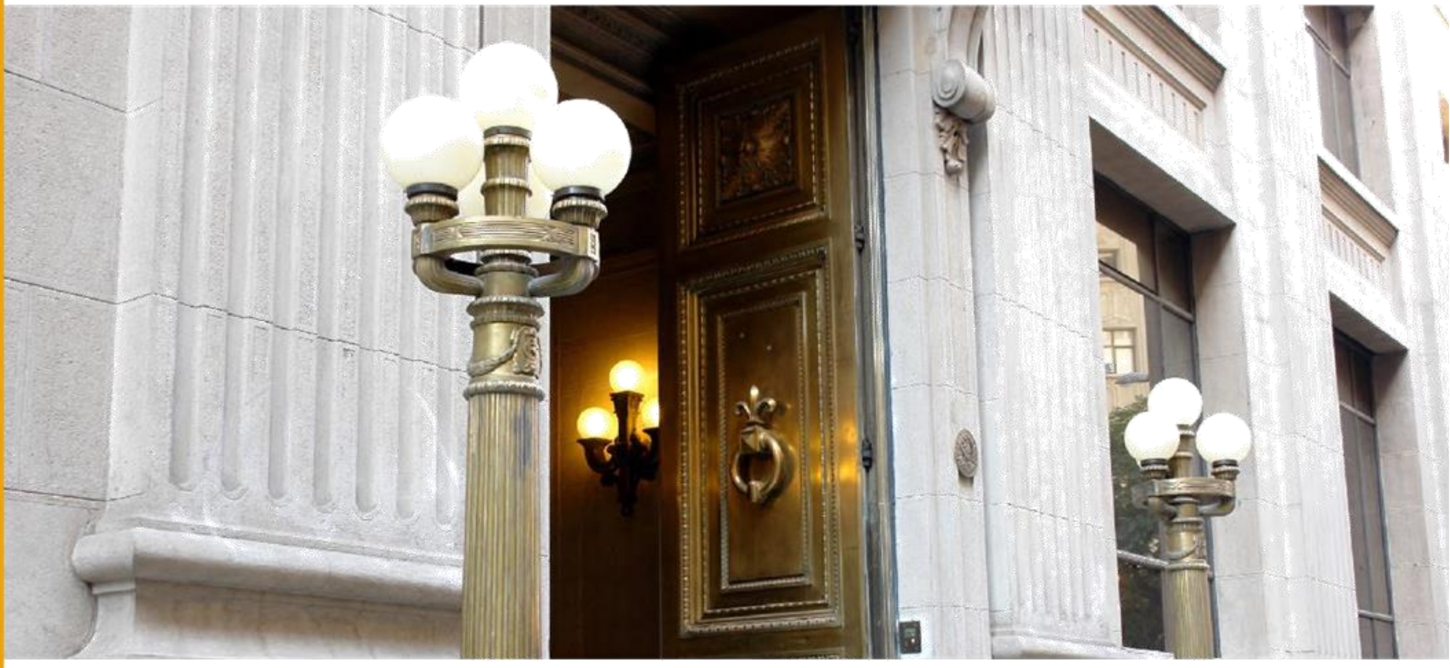
DOCUMENTOS DE TRABAJO

How accurately do consumers report their debts in household surveys?

Carlos Madeira

N° 1069 Marzo 2026

BANCO CENTRAL DE CHILE



Documentos de Trabajo del Banco Central de Chile
Working Papers of the Central Bank of Chile
Agustinas 1180, Santiago, Chile
Teléfono: (56-2) 3882475; Fax: (56-2) 38822311

Documento de Trabajo N° 1069 | **Working Paper N° 1069**

How accurately do consumers report their debts in household surveys?

Carlos Madeira*

Resumen
Este artículo hace una comparación entre los datos financieros de una muestra representativa de individuos prestatarios de la Encuesta Financiera de Hogares (EFH) de Chile y sus registros de préstamos bancarios provenientes de la anterior Superintendencia de Bancos e Instituciones Financieras (SBIF). Se encuentra que la encuesta difiere del registro de crédito, no solo en el número de préstamos reportados, sino también en su monto, con un grado sustancial de heterogeneidad. El estado de morosidad es reportado con precisión por los encuestados. Además, una fracción considerable de las discrepancias puede explicarse por errores de redondeo en las respuestas de la encuesta. Finalmente, encuentro que las discrepancias son mayores cuando los encuestados no son la persona de mayor ingreso en la familia.

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, both in the number of loans reported and the debt amounts, 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.

* Central Bank of Chile, cmadeira@bcentral.cl. I would like to thank Claudio Raddatz, 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 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, Gathergood and Olafsson 2024). Surveys on finances are 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 represents more than 25% of the labor force. However, surveys 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. The analysis assumes that the CMF banking registry dataset can be used as a high quality external measure for validating the survey information. [2]

This microdata 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).

Finally, I show explanations for the differences between survey and administrative records, including: i) whether survey and administrative records differ because of a time lag relative to when the administrative records are updated and the period in which the interview happens; ii) whether households round their answers and this gives a discrepancy relative to the exact value in the administrative records; iii) whether households with several members with an equal share of income may have more mis-reporting of values.[3]

There are several novel contributions in this work. The first contribution is studying the differences between survey and debt registry at the borrower level and not simply in terms of the aggregate sum of all loans (Zinman 2009, Brown et al. 2015, Bhandari et al. 2020). The second contribution is expanding the literature of individual measurement error in surveys, which has focused on unemployment, income and social programs (Abowd and Stinson 2013, Meyer et al. 2015, Meyer and Mittag 2019, Meyer and Mittag 2021), by adding a study for banking debt data. The third contribution is analyzing explanations for the discrepancies between survey and

[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)).

[2] Note that in practice it is difficult to find any dataset that is error-free. However, the CMF registry has instances where initial reporting errors can be corrected in later stages, which should reduce the frequency of errors.

[3] For instance, a working couple could have one member being responsible for the loan, but with the interviewed member reporting the loan as his own.

loan registry, including time lags of records, rounding, retail loans not included in the registry or assuming the loan responsibility of a different household member.[4] The fourth contribution is by showing that the differences between survey and registry may lead to an underestimation of household debt risk.

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.

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,

[4] Note that while rounding has been studied before in the context of other social surveys such as expectations (Giustinelli et al. 2022), the role of rounding for surveys of household debt and their comparison with registry data has not been studied before.

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 through the use of new technologies (such as financial aggregators and eased sharing of information) 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, Meriküll and Rõõm 2020) and more robust measures of interpreting survey responses (Lusardi and Tufano 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. Several countries have the survey and credit registry datasets required for doing an analysis similar to this article, 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 micro-data applications, such as firms (Elosegui et al. 2024), labor markets (Madeira and Salazar 2023), pensions (Cerletti et al. 2025, Madeira 2024) and expectations (Giustinelli et al. 2022). Furthermore, it completes previous research on how financial literacy improves survey reporting (Madeira and Margaretic 2022, Madeira 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 registry 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 in Chile implemented every 3 years, with around 4,000 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).

All interviews are held in person. 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.[5] Therefore, the

[5]Note that this work does not deal with non-bank debts of households. Other works, such as Madeira 2023 and Madeira 2025, study the choice of different debt types of the household, including bank, retail stores, union and cooperatives credit, other non-bank lenders, a mix of banks and retail, while controlling for credit constrained

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. Previous research shows that age, education, income, household size (given by a function of the number of adult members and children of the family) and unemployment risk are significant controls that explain the choices of different loan types (Madeira 2023, 2025). This paper focuses on banking loans and their report accuracy in the survey, therefore non-bank lenders are omitted in the analysis.[6]

The EFH survey asks for the identity of the household member that is responsible for each loan, with the question: "Which member of the household owns this debt?". In the analysis of this article, I only include the survey loans that were identified as being under the responsibility of the interviewed person (who is providing the real ID for match with the banking authority's registry). This allows to compare for the interviewed persons reported loans in the survey and the administrative records.

The EFH survey has a rotating panel structure, where each sample comes from a probabilistic two-stage sampling design. In order to better capture the behavior of households with the highest participation in the financial markets and with the largest share of total assets, the survey oversampled the wealthiest 20% of households in the population. The survey includes homes of all quintiles of taxable value, but there is a higher number of the top quintile of homes (i.e., the most expensive homes). The identification of the richest households is based on the assessed value (by the Chilean Internal Tax Service) of the property they live in, regardless of being renters or owners. Homes which are not interviewed (either by refusal or because they are not home after several attempts of contact) are replaced by homes of similar tax value, which ensures that the sample remains representative and that there is no tendency towards poorer households. This type of sample

household and households with no wish for debt. These works control for demographics such as age, education, household size, and also time-varying economic factors such as income, unemployment risk and year dummies that account for other macroeconomic shocks happening in the survey year. Income and unemployment risk are both heterogeneous factors across households that are also varying over time during the business cycle.

[6] The focus of this paper is only on the reporting of banking loans in the survey and administrative data. Banks in Chile follow rigorous Know your customer (KYC) and Anti money laundering (AML) regulations. Furthermore, informal workers are unlikely to obtain banking loans, since banks in Chile demand an employment contract and proof of income before deciding the loan application. For these reasons, only high income customers get accepted for banking loans (Madeira 2025) and this would exclude most of the low-income self-employed workers. For this reason, I do not analyze whether illegal, black market or informal earnings affect banking debt in the sample data.

design is also used in the Survey of Consumer Finances of the United States and in the European Household Finance and Consumption Survey (Madeira et al. 2022).

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 study the evolution of household indebtedness over time, the Central Bank of Chile and Chilean Banking Authority (SBIF) built a survey-banking registry dataset, where the survey's information is linked to the monthly banking credit information for each month over the period January 2003 to December 2018.[7] 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.[8] 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 and prevents mismatching.[9]

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

[7] This study was developed within the scope of the research agenda conducted by the Central Bank of Chile (CBC) in economic and financial affairs of its competence. The CBC has access to anonymized information from various public and private entities, by virtue of collaboration agreements signed with these institutions.

[8] To secure the privacy of workers and firms, the CBC mandates that the development, extraction and publication of the results should not allow the identification, directly or indirectly, of natural or legal persons. Officials of the Central Bank of Chile processed the disaggregated data. All the analysis was implemented by the authors and did not involve nor compromise the *Comisión para el Mercado Financiero* (CMF) or its predecessor institution, *Superintendencia de Bancos e Instituciones Financieras* (SBIF).

[9] 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.

(whether a loan, a current account or a savings account).[10][11] Table 1 summarizes how many households are observed in each dataset.[12] There are around 13,110 households in the survey dataset (waves 2011, 2014, 2017), with 8,047 respondents having both given a correct national ID number and being matched in the registry banking loan dataset.[13] 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.[14][15]

[10]Note that the EFH survey covers the universe of all urban households, independently of whether these applied for banking products or not. Note also that the banking registry includes all persons who ever applied for a banking product (even if they were refused). Therefore, people who applied for loans or any banking product (loan or account) are a part of the matched registry-survey dataset (even if the loan applications were rejected). This means that there is no bias in the matched registry-survey dataset against discouraged borrowers or credit constrained households.

[11]The EFH survey does ask households whether in the last 12 months: i) they applied for loans and were rejected, ii) they did not apply for loans due to fear of rejection, and iii) whether they did not apply for loans because they simply did not want to have debts. Madeira 2025 studies the household choice of lender types (banks, retail stores, social credit, other lenders, mix of bank and non-bank-lenders), while controlling for households who choose "no wish for debt" and households in "credit constrained" status due to being rejected for loans or fear of being rejected.

[12]Note that the number of observations in the article correspond to different households. All the analysis is made in terms of households, not in terms of the number of loans, since the loans of each category (mortgage, consumer installment loans, credit cards, lines of credit) are aggregated at the level of each household. This implies that all the results in the article are already implicitly clustered by households, since household is the unit of observation.

[13]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, survey companies gained experience about eliciting this sensitive information over time.

[14]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%. 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.

[15]There are some websites online that report real IDs of people. However, the interviews of the EFH survey are

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.

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.[16] 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 possibly a future 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.

always held in-person and it would appear a strange behavior for someone to check a website in order to report the real ID of someone else. Furthermore, there would be no monetary gains or social advantages of engaging in this kind of cheating about a real ID, since households can simply decline to answer the question if they wish so.

[16]Note that complex surveys like the EFH (and other similar household finance surveys in other countries) ask for a large amount of information from households, often more than 2,000 self-reported answers given by families. For this reason, what is examined in this study, is the accuracy of the answers given. That is, each debt answer in the sample is compared in accuracy by itself. There is no dis-regarding of the household observations or discarding of data just because one (or more) of the answers reported by an interviewed person is seen as inaccurate in the matched dataset. For instance, a household can report a debt value that has some discrepancy with the administrative records and yet other values reported by the same household can be accurate. What this article does is to compare the debt reports for different types of banking loans. Both the values with discrepancies and the values that coincide are useful to make an assessment of the factors that may explain the differences between survey and administrative records.

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 differ from more developed economies in terms of the use 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;[17] and iii) the matched survey-banking registry data provide information on the current loan amount, the original loan 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.[18]

Chile has a significant presence of non-bank lenders. In 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 2023, 2025). 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,[19] while 11.1% had union or cooperative debt[20] and

[17]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.

[18]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.

[19]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.

[20]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.

10.6% had other types of debt (vehicle or educational loans).[21]

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.[22]

All the software codes necessary for formatting the matched survey-registry data and replicating the empirical analysis of this article are publicly available in a Mendeley Data link.[23] 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;[24] 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).

[21] This work does not deal with non-bank debts of households. The reason is that only banking debt is part of the records used in the matched survey-registry dataset. Therefore, it is not possible to study the accuracy of non-bank loans. Previous research shows that some non-bank lenders, such as retail stores and social lenders (unions or cooperatives), offer loans of smaller amounts than banks, although retail stores have more customers than banks (Madeira 2023, 2025). However, some other non-bank lenders such as auto-lenders can make loans of sizes even larger than the average bank loan (Madeira 2023). The study of lender type choice, credit constraints, indebtedness and default behavior based on demographics, income, unemployment risk and aggregate factors is shown in Madeira 2023 and Madeira 2025.

[22] 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.

[23] https://data.mendeley.com/datasets/r3xjrx9grj/1

[24] The manual of the banking authority's registry dataset gives extensive definition of the variables requested from the banks about the loans, such as the maturity of the loans in months and interest rates in annual value. The survey data allows for households to report interest rates either in monthly or annual terms. The empirical analysis of this article has formatted the EFH survey to follow the same loan variable definitions of the banking authority's dataset to provide an adequate comparison. The Central Bank of Chile has formatted the EFH survey to have versions of the variables with definitions as close as possible to the ones used by regulators in the registry and other official sources.

Note that this article only deals with measurement error. Sampling error in surveys can be controlled by designing larger samples (Madeira 2019). Groves and Lyberg 2010 list five major sources of nonsampling error: specification error, frame error, nonresponse error, measurement error, and processing error. Nonresponse error can be separated into unit nonresponse (refusal to answer the entire interview) or item nonresponse (refusal to answer certain questions). In the case of the EFH survey, specification error and frame error are studied by conducting field pretests with 200 interviews in order to test the questionnaire. Furthermore, the in-person surveys are also a well established method to reduce specification error and frame error, because the interviewers are intensively trained for one week before the survey and the interviewers can clarify the doubts of the participating households. The unit nonresponse is corrected using postsurvey adjustments to the expansion factors based on statistics well estimated by other larger social surveys in Chile (Madeira 2019). Item nonresponse for loans is small, because the contract of the survey with the interviewers considers that only interviews with demographics, debt amounts, and some of the wealth questions are accepted. Interviews with many missing answers are rejected and unpaid by the Central Bank of Chile to incentivize complete interviews. Processing error in the EFH survey has been improved by using digital tablets to record interviews and giving incentives to survey operational supervisors to check the quality of the field interviewers. This implies that measurement error is maybe the single source of error in the EFH survey which has not been studied yet. Furthermore, the matched survey-registry dataset is ideally suited for the analysis of the measurement error problem for the case of banking debts.

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.[25]

[25]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)

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 loans, credit cards, credit lines) at the borrower level.[26]

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 $k = 0, -1, 1, -2, 2$).[27] I also take into account a large range of

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.

[26]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. Furthermore, if the analysis was made at the loan level, then borrowers with many loans would count much more than the other debtors.

[27]Note that the survey and administrative records may differ because the survey can happen in a different day of the month. Surveys are made on a single date of the month, which can include any day of the month and even weekends or holidays. The administrative records correspond to the last business day of the month. This means that the administrative records can record loans that were not yet taken at the time of the survey interview. There is also the case of a borrower who has already repaid a loan and therefore reports a zero amount, but his repayment is only processed in the following month. In this case the administrative records could still register a loan, but the borrower would correctly report no loan. On the other hand, the survey can also include loans that are not in the

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,[28] 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.[29]

administrative records. For instance, you can take a loan using online banking over the weekend and such a loan is only registered in the following business day which corresponds already to the next month. Another example is if the borrower repays the loan after the survey date and in this case the administrative records would show zero loans, while the survey would correctly report a positive loan amount. All these examples imply that a window of -1,0,+1 months around the survey interview is appropriate. However, the EFH survey started already in 2007, a time in which digital processing of loan information was poorer than today. One must also think that there is the loan information processing by the bank and then there is the information sent by the bank to the banking authority and the processing time lag until it is registered in the banking authority's registry. It was only in 2019 that the CMF forbid the registration of loan information by paper and requested only the digital information format. For this reason, I considered a window of -2,-1,0,+1,+2 around the interview date.

[28]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.

[29]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) - %	Not in survey Not in registry	Not in survey In registry	In survey Not in registry	In survey In registry	Difference in datasets
Mortgage loan	82.0	1.6	9.3	7.0	10.9
Consumer debt	63.1	14.1	5.6	17.2	19.7
Consumer installment loan	79.4	7.7	3.9	9.1	11.6
Credit Card or Credit Line	69.9	14.5	4.2	11.4	18.7
Credit Card	72.1	13.9	3.9	10.1	17.8
Credit Line	88.5	5.4	2.2	4.0	7.6

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

Table 2 summarizes the difference in 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 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). Table A.1 in the appendix shows that the Kolmogorov-Smirnov and Goldman-Kaplan tests do not reject the equality of debt participation binary reports in both datasets across any debt type (mortgage, consumer debt, installment loan, credit card, credit line).

Table 3: Differences between the self-reported survey data and registry of default status (binary variable of arrears of 1 month or more)

Default (Survey, Registry) - %	No default in survey No default in registry	Default only in registry	Default only in survey	Default in both datasets	Difference in datasets
All loans (including loans reported in one dataset only)					
Mortgage loan	97.7	0	0.8	1.5	0.8
Consumer installment loan	84.1	0.2	7.2	8.5	7.4
Including just the sample of borrowers with same debt participation in both datasets					
Mortgage loan	96.3	0	0	3.7	0
Consumer installment loan	87.2	0.2	0.3	12.2	0.6

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

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). The survey data reports several more loan delinquencies 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 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. There is a significant fraction of households reporting default only in the survey for consumer installment loans, around 7.2% of the sample. However, this fraction is reduced to just 0.3% when the analysis includes just the sample of borrowers with debt participation in both survey and the registry. Therefore, the explanation is that borrowers in the survey are reporting default for loans not in the banking registry. Perhaps this is due to confusion on whether the lender is a bank (many corporations own banks and retail store lenders with a very similar corporate brand name) or due to reporting another household member's loan as if it is under their responsibility.

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, this analysis only includes borrowers that have positive debt amounts in the survey or registry. Furthermore, all debt amounts are converted to UF.[30]

[30]UF is a real monetary unit 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 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: CDFs of the original mortgage and consumer debt amounts in the survey and registry

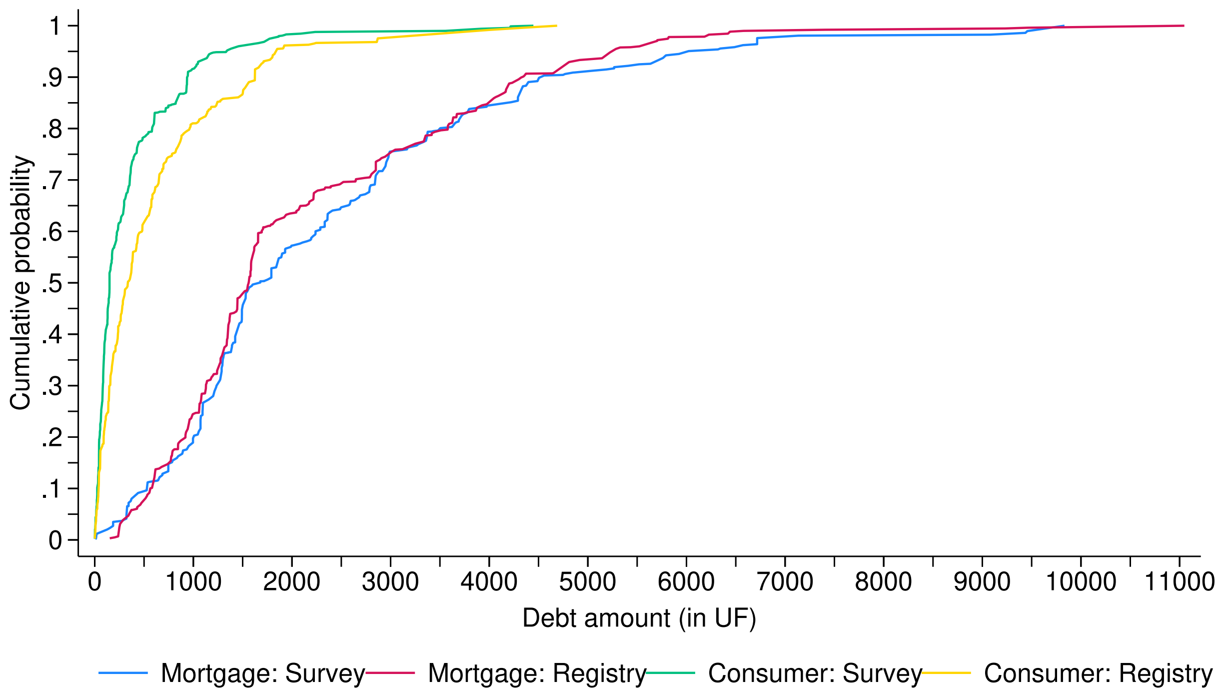
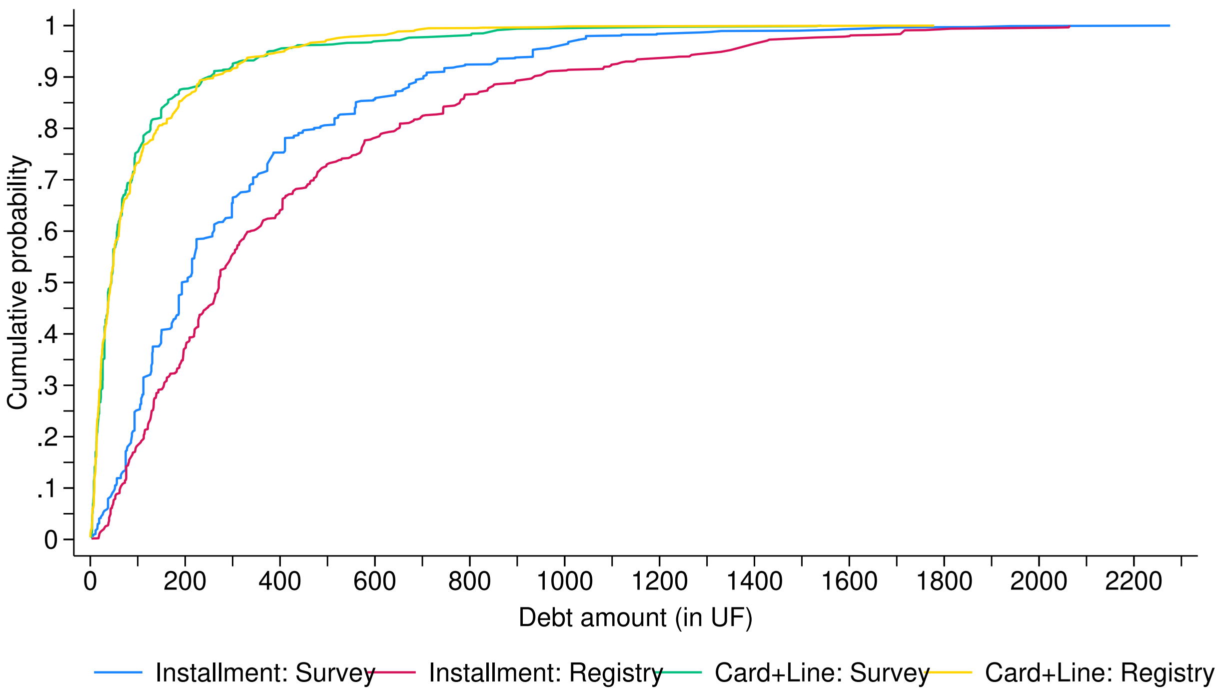


Figure 1 shows the CDFs for the total mortgage and total consumer banking loans of the matched survey-registry borrowers. In the case of mortgages, there is a fairly close match of the debt distribution in the survey and registry until around 1,500 UF (which is roughly the percentile 50 for both survey 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 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. Unlike for mortgages, for consumer loans the registry is much more complete in the inclusion of larger debts.

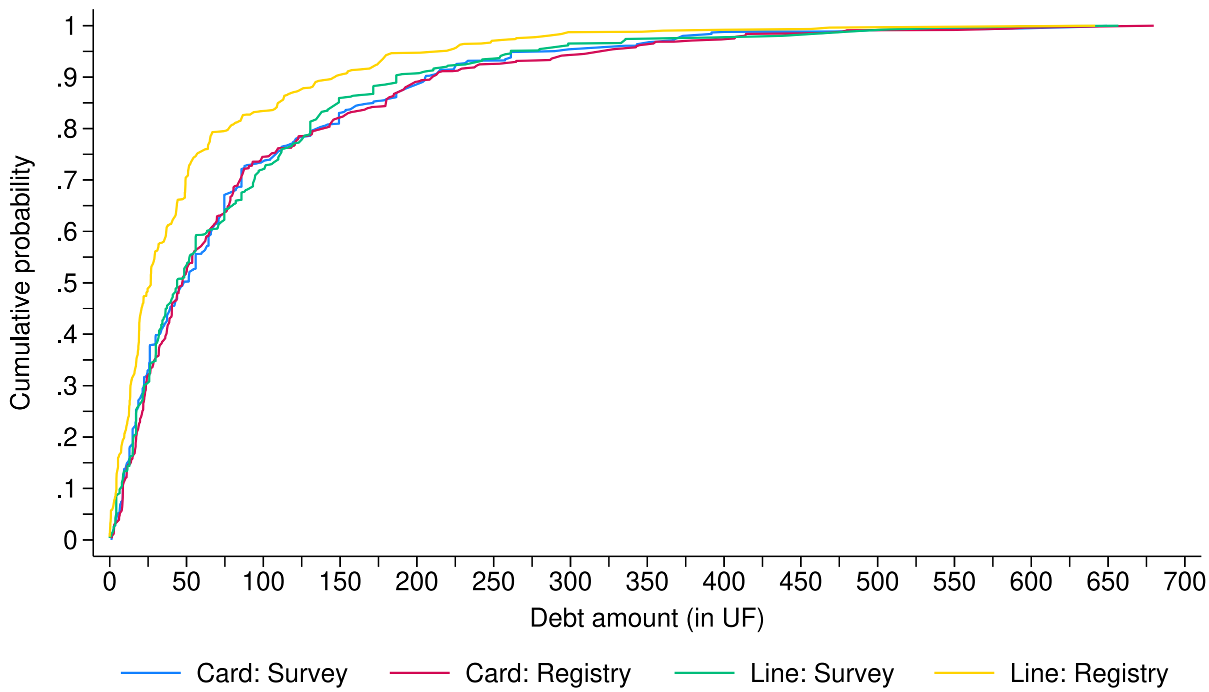
Figure 2: CDFs of the original consumer installment loans and revolving debt instruments (credit cards and lines of credit) in the survey and registry



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

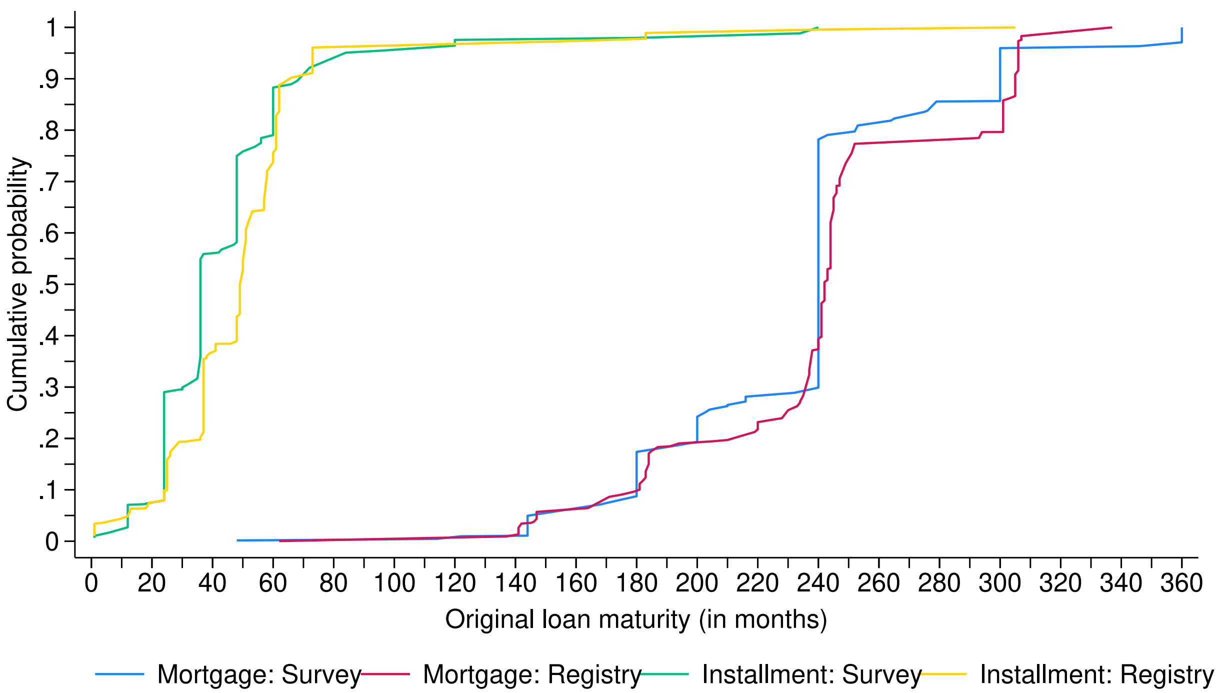
[31]The respondents are asked not to report debt that is due to transactions from credit cards and lines of credit as a means of payment (rather than intentional loans with maturities longer than one month). Note that loans that are meant to be repaid in the current month are still reported in the survey, as long as those loans had original maturities longer than one month or if those are intentional loan contracts with a one month maturity.

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



I then show the differences between survey and registry for 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. Table A.2 shows that both the Kolmogorov-Smirnov and the Goldman-Kaplan tests reject that survey and registry data are similar in the continuous debt distributions of all types of loans (mortgages, installment loans, credit cards, lines of credit) even at the 1% level.

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



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). Table A.3 shows that both the original and residual maturities are rejected to be similar in survey and registry datasets, whether for mortgages or consumer installment loans, even at the 1% statistical significance level.

4 Discrepancies in loan amounts and maturities

4.1 Continuous measure of discrepancy between registry and survey data

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 debt values from either source 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). Furthermore, this specification of the error ratio uses the average between registry and survey in the denominator. This prevents measurement error from either source to lead to unbounded errors and provides robustness to the empirical analysis (Davis et al. 2012).

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:[32][33]

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

[32] Note that the algorithm is only testing which is the minimum amount of difference between the periods -k to +k. This minimum amount can have several solutions (argmin solutions to the minimization problem), as the same loan can be reported across several periods. However, it is not necessary to report which is the argmin period that is providing the solution (as this can lead to multiple possible periods). The exercise is only reporting which is the best outcome for the minimum discrepancy (independent of whether there is a single solution or multiple solutions).

[33] Grace periods and loan restructuring are not part of the banking registry. The banking registry is a panel of borrowers and their banking debts, but it is not a panel of loans since there are no identifiers for loan contracts.

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 debt types and loan characteristics (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 reporting.

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.

The median error ratio in Table 4 shows that mortgage loan amounts tend to be higher in the survey than in the registry, although by a small margin of -2.8%. This could be due to interviewed persons perhaps adding banking fees such as insurance or other loans taken at the same timing of the mortgage in their mental accounting of the total mortgage amount. The median error ratio for consumer debt shows that the survey self reported amounts tend to be smaller than the registry, with median values of 15.8%, 11% and 1.9% for, respectively, overall consumer debt, consumer installment loans and credit cards plus lines of credit. This could be due to consumers rounding lower the values of loans (for instance, reporting values of 1 million pesos for loans above 1 million, such as 1.2 million). However, there is a huge dispersion of the error ratios across each loan category. Across any debt category (mortgages, consumer debt and its sub-categories), there are at least 25% of interviewed persons that report lower amounts in the survey relative to the registry (since all the percentiles 75 for the loan amounts are positive). In a similar way, across any debt category, there are at least 25% of interviewed persons that report higher amounts in the survey relative to the registry (since all the percentiles 25 for the loan amounts are negative).

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.[34] 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. Note that for this table the original debt amount and original maturity are directly available from the survey dataset. However, the residual debt amount, the residual maturity and the debt service must be computed from the reported information about the debt/maturity of the loan, the date of the loan contract relative to the interview and the reported interest rate for the loan. Since the reported interest rate is often missing, these variables often use the average interest rate for each loan category and debt amount range as an imputation.

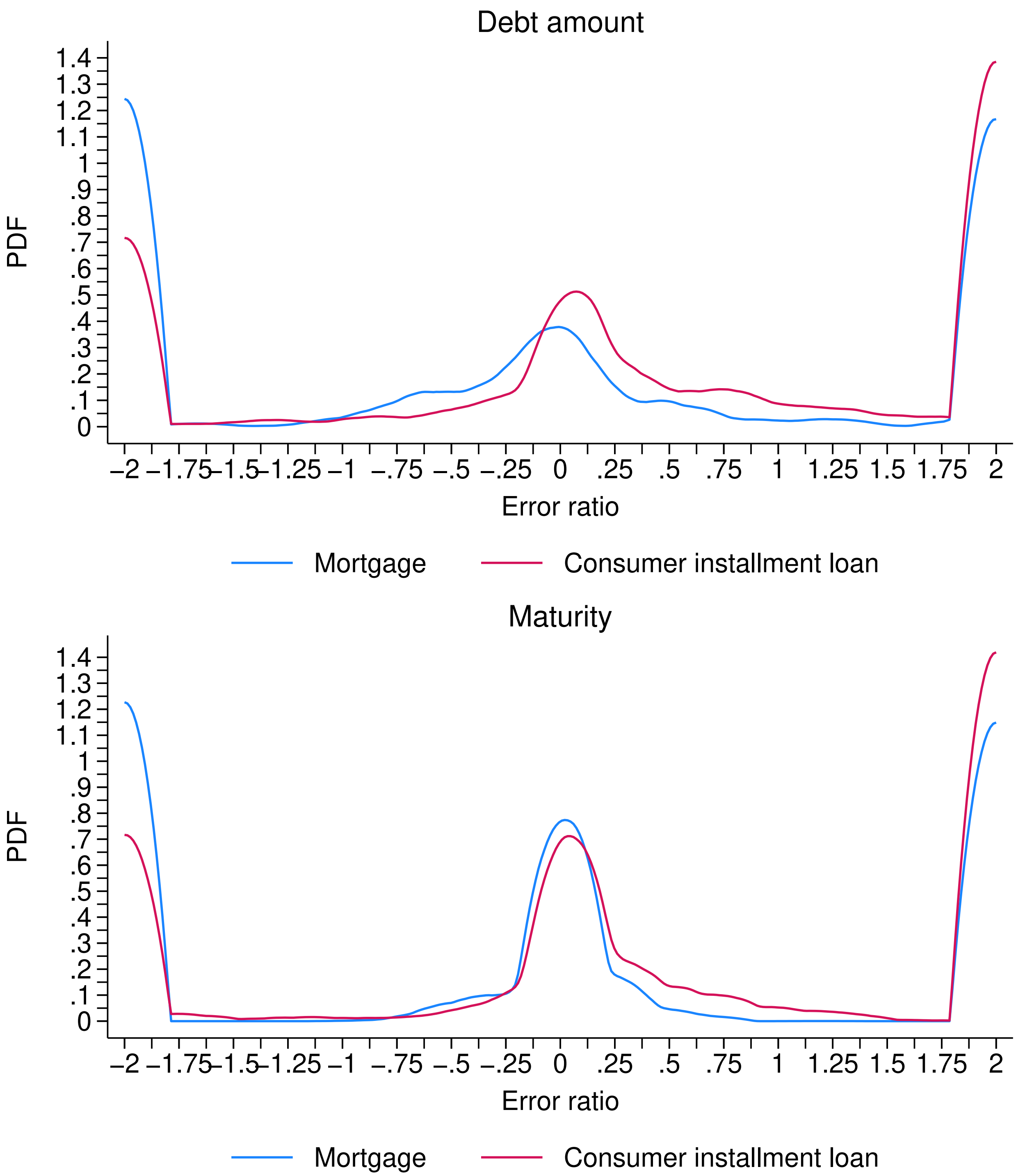
[34] Table 4 is providing descriptive statistics of the percentiles of the distribution for the error report ratios between survey and registry of the borrowers' debts. The percentile numbers belong to specific borrowers, but the borrowers are not necessarily the same across different debt categories. The reason is that the set of people with mortgages is different from the set with consumer installment loans or the set of borrowers with lines of credit. Therefore, the median for the mortgage borrowers is different than the median borrower for consumer debt and its sub-categories (installment loans, credit cards and lines of credit). Therefore, a simple explanation for why the median error for mortgages differs from the median of consumer debt and its sub-categories is that those numbers represent different people. In the online appendix, I estimate selection equations which relates how different demographic characteristics (income, age, education, gender) explain whether certain households are in the group of people with a certain type of debt both in the survey and in the registry, and therefore why different households are in each debt category.

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

Figure 5 shows the differences between survey and registry, as given by the *error ratio*$_{i,t}$. 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 type of error is the mis-reporting of debt participation. In the case of installment loans, the most common error is for a borrower not reporting a bank debt that is registered in the administrative system. For mortgages, borrowers often report debts in the survey that do not appear in the banking system. This is possibly because those mortgages were contracted with 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, survey respondents tend to report larger amounts than they really have in the administrative system.

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



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 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.

4.2 Exercise with a rounding version of the administrative records

One reason for the discrepancies can be due to rounding self-reports in the survey. 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 an approximation of truthful reports. 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:

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

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.[35] 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. However, in most

[35] Some literature, such as Lusardi and Tufano 2015, suggests that borrowers underestimate their debts. However, this algorithm assigns the rounding to the closest value between survey and administrative records. Therefore, it accounts for interviewed persons rounding their answers either below or above the true values.

cases the best amount of rounding is achieved with much lower values than 25%. Therefore the same rounding is not appropriate for all loans and this is why the algorithm searches for the best rounding in a range that is proportional to the value of the loan.[36]

Note this exercise is not meant to find the exact amount of rounding that the respondents are using. The algorithm allows to fit each observation and in this aspect it differs from EM algorithms (which fit parameters of a parametric distribution rather than each observation). The exercise is meant for illustrative purposes, in which an algorithm that allows small, medium and even large amounts of rounding is able to account for some of the differences between registry and survey. However, even after allowing for large amounts of rounding, the exercise still fails to explain most of the differences between registry and survey. Rounding can be an explanation for some of the differences, but it cannot explain the majority of the discrepancies between registry and survey.

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 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.

[36]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

Table 6 shows that most of the observations require no rounding. The reason for this lack of rounding is because the original loans are already a round number. Furthermore, Table 6 shows that even the most extreme percentiles of rounding (percentiles 75 or 90) imply a small amount of rounding, just 16.4% or less of the loan amount.

Table 6: Percentage of rounding of the registry values required for the best possible match in the algorithm (percentiles of distribution)

Rounding required as fraction of the original amount (%)	P10	P25	P50	P75	P90
Mortgage original debt amount	0.0	0.0	0.1	8.2	14.6
Consumer debt original amount	0.0	0.0	0.7	9.4	16.4
Installment Loan original debt amount	0.0	0.0	0.3	7.8	14.5
Credit Lines and Cards' debt amount	0.0	0.0	0.0	6.9	15.5
Credit Cards' debt amount	0.0	0.0	0.0	6.0	14.2
Credit Lines' debt amount	0.0	0.0	0.0	4.6	14.9
Original mortgage maturity	0.0	0.0	0.0	1.6	3.2
Original maturity of installment loans	0.0	0.0	0.0	4.0	11.1

I also analyzed whether excluding a few outliers can help explain the discrepancies between survey data and registry. The motivation for testing outliers is different than for rounding. Rounded numbers can happen through the entire distribution, whether the loan amounts are small (10,000 pesos), medium (100,000 pesos or 1,000,000 pesos) or large (say, 10 million or 100 million pesos). Outliers, however, can happen sometimes through mistakes, such as an interviewer typing one extra zero or one zero less by mistake. Outliers can also happen by other motives, such as an interviewed person deliberately being uncooperative and giving unthoughtful answers. The percentiles 5 and 95 are used, because these are the most standard choices in outlier statistics. The appendix presents a truncated analysis by discarding the values of the $error\ ratio_{i,t}$ at given extremes, with different options: i) below the percentile 5 and above percentile 95, ii) below percentile 2.5 and above 97.5, iii) below percentile 1 and above 99. This analysis is in the online appendix, since it does not change the results substantially. The results show that discarding outliers reduces the measured discrepancies by a minimal amount relative to the analysis in Table 4.

5 Explaining discrepancies

I now test other possible explanations for the differences between 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. The highest income members are defined in the following way: $Y_{i,k} = \max_j Y_{i,j}$, with j denoting any member of the household and k denoting the interviewed member. This allows to define a dummy variable for borrowers who are not among the highest-income members of the household: Dummy for borrower that is not highest-income member$_i = 1(Y_{i,k} < \max_j Y_{i,j})$.

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 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)	Loan in survey only	Loan in both datasets	Loan in survey only	Loan in both datasets
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)	Loan participation only in survey	Loan participation in both datasets
Mortgage loan	1.42	1.20
Installment loan	1.41	1.36

Table 8 tests the same hypotheses in a multivariate logit regression for the event of borrowers reporting a mortgage or installment loan only for the survey. I also 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.[37] Therefore, the regressions add a control for whether the borrower has another loan that is actually reported in the registry. The regressions support all these hypotheses for mortgage reports. For the case of installment loans only the additional control variable of allowing the borrower to confuse loans is significant. Finally, the regressions also control for the age and education of the respondents, which can be seen as proxies for financial literacy (Madeira and Margaretic 2022).

[37]Note that the variable "Dummy for another installment or mortgage loan in registry" does not imply that there is a fully deterministic process in which borrowers confuse loans automatically. The control is applied for testing whether there is a statistical association in which some borrowers could confuse loans of different categories as mortgages. But this relationship is by no means fully deterministic or guaranteed, since in such case the coefficient for the control would be infinite and the logit model would not be estimated.

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

Logit: Loan reported in Survey but unreported in Registry	Mortgage loans Model 1	Mortgage loans Model 2	Installment loans Model 1	Installment loans Model 2
Dummy for borrower that	1.042***	1.027***	0.0484	-0.0638
is not highest-income member$_i$	(0.187)	(0.140)	(0.275)	(0.209)
Number of household members	0.527***	0.544***	0.0316	-0.0259
with the highest-income$_i$	(0.151)	(0.110)	(0.162)	(0.124)
Dummy for borrower that	0.343**	0.443***	0.0847	-0.143
is not the oldest member$_i$	(0.159)	(0.119)	(0.215)	(0.159)
Dummy for another installment	-1.215***	-1.330***	-1.099***	-1.590***
or mortgage loan in registry$_i$	(0.170)	(0.127)	(0.370)	(0.255)
Age (years)		0.000201		-0.00566
		(0.00483)		(0.00564)
Education (years)		-0.0934***		-0.234***
		(0.0274)		(0.0409)
Constant	-0.619***	0.611	-0.769***	2.819***
	(0.211)	(0.480)	(0.235)	(0.688)
Number of Observations	1,627	1,627	1,063	1,063
Pseudo R-squared	0.142	0.148	0.026	0.088

Robust standard errors in ().

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

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 mortgages and credit cards are associated with borrowers who are not the highest-income member of their family. The results are robust to including or not demographic controls such as age and education, as shown in the appendix. Table 9 shows that education (a proxy for financial literacy) reduces the discrepancy of the loan amounts reported for credit cards or lines of credit. Age increases the discrepancy of the loan amounts reported for consumer installment loans.

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

Variables	Mortgage	Consumer debt	Installment loan	Card or line	Credit Card	Credit Line
Dummy: borrower not	19.46***	6.905	8.360	10.91**	17.09***	0.365
highest-income member$_i$	(7.363)	(4.303)	(5.546)	(4.896)	(5.087)	(10.27)
Nr of members with	3.660	1.891	-0.222	5.066	2.763	3.914
the highest-income$_i$	(4.683)	(2.785)	(3.426)	(3.402)	(3.565)	(7.190)
Dummy: borrower not	-7.228	0.945	4.048	2.819	6.219	11.14
the oldest member$_i$	(4.866)	(3.320)	(4.078)	(3.756)	(3.905)	(7.451)
Age (years)	0.337	0.0717	0.300**	0.0294	0.00305	-0.316
	(0.216)	(0.116)	(0.147)	(0.132)	(0.137)	(0.271)
Education (years)	-1.548	1.101	0.251	-1.477*	-1.151	-2.339
	(1.044)	(0.723)	(0.881)	(0.829)	(0.861)	(1.644)
Constant	47.38**	43.19***	30.24*	71.65***	63.90***	107.2***
	(20.18)	(13.19)	(16.36)	(14.86)	(15.38)	(29.68)
Observations	427	1,491	740	1,061	940	370
R-squared	0.037	0.004	0.011	0.017	0.028	0.021

Robust standard errors in ().

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

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 aggregate 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 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, which is riskier. 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.

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.

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. Therefore, in total the survey reports 11.8% of borrowers at risk, while the registry would report 27.3%. This confirms that surveys may underestimate household debt risks. Table 11 shows that the matched survey-registry provides an estimate of 31.9% of borrowers at risk, which is higher than either survey or registry alone.

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

Debt risk (Survey, Registry)	Not in survey Not in registry	In survey Not in registry	Not in survey In registry	In survey In registry
	68.1	4.6	20.1	7.2
Debt risk in each source	In survey		In registry	
All borrowers	11.8		27.3	

This implies that stress tests of banking risk could be improved by using registry and survey data jointly. For instance, Madeira 2018b shows that the EFH survey can be used to calibrate a structural model with heterogeneous households deciding to default based on their liquid assets, shocks to wages and unemployment spells. Such stress test models can be enhanced by using the matched information of survey-registry. Without a matched survey-registry dataset, it is possible that stress tests may underestimate the overall risk of the banking household debt portfolios.

7 Conclusions

Financial information in surveys differs from aggregates in administrative sources. This article advances on the literature by looking at a match of individual borrowers from the Chilean Household Finance Survey (EFH) with the registry of banking loans. This match of individual borrowers survey reports and their registry loan records so far has not been implemented in the previous literature for household debt. This adds value to previous literature on measurement error of household income, unemployment and participation in social programs (Meyer et al. 2015, Meyer and Mittag 2021).

This unique matched dataset allows to study the distribution of differences between the datasets: whether the data differ in terms of the number of borrowers, whether borrowers understate debts and the degree of heterogeneity in the discrepancies. While there are substantial differences in the reports of loan participation, reports of loan delinquency are broadly accurate. The discrepancy between survey and registry in terms of debt participation (a binary indicator of an interviewed borrower having debt) ranges from just 7.6% for lines of credit to 11.6% for consumer installment loans, 17.8% for credit cards and 10.9% for mortgages. For the binary indicator of whether a borrower is in delinquent (in arrears for 1 month or more), the discrepancy between datasets is

0.8% for mortgages and 7.4% for consumer installment loans (although the numbers fall to 0% and 0.6% if considering just the respondents with positive debt participation in both datasets).

There is substantial heterogeneity in the survey-registry discrepancies of individual borrowers, especially for consumer loans. Some of these discrepancies 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.

This study shows that survey self-reported debt may underestimated the true indebtedness and financial risks. However, future researchers may take some clues from the findings of this article. Indebtedness levels are more likely to have measurement error for complex households, families with many members earning equal shares of total income, and households in which the interviewed person is not the highest earning or oldest member. Therefore, analysts of financial risks and stress tests could decide to implement separate analysis for one person households, simpler households with a single earning member, and for complex households with several members. Analysts may also consider in stress tests that survey reports can be rounded and consider slightly higher amounts as a robustness check in those cases. Furthermore, stress tests may study the risk of all debts together in case some borrowers confuse mortgages and installment loans.

Household finance surveys are increasingly important in the analysis of borrowers' behaviors such as delinquency or credit exclusion. Note that surveys may also have significant advantages relative to credit registries, although these advantages may be more difficult to quantify. For instance, the debts that are reported by the interviewed persons may be considered to be the commitments of highest concern for the household. Another example is that while registries can attribute more correctly the legal responsibility of each loan, it is possible that other members of the household also view those loans as being relevant for their overall finances. Surveys also help to measure expectations and attitudes towards risk, which are not captured by administrative records. These are aspects where surveys possibly improve upon pure registry information.

Further research on the strengths and shortcomings of survey financial data is thus warranted. One issue is the measurement of assets, since debt is only one component of wealth (Zinman 2009). Furthermore, the measurement of wealth is probably even more sensitive for interviewed persons. Future studies of household finance may improve their reliability of asset measurement.

References

[1] Abowd, J. and M. Stinson (2013), "Estimating measurement error in annual job earnings: A comparison of survey and administrative data," *Review of Economics and Statistics*, 95(5), 1451-1467.

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

[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] Cerletti, E., M. Cortina, A. Inzunza, F. Martínez and P. Toro (2025), "The finances of Chilean households during the pandemic: an assessment from the 2021 Household Financial Survey," *Latin American Journal of Central Banking*, 100175.

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

[14] 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.

[15] 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..

[16] Davis, S., R. Faberman and J. Haltiwanger (2012), "Labor market flows in the cross section and over time," *Journal of Monetary Economics*, 59(1), 1-18.

[17] 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.

[18] Elosegui, P., M. González, M. Pérez and M. Sangiácomo (2024), "A diffusion index analysis of the Argentinean business economic cycle based on the "Survey of Business Economic Perspectives"," *Latin American Journal of Central Banking*, 5(2), 100108.

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

[20] 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.

[21] Gathergood, J. and A. Olafsson (2024), "The Coholding Puzzle: New Evidence from Transaction-Level Data," *Review of Financial Studies*, hhae016.

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

[23] 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.

[24] 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.

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

[26] 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.

[27] Groves, R. and L. Lyberg (2010), "Total Survey Error: Past, Present, and Future," *Public Opinion Quarterly*, 74(5), 849–879.

[28] 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.

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

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

[31] 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.

[32] 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.

[33] Lusardi, A. and P. Tufano (2015), "Debt literacy, financial experiences, and overindebtedness," Journal of Pension Economics and Finance. 2015

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

[35] 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.

[36] Madeira, C. (2019), "Computing population weights for the EFH survey," *Economía Chilena*, 22(1), 4-26.

[37] Madeira, C. (2023), "Adverse selection, loan choice and default in the Chilean consumer debt market", *Financial Innovation*, 2023, 9, 49.

[38] Madeira, C. (2024), "The effect of the Covid pension withdrawals and the Universal Guaranteed Pension on the income of the future retirees and its fiscal costs", *Latin American Journal of Central Banking*, 5(3), 100122.

[39] Madeira, C. (2025), "Indebtedness and labor risk sorting across consumer lender types in Chile," *Journal of Banking Regulation*, 26, 149–163.

[40] Madeira, C. and Margaretic, P. (2022), "The impact of financial literacy on the quality of self-reported financial information", *Journal of Behavioral and Experimental Finance*, 34.

[41] Madeira, C., Margaretic, P., Martínez, F. and Roje, P. (2022), "Assessing the quality of self-reported debt information in household surveys," *Journal of Survey Statistics and Methodology*, 10(5), 1183–1210.

[42] Madeira, C. and L. Salazar (2023), "The Impact of Monetary Policy on a Labor Market with Heterogeneous Workers: The Case of Chile", *Latin American Journal of Central Banking*, 4(2), 100092.

[43] 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.

[44] Meyer, B. D., Mok, W. K., & Sullivan, J. X. (2015), "Household surveys in crisis," *Journal of Economic Perspectives*, 29(4), 199-226.

[45] Meyer, B. D., & Mittag, N. (2019), "Using linked survey and administrative data to better measure income: Implications for poverty, program effectiveness, and holes in the safety net," *American Economic Journal: Applied Economics*, 11(2), 176-204.

[46] Meyer, B. D., & Mittag, N. (2021), "Combining administrative and survey data to improve income measurement.,"*Administrative records for survey methodology*, 297-322.

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

[48] 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.

[49] 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.

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

[51] 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.

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

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

[54] 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. Section 5 shows several summary statistics of different methods for rounding the dataset. Section 6 shows summary statistics after applying the exclusion of outliers, using different criteria for the percentiles of outliers. Section 7 shows summary statistics of the discrepancies conditional on the borrowers' education. Section 8 shows that summary statistics are robust in terms of selecting a shorter time window for the adjoining periods of the interview. Section 9 shows that the results of the multivariate regressions for the discrepancies between registry and survey are robust to different specifications.

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

[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.

2007 and 2010, the 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 UF,

[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.

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: 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.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

[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.

installment loans, credit cards and credit lines), although their self-reports for the loan maturity tend to be similar to the registry data.

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

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 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).

Figure A.1: 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)

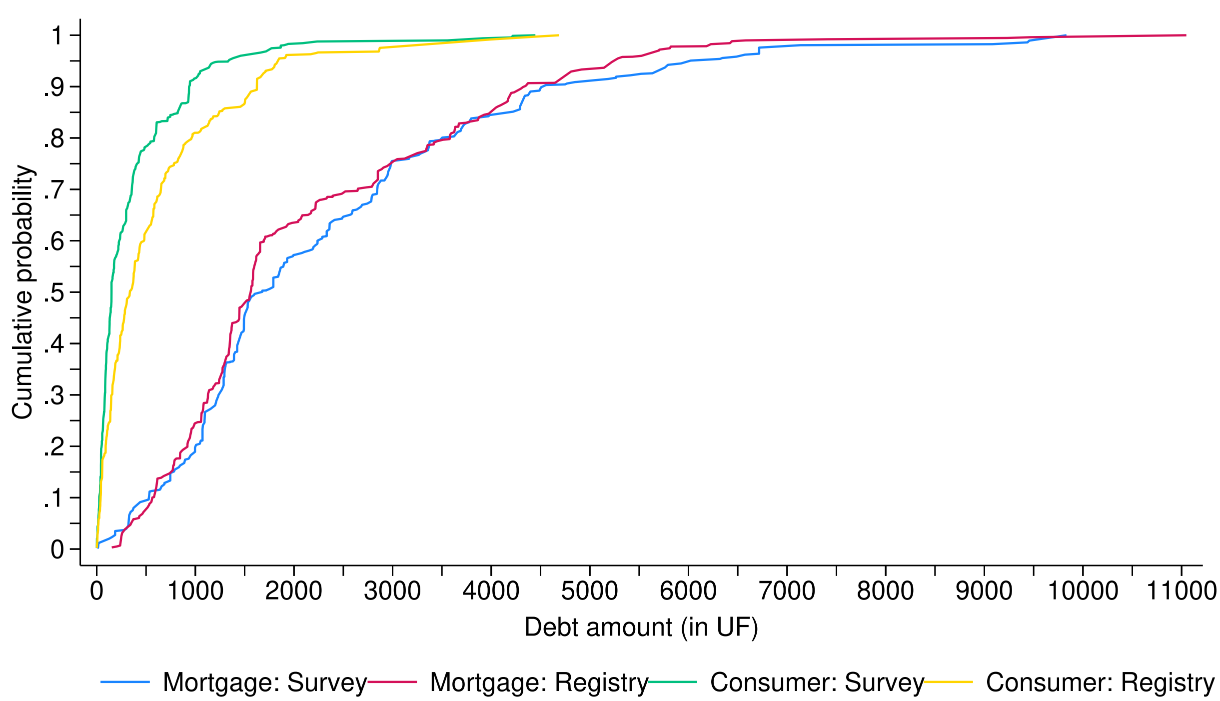


Figure A.2: CDFs of the original consumer installment loans and revolving debt instruments (credit cards and lines of credit) in the survey and registry (when positive amounts of each category of debt are present in both datasets

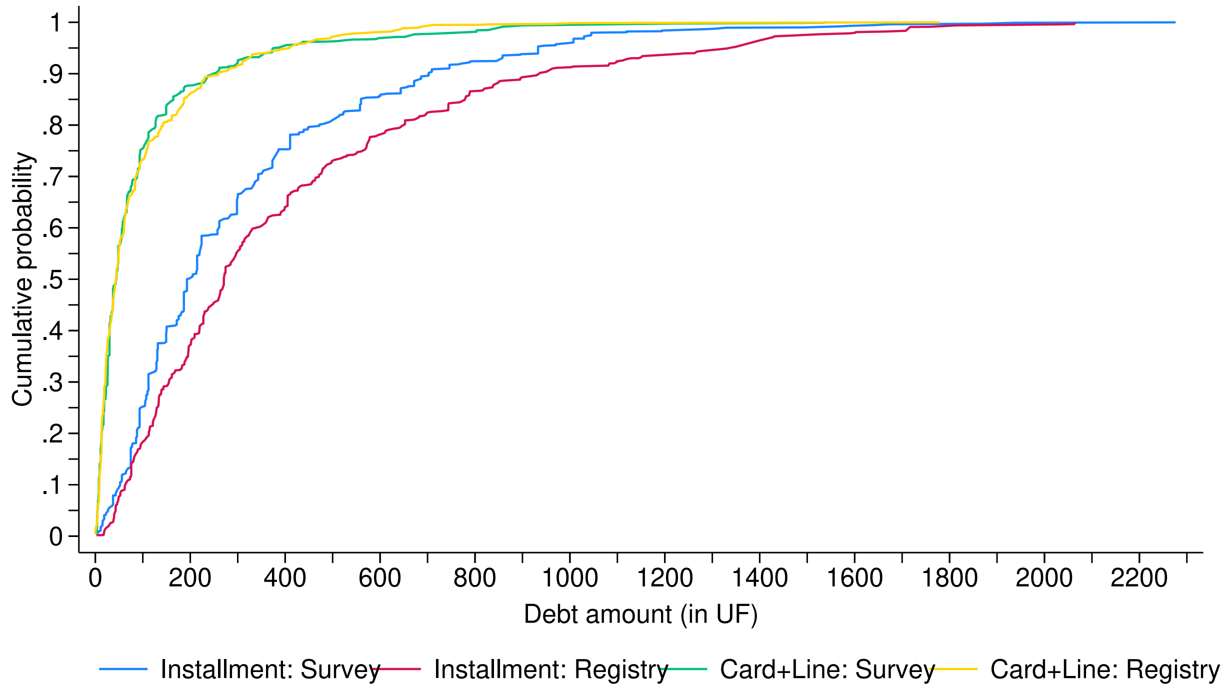
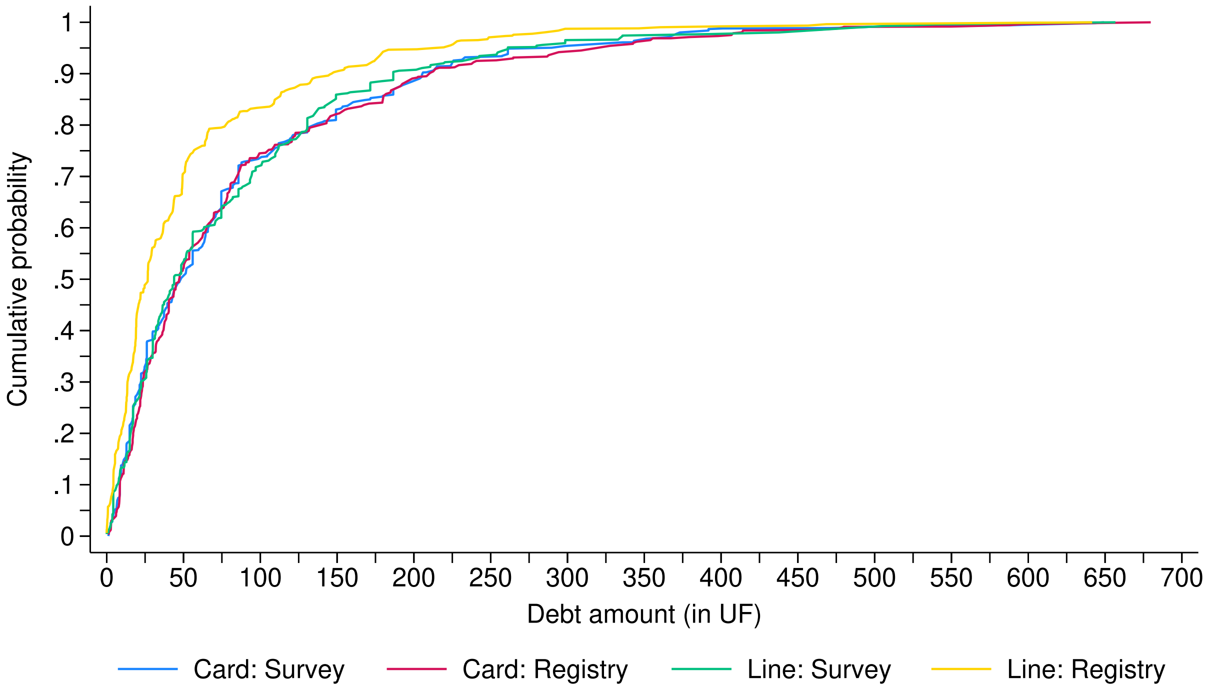


Figure A.3: CDFs of the debt amounts for credit cards and lines of credit in the survey and registry (when positive amounts of each category of debt are present in both datasets)



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.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, 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.

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

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.

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).

5 Exercise with a rounding version of the administrative records

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:

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

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.[9] 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. However, 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.[10]

Note this exercise is not meant to find the exact amount of rounding that the respondents are

[9] Note that some literature, such as Lusardi and Tufano 2015, suggests that borrowers underestimate their debts. However, this algorithm assigns the rounding to the closest value between survey and administrative records. Therefore, it takes into account that some interviewed persons may round their answers below or above the true values.

[10] 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).

using. The algorithm allows to fit each observation and in this aspect it differs from EM algorithms (which fit parameters of a parametric distribution rather than each observation). The exercise is meant for illustrative purposes, in which an algorithm that allows small, medium and even large amounts of rounding is able to account for some of the differences between registry and survey. However, even after allowing for large amounts of rounding (as much as 25% of the amounts), the exercise still fails to explain most of the differences between registry and survey. This implies that rounding can be an explanation for some of the differences, but it cannot explain the majority of the discrepancies between registry and survey.

Table B.1 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 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.

Table B.1: 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

Table B.2 shows that most of the observations require no further rounding by the algorithm (B.1). This algorithm is used in Table B.1 at the beginning of the section. The reason for this lack of further rounding is because the original loans are already a round number. Table B.2 shows that even the most extreme percentiles of rounding (percentiles 75 or 90) imply a small amount of rounding, just 16.4% or less of the loan amount.

Table B.2: Percentage of rounding of the registry values required
for the best possible match in the algorithm (percentiles of distribution)

Rounding required as fraction of the original amount (%)	P10	P25	P50	P75	P90
Mortgage original debt amount	0.0	0.0	0.1	8.2	14.6
Consumer debt original amount	0.0	0.0	0.7	9.4	16.4
Installment Loan original debt amount	0.0	0.0	0.3	7.8	14.5
Credit Lines and Cards' debt amount	0.0	0.0	0.0	6.9	15.5
Credit Cards' debt amount	0.0	0.0	0.0	6.0	14.2
Credit Lines' debt amount	0.0	0.0	0.0	4.6	14.9
Original mortgage maturity	0.0	0.0	0.0	1.6	3.2
Original maturity of installment loans	0.0	0.0	0.0	4.0	11.1

Furthermore, Table B.3 shows that almost 80% of the mortgages and 99.9% of the consumer debt reports in the EFH survey are rounded in terms of 1,000 pesos. This makes sense, because 1,000 pesos is a very accurate rounding (a value similar to 1 USD). Table B.4 shows that rounding in terms of 100,000 pesos happens less often. A rounding of 100,000 pesos (roughly, 100 USD) covers just 66% of the mortgage, 78.5% of the credit card and lines of credit, and 96.7% of the survey reports in the matched EFH-registry dataset.

Table B.3: Survey rounding data for the loan reports (in %)

Rounding data in terms of the fraction of households (one or more loans are rounded in 1,000 pesos)

	Household Survey			Matched Survey-Administrative Data		
Year	Mortgages	Cards & Lines	Consumer Loans	Mortgages	Cards & Lines	Consumer Loans
2011	87.2	99.7	100.0	87.8	99.7	100.0
2014	78.2	99.8	99.7	73.0	100.0	100.0
2017	76.6	99.7	99.7	76.6	99.7	99.7
All years	79.9	99.7	99.9	77.4	99.8	99.9

Rounding data in terms of the fraction of loans

	Household Survey			Matched Survey-Administrative Data		
Year	Mortgages	Cards & Lines	Consumer Loans	Mortgages	Cards & Lines	Consumer Loans
2011	87.2	99.5	100.0	87.7	99.9	100.0
2014	74.3	99.8	99.9	65.4	100.0	100.0
2017	72.7	99.6	99.9	71.2	99.6	99.9
All years	77.0	99.7	99.9	73.4	99.8	100.0

Note: For the purpose of this exercise, survey reports of interval amounts are included as rounding data of the loan amount.

Table B.4: Survey rounding data for the loan reports (in %)

Rounding data in terms of the fraction of households (one or more loans are rounded in 100,000 pesos)

	Household Survey			Matched Survey-Administrative Data		
Year	Mortgages	Cards & Lines	Consumer Loans	Mortgages	Cards & Lines	Consumer Loans
2011	77.9	72.3	96.3	76.6	73.4	96.5
2014	72.1	80.5	97.3	65.9	81.1	97.3
2017	67.4	85.2	97.5	67.4	85.2	97.5
All years	73.1	79.1	96.8	69.9	79.4	97.0

Rounding data in terms of the fraction of loans

	Household Survey			Matched Survey-Administrative Data		
Year	Mortgages	Cards & Lines	Consumer Loans	Mortgages	Cards & Lines	Consumer Loans
2011	78.3	68.7	95.3	77.0	68.6	95.5
2014	68.4	80.3	96.8	58.6	78.9	96.7
2017	66.1	82.1	97.3	65.2	82.3	97.4
All years	70.1	78.7	96.6	66.1	78.5	96.7

Note: For the purpose of this exercise, survey reports of interval amounts are included as rounding data of the loan amount.

I also test another definition of rounding that does not require algorithms as in the expression (B.1). I create a dummy variable for rounding called R1. R1 defines as a rounding value corresponds to observations where either the first digit is 1 or 5 and the other remaining digits are all 0 (with the further restriction that at least three zeros are required). In this definition, 5,543 is not a rounded value, neither is 743,000, neither is 53,000, neither is 12,000 or 1,100,000. I also create a second definition called R2, where rounding corresponds to the above cases plus any digit 1 to 9 followed by at least four zeros. In this second definition, 4,000 is taken as a non-rounded number, but 40,000 or 400,000 can be considered as rounded. Table B.5 shows the fraction of interviewed persons for each loan category which report all their loans of a given category in terms of definitions R1 and R2. The results show that less tha 8% of the mortgages are presented in rounded form, according to either criteria 1 or 2. However, almost 40% of the households present all their consumer debt loans in rounded form R2, with the fraction of borrowers going from 46% for lines of credit to 54% for credit cards. With the exception of lines of credit, this definition of rounding is more prevalent in lower educated households and it is less common in borrowers with technical education.

Table B.5: Households rounding all loan reports of a certain type with criterion 1 and 2 (% of households), according to the education of the interviewed person

Debt type	All households		Secondary or less		Technical education		College education or more	
(of original debt amount)	R1	R2	R1	R2	R1	R2	R1	R2
Mortgage	3.2	7.9	6.8	14.7	0.6	2.3	0.7	3.5
Consumer debt	10.9	39.1	12.6	44.2	9.8	32.9	8.7	35.2
Installment Loan	12.7	48.0	13.7	50.4	10.4	44.3	12.4	45.9
Credit Lines and Cards	12.5	46.1	14.8	52.7	11.3	37.8	10.4	44.7
Credit Cards	14.6	53.9	16.6	59.7	10.7	46.5	15.8	53.4
Credit Lines	14.2	45.6	12.5	43.8	15.4	41.7	15.0	50.8

R1 rounding consists of a dummy, with 1 corresponding to a first digit of 1 or 5, while all the other remaining digits are 0.

R2 rounding consists of a dummy, with 1 corresponding to one in criterion R1 and also to 1 if a number has any first digit between 1 and 9 and is followed by at least four zeros.

6 Exercise with the exclusion of outliers

This section analyzes 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. The motivation for testing outliers is different than for rounding. Rounded numbers can happen through the entire distribution, whether the loan amounts are small (10,000 pesos), medium (100,000 pesos or 1,000,000 pesos) or large (say, 10 million or 100 million pesos). Outliers, however, can happen sometimes through mistakes, such as an interviewer typing one extra zero or one zero less by mistake. Outliers can also happen by other motives, such as an interviewed person deliberately being uncooperative and giving unthoughtful answers. The percentiles 5 and 95 are used, because these are the most standard choices in outlier statistics. Table C.1 shows the summary statistics of the discrepancies between the individual borrower survey reports and their registry. The results show that discarding outliers reduces the measured discrepancies by a minimal amount relative to the analysis in Table 4. Therefore, while accounting for rounding errors goes some way towards explaining the discrepancies, the existence of a small percentage of outliers does not change the results significantly.

Table C.1: Discrepancies between the registry and the survey
loan reports (percentiles and linear correlation coefficients): after truncating outliers

Outlier below percentiles 5 and above 95						
$error\ ratio_{i,t}$ (%)	P10	P25	P50	P75	P90	$\rho_{Y(Registry,Survey)}$
Mortgage original debt amount	-53.2	-25.3	-2.0	11.0	33.9	57.3
Consumer debt original amount	-38.3	1.7	31.5	85.1	133.9	58.4
Installment Loan original debt amount	-23.0	1.7	8.7	71.5	94.9	66.4
Credit Lines and Cards' debt amount	-61.6	-14.6	8.9	49.0	96.8	56.7
Credit Cards' debt amount	-57.2	-12.0	8.2	48.3	84.8	43.1
Credit Lines' debt amount	-104.0	-41.7	0.5	24.2	75.8	51.8
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

Furthermore, I also show the results by using stricter versions of percentiles as outliers. Table C.2 shows the results considering outliers below percentiles 2.5 and 97.5 and an even stricter definition for outliers below percentile 1 and above percentile 99.

Table C.2: Discrepancies between the registry and the survey
loan reports (percentiles and linear correlation coefficients): after truncating outliers

Outlier below percentiles 2.5 and above 97.5					
$error\ ratio_{i,t}$ (%)	P10	P25	P50	P75	P90
Mortgage original debt amount	-76.4	-44.5	-6.1	14.6	56.1
Consumer debt original amount	-86.4	-16.5	12.1	69.3	127.7
Installment Loan original debt amount	-41.3	-2.0	6.8	59.1	123.9
Credit Lines and Cards' debt amount	-140.8	-51.1	0.0	37.2	80.1
Credit Cards' debt amount	-131.2	-30.7	2.9	40.9	83.3
Credit Lines' debt amount	-164.4	-131.6	-46.4	0.7	19.9
Original mortgage maturity	-1.3	0.6	2.1	13.4	25.5
Residual mortgage maturity	-44.6	-6.2	2.8	16.0	45.2
Original maturity of installment loans	-16.8	0.0	5.4	30.6	58.8
Residual maturity of installment loans	-82.4	-34.1	-5.7	20.0	58.8
Outlier below percentiles 1 and above 99					
$error\ ratio_{i,t}$ (%)	P10	P25	P50	P75	P90
Mortgage original debt amount	-60.7	-27.5	0.0	54.9	132.5
Consumer debt original amount	-93.9	-19.1	13.2	56.8	110.9
Installment Loan original debt amount	-48.6	-1.5	15.1	46.3	94.7
Credit Lines and Cards' debt amount	-142.5	-55.1	-0.5	20.9	78.4
Credit Cards' debt amount	-135.2	-36.8	0.6	25.2	85.8
Credit Lines' debt amount	-164.2	-119.7	-4.3	2.4	21.3
Original mortgage maturity	-36.2	-2.5	0.4	2.1	16.5
Residual mortgage maturity	-36.5	-16.9	-3.4	7.4	25.6
Original maturity of installment loans	-16.2	1.7	5.4	28.6	70.3
Residual maturity of installment loans	-52.6	-23.3	-5.7	50.0	82.9

7 Exercises conditional on interviewed person's education level

This section shows the results of the discrepancies between EFH survey and registry, summarized by the educational level of the respondent.

Table D.1: Differences between the self-reported survey data and registry of borrower status (binary loan variable)

Debt participation (Survey, Registry) - %	Not in survey Not in registry	Not in survey In registry	In survey Not in registry	In survey In registry	Difference in datasets
Secondary education or less					
Mortgage loan	87.4	1.4	7.3	3.9	8.7
Consumer debt	69.8	13.2	5.1	12.0	18.3
Consumer installment loan	82.2	7.1	3.9	6.9	11.0
Credit Card or Credit Line	77.4	12.9	3.6	6.2	16.5
Credit Card	79.4	12.1	3.1	5.4	15.3
Credit Line	93.6	3.0	1.6	1.8	4.6
Technical education or two-year college					
Mortgage loan	74.0	1.4	14.2	10.4	15.6
Consumer debt	52.2	15.8	9.4	22.5	25.3
Consumer installment loan	76.1	7.0	6.1	10.8	13.1
Credit Card or Credit Line	58.5	16.9	7.3	17.3	24.2
Credit Card	61.0	16.1	7.5	15.4	23.6
Credit Line	81.8	9.0	2.9	6.3	11.9
College education or post-graduate degree					
Mortgage loan	63.4	3.3	13.6	19.7	16.9
Consumer debt	38.0	16.5	6.7	38.8	23.2
Consumer installment loan	66.7	11.8	2.9	18.6	14.7
Credit Card or Credit Line	42.2	19.4	6.0	32.5	25.3
Credit Card	45.7	20.3	5.7	28.3	26.0
Credit Line	69.3	13.1	5.0	12.7	18.0

Table D.2: Differences between the self-reported survey data and registry of default status (binary variable of arrears of 1 month or more)

Default (Survey, Registry) - %	No default in survey No default in registry	Default only in registry	Default only in survey	Default in both datasets	Difference in datasets
Panel 1: Secondary education or less					
All loans (including loans reported in one dataset only)					
Mortgage loan	97.8	0.0	0.9	1.3	0.9
Consumer installment loan	78.2	0.3	10.5	11.0	10.8
Including just the sample of borrowers with same debt participation in both datasets					
Mortgage loan	95.9	0.0	0.0	4.1	0.0
Consumer installment loan	81.6	0.4	0.6	17.4	1.0
Panel 2: Technical education or two-year college					
All loans (including loans reported in one dataset only)					
Mortgage loan	96.8	0.0	1.4	1.8	1.4
Consumer installment loan	88.9	0.0	5.1	6.1	5.1
Including just the sample of borrowers with same debt participation in both datasets					
Mortgage loan	95.6	0.0	0.0	4.4	0.0
Consumer installment loan	90.5	0.0	0.0	9.5	0.0
Panel 3: College education or post-graduate degree					
All loans (including loans reported in one dataset only)					
Mortgage loan	98.4	0.0	0.1	1.5	0.1
Consumer installment loan	93.9	0.3	2.3	3.5	2.6
Including just the sample of borrowers with same debt participation in both datasets					
Mortgage loan	97.3	0.0	0.0	2.7	0.0
Consumer installment loan	95.6	0.3	0.0	4.1	0.3

Table D.3: Discrepancies between the registry and the survey loan reports (percentiles) by education status

Panel 1: Secondary education or less					
$error\ ratio_{i,t}$ (%)	P10	P25	P50	P75	P90
Mortgage original debt amount	-65.4	-27.5	0.0	54.9	179.3
Consumer debt original amount	-114.3	-19.7	12.2	58.1	118.6
Installment Loan original debt amount	-48.6	-1.8	15.5	49.3	97.0
Credit Lines and Cards' debt amount	-142.6	-55.2	-0.5	23.7	84.1
Credit Cards' debt amount	-137.7	-36.5	0.6	30.0	93.5
Credit Lines' debt amount	-164.6	-124.8	-5.1	2.4	21.3
Original mortgage maturity	-39.6	-2.5	0.4	2.1	16.7
Residual mortgage maturity	-51.3	-16.9	-3.4	7.4	25.6
Original maturity of installment loans	-23.3	0.0	5.4	30.6	72.0
Residual maturity of installment loans	-52.6	-23.3	-5.7	50.0	82.9
Panel 2: Technical education or two-year college					
$error\ ratio_{i,t}$ (%)	P10	P25	P50	P75	P90
Mortgage original debt amount	-76.4	-44.7	-6.1	15.3	56.1
Consumer debt original amount	-93.7	-21.4	12.1	70.3	140.7
Installment Loan original debt amount	-43.1	-2.0	7.0	63.7	129.3
Credit Lines and Cards' debt amount	-144.1	-56.8	0.0	37.3	89.3
Credit Cards' debt amount	-145.3	-34.9	2.9	41.8	94.6
Credit Lines' debt amount	-165.6	-131.6	-46.3	2.4	23.3
Original mortgage maturity	-4.5	0.4	2.1	13.4	29.0
Residual mortgage maturity	-57.1	-6.6	2.8	16.8	50.0
Original maturity of installment loans	-23.3	0.0	6.1	39.0	95.7
Residual maturity of installment loans	-92.7	-46.2	-7.2	20.3	62.9
Panel 3: College education or post-graduate degree					
$error\ ratio_{i,t}$ (%)	P10	P25	P50	P75	P90
Mortgage original debt amount	-63.2	-28.5	-2.2	12.7	40.0
Consumer debt original amount	-69.5	0.0	31.5	91.1	148.1
Installment Loan original debt amount	-33.0	1.4	9.0	74.4	120.3
Credit Lines and Cards' debt amount	-87.8	-18.7	8.2	51.6	106.5
Credit Cards' debt amount	-64.6	-14.7	8.2	50.7	99.8
Credit Lines' debt amount	-120.4	-45.4	1.6	24.2	81.6
Original mortgage maturity	-28.6	-2.4	1.2	7.8	24.2
Residual mortgage maturity	-28.6	-9.6	-0.7	25.3	41.8
Original maturity of installment loans	-52.3	0.0	4.1	27.0	51.5
Residual maturity of installment loans	-74.3	-28.6	0.0	50.0	91.5

8 Exercises that consider only debt matches in the adjoining periods around the interview date

This section considers the debt matches in the adjoining monthly periods -1,0 +1 around the interview date (while the article considers the periods -2,-1,0,+1,+2, as well). The results remain very similar to the article, therefore the conclusions are unchanged by changing the adjoining time frame over which the best match between survey and registry is performed.

Table E.1: Differences between the self-reported survey data and registry of borrower status (binary loan variable)

Debt participation (Survey, Registry) - %	Not in survey Not in registry	Not in survey In registry	In survey Not in registry	In survey In registry	Difference in datasets
Match on registry reports for the monthly periods -1, 0, +1 around interview					
Mortgage loan	82.0	1.6	9.3	7.0	11.0
Consumer debt	62.0	14.6	6.2	17.2	20.8
Consumer installment loan	78.9	7.8	4.3	9.0	12.1
Credit Card or Credit Line	68.9	15.0	4.6	11.5	19.6
Credit Card	71.2	14.4	4.3	10.1	18.7
Credit Line	88.1	5.7	2.3	3.9	8.0

Table E.2: Differences between the self-reported survey data and registry of default status (binary variable of arrears of 1 month or more)

Default (Survey, Registry) - %	No default in survey No default in registry	Default only in registry	Default only in survey	Default in both datasets	Difference in datasets
Match on registry reports for the monthly periods -1, 0, +1 around interview					
All loans (including loans reported in one dataset only)					
Mortgage loan	97.9	0.0	0.8	1.3	0.8
Consumer installment loan	84.0	0.3	7.5	8.3	7.7
Including just the sample of borrowers with same debt participation in both datasets					
Mortgage loan	96.6	0.0	0.0	3.4	0.0
Consumer installment loan	87.1	0.4	0.3	12.2	0.7

Table E.3: Discrepancies between the registry and the survey loan reports (percentiles) by education status

Match on registry reports for the monthly periods -1, 0, +1 around interview					
$error\ ratio_{i,t}$ (%)	P10	P25	P50	P75	P90
Mortgage original debt amount	-67.9	-31.8	-2.7	18.1	77.1
Consumer debt original amount	-109.4	-14.6	16.0	74.1	139.8
Installment Loan original debt amount	-48.3	0.0	11.6	61.4	111.0
Credit Lines and Cards' debt amount	-137.9	-46.9	2.8	51.2	103.6
Credit Cards' debt amount	-126.7	-28.4	4.9	51.9	108.4
Credit Lines' debt amount	-165.6	-113.4	-10.8	11.5	47.0
Original mortgage maturity	-28.0	-1.3	1.2	3.8	24.4
Residual mortgage maturity	-28.6	-12.0	-0.7	15.4	42.4
Original maturity of installment loans	-25.9	0.0	5.4	30.6	70.3
Residual maturity of installment loans	-74.3	-28.6	-2.8	45.2	82.9

9 Multivariate regressions of the discrepancies in the matched survey-registry dataset

Table F.1 tests the several hypotheses for the discrepancies between survey and registry in a multivariate logit regression for the event of borrowers reporting a mortgage only for the survey. The results show that the

Table F.2 tests the several hypotheses for the discrepancies between survey and registry in a multivariate logit regression for the event of borrowers reporting an installment loan only in the survey.

Table F.3 tests 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 mortgages and credit cards are associated with borrowers who are not the highest-income member of their family.

Table F.1: Mortgage loan exists in the survey, but not in the registry (logit)

Control variables	Model 1	Model 2	Model 3	Model 4	Model 5
Dummy for borrower that	1.042***	1.061***	1.033***	1.175***	1.124***
is not highest-income member$_i$	(0.187)	(0.140)	(0.140)	(0.134)	(0.136)
Number of household members	0.527***	0.550***	0.549***	0.514***	0.516***
with the highest-income$_i$	(0.151)	(0.109)	(0.110)	(0.104)	(0.105)
Dummy for borrower that	0.343**	0.440***	0.440***	0.478***	0.482***
is not the oldest member$_i$	(0.159)	(0.116)	(0.119)	(0.111)	(0.115)
Dummy for another installment	-1.215***	-1.390***	-1.341***		-1.099***
or mortgage loan in registry$_i$	(0.170)	(0.132)	(0.132)		(0.370)
Difference between mortgage		-4.85e-10	2.84e-10		
and installment loans		(2.41e-09)	(2.38e-09)		
Rounded loan report (R2)		-0.290	-0.332	-0.108	-0.177
		(0.291)	(0.293)	(0.283)	(0.287)
Age (years)			0.000492		0.00190
			(0.00484)		(0.00466)
Education (years)			-0.0949***		-0.135***
			(0.0275)		(0.0262)
Constant	-0.619***	-0.686***	0.626	-1.083***	0.752
	(0.211)	(0.153)	(0.482)	(0.145)	(0.462)
Number of Observations	1,627	1,627	1,627	1,627	1,627
Pseudo R-squared	0.142				0.026

Robust standard errors in ().

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

R1 rounding consists of a dummy, with 1 corresponding to a first digit of 1 or 5, while all the other remaining digits are 0.

R2 rounding consists of a dummy, with 1 corresponding to one in criterion R1 and also to 1 if a number has any first digit between 1 and 9 and is followed by at least four zeros.

Table F.2: Installment loan exists in the survey, but not in the registry (logit)

Control variables	Model 1	Model 2	Model 3	Model 4	Model 5
Dummy for borrower that	0.0484	-0.0845	-0.0603	-0.0335	-0.0557
is not highest-income member$_i$	(0.275)	(0.207)	(0.209)	(0.198)	(0.203)
Number of household members	0.0316	0.0361	-0.0200	0.0927	0.0262
with the highest-income$_i$	(0.162)	(0.123)	(0.124)	(0.117)	(0.119)
Dummy for borrower that	0.0847	-0.0920	-0.138	-0.0116	-0.0644
is not the oldest member$_i$	(0.215)	(0.153)	(0.159)	(0.145)	(0.152)
Dummy for another installment	-1.099***	-1.793***	-1.605***		
or mortgage loan in registry$_i$	(0.370)	(0.252)	(0.256)		
Rounded loan report (R2)		-0.0529	-0.140	0.0873	-0.0391
		(0.143)	(0.146)	(0.136)	(0.140)
Age (years)			-0.00581		-0.00405
			(0.00564)		(0.00554)
Education (years)			-0.237***		-0.285***
			(0.0412)		(0.0396)
Constant	-0.769***	-0.667***	2.935***	-1.098***	3.153***
	(0.235)	(0.187)	(0.699)	(0.174)	(0.678)
Number of Observations	1,050	1,063	1,063	1,095	1,095
Pseudo R-squared	0.026				

Robust standard errors in ().

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

R1 rounding consists of a dummy, with 1 corresponding to a first digit of 1 or 5, while all the other remaining digits are 0.

R2 rounding consists of a dummy, with 1 corresponding to one in criterion R1 and also to 1 if a number has any first digit between 1 and 9 and is followed by at least four zeros.

References

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

Table F.3: Linear regressions (OLS) of the $error\ ratio_{i,t}$ for the debt amounts of borrowers with positive loans in both survey and registry

Variables	Mortgage	Consumer debt	Installment loan	Card or Line	Credit Card	Credit Line
Dummy: borrower not	11.21	6.733	7.980	10.69**	16.51***	-0.433
highest-income member$_i$	(7.102)	(4.304)	(5.549)	(4.882)	(5.076)	(10.16)
Nr of members with	3.508	1.713	-0.480	5.080	2.791	1.948
the highest-income$_i$	(4.450)	(2.788)	(3.428)	(3.392)	(3.554)	(7.135)
Dummy: borrower not	-4.560	0.924	3.965	2.732	6.034	9.864
the oldest member$_i$	(4.641)	(3.320)	(4.076)	(3.745)	(3.894)	(7.376)
Rounded loan report (R2)	61.13***	3.979	5.214	-9.384***	-9.286***	-20.53***
	(8.997)	(3.181)	(3.735)	(3.450)	(3.562)	(6.638)
Age (years)	0.206	0.0576	0.311**	0.0454	0.00924	-0.291
	(0.206)	(0.117)	(0.148)	(0.132)	(0.137)	(0.268)
Education (years)	-0.329	1.138	0.389	-1.456*	-1.142	-1.845
	(1.008)	(0.723)	(0.886)	(0.827)	(0.858)	(1.632)
Constant	31.90*	42.16***	25.76	74.71***	68.27***	111.3***
	(19.31)	(13.21)	(16.66)	(14.85)	(15.42)	(29.36)
Observations	427	1,491	740	1,061	940	370
R-squared	0.132	0.005	0.013	0.023	0.035	0.046

Robust standard errors in ().

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

R1 rounding consists of a dummy, with 1 corresponding to a first digit of 1 or 5, while all the other remaining digits are 0.

R2 rounding consists of a dummy, with 1 corresponding to one in criterion R1 and also to 1 if a number has any first digit between 1 and 9 and is followed by at least four zeros.

Table F.4: Linear regressions (OLS) of the $error\ ratio_{i,t}$ for the debt amounts of borrowers with positive loans in both survey and registry (regressions with no demographics or rounding dummy)

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.

Documentos de Trabajo **Banco Central de Chile**	**Working Papers** **Central Bank of Chile**
NÚMEROS ANTERIORES	PAST ISSUES
La serie de Documentos de Trabajo en versión PDF puede obtenerse gratis en la dirección electrónica: www.bcentral.cl/esp/estpub/estudios/dtbc. Existe la posibilidad de solicitar una copia impresa con un costo de Ch$500 si es dentro de Chile y US$12 si es fuera de Chile. Las solicitudes se pueden hacer por fax: +56 2 26702231 o a través del correo electrónico: bcch@bcentral.cl.	Working Papers in PDF format can be downloaded free of charge from: www.bcentral.cl/eng/stdpub/studies/workingpaper. Printed versions can be ordered individually for US$12 per copy (for order inside Chile the charge is Ch$500.) Orders can be placed by fax: +56 2 26702231 or by email: bcch@bcentral.cl.

DTBC – 1069
How accurately do consumers report their debts in household surveys?
Carlos Madeira

DTBC – 1068
Riesgo de Crédito Gestionado por Medio de un Modelo de Espacio-Estado Aplicado a un Portafolio Soberano
Pablo Tapia, Diego Vargas

DTBC – 1067
Macroeconomic Effects of Carbon-intensive Energy Price Changes: A Model Comparison
Matthias Burgert, Matthieu Darracq Pariès, Luigi Durand, Mario González, Romanos Priftis, Oke Röhe, Matthias Rottner, Edgar Silgado-Gómez, Nikolai Stähler, Janos Varga

DTBC – 1066
Bank Branches and the Allocation of Capital across Cities
Olivia Bordeu, Gustavo González, Marcos Sorá

DTBC – 1065
Effects of Tariffs on Chilean Exports
Lucas Bertinatto, Lissette Briones, Jorge Fornero

DTBC – 1064
Does Participation in Business Associations Affect Innovation?
Felipe Aguilar, Roberto Álvarez

DTBC – 1063
Characterizing Income Risk in Chile and the Role of Labor Market Flows
Mario Giarda, Ignacio Rojas, Sergio Salgado

DTBC – 1062
Natural Disasters and Slow Recoveries: New Evidence from Chile
Lissette Briones, Matías Solorza

DTBC – 1061
Strategic or Scarred? Disparities in College Enrollment and Dropout Response to Macroeconomic Conditions
Nadim Elayan-Balagué

DTBC – 1060
Quantifying Aggregate Impacts in the Presence of Spillovers
Dave Donaldson, Federico Huneeus, Vincent Rollet

DTBC – 1059
Nowcasting Economic Activity with Microdata
Diego Vivanco Vargas, Camilo Levenier Barría, Lissette Briones Molina

DTBC – 1058
Artificial Intelligence Models for Nowcasting Economic Activity
Jennifer Peña, Katherine Jara, Fernando Sierra

DTBC – 1057
Clasificación de Riesgo de Crédito Bancario, Ventas y Estados Financieros en Base a Información Tributaria de Firmas en Chile
Ivette Fernández D., Jorge Fernández B., Francisco Vásquez L.

DTBC – 1056
Exogenous Influences on Long-term Inflation Expectation Deviations: Evidence from Chile
Carlos A. Medel

DTBC – 907*
Earnings Inequality in Production Networks
Federico Huneeus, Kory Kroft, Kevin Lim

DTBC – 1055
Markup Distribution and Aggregate Dynamics
Mario Giarda, Will Jianyu Lu, Antonio Martner

DTBC – 1054
Decoding Central Banks' Practices: A Closer Look at Inflation Expectations Surveys
Valentina Cortés Ayala, Karlla Muñoz Cáceres, Daniel Pérez Klein

DTBC – 1053
An Assessment of the Effects of Monetary Policy Communication in Chile
Mario González-Frugone, Ignacio Rojas

DOCUMENTOS DE TRABAJO Marzo 2026