# BIS Working Papers

No 1258

# How accurately do consumers report their debts in household surveys?

by Carlos Madeira

Monetary and Economic Department

April 2025

JEL classification: C81, D10, D12, E21, G21

Keywords: household finance surveys, mortgages, consumer credit, default, measurement error

BIS Working Papers are written by members of the Monetary and Economic Department of the Bank for International Settlements, and from time to time by other economists, and are published by the Bank. The papers are on subjects of topical interest and are technical in character. The views expressed in them are those of their authors and not necessarily the views of the BIS.

This publication is available on the BIS website (www.bis.org).

© *Bank for International Settlements 2025. All rights reserved. Brief excerpts may be reproduced or translated provided the source is stated.*

ISSN 1020-0959 (print)
ISSN 1682-7678 (online)

# How accurately do consumers report their debts in household surveys?

Carlos Madeira*

March 2025

**Abstract**

This article advances upon previous studies by using a unique match of a representative sample of individual borrowers from the Chilean Household Finance Survey and their banking loan records. I show that surveys differ from the credit registry, not just in the number of loans reported, but also in their amount, with a substantial degree of heterogeneity. Delinquency status is accurately reported by survey respondents. Furthermore, a substantial fraction of the discrepancies can be explained by rounding error in survey answers. Finally, I find that discrepancies are larger when respondents are not the highest-income member of the family.

**JEL Classification:** C81; D10; D12; E21; G21.

**Keywords:** Household Finance Surveys; mortgages; consumer credit; default; measurement error.

*Bank for International Settlements and Central Bank of Chile, carlos.madeira@bis.org. I would like to thank Jon Frost, Basit Zafar, Sumit Agarwal, Jorge Ponce and seminar participants at the Central Bank of Chile and the Chilean Financial Market Commission (CMF). The views expressed in the article are those of the author and do not necessarily reflect the views of the BIS or the Central Bank of Chile or its board members. All errors are my own.

# 1 Introduction

Household finance surveys, such as the Household Finance Consumption Survey in Europe or the Survey of Consumer Finance in the US, are increasingly used to study families' decisions on savings, investments and borrowing (Dynan and Kohn 2007, Christelis, Georgarakos and Haliassos 2013, Christelis, Ehrmann and Georgarakos 2017, Le Blanc et al. 2015, Bover et al. 2016). Survey information on finances is important (Zinman 2009), especially because many households and small enterprises rely on a diversity of funding sources and credit instruments, which may come from bank and non-bank lenders. Surveys also measure both formal and informal income sources, an advantage in countries where informal employment can represent more than 25% of the labor force. However, surveys across several topics suffer from measurement error, mainly due to intentional mis-reporting (for instance, due to social desirability), lack of memory, cognitive error and rounding by the respondents (Bound et al. 2001, Giustinelli et al. 2022, Stantcheva 2023). Measurement error creates downward bias in coefficient estimates (Bound et al. 2001) and often complicates empirical analysis such as requiring validation samples (Bound et al. 2001), repeated measurement or instrumental variables (Schennach 2016). This makes it important to study the accuracy of survey responses, especially in surveys with complex and difficult questions such as financial surveys.

Comparative studies of the aggregate amounts of household liabilities in survey datasets and national accounts find a significant degree of under-reporting of loans in household surveys (Brown et al., 2015), especially in the case of short-term debt such as consumer loans and credit cards (Karlan and Zinman 2008, Zinman 2009). As a percentage of the total liabilities in national accounts, a review of household finance surveys in Europe and the US between 1998 and 2002 found that surveys underestimated aggregate liabilities by 40% in Germany, 56% in Italy, 72.5% in Portugal, 16% in Finland and 11% in the US (Sierminska et al. 2006, Kavonius and Törmälehto 2010). For this reason, different survey methodologies and interview methods have attempted to improve survey measurement of economic and financial variables (Le Blanc et al. 2015).

This work shows microdata evidence on how households' self-reported debt information differs from their administrative bank loan records. The study takes advantage of a unique matched dataset between the Chilean Household Finance Survey (*Encuesta Financiera de Hogares*, in Spanish, from hence on, EFH) and the Banking Loan administrative records (from the Chilean Banking Authority, in Spanish, *Superintendencia de Bancos e Instituciones Financieras*, hence on SBIF).[1] The matched survey-banking registry dataset improves much on the results of previous studies for other countries. I show at the micro-level how the survey's self-reported loan information differs from the true banking debt records across respondents. This allows to test whether the differences between survey information and loan records are due to the number of non-reported loans or differences in the loan amounts. I also analyze how the quality of survey information differs for households according to education, income levels or demographic complexity (such as the coexistence of several financial decision makers in the same household). Therefore, this is a strong improvement upon previous studies that were limited to a comparison between aggregate debt amounts in survey data and administrative records (Zinman 2009, Brown et al. 2015, Bhandari et al. 2020).

The matched survey-banking registry dataset provides data on all banking loan contracts (including mortgages, consumer installment loans, credit cards and credit lines) for the period 2003 until 2018 and their self-reported cross-sectional information on household's demographic characteristics, income and loans (with either banks and non-banking institutions) for one survey year (with household interviews in 2011, 2014 and 2017). The matched data includes the loan history in the banking system for the interviewed persons plus survey-reported measures of income, age and education for both the interviewed person and household members.

This analysis shows that there are substantial discrepancies between individual borrowers' survey reports and the administrative records. In terms of loan participation, there is a significant percentage of loans that are reported in the survey and not in the registry or vice-versa. However, borrowers report the delinquency status of their loans quite accurately. They also report the maturity of their mortgage and consumer installment loans quite accurately. In terms of loan amounts,there are reasonable discrepancies, with the percentiles 25 to 75 going from a range of -31% to 18% and between -0.7% to 59.3% for mortgage and installment loan amounts, respectively. For credit cards and credit lines, the discrepancies are higher (a result also found for the US; see Zinman 2009). This larger discrepancy for credit cards and lines of credit was to be expected, because the survey interview only asks borrowers to report debts that will last more than one month. Therefore, the survey does not include very short-term revolving debt.

[1] Note that the Chilean Banking Authority was merged with the Insurance and Assets Authority in 2019, therefore creating the Chilean Financial Market Commission (in Spanish, *Comisión del Mercado Financiero*, CMF)).

I then test whether these differences can be explained by rounding errors in the survey reports, since households tend to report round numbers that are multiples of 10. After rounding the registry loans, I find indeed that such a mechanism can go a long way towards explaining the differences in the middle of the debt amount distribution, with the discrepancies between survey and registry becoming zero for a large fraction of the borrowers. For the case of consumer installment loans, more than 25% of the borrowers have a null discrepancy in the debt amount reported between survey and registry after rounding, while for mortgage loans this discrepancy is 7% or less for more than 25% of the borrowers. Finally, I show that borrowers with complex households that have a higher number of members with the highest-income (and possibly more access to loans) and borrowers who are not the highest-income or the oldest members of the family tend to show more discrepancies between their survey report and the registry. The reason is that in such situations the borrower may be reporting loans in the survey that are not his or her own (for instance, the spouse reports a consumer loan or mortgage as his or her debt, but in fact the loan was signed by the highest-income household member, who was not present in the interview).

Household finance surveys are used for the analysis of debt risk and household stress tests (Eurosystem Household Finance and Consumption Network 2009). I show how mismeasurement of borrower debt affects the evaluation of debt service to income ratios. Therefore, matched survey-registry datasets can improve this aspect of policy analysis. Furthermore, improvement of survey design as new technologies (such as financial aggregators and eased sharing of information) are better able to elicit financial values from the survey respondents may help to reduce survey measurement error in the future (Caplin 2025).

This work is closest to studies on the relationship between total debt amounts in household finance surveys and administrative data, such as Zinman 2009 and Brown et al. 2015. Likewise, it relates to the literature on measurement error in surveys using linked survey and administrative information (Bound et al. 2001, Schennach 2016). This paper is also related to microeconomic studies of household debt (Ampudia et al. 2016, Madeira 2018b, 2019a, 2019b) and more robust measures of interpreting survey responses (Madeira and Zafar 2015, Madeira 2018a). Some studies have compared survey and administrative records of debt for a single lender (Karlan and Zinman 2008), although not for all bank lenders at the national level. Other countries that also have survey-administrative matched datasets may also wish to further their analysis in this aspect (see the case of Estonia detailed by Meriküll and Rõõm 2020). Several countries have the survey and credit registry datasets required for doing such an analysis, but often the treatment of household survey data in confidentiality may require special legal treatment.

This study shows the first analysis of how individual reports of debt compare in a nationally representative household finance survey and an administrative registry with full coverage of banking debts. I show that delinquency behavior and maturities are accurately reported in survey data, while loan amounts have a substantial degree of rounding error. This research fits well with the analysis of rounding error in other topics, such as health surveys and expectations (Giustinelli et al. 2022).

This work is organized as follows. Section 2 summarizes the matched survey-banking registry dataset. Section 3 shows the statistical discrepancy in terms of participation of different loan types in the survey data and registry. Section 4 explains the statistical methodology for measuring differences in continuous outcomes, such as debt amounts and maturities, between self-reported and administrative data. I then test how much of the differences between survey and administrative debt values for the same borrower can be explained by rounding error and outliers. Section 5 tests several hypotheses about the causes of the discrepancies in the survey data, including the age profile of the interviewer and income complexity of the household. Section 6 tests the implications of using matched registry and survey data for the measurement of aggregate debt and debt risk. Finally, Section 7 summarizes the conclusions.

## 2 The matched survey-banking registry dataset

### 2.1 Quality of the matching dataset

The EFH is a cross-sectional survey implemented every 3 years to around 4,000 to 4,500 households. The EFH is a representative cross-sectional survey with detailed information on households' assets, debts, income, and financial behavior, being broadly comparable to similar surveys in the US and Europe (Eurosystem Household Finance and Consumption Network 2009, Ampudia et al. 2016). The EFH has a particularly detailed focus of the loans and debt commitments of each household. It asks for the largest three debts that each household has for each type of loan, among a total of

13 categories of loans: banking credit card debt, banking line of credit, banking or financial agency consumer credit loan, retail store credit card, retail store consumer loan, auto loans, union credit, education loans, loans from relatives, loans from usurers, pawn shops, grocery and shopping on credit (i.e., store tabs), and other debts. Therefore, the survey may ask for up to a total of 39 debts that the household has at the moment, although obviously few agents will report having debts with all the possible categories of loans. Like the SCF in the US and its European equivalents, the EFH survey interviews the "person with the highest knowledge of the household's finances", which is labelled the "reference person". The reference person or interviewee is also usually equivalent to the household head. In around 67.5% of the cases the interviewed is also the household head. Furthermore, in 89.6% of the cases the interviewee is either the household head or the partner of the household head.

To obtain a more accurate view of the evolution of each household's indebtedness over time, the Central Bank of Chile and the Chilean Banking Authority (SBIF) decided to build an survey-banking registry dataset, where each survey's information is linked to the monthly banking credit information for each month over the period January 2003 to December 2018. The link between each household's main member in the survey dataset and its entire history of banking debt is made by using Chilean national identity numbers. Chileans make regular use of their national ID to obtain discounts in supermarket chains, apply for loans, or to use the health system. Therefore, participating households are comfortable in providing their information during the survey interview. Furthermore, each national identity number is followed by a validation digit, which allows the surveyor to test whether the stated number is correct. This prevents mismatching in the sense that it allows the surveyor to validate that there was no mistake in the recorded identity number during the interview process.[2]

The national ID numbers of each survey respondent are then matched with the Chilean Banking Authority registry, which include all the people who have ever applied for a banking product (whether a loan, a current account or a savings account). Table 1 summarizes how many households are observed in each dataset. There are around 13,110 households in the survey dataset (waves

[2]Note that the chance of respondents giving a false id and then stating its validation digit correctly by answering a random number are only 10%. Respondents would have to know the fairly complex math behind the validation digits in order to give a false id number with a correct validation digit. This makes it extremely unlikely that the respondents' answers for their id numbers belong incorrectly to someone else's id.

Table 1: Number of total respondents in the survey and matched survey-registry datasets

| Wave | Survey | Survey-banking registry persons | Members with survey debt around time of survey | Members with debt in both sources |
|---|---|---|---|---|
| 2011 | 4,059 | 2,329 | 933 | 521 |
| 2014 | 4,502 | 2,362 | 1,132 | 638 |
| 2017 | 4,549 | 3,356 | 1,790 | 1,033 |
| Total | 13,110 | 8,047 | 3,855 | 2,192 |

Note: The registry data only includes persons that have ever used a loan or applied for a loan. There are more than 8,047 persons reporting their ids in the survey, but some of those persons may have never applied to a loan.

2011, 2014, 2017), with 8,047 respondents having both given a correct national ID number and being matched in the registry banking loan dataset.[3] Notably, not all the survey-banking registry respondents had positive amounts of debt at the time of the survey. This implies that one can only test the discrepancies in non-zero amounts of self-reported loans for 3,855 observations. Positive debt amounts are observed in both survey and registry for 2,192 respondents. The survey dataset has more than 8,047 respondents that reported a correct ID. Those respondents are not a part of the matched survey-banking registry dataset, because the registry dataset only has the persons that have applied for or used a banking product at some point. Therefore, the actual response rate for the personal ID number is greater than 61.4%.[4]

The exclusion for this study of households that never had or applied for a banking loan does not

[3] The respondents' ID disclosure rate changed over time, being much higher in the most recent survey wave of 2017. One reason could be due to changes in field methods, interviewer training, supervision and survey management from the fieldwork companies, which changed from the Social Observatory of University Alberto Hurtado in 2011 to the Ipsos company in 2014 and 2017. Another reason could be from learning by doing. The elicitation of the respondents' ID was implemented through a new interview section with several sensitive questions, which was first added in 2011. Therefore, the survey companies may have gained experience about how to better elicit this sensitive information over time.

[4] The researcher in this article does not know the exact response rate, because the national ID numbers of the survey respondents are not directly available at the Central Bank of Chile. The national ID numbers are deposited for reasons of statistical secrecy with the National Institute of Statistics, the Social Observatory (which implemented the EFH wave of 2011) and Ipsos (which implemented the EFH waves of 2014 and 2017). The Central Bank of Chile only obtained a pseudo-identifier for the matched observations of the EFH-SBIF dataset and at no point had any access to the real national IDs.

imply that such household observations do not provide information. For instance, Madeira 2019b studied the fraction of households that were excluded from credit due to a change in the interest rate ceiling and in this aspect it is informative to have information on households without credit.

The registry of banking loans is considered to be a high-quality dataset. Therefore, this linked dataset can be used to validate the self-reported survey debt values. It would be a serious legal violation if banks failed to report a loan, its amount and the interest rate to the Banking Authority, since: i) banks are obliged to report their loans to regulators for risk assessment purposes; ii) banks are obliged to report each loan, the amount and its interest rate of each loan comply with the usury laws in Chile (the usury interest rate differs by loan amount, therefore both the interest rate and the loan amount must be accurately reported). This implies large fines in the short term for uncooperative banks and over the long term it implies the exclusion of the bank from all its legal activities and corporate charter. Furthermore, the Banking Authority requires all banks to update their information every month, making it unlikely that any loan errors would remain undetected and uncorrected until January 2019, which was when the registry and survey were linked.

Note also that banking products, whether for deposits, loans, insurance or other financial products, operate in a similar way in Chile relative to advanced economies, such as the US, Canada or the EU. Chile has 10 banks currently offering loans to households. The two major banks are Spanish (Santander, BBVA). Scotiabank (a Canadian bank) and Itaú (a Brazillian bank which purchased the Chilean operations of Citibank, a US bank, and HSBC, a British bank) are banks of median size. Chile does have substantial differences from more developed economies in terms of the availability of non-bank debt.

The survey-banking registry matched dataset has a few limitations: i) the universe is limited to individuals who have ever applied or used a banking product (such as a consumer loan, mortgage, credit card, debit card or current account); ii) the monthly loan history is limited to banking loans of different types (consumer installment loans, credit cards, lines of credit, student loans, and mortgages) and therefore does not include loans from non-bank lenders;[5] and iii) the matched survey-banking registry data provide information on the current loan amount, the original loan

[5]Note that this second limitation differs from the first one. If all bank customers had zero non-bank loans, then this second limitation would not matter. However, around 51.6% of the households in the survey-registry dataset had some form of non-banking debt.

amount at the time the contract was made, the total payment due to that loan in a certain month and whether the loan is in delinquency or not, but do not include information on renegotiation of loans, interest rates or on other fees and costs charged.[6]

Note that in Chile there is a significant presence of non-bank lenders. According to the EFH survey wave of 2017, around 49.8% of the Chilean households had some type of non-banking loans, with about 13.2% of the households had both banking and non-banking debt (Madeira 2024). The largest non-bank lender in terms of customers are retail stores. For the case of the matched survey-registry data, around 51.6% of the households had some non-banking debt at the time of the survey. Around 41.8% of the matched survey-registry sample had retail store debt,[7] while 11.1% had union or cooperative debt[8] and 10.6% had other types of debt (vehicle or educational loans).

The population of the survey-banking registry linked dataset is a meaningful population, since it is representative of the users and applicants of loans from all the banks and financial institutions in Chile, which are 61.4% of the urban population in Chile. Furthermore, banks are extremely relevant lenders, with banking loans for households in 2017 representing 34.5% of the GDP (with 24.4% of the GDP in bank mortgages and 10.1% of the GDP in bank consumer loans). Note, however, that the survey-banking registry dataset can only validate the loan reports of the respondents, since the survey does not include the identity numbers of the other household members.[9]

All the software codes necessary for formatting the matched survey-registry data and replicating

[6] The SBIF banking debt registry is not a panel data of loans. It lists all the loans of each individual in a given month, but it is not possible to connect each loan with loans in other periods. If an individual renegotiates a loan, then it is not possible to establish whether the new loan was the result of a renegotiation of a prior loan.

[7] Some companies have both retail stores and banks. Large retail companies like Falabella, Paris and Ripley started by offering retail credit and retail credit cards in their stores. After some years, those companies expanded towards banking businesses, which are often located next to their retail stores.

[8] Note that in Chile all formal workers are registered under a labor union. Labor unions collect membership fees directly from the members' wages. These funds can then be used for a variety of social purposes and also for loans to their members. Furthermore, Chile also has a significant presence of credit cooperatives, which can extend loans to their members.

[9] The EFH survey did some pretests in which the interviewer asked for the identity number of the respondent and of a "second financially relevant person in the household", but respondents were either unaware of the identity number of the other members or reluctant to provide it. For this reason, the matched survey-registry is limited to the IDs of the respondents.

the empirical analysis of this article are publicly available in a Mendeley Data link.[10] The link also includes an online appendix to this article, with extra information and analysis. In particular, the online appendix includes information on: i) the sampling and field operations of the survey; ii) the survey's questionnaire, specifically the questions on debt and loans; iii) additional summary statistics of the data; and iv) the results of Kolmogorov-Smirnov and Goldman-Kaplan tests of the equality of the debt distributions in the survey and registry datasets (Goldman and Kaplan 2018).

## 3 Comparing self-reported information and registry

This section dives deeper into the differences between self-reported survey loans and administrative records. I start by comparing the survey and registry in terms of debt participation (in Table 2) and delinquency (in Table 3). I then complete the section by comparing the cumulative distribution function (CDF) of the debt amounts in the survey and registry. This article will study bank loans classified in 4 products: mortgages, consumer installment loans, credit cards, credit lines.[11] Consumer installment loans are loans that involve repayment over several periods (installments), which can go from a few months (for instance, purchases of clothing or small household items) to several years (such as for payments of vehicles or expensive furniture). This consumer debt category is typically not renewable automatically and therefore it can be contrasted with revolving debt categories (credit cards and lines of credit). Furthermore, I define the overall category of consumer debt, which comprises both consumer installment loans and revolving debt (credit cards, credit lines). Note that only banking debt is considered in the analysis, since the registry does not include non-bank lenders. This comparison is possible because the survey is careful to distinguish each type of lender and loan product (see section 2 of the appendix). Current loan amounts are defined as principal owed, both in survey and registry. The comparison in all the analysis of this article considers the sum of all debts of a certain product type (mortgages, consumer installment

[10] https://data.mendeley.com/datasets/65z482xs9w/1

[11] Note that even if the sample was confined to just households with one loan type, there would still be substantial discrepancies between survey and registry datasets, as shown in Table A.6 (households with one or zero loan types) and Table A.7 (households with only one loan type) of the online appendix. Therefore, the discrepancies between survey data and registry cannot be due to simple confusion between different bank loan types.

loans, credit cards, credit lines) at the borrower level.[12] The reason is that the registry is not a panel dataset of loans, therefore it can include several items belonging to the same loan (insurance, fees, interest) as different transactions.

It is possible that one of the reasons for the disparities between survey data and registry could be a delay in how loan records are updated every month. A borrower could have a new loan that he/she reports for the survey, but a missing record for the loan in the registry. In the same way, a borrower could have finished paying his or her loan this month or last month and report that he/she has no current outstanding debt, but in the administrative credit registry it may take a few days for his or her status to be effectively without debt. For this reason I consider the closest registry report to the survey data around a window of two months from the time of the survey report $t_{Survey}$ ($t_{Registry} = t_{Survey}+k$, with $k = 0, -1, 1, -2, 2$). I also take into account a large range of detailed survey information, such as: whether some loans are from a different household member (for example, a child, parent or spouse) instead of the main borrower, whether the household could have mistaken loan from retail stores and retail banks within the same business group,[13] and whether the household could have confused a consumer installment loan meant for a house related expense with a mortgage. For this reason, for respondents with debt in companies with joint retail stores and retail banks, I consider the debt measurement that is closest to the registry measure.[14]

Table 2 summarizes the difference in terms of the debt participation (a binary variable of 1 for whether the household has a positive loan amount in one dataset) in the matched survey-banking registry dataset. Each pair of binary variables corresponds to whether both datasets coincide in terms of no debt (Survey=0, Registry=0), whether the reports diverge because only the registry

[12] Note that it is also more appropriate for statistical analysis to take the sum of all debts for each borrower. It is easy to assume that debts are independent across different borrowers, but it is unlikely that debts for the same borrower could be taken as independent observations. It would also be the case that if the analysis was made at the loan level, then borrowers with many loans would count much more than the other debtors.

[13] There are 3 companies in Chile that own both a retail store and a bank specialized in consumer credit. These banks and retail stores work as separate institutions for administrative purposes, but are often located in the same commercial branch and under a similar advertising brand name. Therefore it is possible that some borrowers confuse loans from a bank and a retail store of the same name.

[14] This is done with the following algorithm for the survey banking debt amount: $Y^R_{i,t,Survey} \equiv \arg\min_{\tilde{Y}_{i,t+k,Survey}, k\in\{-2,-1,0,1,2\}} \left|\tilde{Y}_{i,t+k,Registry} - Y^R_{i,t,Survey}\right|$, s.to. $Y^R_{i,t,Survey} \in ($ $Y^R_{i,t,Survey}(NonRetailBanks)$, $Y^R_{i,t,Survey}(RetailBank) + Y^R_{i,t,Survey}(NonRetailBanks)$, $Y^R_{i,t,Survey}(RetailBank) + Y^R_{i,t,Survey}(RetailStore) + Y^R_{i,t,Survey}(NonRetailBanks)$ ).

Table 2: Differences between the self-reported survey data and registry of borrower status (binary loan variable)

| Debt participation (Survey, Registry) - % | (0,0) | (0,1) | (1,0) | (1,1) |
|---|---|---|---|---|
| Mortgage loan | 82.0 | 1.6 | 9.3 | 7.0 |
| Consumer debt | 63.1 | 14.1 | 5.6 | 17.2 |
| Consumer installment loan | 79.4 | 7.7 | 3.9 | 9.1 |
| Credit Card or Credit Line | 69.9 | 14.5 | 4.2 | 11.4 |
| Credit Card | 72.1 | 13.9 | 3.9 | 10.1 |
| Credit Line | 88.5 | 5.4 | 2.2 | 4.0 |

Number of observations: 8,047 households (all the rows sum up to 8,047 observations of respondents with or without positive debt).

Table 3: Differences between the self-reported survey data and registry of default status (binary variable)

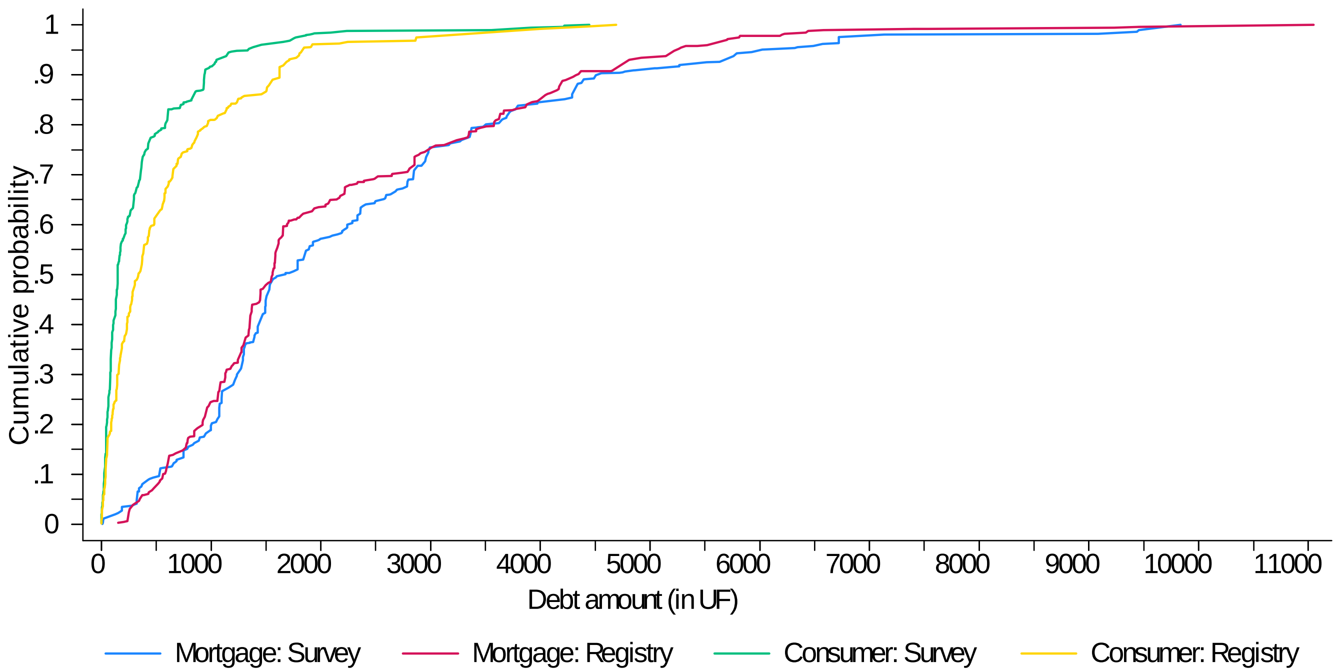
| Default (Survey, Registry) - % | (0,0) | (0,1) | (1,0) | (1,1) |
|---|---|---|---|---|
| All loans (including loans reported in one dataset only) | | | | |
| Mortgage loan | 97.7 | 0 | 0.8 | 1.5 |
| Consumer installment loan | 84.1 | 0.2 | 7.2 | 8.5 |
| Including just borrowers with same debt participation in both datasets | | | | |
| Mortgage loan | 96.3 | 0 | 0 | 3.7 |
| Consumer installment loan | 87.2 | 0.3 | 0.3 | 12.2 |

Number of observations: 8,047 households (all the rows sum up to 8,047 observations of respondents with or without positive debt).

reports a loan (Survey=0, Registry=1), whether the divergence happens because only the survey data reports a loan (Survey=1, Registry=0), or whether both datasets coincide in terms of positive loan amounts (Survey=1, Registry=1).

I also report whether the loans are in default or not in the survey data and registry. I take as default whether the household fulfills one or more of these criteria: it has payments in arrears for 1 month or more, a positive loan amount in arrears or a positive non-performing loan amount. Table 3 reports how much the households diverge in terms of their default reports. I first report the default debt status considering all reports (including reports of a zero loan amount which is treated as a zero default as well). It appears that the survey data reports several more loan events than the registry. However, once I consider only the borrowers in which their debt status (positive amount or not) coincide in both the survey data and registry, then the default reports are almost exactly

Figure 1: CDFs of the original mortgage and consumer debt amounts in the survey and registry

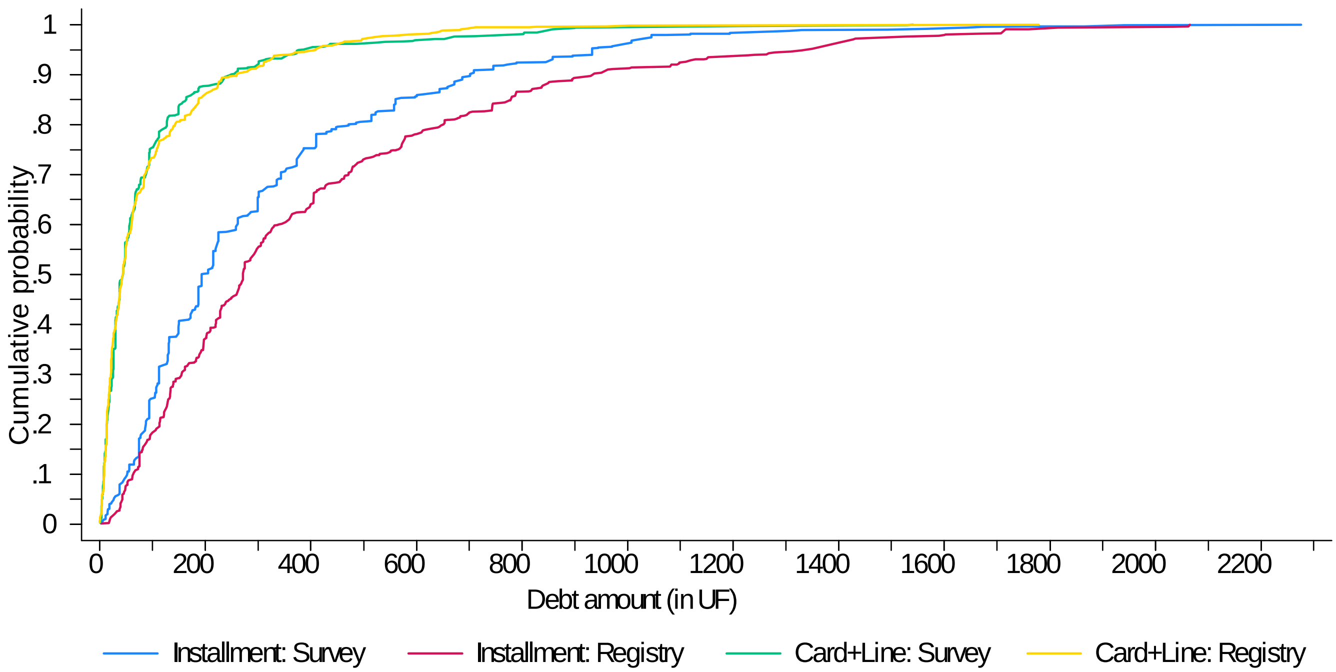



the same in both datasets. Therefore, it appears that the survey differs from the registry only in terms of the loan amount reports, but households report accurately their delinquency status.

Now I show the CDFs of the debt amount in UF of the borrowers' survey self-reported debt and their debts in the banking registry. To avoid a large mass point at zero due to a significant fraction of borrowers with zero debt (as shown in Table 2), this analysis only includes borrowers that have positive debt amounts in the survey or registry. Furthermore, all debt amounts are converted to UF. UF is a real monetary unit applied in Chile, which is updated daily according to the official consumer price inflation (CPI) index in the previous month. This real monetary unit is often used in Chile for long-term contracts, such as mortgages and large consumer loans. One UF was roughly equivalent to 40 USD during the sample period of 2013 to 2019.

Figure 1 shows the CDFs for the total mortgage and total consumer banking loans of the matched survey-registry borrowers. In the case of mortgages, there is a fairly close match of the debt distribution in the survey and registry until around 1,500 UF (which is roughly the percentile 50 for both survey and registry). The survey and registry mortgage distributions then diverge, before approaching again between percentiles 70 to 85 and between percentiles 95 to 100. Afterwards, the

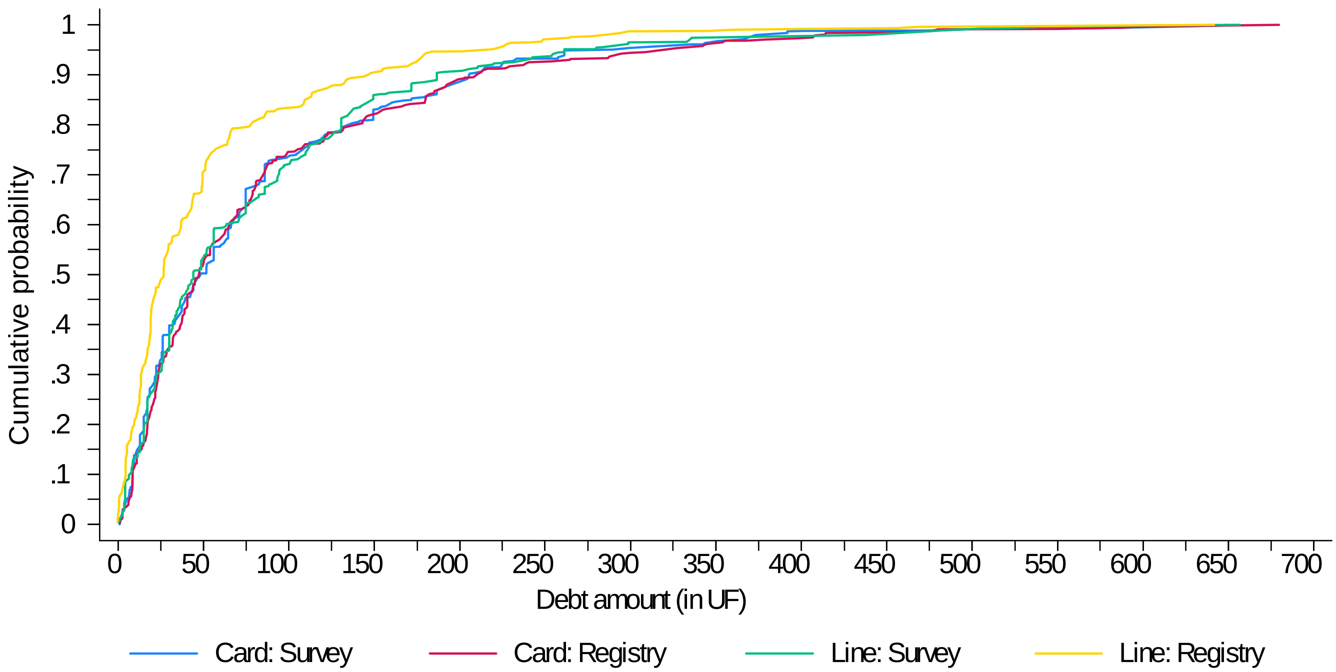
Figure 2: CDFs of the original consumer installment loans and revolving debt instruments (credit cards and lines of credit) in the survey and registry




graph shows that the survey data is much more detailed for mortgage amounts above the percentile 90. Therefore, the survey includes more high-value mortgages than the registry. For the total consumer debt amount, the differences between survey and registry are quite large. The survey and registry CDFs depart after the percentile 18, showing a coincidence only for small consumer loans. Notably, unlike the case of mortgages, for consumer loans the registry is much more complete in the inclusion of larger debt values.

Now I compare the consumer debt distribution for the categories of installment and revolving debt instruments (i.e., the sum of debt in credit cards and lines of credit). It is relevant to note that it is expected that the survey data is less complete for revolving debt, because the respondents are asked not to report debt that is expected to be repaid in less than one month. However, the survey respondents may still report some revolving debt if they expect that the repayment of that debt will last longer than one month. Figure 2 shows that the distribution of revolving debt (card plus line) is quite similar for survey and registry across all percentiles. Therefore, revolving debt seems to be well reported in both survey and registry. Installment loans, except for small debt

Figure 3: CDFs of the debt amounts for credit cards and lines of credit in the survey and registry

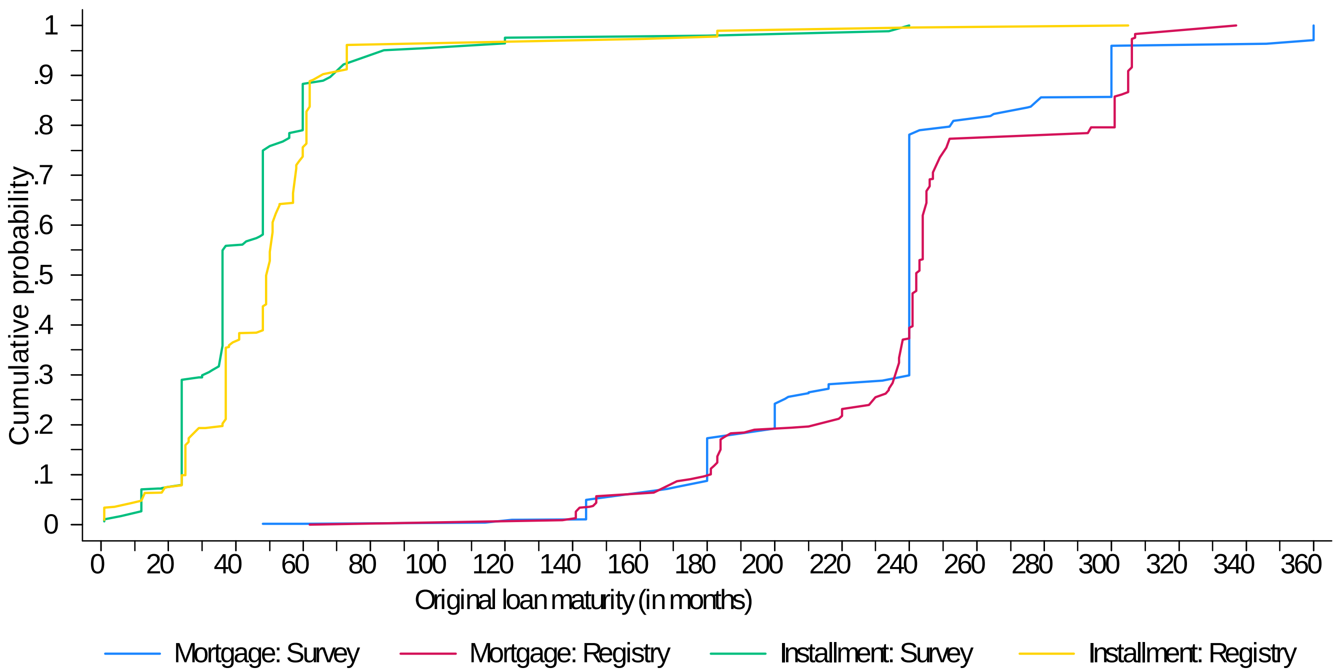



values, are fairly different between survey and registry, departing after percentile 15.

I then show the differences between survey and registry for the distribution of debt in credit cards and lines of credit. Figure 3 shows that the survey and registry distributions are almost the same for credit cards. However, borrowers report higher amounts for lines of credit in the survey dataset. This makes sense, because lines of credit are an expensive form of revolving debt. Therefore, borrowers are likely to repay their line of credit before the amount is recorded in the registry at the end of the month.

Finally, I compare the distributions of loan maturity in the survey and registry. Figure 4 shows that maturities for mortgages are quite similar in the survey and registry datasets. Maturities for mortgages in the survey data, however, have a substantial rounding for 20, 25 and 30 years (i.e., 240, 300 and 360 months). Maturities reported for installment loans differ substantially between survey and registry, with the registry showing significantly longer loan horizons. For installment loans, survey and registry only coincide for the percentiles below 10 and above 90. Again, it is noticeable that the survey respondents tend to round their self-reported maturities, since there are mass points around 24, 36 and 48 months (i.e., 2, 3 and 4 years).

Figure 4: CDFs of the original loan maturities for mortgages and consumer installment loans in the survey and registry



# 4 Discrepancies in loan amounts and maturities

To delve deeper into the discrepancies between survey and administrative reports, I now summarize the discrepancies in terms of continuous loan amounts. For this I use a continuous measure of error, defined as the ratio of the difference between both reports (with variables $Y^R_{i,t,Registry}$ and $Y^R_{i,t,Survey}$ representing the continuous loan outcomes of borrower $i$ at time $t$ in the survey and registry reports) divided by their mean value:

$$1)\ error\ ratio_{i,t} = \frac{Y^R_{i,t,Registry} - Y^R_{i,t,Survey}}{(Y^R_{i,t,Registry} + Y^R_{i,t,Survey})/2}.$$

This ratio has two advantages. The first advantage is that by using both registry and survey measurements in the denominator, then small values from either source (such as a low with a value close to zero) are prevented from influencing the error measure towards extremely large values. This measure is by definition bounded between -2 and 2 (or equivalently, -200 to 200%) and it can include cases in which one of the data reports a zero outcome. The second aspect is that by being bounded, then this error ratio measure does not have a few large values influencing the statistical analysis. Therefore the $error\ ratio_{i,t}$ has become the most standard way of measuring differences between two datasets (Törnqvist, Vartia and Vartia 1985).

There can be disparities between the month in which the registry dataset is updated and the date of the survey report; therefore, I take the closest value of $Y^R_{i,t,Registry}$ in a two-month window:

$$2)\ Y^R_{i,t,Registry} \equiv \arg\min_{\tilde{Y}_{i,t+k,Registry},k\in\{-2,-1,0,1,2\}} \left|\tilde{Y}_{i,t+k,Registry} - Y^R_{i,t,Survey}\right|.$$

Table 4 summarizes the discrepancies between survey and administrative reports at the borrower level, showing the percentiles 10, 25, 50, 75 and 90 of the $error\ ratio_{i,t}$ for different types of loans and different characteristics of each loan type (debt amount of the original loan contract, debt service, original maturity of the loan, and residual maturity of the future debt payments). The results show that differences between survey and administrative reports can be substantial, even if one considers the median measurement error. The median measurement error for debt amounts is fairly low for mortgages, credit lines and credit cards. Also, the median measurement error is fairly low for the maturities of both mortgage and installment loans, representing accurate information in this aspect.

The discrepancies of the self-reported debt amounts for the total sum of consumer debt, the installment loans' original debt amount and the mortgage debt service are above 10%, even for the median error. Yet even if one discards the worst reports (those with percentiles 10 and 90, or more extreme values) there are still reasonable discrepancies around the middle of the distribution, with the percentiles 25 to 75 going from a range of -31% to 18% and between -0.7% to 59.3% for mortgage and installment loan amounts, respectively. The self-reported original and residual mortgage maturity and original installment loan maturity are fairly accurate, with the percentiles 25 to 75 being in the ranges of $[-1.3\%, 4.1\%]$, $[-13\%, 15.4\%]$ and $[0.0\%, 30.6\%]$, respectively.

Also notable in Table 4 is the correlation coefficients between $Y^R_{i,t,Registry}$ and $Y^R_{i,t,Survey}$. The correlation coefficients are all positive with values between 40% to 70% for the debt amounts across different loan types. The correlation coefficients for the installment loan and mortgage debt service are 66% and 79%, respectively. There is also strong correlation of 59% and 67% for the residual and the original mortgage maturity. These high correlation coefficients say that the survey self-reported data can be understood as a strong signal of the real value of the loan and its maturity. Therefore, even if the survey data is not exactly accurate, it is possible for researchers to use it as a valuable source of information, albeit one that suffers from measurement error (Schennach 2016). The maturity of installment loans has a lower correlation between survey self-reports and registry, but the correlation is still positive.

One reason for the discrepancies could be due to rounding errors in survey self-reports. For instance, the respondent may answer 1.5 million pesos as his or her loan value instead of answering 1,647,150 pesos. This could lead to substantial measurement error, although the survey responses can be seen as an approximation of a truthful report. For this reason I compare the survey self-reports with a rounded version of the registry by estimating the rounded discrepancy ratio ($error\ ratio_{i,t}$) using:

3) $Y^R_{i,t,Registry} \equiv \arg\min_{\hat{Y}_{i,t,Registry} \in \{10 \times \mathbb{Z}\}} \left|\hat{Y}_{i,t,Registry} - Y^R_{i,t,Survey}\right|$ s.to. $\left|\hat{Y}_{i,t,Registry} - Y^R_{i,t,Registry}\right| \leq \frac{1}{4} Y^R_{i,t,Registry}$.

Table 4: Summary of the discrepancies between the registry and the survey loan reports: percentiles and linear correlation coefficients

| $error\ ratio_{i,t}$ (%) | P10 | P25 | P50 | P75 | P90 | $\rho_{Y(Registry,Survey)}$ |
|---|---|---|---|---|---|---|
| Mortgage original debt amount | -67.9 | -31.0 | -2.8 | 18.1 | 76.2 | 69.9 |
| Consumer debt original amount | -97.6 | -13.7 | 15.8 | 70.3 | 138.1 | 53.3 |
| Installment Loan original debt amount | -48.3 | -0.7 | 11.0 | 59.3 | 110.7 | 68.5 |
| Credit Lines and Cards' debt amount | -131.6 | -43.3 | 1.9 | 45.6 | 94.4 | 51.5 |
| Credit Cards' debt amount | -123.9 | -23.9 | 3.0 | 46.2 | 94.9 | 39.8 |
| Credit Lines' debt amount | -163.5 | -102.2 | -8.4 | 11.5 | 47.0 | 47.7 |
| Mortgage debt service | -66.5 | -37.0 | -16.4 | -6.4 | 12.2 | 79.3 |
| Original mortgage maturity | -28.0 | -1.3 | 1.4 | 4.1 | 24.8 | 66.6 |
| Residual mortgage maturity | -38.3 | -13.0 | -0.5 | 15.4 | 41.8 | 58.5 |
| Installment Loan debt service | -42.9 | -9.8 | 8.1 | 25.0 | 65.5 | 66.4 |
| Original maturity of installment loans | -25.9 | 0.0 | 5.4 | 30.6 | 70.3 | 23.7 |
| Residual maturity of installment loans | -74.3 | -28.6 | -2.8 | 45.2 | 82.9 | 19.0 |

This rounding function takes the closest rounded value in terms of a number that is an integer multiple of 10 ($\mathbb{Z}$ denotes the set of integer numbers, therefore $10 \times \mathbb{Z}$ denotes the set of integers multiples of 10), but with a rounding error less than one fourth of the original value: this means that 800,000 can be rounded to 1,000,000 but not to 500,000. Note that rounding above 25% of the original value is not allowed. Therefore, 651,000 can be rounded to 500,000 but 667,000 can no longer be rounded to 500,000. The algorithm chooses the amount of rounding that fits best with the survey report, as long as the rounding is below 25% of the amount. Note, however, that 25% is the largest amount of rounding allowed in the algorithm, but in most cases the best amount of rounding is achieved with lower values. The reason is that loans have very wide range of values, therefore the same rounding may not be appropriate for loans of all sizes and this is why the algorithm searches for the best rounding in a range that is proportional to the value of the loan.[15]

Table 5 summarizes the differences between the rounded registry and the survey reports. As expected, the range of the percentiles 25 and 75 for the $error\ ratio_{i,t}$ show discrepancies of much

[15] For instance, in the last few years the value of the Chilean currency has traded close to 1,000 pesos per USD. Therefore, a rounding of around 100,000 pesos could be a very large rounding error in terms of reporting a small credit card debt. However, in terms of reporting a mortgage, a rounding error of 100,000 pesos could be extremely accurate and even errors 10 times above that would still be considered accurate (say, the difference between reporting an amount of 301,000 USD as if it were 300,000 USD).

Table 5: Discrepancies between the registry and the survey loan reports (percentiles and linear correlation coefficients): after "rounding" the registry

| *error ratio*$_{i,t}$ (%) | P10 | P25 | P50 | P75 | P90 | $\rho_{Y(Registry,Survey)}$ |
|---|---|---|---|---|---|---|
| Mortgage original debt amount | -60.9 | -22.2 | 0.0 | 7.0 | 61.0 | 76.0 |
| Consumer debt original amount | -87.6 | -2.8 | 4.7 | 59.7 | 133.3 | 58.4 |
| Installment Loan original debt amount | -33.3 | 0.0 | 0.0 | 50.0 | 103.0 | 72.4 |
| Credit Lines and Cards' debt amount | -122.3 | -33.3 | 0.0 | 28.6 | 85.7 | 56.7 |
| Credit Cards' debt amount | -117.4 | -15.4 | 0.0 | 35.3 | 85.7 | 43.1 |
| Credit Lines' debt amount | -160.0 | -85.7 | 0.0 | 0.0 | 35.3 | 51.8 |
| Mortgage debt service | -58.4 | -27.2 | -6.5 | 0.0 | 4.5 | 81.7 |
| Original mortgage maturity | -22.2 | 0.0 | 0.0 | 0.0 | 22.2 | 69.4 |
| Residual mortgage maturity | -31.4 | -9.5 | 0.0 | 11.8 | 38.9 | 71.1 |
| Installment Loan debt service | -31.2 | -3.9 | 0.0 | 14.3 | 53.2 | 72.8 |
| Original maturity of installment loans | -4.7 | 0.0 | 0.0 | 28.6 | 66.7 | 30.7 |
| Residual maturity of installment loans | -66.7 | -18.2 | 0.0 | 40.0 | 66.7 | 23.9 |

smaller magnitude, with this range falling to just $[-22\%, 7\%]$ and $[0\%, 50\%]$ for mortgages and installment loan debt amounts, respectively. Also, after rounding more than 50% of the original mortgage maturity observations could be fully explained with zero discrepancies relative to the registry. More than 50% of the observations for installment loan maturities have less than 28.6% discrepancy between survey and registry. This means that a simple error in which respondents round their answers to a multiple of 10 can go a long way towards explaining the discrepancies in the survey data. Therefore survey answers can be usefully interpreted as an approximate version of the real indebtedness of the borrowers. However, rounding does not reduce much the discrepancies in terms of the survey reports of debt amounts in credit cards and credit lines. This makes sense, since the survey interview asks only for credit card and credit line loans that are meant to be paid in periods longer than one month. Since most debt in credit cards and credit lines is revolving debt that is paid every month, survey data can differ substantially from the registry.

Finally, I analyze the possibility that excluding a few outliers can help explain the discrepancies between survey data and registry. In particular, I present a truncated analysis by discarding the values of the *error ratio*$_{i,t}$ below the percentile 5 and above the percentile 95. This analysis discards the 10% most extreme values. Table 6 shows the summary statistics of the discrepancies between the individual borrower survey reports and their registry. The results show that discarding

Table 6: Discrepancies between the registry and the survey
loan reports (percentiles and linear correlation coefficients): after truncating outliers

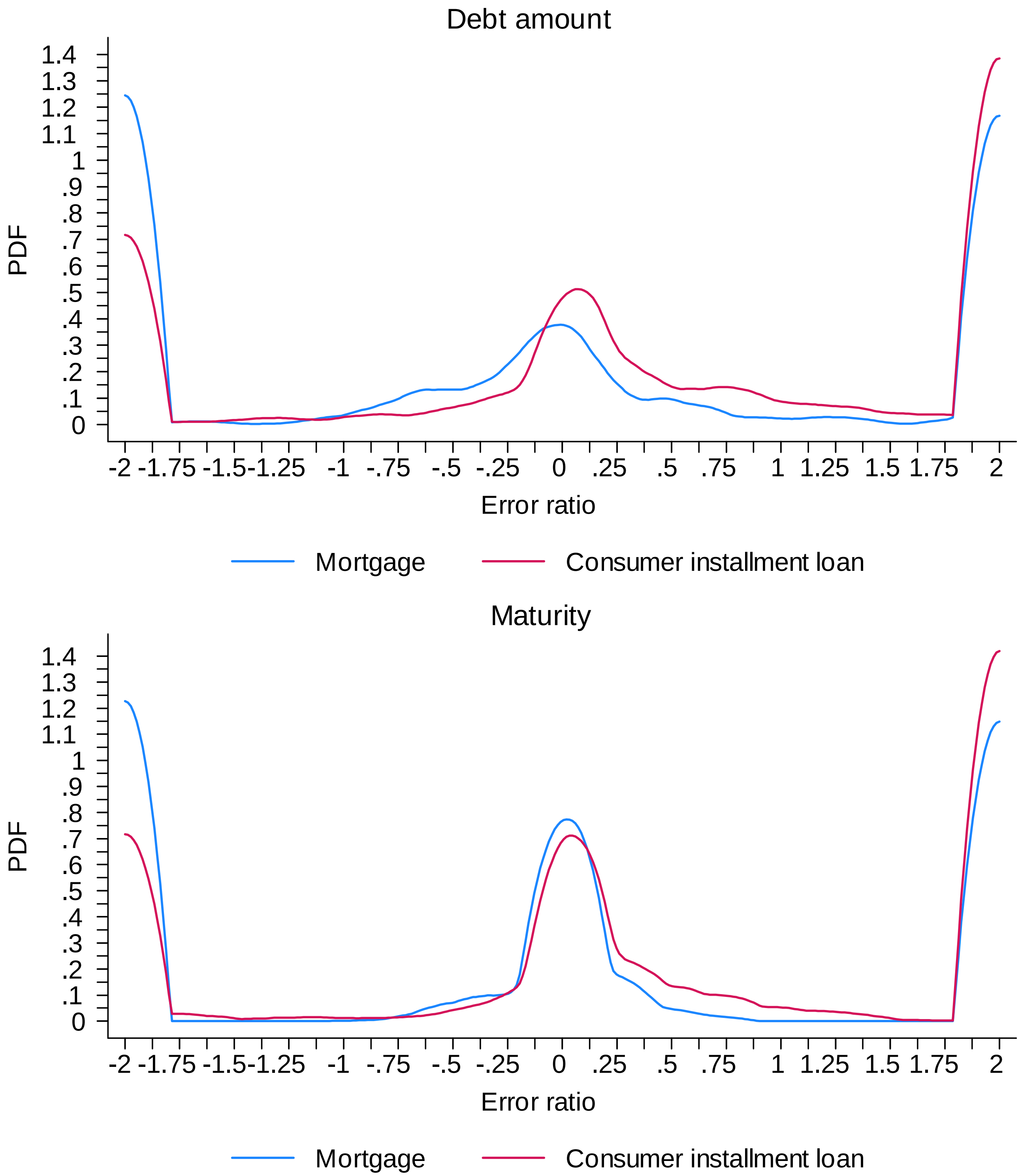
| $error\ ratio_{i,t}$ (%) | P10 | P25 | P50 | P75 | P90 | $\rho_{Y(Registry,Survey)}$ |
|---|---|---|---|---|---|---|
| Mortgage original debt amount | -64.1 | -27.9 | -2.6 | 14.6 | 63.9 | 57.3 |
| Installment Loan original debt amount | -49.2 | -2.2 | 8.0 | 42.6 | 98.6 | 66.4 |
| Mortgage debt service | -55.7 | -27.3 | -14.9 | -5.7 | 8.1 | 75.3 |
| Original mortgage maturity | -28.6 | -1.3 | 1.7 | 3.8 | 24.2 | 69.0 |
| Residual mortgage maturity | -28.6 | -12.9 | -0.8 | 14.3 | 36.8 | 72.7 |
| Installment Loan debt service | -47.1 | -13.4 | 3.3 | 18.0 | 41.6 | 79.9 |
| Original maturity of installment loans | -27.3 | 0.0 | 4.1 | 28.6 | 70.3 | 21.7 |
| Residual maturity of installment loans | -66.7 | -28.6 | -2.8 | 40.0 | 80.0 | 23.5 |

outliers reduces the measured discrepancies by a minimal amount relative to the analysis in Table 4. Therefore, while accounting for rounding errors goes a long way towards explaining the discrepancies, the existence of a small percentage of outliers does not change the results significantly.

Figure 5 shows the distribution of differences between survey and registry, as given by the $error\ ratio_{i,t}$. The figure shows that the distribution probability of errors for either the debt amount or the maturity are centered close to zero, with the probability of large errors falling significantly for values further away from zero. For the debt amount, the main types of error are the mis-reporting of debt participation. In the case of installment loans, the most common error is for a borrower not reporting a banking debt that is actually registered in the administrative system,. For mortgages the borrowers often report mortgages in the survey that do not exist in the banking system (possibly, because those are mortgages from non-bank lenders). The mis-reporting of the continuous debt amounts in installment loans is more balanced towards borrowers reporting lower amounts in the survey. For mortgages it is the opposite and survey respondents tend to report larger amounts than they really have in the administrative system.

The misreporting of maturities differs qualitatively and quantitatively from the debt amounts. Again, in the case of installment loans the most common error is for a borrower not reporting a banking debt that is actually registered in the administrative system. For mortgages the borrowers often report mortgages in the survey that do not exist in the banking system. However, in terms of the continuous error distribution, it is easy to observe that the reporting of maturities is much more accurate than for the loan amounts and the errors for maturities are quite tightly centered

Figure 5: Probability density distribution of the differences between the matched survey-administrative records and the survey for mortgage and consumer installment loans: original debt amounts and loan maturities



around zero. Figure 5, therefore, confirms that the reporting of loan maturities in the survey is quite precise. The main survey errors are mis-reporting of debt participation.

## 5 Explaining discrepancies

Besides rounding errors, I test other possible explanations for the differences between the survey and registry. One hypothesis is that the borrower reports as his or her own some loans that actually belong to another household member, such as a spouse or parent that signed the mortgage. This is more likely to happen in households with more members, especially if there is more than one member of high income and with access to banking loans (Brown et al. 2015). I test this hypothesis by looking at the percentage of borrowers that are not the highest-income member or the oldest member across different loan types, comparing the results if a positive loan is reported in both datasets versus if a loan is reported in the survey only. Table 7 shows that borrowers with reported loans in the survey but not in the registry are much more likely not to be the highest-income or oldest member in the household. Also, some households may have more than one member that is the highest-income member of the household. This can happen in cases where, due to rounding, both members of a couple report the same income value, such as one million pesos. Table 7 shows that borrowers with reported loans only for the survey have more cases of households with multiple members of the highest-income. The average number of highest-income members is 1.4 for households that only report loans in the survey.

Table 8 tests the same hypothesis in a multivariate logit regression for the event of borrowers reporting a mortgage or installment loan only for the survey. As an additional explanation, I test whether households could confuse a mortgage with an installment loan: for instance, the borrower takes a loan for repairs in the house, but it is classified as a consumer installment loan and not as a mortgage. Therefore the regressions also add as a control variable for whether the borrower has another loan that is actually reported in the registry. The regressions in Table 8 support all these hypotheses for the case of mortgage loans, but for the case of installment loans only the additional control variable of allowing the borrower to confuse loans is significant.

Finally, in Table 9 I test whether the same controls can explain the dispersion in the continuous discrepancy ratio for the debt amounts of each loan type. This shows that higher discrepancies in

Table 7: Borrowers that report bank loans in the survey, but have no loans in the registry

| | % of borrowers who are not the highest-income member | | % of borrowers who are not the oldest member | |
|---|---|---|---|---|
| Loan participation (Survey, Registry) | (1,0) | (1,1) | (1,0) | (1,1) |
| Mortgage loan | 35.4 | 10.6 | 49.4 | 29.2 |
| Consumer debt | 31.9 | 14.6 | 39.4 | 31.1 |
| Installment Loan | 16.8 | 14.6 | 35.6 | 33.0 |
| Credit Lines and Cards | 30.5 | 16.9 | 51.6 | 30.1 |
| Credit Cards | 25.6 | 18.7 | 53.7 | 32.7 |
| Credit Lines | 25.5 | 17.2 | 36.9 | 20.0 |

| Number of household members with the highest-income (average): | | |
|---|---|---|
| Loan participation (Survey, Registry) | (1,0) | (1,1) |
| Mortgage loan | 1.42 | 1.20 |
| Installment loan | 1.41 | 1.36 |

Table 8: Loan exists in the survey, but not in the registry (logit)

| $Logit(Survey = 1, Registry = 0)$ | Mortgage loans | Installment loans |
|---|---|---|
| Dummy for borrower that | 1.042*** | 0.0484 |
| is not highest-income member$_i$ | (0.187) | (0.275) |
| Number of household members | 0.527*** | 0.0316 |
| with the highest-income$_i$ | (0.151) | (0.162) |
| Dummy for borrower that | 0.343** | 0.0847 |
| is not the oldest member$_i$ | (0.159) | (0.215) |
| Dummy for another installment | -1.215*** | -1.099*** |
| or mortgage loan in registry$_i$ | (0.170) | (0.370) |
| Constant | -0.619*** | -0.769*** |
| | (0.211) | (0.235) |
| Number of Observations | 1,627 | 1,050 |
| Pseudo R-squared | 0.142 | 0.026 |

Robust standard errors in ().

*, **, ***, statistical significance at the 10%, 5% and 1%, respectively.

Table 9: Linear regressions (OLS) of the *error ratio*$_{i,t}$ for the debt amounts of borrowers with positive loans in survey and registry (Survey=1, Registry=1)

| Variables | Mortgage | Consumer debt | Installment loan | Credit Card | Credit Line |
|---|---|---|---|---|---|
| Dummy: borrower not | 16.70* | 12.96** | 8.505 | 22.15*** | -2.839 |
| highest-income member$_i$ | (9.855) | (5.450) | (7.125) | (7.165) | (12.92) |
| Nr of members with | 3.698 | -0.789 | -2.579 | -1.204 | 14.18* |
| the highest-income$_i$ | (5.992) | (3.914) | (3.899) | (4.590) | (8.371) |
| Dummy: borrower not | -6.815 | 0.896 | 4.625 | 9.917* | 16.63* |
| the oldest member$_i$ | (5.893) | (4.453) | (5.266) | (5.069) | (9.803) |
| Constant | 39.25*** | 64.84*** | 49.30*** | 48.87*** | 41.87*** |
| | (8.195) | (5.558) | (5.595) | (6.047) | (10.45) |
| Observations | 427 | 1,488 | 740 | 940 | 370 |
| R-squared | 0.017 | 0.007 | 0.006 | 0.038 | 0.036 |

Robust standard errors in ().

*, **, ***, statistical significance at the 10%, 5% and 1%, respectively.

mortgages and credit cards are associated with borrowers who are not the highest-income member of their family.

## 6 Implications for the analysis of debt amounts and debt risk

This section analyzes the aggregate implications of survey misreporting and the advantages of the matched administrative-survey dataset for the analysis of banking debt amounts and risk in Chile. Household finance surveys are used for the analysis of debt risk and household stress tests (Eurosystem Household Finance and Consumption Network 2009). This section shows how mismeasurement of borrower debt affects the evaluation of debt service to income ratios. The use of debt service to income ratios is widespread across many countries, being one element used by regulators to assess the risk and provisions of loans (Tiongson et al. 2010).

Table 10 shows that in aggregate terms the measures of bank debt in the survey and registry are quite similar, with a difference of 8.8%. However, there are large differences between mortgage and

Table 10: Aggregate banking debt amounts in the registry and the survey datasets (in thousands of millions of pesos*)

| | Survey | Registry | Difference (in %) |
|---|---|---|---|
| Total bank debt | 45,885 | 42,029 | -8.8 |
| Bank mortgage debt | 35,490 | 28,547 | -21.7 |
| Bank consumer debt | 10,394 | 13,482 | 25.9 |

* Note: In 2024, one USD is roughly equivalent to 950 pesos. Therefore, a thousand of million pesos is roughly similar to a million of US dollars.

Table 11: Fractions of borrowers (in %) considered to be at risk

| Debt risk (Survey, Registry) | (0,0) | (1,0) | (0,1) | (1,1) |
|---|---|---|---|---|
| All borrowers | 68.1 | 4.6 | 20.1 | 7.2 |

consumer debt, with a respective difference of 21.7% and 25.9% between both datasets. Consumer debt has no collateral and is subject to rollover risk due to its short maturity. Therefore, it is regarded as riskier for banks. The registry has a much larger amount of consumer debt, while showing lower mortgage debt. This is an indicator of a riskier banking system in relation to a pure survey analysis.

The use of debt-service-to-income ratios is widespread across many countries, being one element used by regulators to assess the risk and provisions of loans (Tiongson et al. 2010). An international comparison of macroprudential policy frameworks with IMF data (Alam et al. 2019) shows that 55 countries (in a total sample of 135 countries) have some sort of debt-service-to-income regulation. A standard rule of thumb used by banks and regulators is that borrowers with debt-service-to-income ratios above 40% are considered risky. Table 11 shows the differences between the survey-banking registry and the survey borrowers' analysis of risk status. Around 7.2% of bank borrowers are considered to be risky in the matched survey-banking registry. However, 20.1% of borrowers are evaluated as risky in the registry, but not in the survey. The matched administrative and survey dataset, therefore, helps to obtain a more precise image of the risks in the banking system. Around 4.6% of borrowers are classified as risky according to the survey, but not in the registry. This discrepancy is smaller than the previous one. This confirms that surveys may underestimate household debt risks.

## 7 Conclusions

It is well-known that financial information self-reported in surveys differs from aggregates in administrative sources. This article advances on the literature by looking at a unique match of individual borrowers from the Chilean Household Finance Survey (EFH) with the registry of banking loans. This allows to study the distribution of differences between the datasets: whether the data differ in terms of the number of borrowers, whether borrowers understates debts and the degree of heterogeneity in the discrepancies. While there are substantial differences in the reports of loan participation, I find that reports of loan delinquency are broadly accurate. There is substantial heterogeneity in the discrepancy between debt amounts of individual borrowers in survey data and registry, especially for consumer loans. A substantial fraction of these discrepancies, however, can be explained by rounding errors, in which survey respondents reply with values that are multiples of 10.

Another interesting finding is that the loans that are reported in one dataset but not in the other tend to be much smaller than the median, especially in the case of credit cards and lines of credit. Finally, I find that mis-reporting of loans in the survey is strongly associated with households that are complex in the sense that they have several members with the same high income level. Therefore, mis-reporting could be due to loans that are unknown to the interviewee. Discrepancies between survey and registry are also associated with borrowers who are not the highest-income or the oldest member of the family. These borrowers could report in the survey certain loans that they took but which were in fact signed by the spouse, parent or another family member.

Household finance surveys are increasingly important in the analysis of borrowers' behaviors such as delinquency or credit exclusion. Further research on the strengths and shortcomings of survey financial data is thus warranted.

## References

[1] Ampudia, M., H. van Vlokhoven and D. Zochowski (2016), "Financial Fragility of Euro Area Households," *Journal of Financial Stability*, 27, 250-262.

[2] Avery, R., G. Elliehausen and A. Kennickell (1988), "Measuring Wealth with Survey Data: An Evaluation of the 1983 Survey of Consumer Finances." *Review of Income and Wealth*, 34(4), 339–369.

[3] Avery, R., P. Calem, G. Canner and R. Bostic (2003), "An Overview of Consumer Data and Credit Reporting." *Federal Reserve Bulletin*, 89(2), 47-73.

[4] Bertaut, C., M. Haliassos and M. Reiter (2009), "Credit Card Debt Puzzles and Debt Revolvers for Self Control," *Review of Finance*, 13(4), 657–692.

[5] Bhandari, A., S. Birinci, E. McGrattan and K. See (2020), "What Do Survey Data Tell Us about US Businesses?," *American Economic Review: Insights*, 2(4), 443-458.

[6] Bover, O., J. Casado, S. Costa, P. Du Caju, Y. McCarthy, E. Sierminska, P. Tzamourani, E. Villanueva and T. Zavadil (2016), "The Distribution of Debt across Euro-Area Countries: The Role of Individual Characteristics, Institutions, and Credit Conditions," *International Journal of Central Banking*, 12(2), 71-128.

[7] Bound, J., C. Brown and N. Mathiowetz (2001), "Measurement Error in Survey Data." In J. Heckman and E. Leamer, eds., *Handbook of Econometrics*, vol. 5.

[8] Brown, M., A. Haughwout, D. Lee and W. van der Klaauw (2015), "Do We Know What We Owe? Consumer Debt as Reported by Borrowers and Lenders," *Economic Policy Review*, 21(1), 19-44.

[9] Bucks, B. and K. Pence (2008), "Do Borrowers Know Their Mortgage Terms?" *Journal of Urban Economics*, 64(2), 218-33.

[10] Calem, P., M. Gordy and L. Mester (2006), "Switching Costs and Adverse Selection in the Market for Credit Cards: New Evidence," *Journal of Banking and Finance*, 30, 1653–85.

[11] Caplin, A. (2025), "Data Engineering for Cognitive Economics," *Journal of Economic Literature*, 63(1), 164–96.

[12] Christelis, D., D. Georgarakos and M. Haliassos (2011), "Stockholding: Participation, location, and spillovers," *Journal of Banking and Finance*, 35(8), 1918-1930.

[13] Christelis, D., D. Georgarakos and M. Haliassos (2013), "Differences in Portfolios across Countries: Economic Environment versus Household Characteristics," *Review of Economics and Statistics*, 95(1), 220-236.

[14] Christelis, D., M. Ehrmann and D. Georgarakos (2017), "Exploring Differences in Household Debt Across the United States and Euro Area Countries," CSEF Working Papers 465..

[15] Dynan, K., and D. Kohn (2007), "The Rise in U.S. Household Indebtedness: Causes and Consequences." FEDS Working Paper No. 2007-37, Board of Governors of the Federal Reserve System.

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

[17] Fesseau, M., F. Wolff y M. Mattonetti (2013), "A cross-country comparison of household income, consumption and wealth between micro sources and national accounts aggregates," OECD.

[18] Georgarakos, D., M. Haliassos and G. Pasini (2014), "Household Debt and Social Interactions," *Review of Financial Studies*, 27(5), 1404-1433.

[19] Gerardi, K., K. Herkenhoff, L. Ohanian and P. Willen (2018), "Can't Pay or Won't Pay? Unemployment, Negative Equity, and Strategic Default," *Review of Financial Studies*, 31(3), 1098–1131.

[20] Giustinelli, P., C. Manski and F. Molinari (2022), "Tail and center rounding of probabilistic expectations in the Health and Retirement Study," *Journal of Econometrics*, 231(1), 265-281.

[21] Goldman M. and D. Kaplan D. (2018), "Comparing distributions by multiple testing across quantiles or CDF values," *Journal of Econometrics*, 206, 143–166.

[22] Gross, D. and N. Souleles (2002), "Do Liquidity Constraints and Interest Rates Matter for Consumer Behavior? Evidence from Credit Card Data." *Quarterly Journal of Economics*, 117(1), 149–85.

[23] Hurst, E., G. Li and B. Pugsley (2014), "Are Household Surveys Like Tax Forms? Evidence from Income Underreporting of the Self-Employed," *Review of Economics and Statistics*, 96(1), 19-33.

[24] Kaplan, D. (2019), "distcomp: Comparing distributions," *Stata Journal*, 19(4), 832-848.

[25] Karlan, D. and J. Zinman (2008), "Lying About Borrowing." *Journal of the European Economic Association*, 6(2-3), 510–21.

[26] Kavonius, I. and V. Törmälehto (2010), "Integrating Micro and Macro Accounts – The Linkages Between Euro Area Household Wealth Survey and Aggregate Balance Sheets for Households", mimeo, IARIW.

[27] Le Blanc, J., A. Porpiglia, F. Teppa, J. Zhu and M. Ziegelmeyer (2015), "Household Saving Behaviour and Credit Constraints in the Euro Area," ECB Working Paper No. 1790.

[28] Lusardi, A. and P. Tufano (2009), "Debt literacy, financial experiences, and overindebtedness," NBER WP 14808.

[29] Madeira, C. and B. Zafar (2015), "Heterogeneous inflation expectations and learning", *Journal of Money, Credit and Banking*, 47(5), 867-896.

[30] Madeira, C. (2018a), "Testing the rationality of expectations of qualitative outcomes," *Journal of Applied Econometrics*, 33(6), 837-852.

[31] Madeira, C. (2018b), "Explaining the cyclical volatility of consumer debt risk using a heterogeneous agents model: The case of Chile," *Journal of Financial Stability*, 39, 209-220.

[32] Madeira, C. (2019a), "Measuring the covariance risk of consumer debt portfolios," *Journal of Economic Dynamics and Control*, 109, 21-38.

[33] Madeira, C. (2019b), "The impact of interest rate ceilings on households' credit access: evidence from a 2013 Chilean legislation", *Journal of Banking and Finance*, 106, 166-179.

[34] Madeira, C. (2024), "Indebtedness and labor risk sorting across consumer lender types in Chile," *Journal of Banking Regulation*, https://doi.org/10.1057/s41261-024-00250-1.

[35] Meriküll, J. and T. Rõõm (2020), "Stress Tests of the Household Sector Using Microdata from Survey and Administrative Sources," *International Journal of Central Banking*, 16(2), 203-248.

[36] Schennach, S. (2016), "Recent Advances in the Measurement Error Literature," *Annual Review of Economics*, 8(1), 341-377.

[37] Sierminska, E., A. Brandolini and T. Smeeding (2006), "Comparing wealth distribution across rich countries: First results from the Luxembourg Wealth Study." LWS Working Paper 1, Luxembourg Income Study, Luxembourg.

[38] Stantcheva, S. (2023), "How to Run Surveys: A Guide to Creating Your Own Identifying Variation and Revealing the Invisible," *Annual Review of Economics*, 15(1), 205-234.

[39] Stiglitz, J., A. Sen and J. Fitoussi (2009), "Report by the Commission on the Measurement of Economic Performance and Social Progress," IMF/FSB.

[40] Tiongson, E. R., N. Sugawara, V. Sulla, A. Taylor, A. I. Gueorguieva, V. Levin and K. Subbarao (2010), "The Crisis Hits Home: Stress-Testing Households in Europe and Central Asia," Washington DC: World Bank.

[41] Törnqvist, L., P. Vartia and Y. Vartia (1985), "How Should Relative Changes Be Measured?," *The American Statistician*, 39(1), 43-46.

[42] Wooldridge, J. (2010), "Econometric Analysis of Cross Section and Panel Data," MIT Press.

[43] Zinman, J. (2009), "Where Is The Missing Credit Card Debt? Clues And Implications," *Review of Income and Wealth*, 55(2), 249-265.

# How accurately do consumers report their debts in household surveys?

Online appendix

March 2025

**Abstract**

This is an online appendix to the article. Section 1 explains the sampling and field operations of the Chilean Household Finance Survey. Section 2 summarizes the questionnaire in terms of the debt questions. Section 3 shows additional summary statistics of the data. Section 4 shows the results for the tests of the equality of the debt distributions in the survey and registry datasets.

JEL Classification: C81; D10; D12; E21; G21.

Keywords: Household Finance Surveys; mortgages; consumer credit; default; measurement error.

## 1 Sampling and operations of the Household Finance Survey

For this study I use the national waves of the EFH 2011, 2014 and 2017, which covered 13,110 urban households.[1] Each survey sample was collected over roughly an entire calendar year, with the EFH 2011 being collected between July 2011 and May 2012, while the EFH 2014 was collected between July 2014 and March 2015, and the EFH 2017 was collected between June and December 2017. However, for simplicity the survey waves are labelled as EFH 2011 and 2014, which corresponds to the year in which the largest portion of their respective samples was collected. Each sampled household had one member who was selected for the interview, with this member being the household person with the greatest knowledge of the family finances or the highest income. The survey, however, elicits demographic, net wealth, asset, debt and income information for all the household members. The sample selection of the survey was based on an exhaustive list of homes from the Chile Internal Tax Service and is therefore representative of the national urban population after expansion factors are applied to each unit (Madeira 2018b).[2]

The survey questionnaire is very close to the Household Finance Consumption Survey implemented by the national central banks in Europe and coordinated by the European Central Bank (ECB), with the Central Bank of Chile having also participated in several meetings of the ECB's network. The first wave of the EFH at a national level was implemented in 2007. The first EFH questionnaire was largely based on the Spanish Household Finance Survey (which started in 2002) and is, therefore, quite similar to the European household finance surveys. Furthermore, between 2007 and 2010, the

[1] There were EFH waves in 2007, 2008, 2009 and 2010. However, the survey provider in these waves did not record the month of the interview due to privacy and anonimity concerns. Therefore these previous waves cannot be used to test whether self-reported loans in the survey are similar or not to the administrative records in the same monthly period or within a window of one or two months.

[2] Note that the list of homes of the Chilean Internal Tax Service includes all homes in Chile (all homes have a taxable value evaluation) and therefore the survey is representative of the Chilean urban population. The EFH survey sample is a sample of homes, whether its residents own, rent or even illegally occupy the property. It is an exhaustive set of all homes in urban areas of Chile, whether its occupiers are tax payers or not.

Rural homes were not selected because those homes represent a small share of the population and are costlier to reach in terms of the transportation of the interviewers. Therefore, the rural population is outside of the sampling universe. According to estimates by the Institute of National Statistics (INE, from the Spanish, *Instituto Nacional de Estadísticas*), the urban population in Chile for the year 2021 was 17,118,329 people, which is around 87.8% of the total national population.

survey preparation team used training materials quite similar to the Bank of Spain in its field staff capacitation. Between 2014 and 2017, the company Ipsos led the field staff training and survey interview operations, having applied similar training and materials as Ipsos does for its household finance surveys in Canada and other countries.

The questionnaire of the survey was based on an exhaustive study of the questionnaires of the Survey of Consumer Finances in the US and of previous survey experiences in Italy and the Netherlands, as well. Afterwards, there were EFH surveys in the Santiago capital region (which represents more than 40% of the national GDP and population) in the years 2008, 2009 and 2010. These three surveys were limited to the capital region due to budget limitations.[3] The survey waves prior to 2011 are not included in the analysis of this article, because those surveys did not elicit the respondents' national id numbers.

Like the European household finance surveys, the questionnaire is divided into several topic sections.[4] The survey has 12 topic sections: A) household structure, B) education, C) employment, D) payment methods and services, E) real estate assets and mortgages (main home, other properties), F) non-mortgage debts, G) perceptions of debt service and credit restrictions, H) vehicles and other real assets, I) financial assets, pensions and insurance, J) non-labor income and other earnings, K) expectations of future income, L) contact information and willingness to participate in future survey waves. The largest sections are intended for the measurement of household loans, with section E containing around 85 questions (most of them related to mortgages) and section F inquiring about 45 questions.[5]

[3] Conducting a national survey with interviewers sent across a country that stretches for over 4,300 km from north to south is much more expensive than simply covering a capital city. Furthermore, household finance surveys ask several sensitive financial questions, which require all interviewers to undergo several days of training before the field work starts.

[4] The survey questionnaires of the EFH and the datasets for each wave (but not the confidential matched datasets made with the Banking Loan System) are available at the website of the Central Bank of Chile: https://www.bcentral.cl/en/areas/encuestas-economicas/acceso-a-efh.

[5] The F section is particularly long, because several of the 45 questions are repeatedly applied to 13 distinct types of non-mortgage loans, namely: bank credit cards, bank lines of credit, retail store credit cards, bank installment loans, retail store installment loans, labor unions and credit cooperative loans, auto loans, education loans, loans from family and friends, pawnshops, store credit, loan-shark debts, other informal debts.

## 2 Questionnaire description

In the case of home purchase loans (including mortgages and some non-mortgage loans), these were the questions asked for the main home and up to three other real estate properties that the household may own.[6] The questionnaire explicitly starts with all the loans (mortgage and non-mortgage loans) for the main home of the household (that is, where the members of the household live). It then asks the same questions for the three largest other real estate properties of the household, with the order being given from the highest-value real estate property first to the lowest-value one last.

"When your household purchased your home, how was the purchase financed? 1) Self resourced (own savings), 2) House subsidy, 3) Mortgage from a financial institution, 4) Non-mortgage (example: consumer loan or complementary house purchase loan) loan from a financial institution, 5) Transfers corresponding to part or the totality of the property (example: gift or inheritance), 6) Loan from relatives or friends".

"For each one of the 3 different home loan types (options 3, 4, 6), let us talk about your home credit. With which institution did you obtain your mortgage loan at the time of the house purchase? 1) State loan, 2) Bank or Financial Institution (which bank or financial institution?), 3) Labor union, 4) Insurance company, 5) Cooperative or credit union, 6) Another institution (which institution? please specify), 8) Does not answer, 9) Does not know".

"What was the amount of your house loan at the time of the purchase, whether in pesos or in UF?[7] For the cases in which the respondent does not answer or does not know, please show him or her the card with a range of 19 possible intervals: 1) Less than 20 UF, 2) Between 21 and 80 UF, 3) 81 to 150 UF, 4) 151 to 250 UF, 5) 251 to 500 UF, 6) 501 to 1000 UF, 7) 1001 to 1500 UF, 8) 1501 to 2000 UF, 9) 2001 to 2500 UF, 10) 2501 to 3700 UF, 11) 3701 to 5000 UF, 12) 5001 to 6500

[6] Among EFH mortgage borrowers, 92.9% have a mortgage on their main home, 18.5% have a mortgage on another property and 11.5% have mortgages both on their main home plus on another property. Some wealthy households can report up to 4 mortgages. In the EFH sample, 87.9% of the mortgage borrowers have only one mortgage, 10.7% have 2 mortgages, 1.1% have 3 mortgages, and 0.3% have 4 mortgages.

[7] The UF is a real monetary unit that is updated for inflation in Chile and it is widely used for many long term contracts, such as rents, mortgages, consumer loans, and wages. Between 2010 and 2019, the average value of the UF was 42.13 USD and fluctuated between 38.25 and 46.43 USD.

UF, 13) 6501 to 8000 UF, 14) 8001 to 10000 UF, 15) 10001 to 12500 UF, 16) 12501 to 15000 UF, 17) 15001 to 20000 UF, 18) 20001 to 25000 UF, 19) More than 25000 UF."

"Are you still paying this loan? Is your loan still ongoing?"

"For how many years have you asked this loan for?"

"How many years and months of repayment did you complete?"

"How many years and months do you still have to repay?"

It is worth emphasizing that the real estate and mortgage questions appear first, while the non-mortgage loan sections appear afterwards. For non-mortgage loans, the survey asked the following questions:

"For each of these loan types 1) Bank credit cards, 2) Banking lines of credit, 3) Retail store cards or non-bank credit cards, 4) Retail store or non-bank consumer installment loans (example: store cash advancement), 5) Bank or financial institution consumer installment loan (do not include other loans already reported in previous sections, such as the home complementary loans), 6) Cooperatives, labor unions or other similar institutions, 7) Auto loans, 8) Educational debts, 9) Loans from relatives or friends, 10) Credit from money lenders and shark loans, 11) Pawn shops, 12) Tab credit (store tab or store account loan), 13) Debts from other sources (for example: employer). Please specifiy.

Do you or any other member of the household have a credit or debt of the above mentioned types?

How many loans of each type does the household have? Now for each type of loan, please elicit the following information for the three largest household loans (that is, the three loans with the highest debt amount), with the order being given from the largest loan first to the third-largest loan last.

Which member of the household owns this debt?

Which institution is the lender for this debt?

For what amount was this loan for (at the time you contracted the loan)? For the cases in which the respondent does not answer or does not know, please show him the card with a range of 17 possible intervals ($ indicates Chilean pesos): 1) $0 to $30,000; 2) $30,001 to $50,000; 3) $50,001 to $120,000; 4) $120,001 to $250,000; 5) $250,001 to $500,000; 6) $500,001 to $1,000,000; 7) $ 1,000,001 to $ 2,000,000; 8) $ 2,000,001 to $ 3,500,000; 9) $ 3,500,000 to $ 5,000,000; 10)

$ 5,000,001 to $ 7,500,000; 11) $ 7,500,001 to $10,000,000; 12) $10,000,001 to $15,000,000; 13) $15,000,001 to $20,000,000; 14) $20,000,001 to $30,000,000; 15) $30,000,001 to $80,000,000; 16) $80,000,001 to $190,000,000; 17) Over $190,000,000.

For how many years have you asked this loan for?

How many years and months of repayment did you complete?

How many years and months do you still have to repay?

What is the monthly amount that you must pay each month as a dividend (interest plus amortization and other fees) for this loan?"

## 3 Summary statistics of the matched survey-registry data

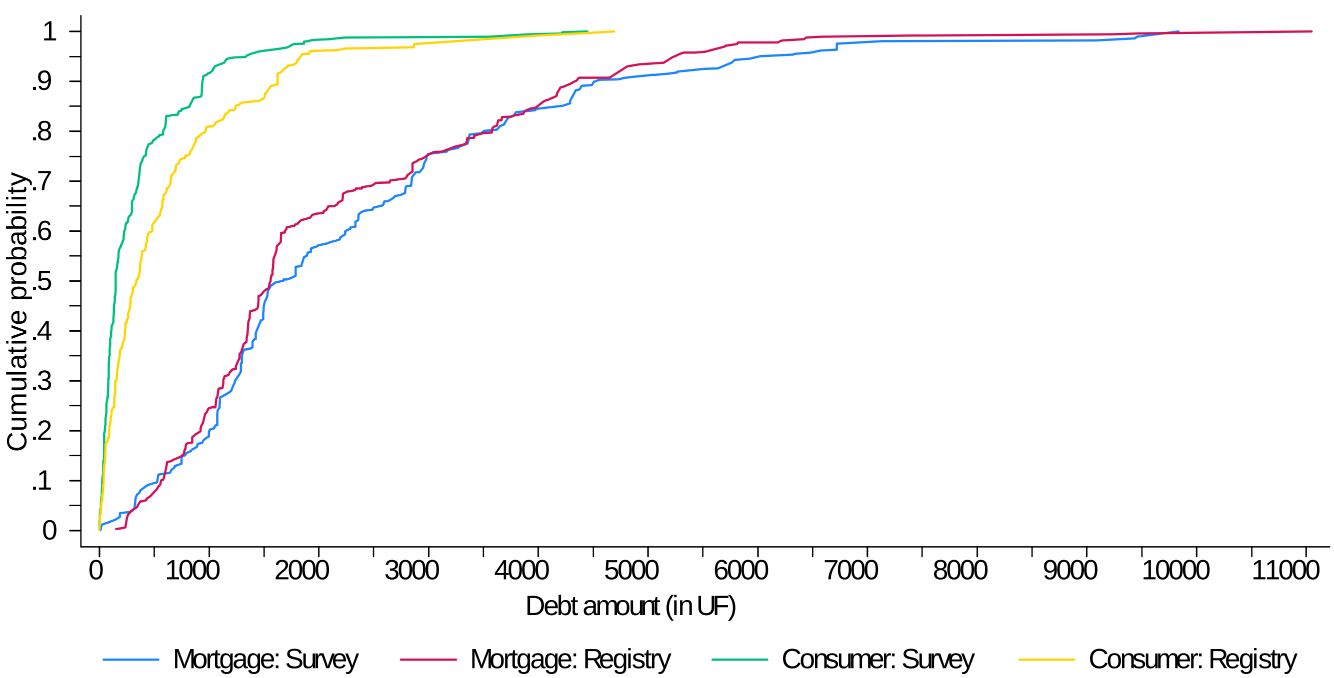
Table A.1 summarizes the univariate distributions of the loans in either the survey and registry datasets (with positive amounts in either dataset) for the household sample of this study. In particular, I look at the original debt amount (which is the question reported by survey respondents), the total monthly debt service and the loan maturity. I look at the total amounts for each debt category: mortgages, installment loans, credit cards, and credit lines.[8] The results show that the survey and registry report similar loan amounts in terms of mortgages, credit cards and credit lines. Both datasets also show similar amounts of mortgage debt service, although the survey borrowers tend to report lower mortgage maturities. The survey borrowers tend to report smaller debt amounts and debt service for consumer installment loans and for their overall consumer debt (which includes installment loans, credit cards and credit lines), although their self-reports for the loan maturity tend to be similar to the registry data.

Finally, I summarize whether it happens that the loan amounts are small when the survey data and registry differ in terms of debt ownership. It may happen that, for instance, households do not report loans that are too small in the survey and therefore both datasets would differ in terms of the number of borrowers, but mostly because small and insignificant loans are ignored. Table A.2 summarizes the mean and median distribution of the loan amounts of each type for the matched survey-banking registry sample, separating the cases with positive debt in both survey

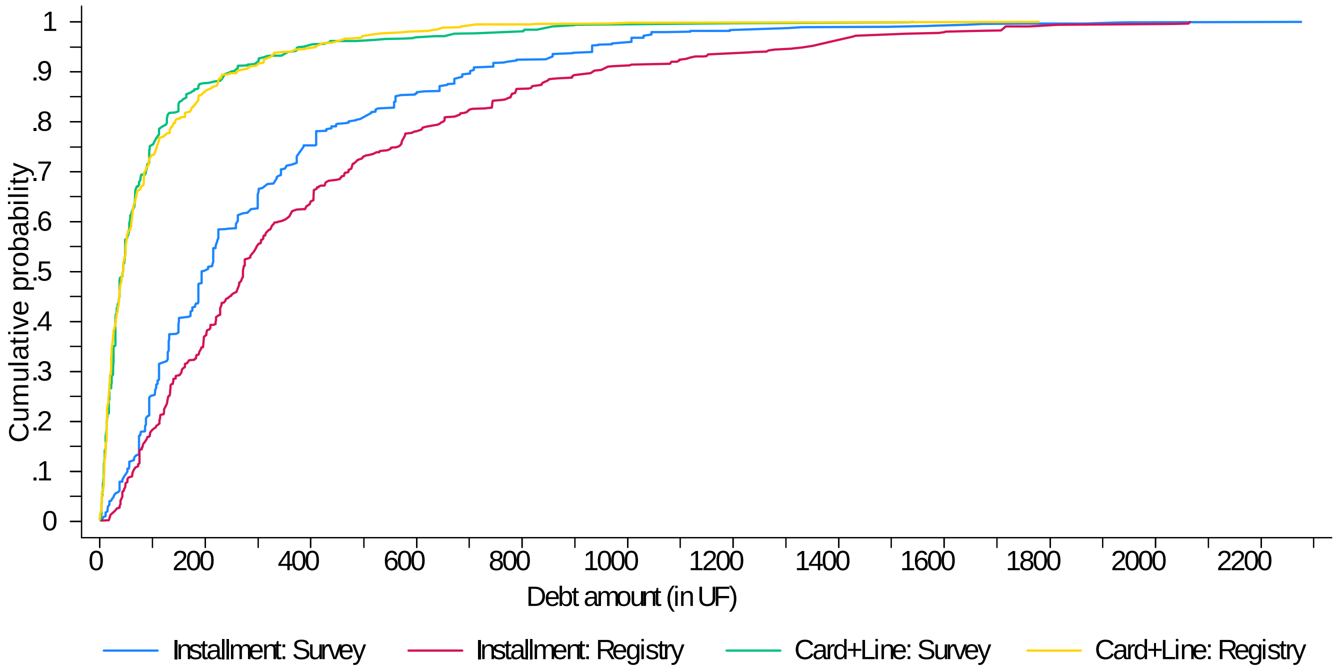
[8] If a borrower has more than one loan in a certain category, then we report the total debt amount and debt service, plus the maturity weighted by the debt amount of each loan.

and registry and the cases with only one datasource reporting a positive loan amount. It is worth noticing that the survey has large mortgages which are not reported in the registry dataset. This could happen because some mortgages in the survey can come from non-banks, such as insurance companies, although borrowers report them as banking loans. However, it is indeed true that when the registry reports a debt that the interviewer did not report in the survey, then the corresponding amount for mortgage loans plus the loan amounts in credit cards and credit lines is much smaller than usual. This lends support to the idea that some households simply do not report loans in their interviews, because they consider those loans as unimportant.

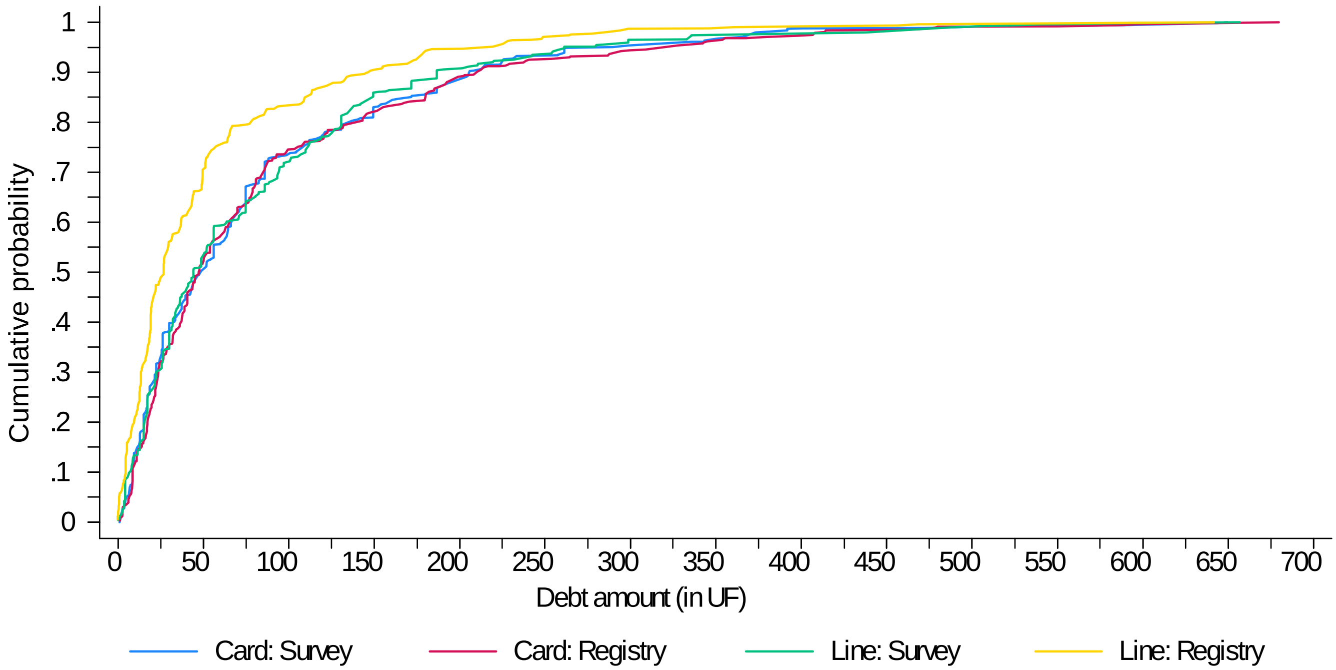
Since Table A.2 is limited to the mean and median, I also show the cumulative distribution function (CDF) of the debt amounts in the survey and registry datasets when the debt ownership coincides. That is, these figures exclude loans which are reported in only one of the datasets. The results are similar to the figures in the article, which include all positive debt values in the survey and registry. Therefore, debt ownership is not the single source of discrepancy between the survey and registry. The datasets differ both in the extensive margin (debt ownership reports) and the intensive margin (loan amounts for each debt category).



CDFs of the original mortgage and consumer debt amounts in the survey and registry (when positive amounts of each category of debt are present in both datasets)




CDFs of the original consumer installment loans and revolving debt instruments (credit cards and lines of credit) in the survey and registry (when positive amounts of each category of debt are present in both datasets)




CDFs of the debt amounts for credit cards and lines of credit in the survey and registry (when positive amounts of each category of debt are present in both datasets)

## 4 Tests of the equality of the survey and registry loan distributions

This section shows the results of two non-parametric tests of the equality of the cumulative probability distributions (CDFs) of the debt categories in the survey and registry datasets. The first one is the Kolmogorov–Smirnov test for whether two underlying one-dimensional probability distributions differ, which is based on the supremum of the difference between the two CDFs. Goldman and Kaplan (2018) provide an alternative procedure that distributes power across the distribution more evenly than the Kolmogorov-Smirnov test, which suffers low sensitivity to tail deviations. I show the results of these two tests for the equality of the debt distributions in the survey and registry datasets. Table A.3 shows the test results for the debt ownership distributions across each loan category. Table A.4 shows the test results for the positive debt distributions across each loan category. Finally, Table A.5 shows the test results for the equality of the distribution of loan maturities in the survey and registry datasets.

Table A.3 shows that the data does not reject that the debt ownership of mortgages has the same distribution for survey and registry, according to either test. However, for the other debt categories, the two tests differ, with the Kolmogorov-Smirnov test rejecting that the distributions of debt ownership are similar, while the Goldman-Kaplan test does not reject that debt ownerships are the same.

Table A.4, however, shows that the continuous distributions of positive debt values are rejected to be the same at the 5% level or lower across all debt categories. Therefore, both the Kolmogorov-Smirnov and the Goldman-Kaplan tests reject that survey and registry data are similar in the continuous debt distributions.

Finally, Table A.5 shows that both the original maturity (that is, the maturity at the beginning of the loan contract) and the residual maturity (the months still left to pay at the time of the survey) are rejected to be similar in survey and registry datasets, whether for mortgages or consumer installment loans.

Goldman M. and D. Kaplan D. (2018), "Comparing distributions by multiple testing across quantiles or CDF values," *Journal of Econometrics*, 206, 143–166.

Table A.1: Univariate distributions (percentiles 25, 50, 75) of the loan amounts, debt service and maturities for the borrowers in the survey-banking registry data

| Loan variable (positive amounts) | Survey | | | | Registry | | | |
|---|---|---|---|---|---|---|---|---|
| Registry>0, Survey>0 | P25 | P50 | p75 | Obs | P25 | P50 | p75 | Obs |
| Mortgage original debt amount* | 16.7 | 17.3 | 18.0 | 866 | 16.7 | 17.3 | 17.8 | 784 |
| Consumer debt original amount* | 13.5 | 14.6 | 15.6 | 2,054 | 13.8 | 15.0 | 15.9 | 2,786 |
| Installment Loan original debt amount* | 14.4 | 15.0 | 15.9 | 1,046 | 14.6 | 15.4 | 16.1 | 1,412 |
| Credit Lines and Cards' debt amount* | 12.6 | 13.8 | 14.7 | 1,511 | 12.7 | 13.7 | 14.7 | 2,378 |
| Credit Cards' debt amount* | 12.6 | 13.5 | 14.5 | 1,309 | 12.6 | 13.4 | 14.4 | 2,222 |
| Credit Lines' debt amount* | 12.9 | 13.8 | 14.9 | 562 | 12.6 | 13.4 | 14.4 | 941 |
| Mortgage debt service* | 11.7 | 12.3 | 12.9 | 1,774 | 11.5 | 12.1 | 12.6 | 786 |
| Original mortgage maturity (months) | 212 | 240 | 240 | 1,323 | 232 | 244 | 301 | 789 |
| Residual mortgage maturity (months) | 96 | 156 | 216 | 1,324 | 101 | 167 | 228 | 789 |
| Installment Loan debt service* | 11.4 | 12.0 | 12.6 | 1,019 | 11.5 | 12.1 | 12.8 | 1,420 |
| Original maturity of installment loans (months) | 24 | 36 | 48 | 1,112 | 31 | 44 | 51 | 305 |
| Residual maturity of installment loans (months) | 12 | 24 | 36 | 1,095 | 14 | 26 | 40 | 239 |

* Logarithm of the amount in Chilean pesos.

Table A.2: Univariate loan amounts (mean and medians) in the survey and registry datasets when binary loan status coincides or diverges

| * Loan amounts in logarithm | Survey | | | Registry | | |
|---|---|---|---|---|---|---|
| Ownership debt status in Survey, Registry | (1,0) | (1,1) | Obs | (0,1) | (1,1) | Obs |
| | | | Mean | | | |
| Mortgage original debt amount* | 17.0 | 17.2 | 866 | 16.7 | 17.2 | 784 |
| Consumer debt original amount* | 14.4 | 14.5 | 2,054 | 12.8 | 14.8 | 2,786 |
| Installment Loan original debt amount* | 15.0 | 15.1 | 1,046 | 15.0 | 15.4 | 1,412 |
| Credit Lines and Cards' debt amount* | 13.7 | 13.7 | 1,511 | 11.9 | 13.6 | 2,378 |
| Credit Cards' debt amount* | 13.6 | 13.5 | 1,309 | 11.9 | 13.5 | 2,222 |
| Credit Lines' debt amount* | 13.2 | 13.8 | 562 | 10.4 | 13.2 | 941 |
| | | | Median | | | |
| Mortgage original debt amount* | 17.1 | 17.3 | 866 | 16.5 | 17.2 | 784 |
| Consumer debt original amount* | 14.5 | 14.6 | 2,054 | 13.1 | 15.0 | 2,786 |
| Installment Loan original debt amount* | 15.1 | 14.9 | 1,046 | 15.0 | 15.4 | 1,412 |
| Credit Lines and Cards' debt amount* | 13.6 | 13.7 | 1,511 | 12.2 | 13.7 | 2,378 |
| Credit Cards' debt amount* | 13.5 | 13.5 | 1,309 | 12.2 | 13.4 | 2,222 |
| Credit Lines' debt amount* | 13.3 | 13.8 | 562 | 11.0 | 13.4 | 941 |

Table A.3: P-values (in %) for the equality of the survey and registry distributions of debt ownership (binary variable)

| Debt type | Kolmogorov-Smirnov | Goldman-Kaplan |
|---|---|---|
| Mortgages | 78.3 | 100 |
| Consumer debt | 0.0 | 100 |
| Installment loans | 0.0 | 100 |
| Cards and lines of credit | 0.0 | 100 |
| Lines of credit | 0.0 | 100 |
| Credit cards | 0.0 | 100 |

Table A.4: P-values (in %) for the equality of the survey and registry distributions of positive debt values

| Debt type | Kolmogorov-Smirnov | Goldman-Kaplan |
|---|---|---|
| Mortgages | 2.1 | 0.0 |
| Consumer debt | 0.0 | 0.0 |
| Installment loans | 0.0 | 0.1 |
| Cards and lines of credit | 0.0 | 0.0 |
| Lines of credit | 0.0 | 0.0 |
| Credit cards | 0.0 | 0.0 |

Table A.5: P-values (in %) for the equality of the survey and registry distributions of loan maturities

| Maturity | Kolmogorov-Smirnov | Goldman-Kaplan |
|---|---|---|
| Mortgage: original | 0.0 | 0.0 |
| Mortgage: residual | 0.1 | 0.2 |
| Installment: original | 0.0 | 0.0 |
| Installment: residual | 0.0 | 0.0 |

Table A.6: Differences between the self-reported survey data and registry of borrower status (binary loan variable): Sample of respondents with none or only one banking loan type in the survey or in the registry

| Debt participation (Survey, Registry) - % | (0,0) | (0,1) | (1,0) | (1,1) |
|---|---|---|---|---|
| Mortgage loan | 83.7 | 1.6 | 10.5 | 4.2 |
| Consumer debt | 64.5 | 17.5 | 7.2 | 10.8 |
| Consumer installment loan | 84.4 | 7.3 | 4.2 | 4.2 |
| Credit Card or Credit Line | 71.7 | 16.9 | 5.4 | 5.9 |
| Credit Card | 73.9 | 15.9 | 5.1 | 5.1 |
| Credit Line | 91.5 | 5.5 | 2.3 | 0.7 |

Number of observations: 7,016 households (all the rows sum up to 7,016 observations of respondents with or without positive debt).

Table A.7: Differences between the self-reported survey data and registry of borrower status (binary loan variable): Sample of respondents with only one banking loan type in the survey or in the registry

| Debt participation (Survey, Registry) - % | (0,0) | (0,1) | (1,0) | (1,1) |
|---|---|---|---|---|
| Mortgage loan | 62.3 | 2.7 | 23.7 | 11.4 |
| Consumer debt | 24.2 | 36.7 | 11.0 | 28.2 |
| Consumer installment loan | 69.1 | 12.9 | 6.9 | 11.0 |
| Credit Card or Credit Line | 41.8 | 35.2 | 7.4 | 15.5 |
| Credit Card | 46.4 | 33.5 | 6.8 | 13.3 |
| Credit Line | 86.5 | 9.1 | 2.7 | 1.8 |

Number of observations: 2,682 households (all the rows sum up to 2,682 observations of respondents with or without positive debt).

## Previous volumes in this series



All volumes are available on our website www.bis.org.