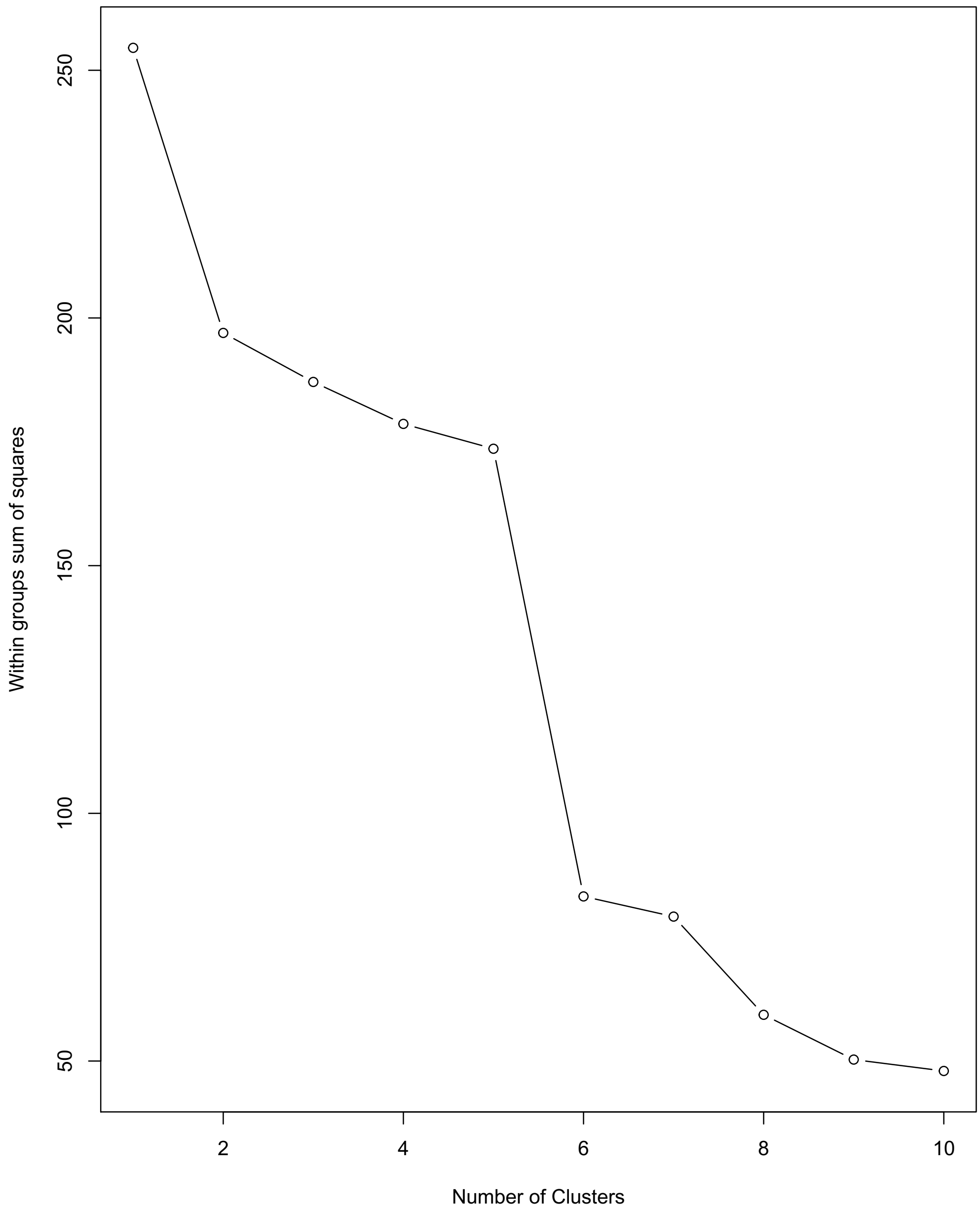

Fig. 1. Within cluster sum of squares against k.
Source: 2014 EFH data.

Fig. 1 shows, as expected, that as we increase the number of clusters, starting from $k = 1$, the dissimilarity within each cluster and therefore the within cluster SS, reduce. However, it does so at the expense of a description of the data which is less parsimonious. From Fig. 1, we prefer $k = 6$, as it is the minimum number of clusters, such that the reduction in the within cluster SS is substantial. Note that increasing the number of clusters from 6 to 7 only involves a minor reduction in the within cluster SS. Interestingly, Frey and Van Groenewoud (1972)'s index also selects $k = 6$ as the optimal number of clusters.[12]

As a first step towards characterizing the six identified clusters or groups of households, we apply a non-parametric chi-square test

[12] To assess the validity of our assumption of groups being mutually exclusive, we implement a Fuzzy C-means algorithm, which allows observations to potentially be a member of more than one cluster (Bezdeck, 1981), with six clusters. We then compare the results of our K-means with the Fuzzy C-means, by computing an adjusted rand index (ARI). Being a measure of the similarity between two clusterings, the ARI can take values between 0 and 1, with 0 indicating that the two clusterings do not agree on any pair of points and 1 indicating that they are exactly the same. We find that the ARI equals 0.76, thus indicating that the two clusterings are similar. We thus conclude that the assumption of mutually exclusive clusters is not a restrictive one.

to examine whether the contextual factors we consider (namely, households' motives for holding consumer debt, access to credit and exposure to a negative shock) statistically and significantly vary across the identified groups.[13] Table A2, in the appendix, reports the chi-square test statistic, the degrees of freedom and the P-value for each discrete variable.

Table A2 provides strong evidence in favor of our hypothesis that accounting for these contextual factors results in meaningfully different types of consumer credit users. This is because the values of the chi-square test statistics are high and, in all cases except one, we reject the null hypothesis that there are no statistically significant differences between the identified groups of households, in the variables reported in Table A2, at 99% confidence level.[14]

To characterize the identified groups of households, Table 1 starts by reporting the mean values of the variables we use in the cluster analysis, by credit user type, together with the coefficients of variation.

Second, regarding group II, the most noticeable features are that all households in this group report holding consumer debt to finance business activities; it is the group with the smallest proportion of households having been exposed to a negative shock over the last 12 months (0.25); it is one of the two groups with the largest fraction of households having at least one credit denied (0.10), while it is the group with the smallest fraction of households having at least one credit partially granted (0.01). Based on the previous elements and given that group II is strongly identified by the motive finance business activities, we label it as the “entrepreneur credit user type”.[15]

Third, all households in group III report using consumer debt to finance their vacations; it is the group with the smallest proportion of households with at least one credit denied (0.01); it is the group with the largest average income strata (2.50) and almost half of them have liquid assets (0.49). The way we interpret the evidence in Table 1 is that households in group III use consumer debt because they value liquidity and/or for transactional motives (for instance, to profit from installments at zero interest, a common practice among Chilean credit card providers).[16] Based on the previous elements, we label group III as the “transactional consumer credit user type”.

Fourth, group IV is the largest group, accounting for 40% of total households with consumer debt. In addition, a large proportion of households in group IV report holding consumer debt to purchase household articles (0.64); it is the group with the smallest fraction of households using consumer debt to finance medical treatments; it is the group with the largest proportion of households holding consumer debt to finance home renovations (0.12). In addition, group IV registers the largest average household size (3.95 persons) and the lowest average respondents' age (40.39 years), with the latter suggesting that respondents in this group tend to be relatively young. Considering its characteristics, we refer to this group as the consumer credit user type with the motive purchase of household-related goods.

Fifth, group V is the group with the largest fraction of households using consumer debt to finance medical treatment (0.21) and with the smallest proportion of households using consumer debt to purchase cars or finance home-renovations; it is the group with the smallest proportion of households having asked for credit and having at least one credit partially granted. Furthermore, group V is strongly identified by socio-economic variables: It registers the smallest average income strata and household size (1.47 and 2.44 persons, respectively); it is the group with the smallest proportion of households holding liquid assets (0.20) and it exhibits the largest average respondents' age (61.17). For all the previous reasons, we call this group as the pensioner credit user type.

Finally, while the number of households belonging to group VI is small, it appears to be the group with the smallest proportion of households holding consumer debt to purchase household articles and with the largest fraction using consumer debt to purchase a car (0.12). Also, group VI is strongly characterized by the motive purchase of financial assets, since all households in this group report holding consumer debt for this reason. On top of that, it is the group with the largest proportion having asked for credit over the last 12 months (0.35) and having at least one credit partially granted (0.08). Since its most salient feature is the motive purchase of financial assets, we label this group as the financial credit user type.

Wrapping up, the evidence in Table 1 confirms our initial hypothesis that there are meaningfully different types of consumer credit users, who differ in the way and/or the motive they enter into the consumer credit contract. Indeed, the fact that we can give different labels to the six identified groups of households, based on the internal information we rely on, reinforces the idea that the identified consumer credit user types have distinct characteristics, and at the same time, it is an indication of the validity of our cluster analysis (internal criterion).

[13] Researchers use the non-parametric chi-square test to examine the relation between discrete variables. By comparing the relative frequencies among two or more groups, it tests the null that there is no statistically significant difference between groups in the variable of interest. In any chi-square analysis, the null hypothesis generates expected frequencies against which observed frequencies are tested. If the observed frequencies are similar to the expected frequencies, then the value of the chi-square test statistic is low and the null hypothesis should be retained. On the contrary, if the groups are sufficiently different, then the value of the chi-square test statistic is high and the null hypothesis may be rejected. See Bahovec et al. (2015) for an application of this test, with clustered data.

[14] For completeness, for the continuous variables, we compute statistical tests to compare, by pairs of clusters, the mean of household size (or the age of the household head) of any two given groups of households. In all cases, we reject the null hypothesis that there are no statistically significant differences between the mean values of any two given groups.

[15] Supporting the label, the share of self-employed respondents on the population of households belonging to this type is the largest and equals to 61%, well above the overall 22% that self-employed respondents represent on the population of Chilean households with consumer debt.

[16] To assess the latter, Table A3, in the appendix, displays the use of bank and/or retail credit cards to pay without installments, in installments at zero interest and in installments with interests, by group. Interestingly, Table A3 shows that group III exhibits the largest fraction of households using bank or retail credit cards to pay in installments at zero interest. Therefore, this evidence is consistent with our interpretation that households in this group use consumer debt because they value liquidity.

Table 1
Mean values of the variables used in the cluster analysis, by type of consumer credit user.

Group Id	I	II	III	IV	V	VI	Mean	CV
							All	
Motives for holding consumer debt[a]								
Purchase of household articles	0.46	0.52	0.45	0.64	0.57	0.41	0.56	0.88
Purchase of cars	0.08	0.07	0.11	0.11	0.03	0.12	0.08	3.37
Vacations	0.00	0.08	1.00	0.00	0.00	0.14	0.07	3.53
Medical treatment	0.10	0.05	0.07	0.02	0.21	0.06	0.09	3.15
Pay other debts	1.00	0.18	0.20	0.00	0.02	0.16	0.21	1.93
Finance business activities	0.00	1.00	0.00	0.00	0.00	0.06	0.06	3.93
Purchase of financial assets	0.00	0.00	0.00	0.00	0.00	1.00	0.02	7.51
Renovations	0.10	0.08	0.09	0.12	0.06	0.10	0.10	3.03
Shocks and access to credit								
Unexpected negative shocks[a]	0.41	0.25	0.28	0.28	0.34	0.37	0.32	1.47
Did the household ask for credit?[a]	0.32	0.29	0.28	0.17	0.11	0.35	0.20	2.03
Households with 1+ credits denied[b]	0.10	0.10	0.01	0.03	0.02	0.02	0.04	8.11
Households with 1+ credits p. granted[c]	0.04	0.01	0.05	0.03	0.01	0.08	0.03	7.61
Socio-economic variables								
Total income strata[d]	2.01	2.38	2.50	2.05	1.47	2.18	1.94	0.42
Liquid assets holdings[a]	0.28	0.39	0.49	0.33	0.20	0.27	0.30	1.54
Household size (persons)	3.76	3.69	3.36	3.95	2.44	3.78	3.45	0.48
Age of the household head (years)	46.25	48.46	46.92	40.39	61.17	49.00	48.15	0.31
Observations	507	168	188	1136	761	49	2809	
Population represented	578,826	132,872	151,005	1,253,352	819,349	43,545	2,978,949	

Source: 2014 EFH data.
Notes: Internal variables reported, that is, variables that are in the cluster analysis. Underlined numbers indicate the row-minimum values. Numbers in boxes refer to the row-maximum values.

[a] The variable(s) is(are) expressed as proportion(s), corresponding to the fraction of households with a certain characteristic over the total number of households in a group.

[b] Instead of the number of households with 1 or more credits denied, used in the cluster analysis, for expositional convenience, this table reports the proportion of households with 1 or more credits denied.

[c] Instead of the number of households with 1 or more credits partially granted, used in the cluster analysis, for expositional convenience, this table reports the proportion of households with 1 or more credits partially granted.

[d] We classify household total income into three strata. To begin with, Table 1 shows that all households in group I report holding consumer debt to pay other debts (1.00); almost half of them have been exposed to an unexpected negative shock in the last 12 months, prior to the interview date (0.41), and it is one of the two groups exhibiting the largest fraction of households with at least one credit denied over the last 12 months (0.10). Based on the previous elements, we label group I as a “credit constrained user type”.

To dive deeper into the characteristics of the identified groups of households, we now compare the identified groups in terms of their financial situation. Specifically, while Table A4 compares the consumer credit user types in terms of various consumer and mortgage debt holdings, and wealth, Table A5 focuses on the DSR, the debt to income ratio or DIR and the proportion of households with payment arrears or default. Complementing the latter, Table A6 presents the motives for holding consumer debt, distinguishing by credit user type and consumer debt type, whereas Table A7 exhibits the median maturity of the several consumer debt types, by credit user type. Tables A4–A7 are all in the appendix.

To begin with, Tables A4–A7 provide support for our label of the credit constrained user type. This is because the evidence shows that type I exhibits the largest fraction of households holding four of the ten debt types that Table A4 reports; households in this type use both short-term and long-term loans to pay other debts (Tables A6 and A7)[17]; type I exhibits the smallest average wealth (Table A4) and finally, households in this type are experiencing difficulties to pay back their debts (Table A5). Supporting the latter, type I registers the largest mean DIR (relative to the other credit user types), both for consumer debt and total debt, the largest DIR (total debt) at the 90th percentile and the largest fraction of households with payment arrears or default in six out of the eight debt types we consider in Table A5.

Interestingly, using the SHF, Madeira (2015) finds that using consumer debt to pay past debts is correlated with the rates of loan delinquency. The fact that households in type I may be experiencing payment difficulties is hence consistent with the author's result.

[17] To classify a debt type into short-term or long-term, we look at the information on median maturity. More precisely, we consider a debt type as short-term (long-term), if its median maturity is below (above) one year. For other debts though, it is not possible to compute its median maturity, since the SHF does not ask for this information. From table A6, we observe that type I always exhibits the largest fraction of households using a debt type to pay other debts. Concerning the maturities, table A7 shows that while bank credit cards, credit lines, retail credit cards and installment loans of retail stores have median maturities of 9, 9, 6 and 11 months, respectively, union loans and educational loans have median maturities of 17 and 64 months, respectively.

Also, Madeira shows that having a low income increases the probability of using consumer debt to pay other debts. As a matter of fact, our credit constrained user type exhibits the second smallest average income strata (Table 1) and the smallest average wealth (Table A4), relative to the other credit user types.

Second, Tables A4, A6 and A7 show that type II is the group with the largest fraction of households using both short-term (bank and retail credit cards, and credit lines) and long-term (bank and union loans) consumer debts to finance business activities (long term motive, Table A6); that type II registers the largest debt service to income ratios and the largest proportion of households with payment arrears or default in the case of bank loans. Therefore, the evidence reinforces the finding that the entrepreneur credit user type may be facing some borrowing constraints to finance their business activities.

Third, the evidence that credit user type III exhibits the largest proportion of households with mortgages and bank credit cards and the lowest fraction of households with other debts, with respect to the other credit user types (A4); that it has the largest average wealth (Table A4); that only a minor fraction of households in this type have undergone into payment difficulties (Table A5); together with the finding that more than half of them have liquid assets (Table 1) speak to the so-called Credit Card Puzzle (Gross and Souleles, 2002; Bertaut and Haliassos, 2006; Ponce et al., 2017).[18] Interestingly, our results for type III suggest that households in this type may be having this puzzled behavior.

Fourth, Table A5 indicates that using consumer debt to purchase home-related items does not seem to represent a heavy debt burden for households belonging to type IV. This is because type IV is the one registering the smallest debt service to income ratios and the smallest mean consumer debt to income ratio, while it is one of the types with the smallest fraction of households with payment arrears or default on bank credit cards, installment loans of retail stores and auto loans (Table A5).

Fifth, our pensioner credit user type is the one exhibiting the smallest debt to income ratios, with one exception, and it is one of the types with the smallest fraction of households with payment arrears or default on five out of the seven debt types that Table A5 reports. Finally, type VI exhibits the largest proportion of households holding installment loans of retail stores and auto loans (Table A4); the use of consumer debt to purchase financial assets is widespread across debt types (Table A6); and last, although type VI exhibits the largest 75^{th} and 90^{th} percentiles for consumer DIR, it only registers the largest fraction of households with payment arrears or default only in the case of bank credit cards (0.04, Table A5).

Summing up, the results in this section show that the identified credit users are meaningfully associated with different forms of accessing the credit market. Therefore, it is important to take this heterogeneity into account when assessing Chilean households' consumer indebtedness. Furthermore, the external pieces of information we consider (Tables A3–A6 and A7) confirm the distinct labels we give for the credit user types identified in the cluster analysis.

4.2. Determinants of households' probability of being highly or excessively indebted

To study the determinants of households' self-assessed debt burden, we start by comparing the identified consumer credit user types in terms of their debt perceptions. Table 2 reports the proportion of households perceiving themselves as low, medium, highly or excessively indebted, by type of consumer credit user. Fig. A.1, in the appendix, depicts the percentage of households reporting highly or excessively indebted, by type of consumer credit user.

There are several findings to extract from Table 2 and Fig. A.1. The first one is that between 40 and 52 percent of households in each type report themselves as being moderately indebted (category medium). The second one is that types I and II are the consumer credit user types with the largest proportions of households reporting themselves as highly or excessively indebted (0.53 and 0.45, respectively), which is consistent with the findings in the previous section, namely, that type I is credit constrained and that type II, the entrepreneur type, may also be facing some difficulties to finance her business activities and to pay back some of her debts. The third element is that type V is the one exhibiting the smallest fraction of households perceiving themselves as highly or excessively indebted (0.27).

To model households' self-assessed debt burden, we group the categories high and excessive together. One advantage of grouping the categories is that it reduces the problem of interpersonal comparability of households' perceptions, that is, households might have different perceptions of their level of indebtedness depending on their attitudes towards risk and personality traits, which we can not observe (Betti et al. (2007) and Cifuentes and Martínez (2019) are two references following the same approach).[19]

To determine the factors determining the probability that households report themselves as being highly or excessively indebted, we estimate the same model specification for each subsample of households belonging to the identified types. Allowing for heterogeneous (type-specific) parameter estimates is justified by our previous finding, that is, that there are meaningfully different types of consumer credit users that differ in the way and/or the motives they enter into the consumer credit contract. Also, it is supported by the statistical tests we conduct, which show that in all cases except six (out of the 46 tests we conduct), the determinants we consider are significantly different across consumer credit user types.[20] Finally, the empirical analysis excludes credit user type VI, since the number of households belonging to this type is too small.

It is important to stress that the aim of this regression analysis is prediction, not causal analysis. Table 3 reports the marginal

[18] According to this strand of literature, the puzzle arises because by simultaneously having expensive credit card debt and low yielding savings in deposit accounts, households forgo profitable investment. Telyukova and Wright (2008) and Telyukova (2013) explain this puzzle with the observation that money is more liquid than credit cards.

[19] Supporting our grouping choice, Cifuentes and Martínez (2019) provide evidence of significant differences in the empirical distributions of the "low" and "moderate" responses on the one hand, and "high" and "excessive" on the other.

[20] Specifically, for the continuous variables, we conduct mean comparison tests, by pairs of credit user types, whereas for the discrete variables, we consider non-parametric Pearson Chi-square and Kendall's tau tests.

Table 2
Self-assessed debt burden, by type of consumer credit user.

Group Id	I	II	III	IV	V	VI	All
Low	0.08	0.11	0.15	0.19	0.23	0.09	0.17
Medium	0.40	0.44	0.46	0.52	0.51	0.52	0.49
High	0.38	0.33	0.26	0.22	0.17	0.24	0.24
Excessive	0.15	0.12	0.12	0.06	0.10	0.16	0.10
Obs.	507	168	188	1135	757	49	2804

Source: 2014 EFH data.
Notes. Variable description: The categories for the self-assessed debt burden correspond to low, medium, high and excessive. Proportions incorporate population weights. Underlined numbers indicate the row-minimum values. Numbers in boxes refer to the row-maximum values. The reduction from 2809 to 2804 observations is due to some missing values in households' self-assessed debt burden. Labels for the consumer credit user types: I = credit constrained type; II = entrepreneur type; III = transactional type; IV = type with motive purchase of household goods; V = pensioner type; VI = financial type.

effects at means of the model specification in equation 1, for each credit user type. Table A8, in the appendix, exhibits the coefficient estimates of the same model specification, whereas Table A9, also in the appendix, compares the mean predicted probability and the proportion of households perceiving themselves as highly or excessively indebted, by consumer credit user type.

A key conclusion we can draw from Table 3 is that there is heterogeneity across consumer credit user types, both in the level and the significance of the marginal effects of the factors driving the (conditional) probability of households being highly or excessively indebted. In what follows, we detail in which sense the determinants are different across credit user types.

First, regarding the socio-demographic variables, we observe that age and age-squared exhibit statistically significant effects on the probability of households being highly or excessively indebted, only for types I and IV, while there is no statistically significant effect for the other credit user types. Notably, credit user types I and IV have in common that they are the types exhibiting the smallest averages of respondents' age (relative to the other credit user types).

Second, years of education has a statistically significant negative effect on the probability of households being highly or excessively indebted only in the case of type III. Interestingly, respondents belonging to type III are the most educated (relative to the other credit user types), as measured by years of education. In turn, a respondent being male presents a statistically significant negative effect on type IV probability, whereas the number of persons in the household exhibits a positive and significant marginal effect on the same probability for types I, IV and V. Among them, type V is the credit user type with the smallest average of number of persons in the household.

Third, concerning the dummy variable for being married, we find that it exhibits a statistically significant negative effect on households' probability of being highly or excessively indebted, only for type I. Since being married makes more likely that there is at least a second income earner in the household, the way we interpret the significantly negative coefficient for type I is that this effect is stronger for households who are at least to some extent financially constrained.

Fourth, in relation to the geographical controls, we observe that the Metropolitan area and the South of Chile have a statistically significant effects on the same probability, for types I and II. The latter reveals a limited geographical heterogeneity for households' self-assessed debt burden. Fifth, regarding the economic covariates, we find that the total number of employed persons in the household has no statistically significant effect on the probability of being highly or excessively indebted, whereas self-employment, which is a proxy for household income instability, increase type III and type IV probability of being highly or excessively indebted.

Sixth, the dummy variables for housing regimes exert a statistically significant and positive effect on the probability of being highly or excessively indebted mainly for type III. Interestingly, the fact that credit user type III is the type with the largest proportion of households holding mortgages (0.40, Table A4) may contribute to explain why paying the mortgage or a rent has a positive effect on the probability of households belonging to this type and perceiving themselves as highly or excessively indebted.

Seventh, Table 3 shows that an increase in the financial asset holdings, as a proxy for financial wealth, decreases the probability of households reporting high or excessive indebtedness in the case of type IV. Notably, type IV is a credit user type being strongly identified by the motive purchase of household items, it exhibits the largest household size and the smallest average respondents' age, relative to the other credit user types (Table 1). Intuitively, holding some financial assets improves the household's perception of her financial situation.

Finally, an increase in the DSR (total) has a positive and significant marginal effect on the conditional probability of households being highly or excessively indebted, in all cases, except for credit user type II. As expected, the larger the fraction of a household's income spent to pay total financial obligations, the more likely it is that the household perceives herself as highly or excessively indebted. Importantly, the level of the marginal effects for DSR (total) considerably varies across credit user types, with type I presenting the largest marginal impact.

We conduct several robustness checks. First, we examine whether there is a sample selection problem in the model estimates reported in Table 3. To do so, for each credit user type, we fit maximum-likelihood probit models with sample selection. For identification, we follow closely the identification strategy of Ruiz-Tagle and Vella (2016), which exploits the geographical variation in the access to credit, through the availability of bank and retail stores within the Chilean municipalities. In addition, regarding the

Table 3
Marginal effects at means of the credit user-specific probit models.

	I	II	III	IV	V
Age	0.02*	−0.01	0.01	0.02*	−0.01
	(0.01)	(0.03)	(0.02)	(0.01)	(0.01)
Age^2	0.00*	0.00	0.00	0.00**	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Years of education	0.01	−0.02	−0.06***	0.00	−0.01
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
Male = 1	0.00	−0.04	−0.04	−0.11***	0.02
	(0.05)	(0.12)	(0.08)	(0.04)	(0.05)
Number of persons in the household	0.04*	0.05	0.03	0.05***	0.06**
	(0.02)	(0.04)	(0.03)	(0.01)	(0.02)
Civil status: married	−0.21***	−0.16	−0.09	−0.07	−0.03
	(0.06)	(0.12)	(0.08)	(0.04)	(0.06)
Center of Chile	−0.01	−0.10	−0.15	−0.05	0.11
	(0.07)	(0.20)	(0.21)	(0.05)	(0.08)
Metropolitan region	−0.14*	−0.13	−0.13	−0.08	0.03
	(0.07)	(0.18)	(0.20)	(0.05)	(0.08)
South of Chile	0.09	−0.33*	−0.09	0.00	0.06
	(0.10)	(0.21)	(0.21)	(0.07)	(0.11)
Number of employed persons in the household	−0.01	−0.11	−0.07	−0.02	0.04
	(0.03)	(0.07)	(0.05)	(0.02)	(0.03)
Self-employed	0.04	0.19	0.16*	0.14***	0.05
	(0.07)	(0.12)	(0.08)	(0.04)	(0.07)
Owner with a mortgage	−0.05	0.20	0.54***	0.03	0.01
	(0.08)	(0.14)	(0.11)	(0.05)	(0.07)
Rental home	0.00	−0.02	0.49***	−0.09	0.13*
	(0.08)	(0.15)	(0.13)	(0.05)	(0.07)
Other housing regimes	0.09	−0.24	0.48***	−0.04	0.08
	(0.08)	(0.19)	(0.17)	(0.06)	(0.06)
Holding of financial assets	−0.09	−0.05	−0.07	−0.08**	−0.08
	(0.06)	(0.13)	(0.08)	(0.04)	(0.06)
DSR (total debt)	0.42***	0.12	0.25**	0.17***	0.20***
	(0.09)	(0.13)	(0.10)	(0.05)	(0.05)
Observations	505	168	188	1131	757
Pseudo R^2	0.17	0.14	0.41	0.12	0.09

Notes: Constant not reported. DSR (total debt): Total debt service to income ratio, winsorized 1% in each tail. Bootstrapped standard errors computed with 1000 replicates, in parentheses. Level of significance: *10%, **5%, ***1%. Labels for the consumer credit user types: I = credit constrained type; II = entrepreneur type; III = transactional type; IV = household goods purchase type; V = pensioner type.

other exclusion restrictions the authors impose, we use, whenever possible, the same variables they choose to model households' access to consumer credit. The latter is for comparability.[21]

Table 4 presents the coefficient estimates for the first stage, whereas Table 5 reports the coefficient estimates for the second stage, including the estimate for *atanh* ρ, with ρ being the correlation coefficient of the residuals in the selection and choice equations. In addition, Table A10, in the appendix, exhibits the descriptive statistics of the variables used in the first stage.

Second, we investigate whether bank and retail debt holdings, as defined in Section 3, have some explanatory power for households' probability of being highly or excessively indebted. Because we find that the identified credit user types have, on average, distinct portfolios of consumer debt (Section 4.1), the objective of this exercise is to evaluate whether these distinct patterns also affect households' perception of their debt burden. To do so, we augment the model specification in Table 3, by consumer credit user type, with the dummy variables for bank and retail debt. Table A11, in the appendix, reports the marginal effects at means of this alternative model specification.

Third, we consider some alternative and additional covariates to the ones reported in Table 3. For example, instead of the dummy variables for home ownership, we include home value and household wealth; in place of the financial asset holdings, we consider the liquid asset holdings; as a substitute for years of education, we use dummy variables for the educational attainment of the respondent. In addition, we incorporate two dummy variables which take the value of 1 if the household uses the bank or retail credit card to pay in installments at zero interest.

Fourth, instead of estimating a probit model by consumer credit user type, we estimate a single multinomial probit model, allowing for correlation in the error term. We then compare our baseline probit model estimates with the multinomial probit model estimates, in terms of their predictive performance. Table A12, in the appendix, defines the metrics we use to compare the two model estimates, whereas Table A13, also in the appendix, reports the values of these metrics for the two alternatives model estimates.

Finally, instead of grouping the self-assessed indebtedness categories high and excessive into one category, as a robustness check,

[21] In the second stage, Ruiz-Tagle and Vella (2016) have as dependent variables the demand for consumer, mortgage and total debt. Therefore, our and their second stages are not comparable.

Table 4
Coefficient estimates: robustness check 1, first stage.

	I	II	III	IV	V
Age	0.02	0.03	0.02	0.03	0.20***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Age2	0.00	0.00	0.00	0.00***	0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Years of education	0.00	0.03	0.05***	0.00	−0.01
	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
Male = 1	−0.08	0.18	−0.02	−0.05	0.11
	(0.09)	(0.13)	(0.13)	(0.08)	(0.09)
Number of persons in the household	0.04	−0.06	−0.07*	0.18***	−0.23***
	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)
Spouse present	0.13	0.35**	−0.23	0.11	−0.24*
	(0.10)	(0.16)	(0.15)	(0.09)	(0.12)
Signed contract	0.26*	−0.01	−0.08	0.04	0.07
	(0.13)	(0.14)	(0.22)	(0.12)	(0.13)
Current account savings	−0.18*	−0.04	0.17	−0.05	−0.08
	(0.10)	(0.14)	(0.14)	(0.09)	(0.18)
Internet banking	0.22**	0.27**	0.32**	0.31***	0.07
	(0.10)	(0.13)	(0.15)	(0.09)	(0.18)
Pension system	0.21**	−0.10	0.12	0.20**	−0.13
	(0.10)	(0.16)	(0.13)	(0.09)	(0.10)
Insurance holding	0.27***	0.06	0.17	0.13	−0.19*
	(0.08)	(0.13)	(0.14)	(0.08)	(0.11)
Center of Chile	−0.62***	0.01	0.23	−0.05	0.02
	(0.16)	(0.29)	(0.28)	(0.14)	(0.17)
Metropolitan region	−3.58***	1.23	3.29**	−0.16	−1.07
	(0.88)	(1.60)	(1.50)	(0.76)	(0.94)
South of Chile	−0.60***	−0.06	0.74***	0.20	0.54***
	(0.15)	(0.21)	(0.23)	(0.15)	(0.16)
Number of employed persons in the household	0.09*	0.15**	0.13**	0.07*	0.00
	(0.05)	(0.07)	(0.07)	(0.04)	(0.06)
At least one unemployed person in the house	0.36***	0.33	0.01	−0.14	−0.20
	(0.14)	(0.32)	(0.44)	(0.17)	(0.21)
Self-employed	0.08	0.81***	−0.21	−0.17	−0.22*
	(0.12)	(0.13)	(0.17)	(0.10)	(0.12)
Pension savings	0.00**	0.00	0.00	0.00**	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Real estate assets (std)	−0.09*	0.04	0.08*	−0.03	−0.14**
	(0.06)	(0.05)	(0.05)	(0.04)	(0.07)
Total regional income (std)	1.54***	−0.89	−1.47**	0.14	0.60
	(0.41)	(0.73)	(0.69)	(0.35)	(0.45)
Number of banks per inhabitant	−0.40	0.56*	0.22	−0.13	−0.12
	(0.25)	(0.31)	(0.30)	(0.23)	(0.42)
Number of retail stores per inhabitant	−0.53	−8.04**	−6.44*	4.46**	0.96
	(2.29)	(3.53)	(3.63)	(2.27)	(2.96)
Observations	4358	4358	4358	4358	4358

Notes: Constant not reported. Real estate assets (std): Standardized values of real estate assets. Total regional income (std): Standardized values of total regional income. Not having consumer debt is the base outcome. Level of significance: *10%, **5%, ***1%. Labels for the consumer credit user types: I = credit constrained type; II = entrepreneur type; III = transactional type; IV = type with motive purchase of household goods; V = pensioner type; VI = financial type.

we estimate ordered probit models by credit user type. We then compare our baseline probit model estimates with the ones resulting from estimating ordered probit models, in two manners: On the one hand, relying on information criteria (the Akaike Information Criterion or AIC, and the Bayesian Information Criterion or BIC); on the other hand, by examining their predictive performance. Table A14, in the appendix, compares the binary and the ordinal probit model estimates in terms of the information criteria, whereas Table A15, also in the appendix, compares them in terms of their predictive performance.

In what follows, we examine the main findings of robustness checks one to five. Regarding the third robustness check, we do not report the model estimates for each of the alternative model specifications.

Starting with the model estimates of the first stage in Table 4, it is possible to draw the following conclusions. First, in relation to the impact of the primary exclusion restrictions, that is, the level of financial depth, as measured by the number of banks per inhabitant and the number of retail stores per inhabitant, both at the municipality level, we find that the number of banks per inhabitant increases households' access to consumer debt for type II, whereas the number of retail stores per inhabitant exhibits

Table 5
Coefficient estimates: robustness check 1, second stage.

	I	II	III	IV	V
Age	0.04	0.00	0.02	0.07**	−0.07
	(0.04)	(0.09)	(0.10)	(0.03)	(0.14)
Age^2	0.00	0.00	0.00	0.00**	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Years of education	0.01	−0.03	−0.29**	0.01	−0.02
	(0.03)	(0.06)	(0.10)	(0.02)	(0.03)
Male = 1	0.08	−0.13	−0.32	−0.40***	0.03
	(0.16)	(0.40)	(0.38)	(0.13)	(0.18)
Number of persons in the household	0.11	0.09	0.15	0.21***	0.20
	(0.06)	(0.15)	(0.15)	(0.06)	(0.15)
Civil status: married	−0.58***	−0.38	−0.41	−0.18	−0.04
	(0.18)	(0.39)	(0.36)	(0.13)	(0.16)
Center of Chile	0.03	−0.39	−0.31	−0.19	0.25
	(0.21)	(0.63)	(1.01)	(0.17)	(0.38)
Metropolitan region	−0.26	−0.41	−0.56	−0.23	0.03
	(0.24)	(0.63)	(0.91)	(0.17)	(0.31)
South of Chile	0.49	−0.94	−0.53	0.08	0.04
	(0.31)	(0.76)	(0.97)	(0.23)	(0.58)
Number of employed persons in the household	−0.08	−0.26	−0.40	−0.01	0.12
	(0.10)	(0.25)	(0.25)	(0.08)	(0.09)
Self-employed	0.15	0.55	0.59	0.34*	0.17
	(0.20)	(0.55)	(0.41)	(0.18)	(0.22)
Owner with a mortgage	−0.08	0.47	2.12***	0.10	0.02
	(0.22)	(0.49)	(0.74)	(0.15)	(0.20)
Rental home	0.04	−0.11	2.02**	−0.21	0.37
	(0.21)	(0.50)	(0.75)	(0.16)	(0.27)
Other housing regimes	0.30	−0.76	1.74	−0.09	0.24
	(0.24)	(0.85)	(1.04)	(0.18)	(0.20)
Holding of financial assets	−0.36**	−0.07	−0.29	−0.23*	−0.22
	(0.17)	(0.42)	(0.41)	(0.13)	(0.25)
DSR (total debt)	1.08***	0.38	0.89	0.50***	0.59**
	(0.31)	(0.47)	(0.58)	(0.20)	(0.22)
Selection correction	−0.60	0.27	−0.53	0.75	−0.76
	(0.46)	(1.19)	(1.16)	(0.89)	(2.28)
Observations	4358	4358	4358	4358	4358

Notes: Constant not reported. DSR (total debt): Total debt service to income ratio, winsorized 1% in each tail. Selection correction: Correction for positive consumer debt. $Sigma^2$:. Rho_1:. Bootstrapped standard errors computed with 1000 replicates, in parentheses. Level of significance: *10%, **5%, ***1%. Labels for the consumer credit user types: I = credit constrained type; II = entrepreneur type; III = transactional type; IV = type with motive purchase of household goods; V = pensioner type; VI = financial type.

heterogeneous coefficient estimates across credit user types. Precisely, while the number of retail stores per inhabitant has statistically significant negative estimates for types II and III, it shows a significant and positive coefficient estimate in the case of type IV.[22]

Second, regarding the other exclusion restrictions that appear in the first stage, but not in the second one, Table 4 shows that, with one exception, using internet banking; being affiliated to the pension system and having at least one voluntary insurance statistically and significantly increase the probability of households holding consumer debt. Oppositely, having a current account and real estate assets tend to exhibit statistically significant negative coefficient estimates, with one exception. Lastly, while the estimated coefficients for pension savings are significant only for types I and IV, total regional income exhibits unstable estimates. The latter may be due to the correlation between total regional income and the geographical controls (Centre, Metropolitan and South areas).

In particular, to interpret the significantly negative coefficient estimates for real estate assets, it is worth considering some features of the data: In Chile, the proportion of households holding real assets represents three times the fraction of households having financial assets (0.79 and 0.26, respectively; Survey of Household Finances, 2015a). Moreover, more than three-quarters of the households holding real assets (0.78) have real estate assets. Therefore, knowing that the majority of household wealth in Chile is real estate, the results in Table 4 are indicating that the wealthier the household is, the less likely it is to hold consumer debt. Interestingly, Ruiz-Tagle and Vella (2016) find similar results for real estate assets.

Third, concerning the other control variables appearing in the first stage of Table 4, we find that age, age squared, spouse present, the number of persons in the household, the number of working persons in the household and self-employment exhibit statistically significant coefficient estimates. The same is true for the geographical controls, confirming that spatial heterogeneity matters to explain households' access to consumer debt. Summing up, our identifying variables appear to collectively explain sufficient variation

[22] The identification strategy also requires sufficient variability of the variables acting as primary exclusion restrictions; in particular, those related to access to credit. Table A10, in the appendix, supports this requirement. In particular, the table shows that the number of banks per municipality ranges from 1 to 21, whereas the number of retail stores, at the same municipality level, from 0 to 24.

of the probability of observing consumer debt, thus suggesting that the first stage model specification is identified.

If we now examine the credit user-specific model estimates of the second stage, there are two conclusions to highlight. On the one hand, when correcting for self-selection, we find that the estimate for *atanh* ρ is never statistically significant. We thus conclude that there does not seem to exist a sample selection problem, implying that households' self-assessed debt burden takes into account the overall financial situation of the households, being independent of households' access to a certain debt type.

On the other hand, Table 5 confirms the fundamental results obtained in Tables 3 and A8, namely, that there is heterogeneity across consumer credit user types in the significant variables affecting the (conditional) probability of households perceiving themselves as highly or excessively indebted. Therefore, since there does not seem to exist a sample selection problem and given that the fundamental results from Table 3 remain unchanged when correcting for self-selection, we keep our model specification of Table 3 as the baseline model specification.

Regarding the second robustness check, the model estimates in Table A11, in the appendix, confirm the main conclusions we draw from Table 3: The socio-economic, demographic and regional variables we consider have sufficient explanatory power to model the probability of households being highly or excessively indebted. Furthermore, Table A11 shows that bank debt holdings exhibit statistically significant positive effects on households' probability of being highly or excessively indebted, in the case of type III, IV and V. The retail debt holding only has a statistically significant positive marginal effect on type I probability.

Interestingly, the statistically significant marginal effects are reflecting two different situations: On the one hand, there is the credit user type I, which appears to be credit constrained or at least to be facing some financial difficulties. For this type, having a retail loan makes her more likely to perceive herself as being highly or excessively indebted. On the other hand, there are types III, IV and V. In particular, types IV and V are the ones exhibiting the smallest DSR and the lowest DIR, respectively. For them, the intuition would be that because their indebtedness is relatively low (relative to the other credit user types), an increase in their bank debt holdings augments the probability that they perceive themselves as highly or excessively indebted.

Concerning robustness check three, we find that the results do not change when considering the alternative covariates, namely, home value and household wealth, liquid asset holdings, the dummy variables for the educational attainment of the household head and the dummy variables for the use of bank or retail credit card to pay in installments at zero interest.

Finally, in relation to robustness checks four and five, the analyses in appendices A.3 and A.4 show that while both alternative models (that is, the multivariate probit model and the ordered probit models) and our baseline probit models exhibit similar overall predictive performance, our baseline model is much better at correctly identifying those households who are financially vulnerable. This is because in both robustness checks, type II errors (false negative rate) in our baseline probit model estimates are considerably smaller than in the two alternative model estimates, with one exception, while the sensitivity rates (which measures how often the model predicts a household as financially vulnerable, when she actually reports highly or excessively indebtedness) tend to be larger in the case of our probit model estimates, relative to the two alternative model estimates. Therefore, since the aim of the paper is to identify and measure the importance of FVH, we prefer our baseline specification. Appendices A.3 and A.4 contain the details of the two robustness checks.

Summing up, because the estimates in Table A11 and the findings from the third robustness check continue to confirm the conclusions we draw from Table 3, together with the result that our baseline model is the best in terms of its predictive performance of FVH (relative to the multivariate probit model and the ordered probit models), we keep the binary probit models, by type, as the baseline. In the next section, we use the model estimates reported in Table 3 to determine thresholds of financial vulnerability, from which we measure the importance of financially vulnerable households' consumer indebtedness.

4.3. Measuring the importance of financially vulnerable households

In this section, we conduct several analyses: i) We quantify the importance of financially vulnerable households' indebtedness on the population of households with consumer debt and on the total consumer debt; ii) we compare the findings in i) (assuming type-specific thresholds of FV) to the ones obtained supposing a homogenous threshold; iii) we compare our measure to identify FVH based on self-assessed debt burden with alternative financial measures used in the literature, relying on the DSR and the prevalence of payment arrears or default, among others.

Specifically, Fig. 2 starts by illustrating the way we determine the thresholds of financial vulnerability and identify the FVH. Each subfigure depicts, by consumer credit user type, the mean proportion of households reporting themselves as being highly or excessively indebted, together with the predicted probability, against the level of DSR (total debt). Table 6, in turn, reports the DSR (total debt) at the type-specific thresholds of FV, together with three measures of the importance of the FVH, namely, the share that FVH represent on each group population, on the total population of households holding consumer debt and finally, on total consumer debt.

To begin with, Fig. 2 shows, as expected, that the probability of a household being highly or excessively indebted increases with the level of DSR (total debt). Intuitively, the largest the fraction of the household income that is spent to pay the debt service, the more likely it is that the household perceives herself as highly or excessively indebted.

Second, Fig. 2 and Table 6 show that the smallest DSR (total debt) at the threshold of FV corresponds to type I. No matter the measure we consider, type I is also the type where the importance of financially vulnerable households is the largest. In particular, the consumer debt of financially vulnerable households being part of type I accounts for 22% of total consumer debt. Furthermore, this is consistent with our previous characterization of households belonging to type I and it confirms our label of the credit constrained user type.

Third, when examining type II, we find that while financially vulnerable households represent almost 30% of the group

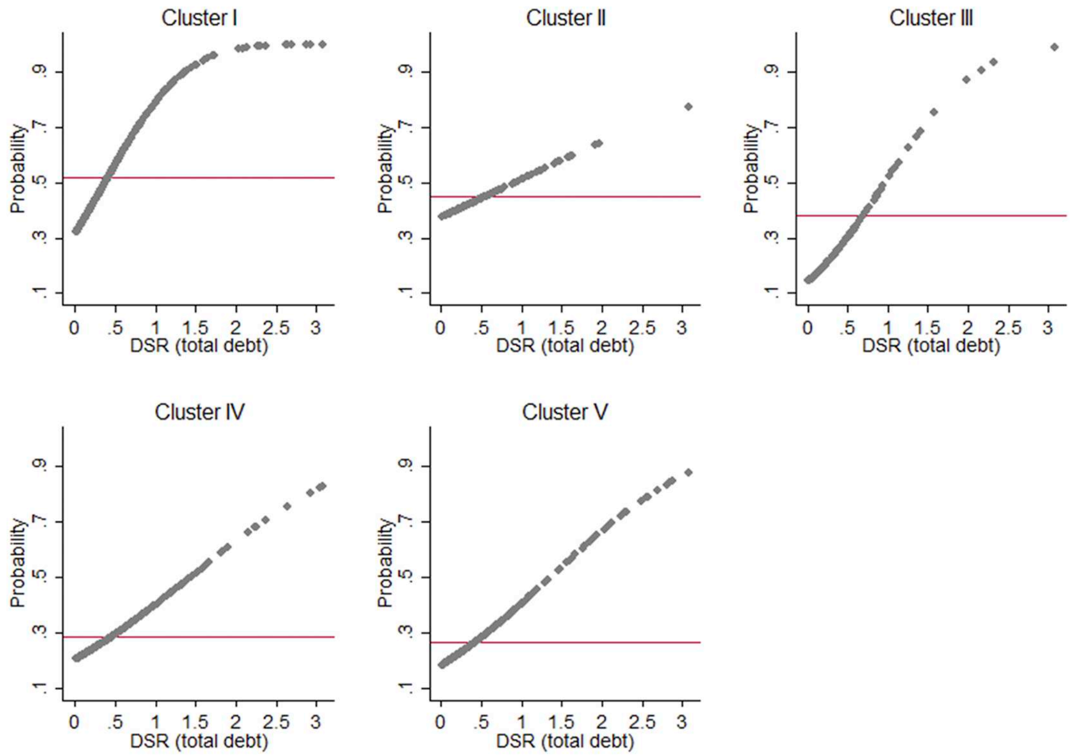

Fig. 2. Identification of FVH, by type of consumer credit user.
Source: 2014 EFH data. Notes. DSR (total debt): Total debt service to income ratio, winsorized 1% in each tail. Red line corresponds to the credit user-specific proportion of households reporting themselves as being highly or excessively indebted. Grey curve depicts the predicted probability of being highly or excessively indebted. Labels for the consumer credit user types: I = credit constrained type; II = entrepreneur type; III = transactional type; IV = type with motive purchase of household goods; V = pensioner type.

Table 6
DSR (total debt), population of FVH on group population, on total population with consumer debt and on total consumer debt, by consumer credit user type.

Group Id	Obs.	DSR (Total debt)	Population of FVH on the group population (proportion)	Population of FVH on population with consumer debt (proportion)	FVH consumer debt over total consumer debt (proportion)
I	505	0.42	0.39	0.07	0.22
II	168	0.54	0.29	0.01	0.07
III	188	0.69	0.15	0.01	0.05
IV	1131	0.48	0.16	0.07	0.14
V	757	0.42	0.22	0.06	0.05

Source: 2014 EFH data.
Notes. DSR (total debt): Total debt service to income ratio, winsorized 1% in each tail. FVH stands for financially vulnerable households'. Proportions incorporate population weights. Labels for the consumer credit user types: I = credit constrained type; II = entrepreneur type; III = transactional type; IV = type with motive purchase of household goods; V = pensioner type.

population (the second largest proportion), they account for only 1% of the total population with consumer debt and for 7% of the total consumer debt. The reason for this apparent contradiction is that although type II registers the second largest proportion of households perceiving themselves as highly or excessively indebted, the number of observations belonging to this type is small.

Fourth, regarding type III, Fig. 2 and Table 6 indicate that financially vulnerable households represent a small fraction of the group population (0.15) and of the total population with consumer debt (0.01). On top of that, the importance of financially vulnerable households on total consumer debt is the smallest (0.05), whereas the level of DSR (total debt) at the type-specific threshold of FV is the largest. Therefore, the low importance of type III financially vulnerable households reinforces the previous conclusions we extract for this type, namely, that it is a credit user type that uses consumer debt, because it values liquidity.

Fifth, although type IV has not shown signs of over-indebtedness (moderate DSR and DIR, together with a low proportion of households experiencing payment arrears or default, from Table A5), the fact that it is the largest group in terms of observations, explains why it is the second largest group in respect of the importance of financially vulnerable households on the total population with consumer debt and on the total consumer debt. In contrast, in terms of the group population, FVH exhibit the second lowest fraction.

We can obtain a similar conclusion for type V: While it is the type with the third smallest proportion of financially vulnerable households, the number of observations belonging to this type is the second largest. The latter explains why FVH belonging to type V account for 5% of the total population with consumer debt.

Summing up, allowing for heterogeneous thresholds of financial vulnerability results in the financially vulnerable households representing 22% of the population of households with consumer debt. Instead, if we were to compute a single threshold of FV, for example, at the overall mean proportion of households reporting themselves as being highly or excessively indebted, the share of FVH on the total population with consumer debt would be 30%. In other words, if using the second metric, we would be considering too many households as financially vulnerable and this way, we would be overestimating the importance of FVH and the risk they could entail on the financial system.[23]

4.3.1. Comparison between alternative measures to identify FVH

While there is no universal agreement on which indicator best identifies financially vulnerable households and over-indebtedness, studies have tended to converge on a common set of indicators to identify FV, three of them relying on financial metrics and one of them on self-assessment. The aim of this section is to provide an additional robustness check and compare our measure of FVH based on self-assessed debt burden with these three alternative measures used in the literature.

Specifically, we consider the following alternative criteria: i) Households spending more than 50% of their monthly income on total borrowing repayments, that is, DSR (total debt) >50% (D'Alessio and Iezzi, 2013); ii) households having four or more credit commitments (Kempson, 2002); iii) at least one episode of payment arrears or default in any debt category, over the last 12 months (Betti et al., 2007).

Table 7 first examines the overlap in the proportion of households identified as being financially vulnerable, according to the four criteria. Precisely, each row in Table 7 reports the proportion of households who are identified as financially vulnerable both by the row and the column indicator, relative to the total number of financially vulnerable households identified by the row measure. In turn, Table 8 exhibits the mean DSR (total debt), the average total number of loans and the mean prevalence of payment arrears or default, distinguishing between the identified financially vulnerable households and those who are not according to our methodology, and between consumer credit user types. In addition, Table 8 reports the same mean values computed over the whole population of households.

When looking at Table 7 by column, we find that our measure based on self-assessed debt burden shows the largest overlap with the other indicators. This is because it exhibits the three largest proportions of vulnerable households identified by a different measure who are also FVH according to our methodology, with this proportions ranging between 0.53 and 0.73. We conclude that our measure encompasses much of the information captured in the other indicators, thus explaining why it has the greatest overlap.

Furthermore, the evidence in Table 7 shows that the measure based on payment arrears or default is the criteria with the smallest overlap with the remaining measures. This is to be expected as this indicator considers a household as financially vulnerable when she is already in arrears or default, not accounting for households in a weak budgetary situation and thus being vulnerable to budget shocks (D'Alessio and Iezzi, 2013), but who still manage to pay their debt duties in time.

Table 8 shows, as expected, that FVH (according to our self-assessed debt burden measure) are also more likely to exhibit large debt service to income ratios, to hold a larger number of loans and to experience payment arrears or default, relative to those who are not financially vulnerable. The only exception is the measure total number of loans for type II, as there is no significant difference between the two subgroups, FVH =1 and FVH =0. The way we read this finding is that, regardless of being financially vulnerable or not, the entrepreneur credit user type is facing some borrowing constraints to finance her business activities. Recall that type II was using both short-term and long-term consumer loans to finance business activities and that it was one of the two groups with the largest fraction of households having at least one credit denied (Tables 1, A4 and A6).

On top that, the evidence in Table 8 shows a considerable disparity across consumer credit user types, in the dimensions reported in the table. The latter thus reinforces the relevance of having a methodology that allows for heterogeneity. As an illustration of the latter, 40% of the households that have been identified as financially vulnerable and belong to type I (the credit constrained credit user type) have experienced some payment arrears or default over the previous 12 months, well above the 25% that households with payment arrears or default represent in the subpopulation of households not being financially vulnerable.

Wrapping up, by comparing our measure to identify FVH, with alternative measures used in the literature, we show that our measure is not only consistent with these alternative criteria but, on top of that, it encompasses much of the information captured in the other indicators. We thus conclude that relying on households' self assessed debt burden permits a more comprehensive assessment of households' financial conditions, which is crucial to correctly measure the risks that FVH could entail to the financial system. The evidence in this section also reinforces the importance of allowing for households' heterogeneities.

5. Conclusions and policy implications

In this paper, we propose a methodology to identify and measure Chilean households' financial vulnerabilities. To do so, we start with the hypothesis that accounting for the motives and contexts in which households go into debt consumer contracts, results in

[23] The levels of DSR (total debt) at the homogeneous threshold of FV for each credit user type are available upon request.

Table 7
Comparison between alternative measures to identify FVH.

	FVH =1	DSR (total debt) >50%	Total loans>4	Default =1 (total debt)
FVH =1	–	0.36	0.33	0.27
DSR (total debt)>50%	0.73	–	0.52	0.32
Total loans>4	0.53	0.41	–	0.27
Arrears or def total debt	0.61	0.37	0.39	–

Source: 2014 EFH data.
Notes. Each row in the table reports over the total number of vulnerable households identified by the row indicator, the proportion of households who are also vulnerable according to the indicator in the column. FVH =1 stands for financially vulnerable households, according to the subjective measure. DSR (total debt): Total debt service to income ratio, winsorized 1% in each tail. Default =1 identifies financially vulnerable households as those households with at least one episode of payment arrears or default in any debt category, over the last 12 months. Proportions incorporate population weights.

Table 8
Mean DSR (total debt), total number of loans and prevalence of payment arrears or default, distinguishing between FVH =1 and FVH =0 and between consumer credit user types.

DSR (total debt)		Total loans	Default rate
Whole population			
All	0.25	2.62	0.13
FVH =1			
I	0.75	5.33	0.40
II	0.74	4.75	0.36
III	0.57	5.67	0.31
IV	0.45	3.79	0.26
V	0.48	2.83	0.17
FVH =0			
I	0.27	4.05	0.25
II	0.42	4.50	0.22
III	0.35	4.70	0.09
IV	0.19	2.84	0.11
V	0.20	2.65	0.04

Source: 2014 EFH data.
Notes. DSR (total debt): Total debt service to income ratio, winsorized 1% in each tail. Default rate corresponds to the proportion of households with at least one episode of payment arrears or default in any debt category, over the last 12 months. FVH =1 stands for financially vulnerable households, according to our subjective measure. FVH =0 corresponds to the subpopulation of households not being identified as financially vulnerable. Proportions incorporate population weights. Labels for the consumer credit user types: I = credit constrained type; II = entrepreneur type; III = transactional type; IV = type with motive purchase of household goods; V = pensioner type.

meaningfully different types of consumer credit users. By making use of cluster analysis techniques, we confirm this hypothesis.

The cluster analysis delivers six distinct types of consumer credit users: Those who have assets but value keeping precautionary liquid balances, those who finance the purchase of durable goods with consumer loans, those who use consumer debt to finance business activities or make financial investments, those who are exposed to a negative shock (for instance, those who take a consumer loan to pay medical treatments) and, finally, those who are re-financing previous debts.

Second, we show that there are different determinants driving the probability of households being highly or excessively indebted, with the level and significance of the estimated parameters considerably varying across types. Finally, by setting heterogeneous thresholds of FV, we find the financially vulnerable households represent 22% of the population of households with consumer debt.

To sum up, this paper contributes to the literature by proposing a way to better quantify the financial risks that financially vulnerable households could entail on the financial system. One way to see it is that if we were to compute a single threshold of FV, as it is typically done in the literature, the share of FVH on the total population with consumer debt would be 30%. In other words, if using the second metric, we would be considering too many households as financially vulnerable and this way, we would be overestimating the importance of FVH and the risk they could entail on the financial system.

We envision two possible applications for our methodology. First, accounting for contextual factors to classify households and identify different consumer credit users, should be useful for financial institutions, to improve their credit assessment models, and also for regulators and policy makers with a mandate on financial stability. In particular, policy makers could use these techniques to identify FVH, monitor their evolution through time and eventually, tailor specific policies to them. A second application, which could be of interest to regulators, is to use our approach as a stress test and simulate the impact of a policy change or a negative shock to the economy on the predicted probabilities of households being highly or excessively indebted. One could then use these simulated probabilities due to the shock, to compute the change in the share of the FVH on the household debt market.

Acknowledgements

We would like to thank Pedro Roje, for excellent research assistance; Michael Haliassos, Markus Kirchner, Felipe Martinez, Juan Francisco Martinez, Carlos Medel, Carola Moreno, Claudio Raddatz, Marcela Svarc and Walter Sosa Escudero for very interesting discussions and the participants of seminars at Central Bank of Chile for their very useful comments. Remaining errors are naturally ours.

Appendix A. Appendix

A.1. The SHF: Methodology and questionnaire

The SHF, conducted by the Central Bank of Chile, is the first survey that provides a comprehensive overview of households' balance sheets in Chile. The survey has a rotating panel structure, where each sample comes from a probabilistic two-stage sampling design. In order to better capture the behavior of households with the highest participation in the financial markets and with the largest share of total assets, the SHF oversampled the wealthiest 20% of households in the population. The identification of the richest households is based on the assessed value (by the Chilean Internal Tax Service) of the property they live in, regardless of being renters or owners (Survey of Household Finances, 2015b). This type of sample design is also used in the Survey of Consumer Finances or SCF, United States (Kennickell and Woodburn, 1999), and in the European Household Finance and Consumption Survey or HFCS (European Central Bank, 2013).

The SHF is answered by one respondent, who responds questions about the whole household or when corresponding, about all members of the household. The respondent is the head of household (in 67% of the cases) or the person who has the greatest knowledge of the financial situation of the household.

Regarding the structure of the SHF questionnaire, the survey begins collecting the basic demographics of the households regarding household composition, age and education of the household members (Modules A and B). Then it asks about the employment situation of household members (Module C). In turn, Module D asks about the use of means of payments, whereas Module E questions about real assets, specifically housing, and mortgages. Module F asks about non-mortgage debts held by all household members. Module G then asks about self-assessed debt burden and credit constraints. The self-assessed debt burden is the first question of Module G. The survey continues with questions on vehicles and other real assets (Module H); financial assets, pensions and insurances (Module H); income; expectations of future income and other incomes (Modules J, K and L, respectively).

Finally, it is worth mentioning that regarding missing data, the principal imputation method the SHF uses is the chained equations. Furthermore, the SHF computes and provides to the survey users, 30 imputed datasets.

A.2. Tables

A.2.1. Identifying types of consumer credit users

Table A1
Response rates of the variables we use.

	Response rate (proportion)	Included in the cluster?
Motives for holding consumer debt	1.00	Yes
Self-assessed debt burden	1.00	No
Shocks and access to credit		Yes
Unexpected negative shocks	1.00	
Did the household ask for credit?	1.00	
Households with 1 + credits denied	1.00	
Households with 1 + credits partially granted	1.00	
Debt holdings		No
Bank credit card	1.00	
Bank credit line	1.00	
Retail credit card	1.00	
Installment loans of retail stores	1.00	
Bank loan	1.00	
Union loan	1.00	
Auto loan	1.00	
Educational loan	1.00	
Mortgage	1.00	
Other debts	1.00	
Bank loan holding	1.00	
Retail debt holding	1.00	
Socio-demo and geog. variables		
Number of persons in the household	1.00	Yes
Age of the head of the household (years)	1.00	Yes
Years of education	1.00	No
Civil status	1.00	No

(*continued on next page*)

Table A1 (*continued*)

	Response rate (proportion)	Included in the cluster?
Male = 1	1.00	No
Spouse present	1.00	No
Geographic controls	1.00	No
Eco. and fin. variables		
Income (thousands)	0.86	No
Income strata	1.00	Yes
Liquid asset holdings	0.94	Yes
DSR (total debt)	0.78	No
Number of employed persons/house	1.00	No
Number of unemployed persons/house	1.00	No
Self-employed	1.00	No
Home ownership	1.00	No
Holding of financial assets	1.00	No
Signed contract	1.00	No
Current account	1.00	No
Internet banking	1.00	No
Pension system	1.00	No
Insurance holding	1.00	No
Pension savings	0.97	No
Real estate assets (std)	0.97	No
Total regional income (std)	1.00	No
Number of inhabitants per bank	1.00	No
Number of retail stores	1.00	No

Source: 2014 EFH data. Socio-demo. and geog. variables stands for socio-demographic and geographic variables. Eco. and fin. variables stands for economic and financial variables.

Table A2
Chi-square test results for each discrete variables we use in the cluster analysis.

	Chi-sq stat	DF	P-value
Motives for holding short-term consumer debt[a]			
Purchase of household articles	67.63	5	0.000
Purchase of cars	46.48	5	0.000
Vacations	2535.53	5	0.000
Medical treatment	201.99	5	0.000
Pay other debts	2347.41	5	0.000
Finance business activities	2759.74	5	0.000
Purchase of financial assets	2809.00	5	0.000
Renovations	20.68	5	0.001
Shocks and access to credit[a]			
Unexpected negative shocks	35.27	5	0.000
Did the household ask for credit?	115.37	5	0.000
Households with 1 or more credit denied	42.68	5	0.062
Households with 1 or more credit partially granted	30.06	5	0.012
Socio-economic variables			
Income strata[b]	439.76	5	0.000
Liquid assets holdings (proportion)	77.88	5	0.000

Source: 2014 EFH data.
Notes: Chi-sq stat stands for chi-square test statistic; DF stands for degrees of freedom and P-value, for probability value.
[a] The variable(s) is(are) expressed as proportion(s), corresponding to the fraction of households with a certain characteristic over the total number of households in a group.
[b] We classify household total income into three strata.

Table A3
Use of bank and/or retail credit cards.

Group id	I	II	III	IV	V	VI	All
To pay without installments	0.06	0.15	0.16	0.14	0.11	0.18	0.12
To pay in installments without interest	0.39	0.39	0.41	0.35	0.32	0.23	0.35
To pay in installments with interest	0.38	0.34	0.32	0.38	0.41	0.38	0.38
Missing values	0.17	0.12	0.11	0.14	0.16	0.23	0.15

Source: 2014 EFH data.
Notes. Underlined numbers indicate the row-minimum values. Numbers in boxes refer to the row-maximum values. Labels for the consumer credit user types: I = credit constrained type; II = entrepreneur type; III = transactional type; IV = type with motive purchase of household goods; V = pensioner type; VI = financial type.

Table A4
Mean values of various debt holdings and wealth, by type of consumer credit user.

Group Id	I	II	III	IV	V	VI	All
Debt holdings[a]							
Mortgage	0.29	0.28	0.40	0.25	0.09	0.24	0.23
Bank credit card	0.44	0.60	0.70	0.32	0.18	0.65	0.35
Bank credit line	0.20	0.37	0.31	0.09	0.05	0.29	0.13
Retail credit card	0.65	0.63	0.60	0.75	0.76	0.57	0.71
Installment loans of retail stores	0.45	0.52	0.36	0.20	0.12	0.55	0.26
Bank loans	0.22	0.10	0.13	0.06	0.09	0.16	0.11
Union loans	0.34	0.10	0.22	0.10	0.16	0.12	0.17
Auto loans	0.06	0.08	0.09	0.06	0.02	0.10	0.05
Educational loans	0.14	0.10	0.12	0.12	0.05	0.10	0.10
Other debts[b]	0.13	0.05	0.05	0.05	0.05	0.10	0.07
Total consumer debt	0.28	0.17	0.14	0.28	0.10	0.04	0.21
Wealth (thousands)							
Wealth	42,724,382	125,377,377	128, 192, 493	51,431,442	56,906,350	84,023,430	61,471,644
Observations	507	168	188	1136	761	49	2809
Population represented	578,826	132,872	151,005	1,253,352	819,349	43,545	2,978,949

Source: 2014 EFH data.
Notes: External variables reported, that is, variables which are not in the cluster analysis. Underlined numbers indicate the row-minimum values. Numbers in boxes refer to the row-maximum values. Labels for the consumer credit user types: I = credit constrained type; II = entrepreneur type; III = transactional type; IV = type with motive purchase of household goods; V = pensioner type; VI = financial type.

[a] The variable(s) is(are) expressed as proportion(s), corresponding to the fraction of households with a certain characteristic over the total number of households in a group.

[b] Other debts include loans from relatives and friends, from informal lenders and from warrantable house.

Table A5
Mean values of various DSR and DIR measures and the proportion of households with payment arrears or default, by type of consumer credit user.

Group Id	I	II	III	IV	V	VI	All
DSR (consumer debt) mean	0.44	0.53	0.37	0.30	0.36	0.42	0.36
DSR (consumer debt) p75	0.49	0.65	0.47	0.31	0.35	0.60	0.39
DSR (consumer debt) p90	0.91	1.12	0.76	0.61	0.67	1.00	0.74
DSR (total debt) mean	0.49	0.57	0.42	0.35	0.38	0.46	0.40
DSR (total debt) p75	0.55	0.73	0.52	0.38	0.38	0.67	0.44
DSR (total debt) p90	1.04	1.24	0.87	0.68	0.74	1.04	0.83
DIR (consumer debt) mean	0.40	0.39	0.24	0.18	0.21	0.39	0.25
DIR (consumer debt) p75	0.43	0.44	0.26	0.17	0.15	0.47	0.24
DIR (consumer debt) p90	0.89	0.84	0.56	0.44	0.41	1.24	0.55
DIR (total debt) mean	0.87	0.77	0.74	0.66	0.41	0.79	0.64
DIR (total debt) p75	1.07	0.78	1.09	0.59	0.22	1.06	0.60
DIR (total debt) p90	2.45	1.87	2.08	1.84	0.86	1.74	1.79
Payment arrears or default on mortgage	0.02	0.01	0.01	0.01	0.00	0.02	0.01
Payment arrears or default on bank credit card	0.02	0.02	0.01	0.01	0.01	0.04	0.01
Payment arrears or default on retail credit card	0.10	0.04	0.03	0.06	0.05	0.06	0.06

(*continued on next page*)

Table A5 (*continued*)

Group Id	I	II	III	IV	V	VI	All
Payment arrears or default on installment loans	$\boxed{0.05}$	$\underline{0.01}$	0.02	$\underline{0.01}$	$\underline{0.01}$	0.02	0.01
Payment arrears or default on bank loans	$\boxed{0.07}$	$\boxed{0.07}$	0.02	0.02	0.01	$\underline{0.00}$	0.03
Payment arrears or default on union loans	$\boxed{0.04}$	0.02	0.01	0.01	$\underline{0.00}$	0.02	0.01
Payment arrears or default on auto loans	$\boxed{0.01}$	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	0.00
Payment arrears or default on educational loans	0.01	0.01	$\boxed{0.02}$	0.01	$\underline{0.00}$	$\underline{0.00}$	0.01
Observations	507	168	188	1136	761	49	2809
Population represented	578,826	132,872	151,005	1,253,352	819,349	43,545	2,978,949

Source: 2014 EFH data.

Notes. Underlined numbers indicate the row-minimum values. Numbers in boxes refer to the row-maximum values. DSR (consumer debt): Consumer debt service to income ratio. DSR (total debt): Total debt service to income ratio. DIR (consumer debt): Consumer debt to income ratio. DIR (total debt): Total debt to income ratio. Payment arrears or default: For credit cards, payment below the minimum or no payment, over the last 12 month. For the remaining debt types reported in the table, payment arrears of at least 3 month or no payment, over the last 12 months. For mortgages, it refers to the principal residence. It is not possible to compute the payment arrears or default for bank credit lines and other debts, since the SHF does not ask for this information. Labels for the consumer credit user types: I = credit constrained type; II = entrepreneur type; III = transactional type; IV = type with motive purchase of household goods; V = pensioner type; VI = financial type.

Table A6
Motives for holding consumer debt, by type of consumer credit user.

Type of debt	Motives for holding consumer debt	I	II	III	IV	V	VI	All
Bank credit card	Purchase of household articles	0.23	$\underline{0.22}$	0.26	$\boxed{0.54}$	0.48	0.25	0.38
	Purchase of cars	0.03	0.04	$\underline{0.02}$	0.07	0.04	$\boxed{0.13}$	0.05
	Vacations	$\underline{0.00}$	0.05	$\boxed{0.70}$	$\underline{0.00}$	$\underline{0.00}$	0.19	0.10
	Medical treatment	0.06	0.04	0.03	0.02	$\boxed{0.17}$	$\underline{0.00}$	0.05
	Pay other debts	$\boxed{0.62}$	0.16	0.05	$\underline{0.00}$	0.01	0.06	0.16
	Finance business activities	$\underline{0.00}$	$\boxed{0.62}$	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	0.06	0.07
	Purchase of financial assets	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	$\boxed{0.41}$	0.01
	Renovations	0.03	$\boxed{0.07}$	0.03	0.06	0.03	$\underline{0.00}$	0.04
Bank credit line	Purchase of household articles	0.09	0.19	0.17	$\boxed{0.39}$	0.18	$\underline{0.00}$	0.21
	Purchase of cars	0.01	0.02	0.02	$\boxed{0.08}$	$\underline{0.00}$	$\underline{0.00}$	0.03
	Vacations	$\underline{0.00}$	0.02	$\boxed{0.47}$	$\underline{0.00}$	$\underline{0.00}$	0.07	0.08
	Medical treatment	0.04	$\underline{0.00}$	0.02	0.07	$\boxed{0.43}$	$\underline{0.00}$	0.08
	Pay other debts	$\boxed{0.71}$	0.06	0.12	$\underline{0.00}$	0.03	0.29	0.23
	Finance business activities	$\underline{0.00}$	$\boxed{0.74}$	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	0.12
	Purchase of financial assets	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	$\boxed{0.50}$	0.02
	Renovations	0.04	$\underline{0.03}$	0.07	$\boxed{0.09}$	$\underline{0.03}$	0.07	0.06
Retail credit card	Purchase of household articles	0.59	0.60	0.54	$\boxed{0.68}$	0.63	$\underline{0.46}$	0.63
	Purchase of cars	0.01	$\boxed{0.04}$	$\underline{0.00}$	0.02	0.01	$\boxed{0.04}$	0.01
	Vacations	$\underline{0.00}$	0.05	$\boxed{0.22}$	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	0.01
	Medical treatment	0.06	0.03	0.03	$\underline{0.00}$	$\boxed{0.08}$	0.07	0.04
	Pay other debts	$\boxed{0.23}$	0.05	0.09	$\underline{0.00}$	0.01	0.04	0.05
	Finance business activities	$\underline{0.00}$	$\boxed{0.21}$	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	0.01
	Purchase of financial assets	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	$\boxed{0.32}$	0.00
	Renovations	0.03	$\boxed{0.04}$	$\boxed{0.04}$	0.03	0.03	$\underline{0.00}$	0.03
Installment loans of retail stores	Purchase of household articles	0.11	$\boxed{0.29}$	0.08	0.28	0.23	0.25	0.19
	Purchase of cars	0.03	$\underline{0.00}$	0.04	0.07	0.05	$\boxed{0.13}$	0.04
	Vacations	$\underline{0.00}$	$\underline{0.00}$	$\boxed{0.63}$	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	0.05
	Medical treatment	0.04	0.12	0.04	0.03	$\boxed{0.33}$	$\underline{0.00}$	0.10
	Pay other debts	$\boxed{0.72}$	0.18	0.17	$\underline{0.00}$	0.05	$\underline{0.00}$	0.31
	Finance business activities	$\underline{0.00}$	$\boxed{0.47}$	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	0.13	0.03
	Purchase of financial assets	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	$\boxed{0.63}$	0.02
	Renovations	0.08	$\underline{0.06}$	0.13	$\boxed{0.22}$	$\underline{0.06}$	0.13	0.11
Bank loans	Purchase of household articles	0.03	$\underline{0.01}$	0.06	$\boxed{0.13}$	0.11	0.07	0.07
	Purchase of cars	0.10	$\underline{0.03}$	0.21	$\boxed{0.30}$	0.07	0.04	0.16
	Vacations	$\underline{0.00}$	0.01	$\boxed{0.32}$	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	0.03
	Medical treatment	0.04	$\underline{0.00}$	0.04	0.02	$\boxed{0.35}$	0.04	0.07
	Pay other debts	$\boxed{0.72}$	0.09	0.26	$\underline{0.00}$	0.03	0.11	0.27
	Finance business activities	$\underline{0.00}$	$\boxed{0.84}$	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	0.04	0.10
	Purchase of financial assets	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	$\boxed{0.74}$	0.03
	Renovations	0.09	$\underline{0.03}$	0.06	$\boxed{0.27}$	0.10	0.07	0.14

(*continued on next page*)

Table A6 (*continued*)

Type of debt	Motives for holding consumer debt	I	II	III	IV	V	VI	All
Union loans	Purchase of household articles	0.05	0.13	$\underline{0.00}$	0.14	$\boxed{0.20}$	$\underline{0.00}$	0.11
	Purchase of cars	0.04	$\underline{0.00}$	0.02	0.15	0.02	$\boxed{0.17}$	0.06
	Vacations	$\underline{0.00}$	0.13	$\boxed{0.80}$	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	0.07
	Medical treatment	0.07	0.06	$\underline{0.05}$	$\underline{0.05}$	$\boxed{0.35}$	0.17	0.14
	Pay other debts	$\boxed{0.71}$	0.13	0.12	$\underline{0.00}$	0.04	$\underline{0.00}$	0.29
	Finance business activities	$\underline{0.00}$	$\boxed{0.50}$	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	0.02
	Purchase of financial assets	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	$\underline{0.00}$	$\boxed{0.17}$	0.00
	Renovations	0.06	0.06	$\underline{0.02}$	$\boxed{0.31}$	0.11	0.17	0.13

Source: 2014 EFH data.
Notes. Underlined numbers indicate the row-minimum values. Numbers in boxes refer to the row-maximum values. Labels for the consumer credit user types: I = credit constrained type; II = entrepreneur type; III = transactional type; IV = type with motive purchase of household goods; V = pensioner type; VI = financial type.

Table A7
Debt maturities (months), by type of consumer credit user

Group Id	I	II	III	IV	V	VI	All
Mortgage[a]	154.91	$\boxed{172.80}$	147.33	160.18	$\underline{131.93}$	160.8	157.70
Bank credit card	$\boxed{8.60}$	6.29	5.50	4.79	$\underline{4.66}$	6.73	5.94
Bank credit card	$\boxed{8.60}$	6.29	5.50	4.79	$\underline{4.66}$	6.73	5.94
Bank credit line	$\boxed{8.73}$	7.21	4.56	4.57	3.73	$\underline{3.44}$	5.96
Retail credit card	$\boxed{6.01}$	4.49	4.44	3.85	4.15	$\underline{3.79}$	4.36
Installment loans of retail stores	11.44	$\underline{9.45}$	$\boxed{12.92}$	10.47	10.04	10.94	10.90
Bank loans	22.59	24.70	20.87	$\underline{20.75}$	21.19	$\boxed{26.32}$	22.07
Union loans	16.96	22.99	14.98	$\underline{14.85}$	17.40	$\boxed{29.21}$	16.76
Auto loans	20.26	24.58	$\boxed{25.90}$	22.03	21.01	$\underline{16.72}$	22.04
Educational loans	67.41	$\underline{31.77}$	$\boxed{106.42}$	54.66	70.76	45.53	63.31

Source: 2014 EFH data.
Notes. Underlined numbers indicate the row-minimum values. Numbers in boxes refer to the row-maximum values. Labels for the consumer credit user types: I = credit constrained type; II = entrepreneur type; III = transactional type; IV = type with motive purchase of household goods; V = pensioner type; VI = financial type.
[a] It refers to the mortgage debt maturity of the principal residence.

A.2.2. Determinants of households' probability of being highly or excessively indebted

Table A8
Coefficients estimates of the credit user-specific probit models.

	I	II	III	IV	V
Age	0.07*	−0.02	0.03	0.07*	−0.05
	(0.04)	(0.10)	(0.11)	(0.04)	(0.05)
Age^2	−0.00*	0.00	0.00	−0.00*	0.00
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Years of education	0.02	−0.04	−0.28***	0.00	−0.03
	(0.03)	(0.07)	(0.09)	(0.02)	(0.02)
Male = 1	0.00	−0.11	−0.19	−0.38***	0.09
	(0.18)	(0.45)	(0.43)	(0.14)	(0.18)
Number of persons in the household	0.13*	0.15	0.12	0.18***	0.19**
	(0.07)	(0.16)	(0.17)	(0.05)	(0.09)
Civil status: married	−0.64***	−0.48	−0.39	−0.24	−0.09
	(0.20)	(0.46)	(0.43)	(0.14)	(0.20)
Center of Chile	−0.03	−0.31	−0.66	−0.16	0.38
	(0.24)	(0.74)	(1.19)	(0.19)	(0.30)
Metropolitan region	−0.42	−0.37	−0.56	−0.27	0.10
	(0.24)	(0.67)	(1.14)	(0.18)	(0.29)
South of Chile	0.29	−0.97	−0.42	−0.02	0.22
	(0.33)	(0.84)	(1.18)	−0.25	(0.39)
Number of employed persons in the household	−0.05	−0.33	−0.29	−0.08	0.13
	(0.11)	(0.27)	(0.27)	(0.08)	(0.11)
Self-employed	0.12	0.57	0.72*	0.49***	0.17
	(0.22)	(0.44)	(0.46)	(0.16)	(0.24)

(*continued on next page*)

Table A8 (*continued*)

	I	II	III	IV	V
Owner with a mortgage	−0.15	0.58	2.38***	0.09	0.03
	(0.26)	(0.54)	(0.68)	0.17	(0.26)
Rental home	0.00	−0.05	2.18***	−0.29	0.44*
	(0.24)	(0.56)	(0.75)	(0.18)	(0.24)
Other housing regimes	0.29	−0.72	2.13***	−0.15	0.27
	(0.26)	(0.74)	(0.94)	(0.20)	(0.23)
Holding of financial assets	−0.28	−0.15	−0.29	−0.29**	−0.28
	(0.19)	(0.47)	(0.48)	(0.14)	(0.22)
DSR (total debt)	1.29***	0.35	1.11**	0.58***	0.67***
	(0.31)	(0.53)	(0.64)	(0.19)	(0.18)
Pseudo R^2	0.17	0.14	0.41	0.12	0.09

Notes: Constant not reported. DSR (total debt): Total debt service to income ratio, winsorized 1% in each tail. Bootstrapped standard errors computed with 1000 replicates, in parentheses. Level of significance: *10%, **5%, ***1%. Labels for the consumer credit user types: I = credit constrained type; II = entrepreneur type; III = transactional type; IV = type with motive purchase of household goods; V = pensioner type; VI = financial type.

Table A9

Proportion of households and mean probability of being highly or excessively indebted, by type of consumer credit user.

Group Id	Observed	Predicted	Observations
I	0.52	0.50	505
II	0.41	0.42	168
III	0.29	0.34	188
IV	0.26	0.28	1131
V	0.23	0.24	757
Total	0.31	0.32	2749

Source: 2014 EFH data.

Notes: Mean values without population weights. Labels for the consumer credit user types: I = credit constrained type; II = entrepreneur type; III = transactional type; IV = type with motive purchase of household goods; V = pensioner type; VI = financial type.

A.2.3. Robustness checks 1 and 2

Table A10

Descriptive statistics of the variables in the selection equation – robustness check 1.

	Mean	Std. Dev	Min	Max
Spouse present	0.22	0.41	0.00	1.00
Signed contract	0.85	0.36	0.00	1.00
Current account savings	0.17	0.37	0.00	1.00
Internet banking	0.44	0.50	0.00	1.00
Pension system	0.55	0.50	0.00	1.00
Insurance holding	0.42	0.49	0.00	1.00
Pension savings (thousands)	4527.61	19,047.95	0.00	300,000.00
Real estate assets (thousands)	62,759.41	111,123.60	0.00	1,376,364.00
Total regional income (millions)	1338.31	1009.29	37.40	2336.99
Number of banks (municipality)	9.94	4.20	1.00	21.00
Number of retail stores (municipality)	6.16	4.85	0.00	24.00
Inhabitants (municipality)	217,060.30	125,939.10	5,162.00	610,118.00

Source: 2014 EFH data. Pension savings and real estate assets are in thousands of Chilean pesos. Total regional income is in miilions of Chilean pesos. Number of banks, retail stores and inhabitants are at municipality level.

Table A11
Marginal effects at means: robustness check 2.

	I	II	III	IV	V
Age	0.02*	0.00	0.02	0.02*	−0.01
	(0.01)	(0.03)	(0.02)	(0.01)	(0.01)
Age^2	0.00*	0.00	0.00	0.00**	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Years of education	0.00	−0.02	−0.07***	0.00	−0.01*
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
Male = 1	−0.01	−0.01	−0.07	−0.12***	0.02
	(0.05)	(0.12)	(0.07)	(0.04)	(0.05)
Number of persons in the household	0.05**	0.04	0.04	0.05***	0.06***
	(0.02)	(0.04)	(0.03)	(0.01)	(0.02)
Civil status: married	−0.21***	−0.17	−0.11	−0.07*	−0.03
	(0.06)	(0.12)	(0.07)	(0.04)	(0.06)
Center of Chile	−0.02	−0.03	−0.07	−0.04	0.11
	(0.08)	(0.19)	(0.21)	(0.05)	(0.08)
Metropolitan region	−0.14	−0.07	−0.07	−0.08	0.02
	(0.07)	(0.18)	(0.19)	(0.05)	(0.08)
South of Chile	0.09	−0.29	−0.05	−0.01	0.07
	(0.10)	(0.21)	(0.19)	(0.07)	(0.11)
Number of employed persons in the household	−0.02	−0.10	−0.06	−0.03	0.03
	(0.03)	(0.07)	(0.05)	(0.02)	(0.03)
Self-employed	0.02	0.15	0.15*	0.14***	0.06
	(0.07)	(0.12)	(0.07)	(0.04)	(0.07)
Owner with a mortgage	−0.06	0.22	0.49***	0.02	0.01
	(0.08)	(0.13)	(0.10)	(0.05)	(0.07)
Rental home	0.01	0.00	0.48***	−0.09	0.13**
	(0.07)	(0.15)	(0.13)	(0.05)	(0.06)
Other housing regimes	0.08	−0.24	0.42***	−0.04	0.08
	(0.08)	(0.19)	(0.15)	(0.06)	(0.06)
Holding of financial assets	−0.10*	−0.07	−0.10	−0.08**	−0.10
	(0.06)	(0.12)	(0.07)	(0.04)	(0.06)
DSR (total debt)	0.36***	0.09	0.22**	0.15***	0.15***
	(0.09)	(0.13)	(0.09)	(0.05)	(0.05)
Bank loan holding	0.10	0.27	0.29**	0.06*	0.14**
	(0.06)	(0.20)	(0.12)	(0.04)	(0.06)
Retail loan holding	0.13**	0.05	0.06	0.05	−0.02
	(0.06)	(0.13)	(0.09)	(0.05)	(0.06)
Observations	505	168	188	1131	757
Pseudo R^2	0.19	0.16	0.46	0.13	0.11

Notes: Constant not reported. DSR (total debt): Total debt service to income ratio, winsorized 1% in each tail. Bootstrapped standard errors computed with 1000 replicates, in parentheses. Level of significance: *10%, **5%, ***1%. Labels for the consumer credit user types: I = credit constrained type; II = entrepreneur type; III = transactional type; IV = type with motive purchase of household goods; V = pensioner type; VI = financial type.

A.3. Robustness check 4

Instead of estimating a probit model by consumer credit user type, as a robustness check, we estimate a multinomial probit model, allowing for correlation in the error term. We then identify a new set of financially vulnerable households relying on the multinomial probit model estimates. Finally, we compare our baseline probit model estimates with the multinomial probit model estimates, in terms of their predictive performance.

Let the dependent variable in this robustness check be S_{ik}^{MP}. Subindices i and k continue to refer to household i and consumer credit user type k. Also, we keep excluding cluster VI, since the number of observations is too small. The new dependent variable S_{ik}^{MP} can take values from 0 to 5, with 0 corresponding to household i holding consumer debt and reporting low or medium indebtedness. In turn, values 1–5 in S_{ik}^{MP} correspond to households reporting high or excessive indebtedness and belonging to clusters I to V, respectively.

To identify the FVH in this robustness check, we follow the same procedure: A household is financially vulnerable when she has a predicted probability of being highly or excessively indebted (now based on the multinomial probit model estimates) above the corresponding threshold in her credit user type. The type-specific thresholds continue to be the mean proportion of households reporting themselves as being highly or excessively indebted in each type.

In order to compare our baseline model specification with this alternative multinomial probit model, we examine them in terms of their predictive performance, by using a confusion matrix. A confusion matrix shows the number of correct and incorrect predictions (in our case, households who are identified as financially vulnerable and those who are not) made by the classification model (the procedure for identification of the FVH, based on the binary probit or the multinomial probit model estimates) compared to the actual outcomes in the data (households reporting high or excessive). We consider that a household who is identified as financially vulnerable by one classification model, but who reports low or medium indebtedness, is a classification mistake. We also consider as a mistake a household who is not identified as financially vulnerable by a given classification model, but who reports high or excessive indebtedness.

Table A12 outlines a theoretical confusion matrix. It then defines the five metrics we rely on to compare the classification performance of the two alternative model estimates. In turn, table A13 reports the five metrics for our baseline probit model and for the multinomial probit estimates, that is, type I and type II errors, sensitivity, specificity and accuracy.

Table A12
The confusion matrix.

Self-assessed debt burden = High or Excessive		
Model	0	1
FVH = 0	True Negative (TN)	False Negative (FN)
FVH = 1	False Positive (FP)	True Positive (TP)

Source: 2014 EFH data.

Type I error or false positive rate: $\frac{FP}{TN+FP}$.

Type II error or false negative rate: $\frac{FN}{TP+FN}$.

Sensitivity rate: $\frac{TP}{TP+FN}$. It measures how often the model predicts a household as financially vulnerable, when she is actually highly or excessively indebted.

Specificity rate: $\frac{TN}{TN+FN}$. It measures how often the model predicts a household as not being financially vulnerable, when she perceives herself as low or moderately indebted.

Accuracy: $\frac{TN+TP}{TN+TP+FN+FP}$. It corresponds to the overall share of correct predictions.

Table A13
Comparison of the binary probit and the multinomial probit model estimates in terms of their predictive performance.

	Binary probit	Multinomial probit
Error I	31.7	14.9
Error II	42.9	71.1
Sensitivity	57.1	28.9
Specificity	68.3	85.1
Accuracy	64.8	67.5

Source: 2014 EFH data. Error I stands for type I error, which corresponds to the false positive rate. Error II refers to type II error, which corresponds to the false negative rate. Sensitivity rate measures how often the model predicts a household as financially vulnerable, when she actually reports highly or excessively indebted. Specificity rate measures how often the model predicts the household as not being financially vulnerable, when she does not perceive herself as highly or excessively indebted. Accuracy rate corresponds to the overall share of correct predictions.

Table A13 shows that while both models tend to exhibit similar accuracy rates (which assess the overall predictive performance of the model), our baseline model is better at correctly identifying those households who are financially vulnerable. The two ways to see this is by looking at type II error and at the sensitivity rate, with the former being considerably smaller and the latter, being larger in the case of the binary probit model estimates, relative to the multivariate probit model specification. Since the aim of the paper is to identify and measure the importance of FVH, we prefer our baseline specification.

A.4. Robustness check 5

In robustness check 5, we estimate ordered probit models, by credit user type, using the four self-assessed indebtedness categories, that is, “high”, “excessive”, “low” and “moderate”. We then compare the estimates of our baseline probit model with the ones resulting from estimating ordered probit models, in two manners: On the one hand, using information criteria, that is, the Akaike Information Criterion or AIC, and the Bayesian Information Criterion or BIC; on the other hand, by examining their predictive performance.

In the case of the ordered probit models, the classification rule to identify the FVH is somewhat different: Since the ordered probit model produces predicted probabilities for each outcome of the ordinal variable, the identification of FVH will now depend on two predicted probabilities, that is, the probability of being highly indebted and the probability of being excessively indebted.

Precisely, a household is now financially vulnerable when she has a predicted probability of being highly indebted (based on the ordered probit model estimates) above the type-specific mean proportion of households perceiving themselves as being highly indebted; and when she has a predicted probability of being excessively indebted above the type-specific mean proportion of households reporting excessive indebtedness. Importantly, the reason for having two thresholds for each credit user type is to maximize comparability with the classification rule in our baseline model.

Table A14 compares our baseline probit model estimates with the ordered probit model estimates, relying on the AIC and the BIC information criteria. The aim of this analysis is to assess the loss of information, if any, resulting from aggregating the four categories of household's self-assessed debt burden into two (low or medium and high or excessive).

Table A14
Comparison of the binary probit and ordered probit model estimates in terms of information criteria.

	Binary probit for each cluster					Ordered probit for each cluster				
	I	II	III	IV	V	I	II	III	IV	V
Obs.	505	168	188	1131	757	505	168	188	1131	757
Param.	17	17	17	17	17	19	19	19	19	19
Log-Lik.	−327,541	−78,316	−59,195	−659,065	−425,230	−642,637	−149,826	−150,372	−1,347,168	−939,875
AIC	653,649	156,666	118,424	1,291,208	850,495	1,271,724	299,689	300,782	2,694,374	1,879,787
BIC	653,721	156,719	118,479	1,291,294	850,573	1,271,804	299,749	300,843	2,694,469	1,879,875

Notes: Obs. stands for observations. Param. refers to number of parameters. Log-Lik stands for Log-likelihood. AIC and BIC refer to the Akaike Information Criterion and the Bayesian Information Criterion, respectively. Labels for the consumer credit user types: I = credit constrained type; II = entrepreneur type; III = transactional type; IV = type with motive purchase of household goods; V = pensioner type; VI = financial type.

Both the AIC and the BIC criteria penalize the number of model parameters but reward goodness of fit; therefore, the best model is the one with lowest AIC or BIC. Note that since the ordinal dependent variable has four categories, three intercepts need to be estimated in the case of the ordered probit models, instead of one as in the case of the probit models. This results in an increase in the model parameters from 17 to 19, as table A14 indicates.

Table A14 shows that, for all types, the binary probit models exhibit lower values of both the AIC and the BIC, suggesting that there is no loss of information when collapsing the categories high and excessive into a single category. The latter also suggests that households with high and excessive self-assessed debt burden do not significantly differ in observable characteristics. Supporting the latter, we conduct difference in mean tests and find that households perceiving themselves as excessively indebted are not significantly different from households perceiving themselves as highly indebted, in most of the observable characteristics included in the model specification for households' self-assessed debt burden. Indeed, we find only three exceptions, over a total of 16 difference in mean tests.

Table A15 now compares our baseline binary probit and the ordered probit model estimates in terms of their predictive performance. For the comparison, we examine the same five metrics that we consider in the previous robustness check, that is, type I and type II errors, sensitivity, specificity and accuracy.

Table A15
Comparison of the binary probit and ordered probit model estimates in terms of their predictive performance.

	Binary probit for each cluster					Ordered probit for each cluster				
	I	II	III	IV	V	I	II	III	IV	V
Error I	30.2	22.2	26.3	34.4	29.0	16.7	27.3	15.0	22.7	26.4
Error II	43.5	49.3	27.3	43.8	42.9	66.4	44.9	45.5	55.9	45.8
Sensitivity	56.5	50.7	72.7	56.2	57.1	33.6	55.1	54.5	44.1	54.2
Specificity	69.8	77.8	73.7	65.6	71.0	83.3	72.7	85.0	77.3	73.6
Accuracy	63.2	66.7	73.4	63.3	67.8	57.8	65.5	76.1	68.8	69.1

Notes: Error I stands for type I error, which corresponds to the false positive rate. Error II refers to type II error, which corresponds to the false negative rate. Sensitivity rate measures how often the model predicts a household as financially vulnerable, when she actually reports highly or excessively indebted. Specificity rate measures how often the model predicts the household as not being financially vulnerable, when she does not perceive herself as highly or excessively indebted. Accuracy rate corresponds to the overall share of correct predictions. Labels for the consumer credit user types: I = credit constrained type; II = entrepreneur type; III = transactional type; IV = type with motive purchase of household goods; V = pensioner type; VI = financial type.

Table A15 shows that while both model estimates tend to exhibit similar accuracy rates, our baseline probit model is better at correctly identifying those households who are financially vulnerable. This is because, for all types except one, type II error is smaller and the sensitivity rate is larger in our baseline probit model estimates, relative to the ordered probit model estimates.

Since the binary probit models are better at correctly identifying those households who are financially vulnerable (relative to the ordered probit models), coupled with the finding that there is no loss of information when collapsing the categories high and excessive into a single category, we continue to prefer our baseline specification.

A.5. Figure

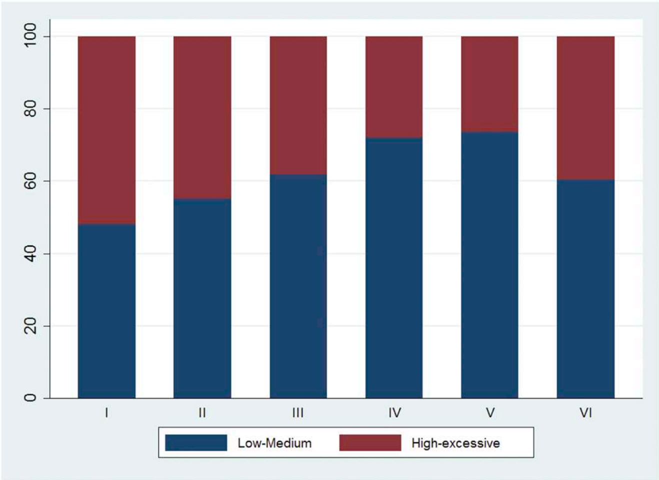

Fig. A.1. Percentage of households reporting themselves as highly or excessively indebted, by type of consumer credit user.
Source: 2014 EFH data. Variable description: We group the categories for the self-assessed debt burden low and medium together and high and excessive together. Labels for the consumer credit user types: I = credit constrained type; II = entrepreneur type; III = transactional type; IV = type with motive purchase of household goods; V = pensioner type; VI = financial type.

References

Albacete, N., Fessler, P., 2010. Stress testing Austrian households. Financ. Stab. Rep. 19, 72–91.
Anderberg, M., 1973. Measures of association among data units. In: Anderberg, M. (Ed.), Cluster Analysis for Applications, pp. 98–130 New York.
Bahovec, V., Barbic, D., Palic, I., 2015. Testing the effects of financial literacy on debt behavior of financial consumers using multivariate analysis methods. Croat. Oper. Res. Rev. 6 (2), 361–371.
Bertaut, C., Haliassos, M., 2006. Credit cards: facts and theories. In: Bertola, G., Disney, R., Grant, C. (Eds.), The Economics of Consumer Credit Demand and Supply, pp. 181–238 Cambridge, MA.
Betti, G., Dourmashkin, N., Rossi, M., Yin, Y., 2007. Consumer over-indebtedness in the EU: measurement and characteristics. J. Econ. Stud. 34 (2), 136–156.
Bezdeck, J., 1981. Pattern recognition with fuzzy objective function algorithms. In: Bezdeck, J. (Ed.), Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York.
Chiang, M., Mirkin, B., 2010. Intelligent choice of the number of clusters in k-means clustering: an experimental study with different cluster spreads. J. Classif. 27 (1), 3–40.
Cifuentes, R., Martínez, F., 2019. Sobre-endeudamiento de las familias chilenas: Medición y determinantes Unpublished manuscript.
D'Alessio, G., Iezzi, S., 2013. Household over-indebtedness: definition and measurement with Italian data. In: Bank of Italy Occasional Paper 149.
Del Rio, A., Young, G., 2008. The impact of unsecured debt on financial pressure among British household. Appl. Financ. Econ. 18 (15), 1209–1220.
Dey, S., Djoudad, R., Terajima, Y., 2008. A tool for assessing financial vulnerabilities in the household sector. Bank Can. Rev. Summer, 45–54.
Drozd, L., Nosal, J., 2011. Competing for customers: a search model of the market for unsecured credit. In: 2011 Meeting Papers 1033. Society for Economic Dynamics.
Durkin, T., Elliehausen, G., Staten, M., Zywicki, T., 2014. Consumer credit and the American economy. In: Financial Management Association Survey and Synthesis. Oxford University Press.
European Central Bank, 2013. The Eurosystem household finance and consumption survey results from the first wave. In: Statistics Paper Series 2.
Fraley, C., Raftery, A., 1998. How many clusters? Which clustering method? Answers via model-based cluster analysis. Comput. J. 41 (8), 578–588.
Frey, T., Van Groenewoud, H., 1972. A cluster analysis of the D-squared matrix of white spruce stands in Saskatchewan based on the maximum-minimum principle. J. Ecol. 60 (3), 873–886.
Gough, O., Sozou, P., 2005. Pensions and retirement savings: cluster analysis of consumer behaviour and attitudes. Int. J. Bank Mark. 23 (7), 558–570.
Gross, D., Souleles, N., 2002. Do liquidity constraints and interest rates matter for consumer behavior? Evidence from credit card data. Q. J. Econ. 117 (1), 149–185.
Herrala, R., Kauko, K., 2007. Household loan loss risk in Finland – estimations and simulations with micro data. In: Bank of Finland Discussion Papers 5.
Jappelli, T., Pagano, M., Di Maggio, M., 2013. Households' indebtedness and financial fragility. J. Finan. Manag. Mark. Inst. 1 (1), 23–46.
Karasulu, M., 2008. Stress testing household debt in Korea. In: IMF Working Paper 08/255.
Keese, M., 2012. Who feels constrained by high debt burdens? Subjective vs. objective measures of household indebtedness. J. Econ. Psychol. 33 (1), 125–141.
Kempson, E., 2002. Over-indebtedness in Britain. In: A report to the Department of Trade and Industry. Personal Finance Research Centre.
Kennickell, A., Woodburn, L., 1999. Consistent weight design for the 1989, 1992, and 1995 SCFs, and the distribution of wealth. Rev. Income Wealth 45 (2), 193–215.
Livshits, I., MacGee, J., Tertilt, M., 2010. Accounting for the rise in consumer bankruptcies. Am. Econ. J. Macroecon. 2 (April), 165–193.
Livshits, I., MacGee, J., Tertilt, M., 2016. The democratization of credit and the rise in consumer bankruptcies. Rev. Econ. Stud. 83, 1673–1710.
Madeira, C., 2015. Chilean households' motives for holding consumer debt (Motivaciones del endeudamiento de las familias chilenas). Econ. Chil. 18 (1), 90–106.
Madeira, C., 2018. Explaining the cyclical volatility of consumer debt risk using a heterogeneous agents model: the case of Chile. J. Financ. Stab. 39, 209–220.
Madeira, C., 2019. A cross-country analysis of Households' Consumer and Mortgage Debt Unpublished manuscript.
Martínez, F., Cifuentes, R., Madeira, C., Poblete-Cazenave, R., 2013. Measurement of household financial risk with the survey of household finances. In: Central Bank of Chile Working Paper 682.
Martínez, F., Roje, P., Saavedra, T., 2019. Multiple imputation in the Chilean survey of household finances Unpublished manuscript.
Ponce, A., Seira, E., Zamarripa, G., 2017. Borrowing on the wrong credit card? Evidence from Mexico. Am. Econ. Rev. 107 (4), 1335?1361.
Ruiz-Tagle, J., Vella, F., 2016. Borrowing constraints and credit demand in a developing economy. J. Appl. Econ. 31, 865–891.
Sanchez, J., 2018. The Information technology revolution and the unsecured credit market. Econ. Inq. 56 (2), 914–930 April.
Shao, J., Sitter, R., 1996. Bootstrap for imputed survey data. J. Am. Stat. Assoc. 91, 1278–1288.

Survey of Household Finances, 2015a. Encuesta Financiera de Hogares 2014: Principales Resultados. In: Central Bank of Chile Reports.
Survey of Household Finances, 2015b. Documento MetodolÃ³gico EFH 2014. In: Central Bank of Chile Reports.
Telyukova, I., 2013. Household need for liquidity and the credit card debt puzzle. Rev. Econ. Stud. 80 (3), 1148–1177.
Telyukova, I., Wright, R., 2008. A Model of money and credit, with application to the credit card debt puzzle. Rev. Econ. Stud. 75 (2), 629–647.
Theodoridis, S., Koutroumbas, K., 2009. Pattern recognition. In: Theodoridis, S., Koutroumbas, K. (Eds.), Pattern Recognition, Boston.
Trivedi, S., Pardos, Z., Heffernan, N., 2015. The utility of clustering in prediction tasks. Comput. Therm. Sci. 1–11.