BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

Households' portfolio holdings and their banks' characteristics: is there a link?

by Massimiliano Affinito, Ginette Eramo, Romina Gambacorta and Marco Langiulli

July 2025

Number 950

BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

Households' portfolio holdings and their banks' characteristics: is there a link?

by Massimiliano Affinito, Ginette Eramo, Romina Gambacorta and Marco Langiulli

Number 950 – July 2025

The series Occasional Papers *presents studies and documents on issues pertaining to the institutional tasks of the Bank of Italy and the Eurosystem. The* Occasional Papers *appear alongside the* Working Papers *series which are specifically aimed at providing original contributions to economic research.*

The Occasional Papers *include studies conducted within the Bank of Italy, sometimes in cooperation with the Eurosystem or other institutions. The views expressed in the studies are those of the authors and do not involve the responsibility of the institutions to which they belong.*

The series is available online at www.bancaditalia.it .

ISSN 1972-6643 (online)

Designed by the Printing and Publishing Division of the Bank of Italy

HOUSEHOLDS' PORTFOLIO HOLDINGS AND THEIR BANKS' CHARACTERISTICS: IS THERE A LINK?

by Massimiliano Affinito*, Ginette Eramo**, Romina Gambacorta*** and Marco Langiulli*

Abstract

This paper analyses the relationship between households' portfolio decisions and the characteristics of the banks where they hold their wealth. We use a unique bank-client level dataset covering the period 2004-20, which combines information from Banca d'Italia's Survey on Household Income and Wealth with supervisory reports that banks in Italy are required to submit to Banca d'Italia, matching the two sources by the identity of each household's primary bank. Our results indicate that banks play a crucial role in reducing households' non-participation in financial markets, thereby helping to mitigate the associated welfare costs. Moreover, we find that, even after controlling for all individual household characteristics commonly used in the literature and taking into account potential sources of self-selection and sorting between households and banks, certain bank characteristics remain significantly associated with households' portfolio decisions. These links vary depending on the class of households' asset holdings and the category of banks.

JEL Classification: D14, G21.

Keywords: household financial investments, bank-customer relationship, bank factors, self-selection, sorting.

DOI: 10.32057/0.QEF.2025.950

---

* Banca d'Italia - Directorate General for Economics, Statistics and Research.

** Banca d'Italia - Directorate General for Financial Supervision and Regulation.

*** Banca d'Italia - Directorate General for Human Resources, Information and Banknote Production.

1 Introduction[1]

This paper examines the relationship between portfolio holdings of households and the characteristics of the banks where they are clients. The literature has shown that bank characteristics influence household decisions with regard to borrowing (e.g., Foà et al., 2019; Damar et al., 2020; Guiso et al., 2022; Affinito et al., 2025). However, when it comes to financial portfolio decisions, the literature (e.g., Guiso et al., 2002; Cardak & Wilkins, 2009; Guiso & Sodini, 2013; Ampudia Fraile et al., 2016) tends to focus solely on demographic and socio-economic traits of households (age, income, wealth), largely overlooking the possible role of bank characteristics (size, business models, profitability). Yet, banks can play a crucial role in the broader financial lives of their customers, and the close relationship between banks and their retail clients may also be reflected in investment decisions.

First, today, particularly in developed countries, households take on more financial risk than in the past in response to demographic changes and increased responsibility for retirement provision, and financial markets offer a wider range of financial investments. Second, household financial knowledge is often limited (e.g., Lusardi & Mitchell, 2007; van Rooij et al., 2011), and retail investors need to rely on professional support when purchasing financial assets. Third, banks are often the primary entry point for households into the financial markets, and clients may turn to bank employees with their questions and seek support before making investment decisions (e.g., Guiso et al., 2004; Guiso et al., 2008; Georgarakos & Inderst, 2011; Gennaioli et al., 2015).[2] When bank staff provide advice to clients, their perspectives are likely shaped by the bank's corporate culture, experience, characteristics, and business model. As a result household financial investment decisions may be associated (also) with the characteristics of their bank, leading to variations in portfolios across different banks. The question is potentially relevant as, if the relationship between households and their bank influences portfolio holdings, this can

[1]The authors wish to thank Laura Bartiloro, Silvia Fabiani, Andrea Neri and Alfonso Rosolia for the helpful comments and suggestions. The views expressed are those of the authors and do not necessarily reflect those of the Bank of Italy. Contact emails: Massimiliano.Affinito@bancaditalia.it, Ginette.Eramo@bancaditalia.it, Romina.Gambacorta@bancaditalia.it and Marco.Langiulli@bancaditalia.it.

[2]In Italy, according to the Bank of Italy Survey of Household Income and Wealth (SHIW), 96 per cent of Italians who have made a financial investment have purchased it through a commercial bank, and 85 per cent of them have relied on advice from its staff. The situation is similar in other systems where the role of banks is pivotal. In Germany, for example, more than 80 per cent of investors consult a financial advisor before making a financial investment decision, and this is a bank for more than two thirds of German investors (Bluethgen et al., 2008).

have an impact on the overall level and quality of an economy's savings and investments.[3]

The main reason why this issue, despite its potential relevance, has been largely neglected in the empirical literature is probably the challenge of matching the identity of retail clients, and their corresponding financial decisions, with the identity of the banks they use, and the respective characteristics. In this paper, we have the opportunity to utilise a unique dataset, which merges for the period 2004-2020 data from the Bank of Italy's Survey of Household Income and Wealth (SHIW) with individual bank data of each household's primary bank from the Bank of Italy's supervisory reports. Our dataset encloses: (i) whether the household has a reference bank ("banked household"), or no reference financial institution ("non-banked household") or another kind of reference institution; (ii) in the case of "banked" households, the exact match of each household with each bank; (iii) the financial assets held by each household, which we group (looking at the underlying asset's risk profile) in medium-risk and riskier assets (and analyse bank bonds also separately, as a relevant subset of the medium-risk assets); (iv) a comprehensive set of demographic and socio-economic characteristics of households, so that we can explore our bank-side variables of interest while controlling for all the main household-side drivers found in the literature; and (v) information on each bank's business model and balance sheet conditions.

Our empirical strategy relies on different estimation models. In particular, we rely on two selection bias correction methods (the former proposed by Dubin & McFadden, 1984; and the latter by Dahl, 2002), which allow us to correct for sample selection biases arising from self-selection (or sorting) of households across banks. In our framework there are potentially two sources of sample selection bias. Households in the survey could report in a non-random way whether they have a bank, for example, because of their particular investments, conditions or preferences, and, even regardless of the survey, households might choose their bank non-randomly, in the sense that households with certain characteristics could be more likely to choose banks with specific characteristics. In both cases, estimates without bias correction would be biased and inconsistent. Both correction models we use follow a two-step procedure: the first step estimates the selection process between households and their reference institution, while the second step estimates the households' financial asset holdings. This approach is similar to the seminal correction method introduced by Heckman (1979). However, while Heckman's model considers a binary choice in the first step, the Dubin & McFadden and Dahl models account for more complex

[3]It should be noted that comparing household investment decisions with optimal portfolio allocation is outside the scope of our paper. To investigate this issue, we would also need specific information, not available in our data, on the financial assets purchased and sold by individual families and the exact timing of these transactions to estimate returns and to compare them with a benchmark.

selection structures, where the selectivity involves a larger set of options. Specifically, in the first step, the selection process assumes that households maximize their welfare under perfect knowledge, and their matching with the reference financial institution reflects all relevant characteristics. As a result, the correction estimations imply that the decisions made by households are optimal. Moreover, we exploit our rich dataset by saturating regressions with a large set of fixed effects defined at time-, bank-, and geographical-level, which contribute to our identification strategy by absorbing residual, observable and unobservable, factors that can affect financial investment decisions.

Our main results can be summarized as follows. First, our results confirm the existence of sorting, that is, show that households single out their reference financial institution, which also confirms that it is appropriate to apply a method that takes into account and allows for sample selection. Second, being a household with a bank ("banked household") - after controlling for the endogenous nature of the choice - favours participation in financial markets. This may be a relevant contribution of banks to growth and economic welfare of their clients. Non-participation in risky asset markets and under diversification of risky portfolios are well-known causes of high welfare costs (Haliassos & Bertaut, 1995; Guiso & Jappelli, 2005; Campbell, 2006; Calvet et al., 2007; Bolton et al., 2007). Third, even after controlling for the household characteristics and their sorting across banks, bank characteristics are significantly associated to financial decisions of their retail clients. Households are more likely to invest in medium-risk assets when their bank is specialized in proprietary financial investments rather than traditional lending, relies more on bond issuance, has lower profits. Households are more likely to hold riskier financial assets when their bank has a larger share of non-performing loans.

To the best of our knowledge, there are no empirical studies on the links between bank characteristics and households' financial investments. However, despite the novelty of our analysis, our paper is connected to three major strands of research.

First, as mentioned above, our perspective of examining the links between household decisions and their bank characteristics is already utilized in the literature on household borrowing. For example, Damar et al. (2020) study whether banks' exposure to different funding sources and shocks affects the credit and consumption expenditures of Canadian households. Foà et al. (2019) and Guiso et al. (2022) investigate bank lending for mortgages in Italy and the choice between adjustable and fixed rates, and find that it is significantly affected by changes in banks' supply factors. Affinito et al. (2025) show that household debt growth in Italy is related to both bank and household characteristics, and in different ways for mortgages and consumer loans. We build upon this literature by selecting the bank characteristics that may be more influential in household decisions, and we extend it by delving into a different area of the financial relationship between banks and their retail customers.

Second, our paper is related to the literature that examines the characteristics of financial advisors and the impact they have on composition and performance of portfolios. This literature emphasizes that consumers and investors often make mistakes due to a lack of financial education, limited information, and cognitive biases (Campbell, 2006; Calvet et al., 2007; Inderst & Ottaviani, 2009; Uppal & Bhamra, 2017; Affinito et al., 2024a). It also suggests that financial institutions can help mitigate these failures by acting as financial advisors (Haliassos & Bertaut, 1995; Bluethgen et al., 2008; Guiso & Jappelli, 2005; Bolton et al., 2007). Our contribution is to show that banks play a central role in reducing households' non-participation in financial markets.

Third, our paper is related to the vast literature on conflicts of interest within financial institutions. This stream of research argues that financial institutions may prioritize their interests over those of their clients. In the case of banks, in particular, the combined exercise of production and distribution of financial products, along with the market knowledge gained through proprietary trading, could urge them to steer clients toward specific products, particularly when these generate higher brokerage fees, and to facilitate the placement of financial assets issued by the banks themselves, their issuing clients, other institutions in the same banking group, or from their own portfolios (e.g., Crockett et al., 2003; Massa & Rehman, 2008; Acharya & Johnson, 2007; Inderst & Ottaviani, 2009; Ivashina & Sun, 2011; Hackethal et al., 2012; Fecht et al., 2018; Albareto et al., 2020). However, this literature has led to mixed conclusions, generally acknowledging that, even when a conflict is identified, it does not necessarily entail to harm clients, as a negative impact on clients' outcomes would only be evident if the quality of bank services or investor returns were compromised, which is far more challenging to prove (e.g., Focarelli & Pozzolo, 2005; Cabral & Santos, 2001; Drucker & Puri, 2006; Mehran & Stulz, 2007; Egan, 2019; Albareto, 2019). Some of our results echo the conflict-of-interest literature, particularly since we find that households are more likely to purchase bank-issued bonds when they are customers of banks that issue a larger volume of such bonds. However, even in our case, not only can this positive correlation not imply a causal relationship (since we cannot determine whether it is the households driving the demand for bank bonds and prompting the banks to issue more, or whether the reverse is true), but moreover, exactly in line with the arguments from this literature, even if banks were to drive the demand for their bonds, our results do not imply any detrimental effect on investors' returns.

The paper is organized as follows. Section 2 summarizes our baseline empirical strategy. Section 3 describes our dataset. Section 4 discusses the baseline results. Section 5 and Section 6 allow for self-selection and sorting of households across banks. Section 7 summarizes other robustness checks. Section 8 draws the main conclusions.

2 Empirical strategy

To explore household portfolio composition we group all financial assets into two main classes (medium-risk and riskier assets) according to their underlying risk, and focus our attention to whether the household holds these asset classes. As we clarify more in detail in the next Section, we exclude from our analysis safe assets (i.e, deposits, saving accounts, etc.) as these are held by nearly all households in our dataset. Instead, we also separately study bank bonds, which are a relevant subset of the medium-risk asset class and have been traditionally held by households in Italy more than in other countries (Bank of Italy, 2011; Bank of Italy, 2015; Coletta & Santioni, 2016; Albareto, 2019).

Our estimation model is specified as follows:

$$\begin{aligned} Prob(y_{ibt}=1) = \alpha_1 \,+\, \beta_1^\tau B_{bt} \,+\, \eta_1^\tau H_{it} \,+\, \mu_1^\tau M_{ibt} \,+ \\ +\, \tau_1^\tau d_t \,+\, \gamma_1^\tau d_g \,+\, \phi_1^\tau d_b \,+\, \epsilon_{ibt}, \end{aligned} \tag{1}$$

where y_{ibt} is equal to 1 if household i customer of bank b at time t holds a specific class of assets (alternatively, medium-risk assets, riskier assets, and bank bonds). The covariates are defined at bank-, household-, and bank household pairwise- level: B_{bt} is a matrix of bank b characteristics at time t; H_{it} is a matrix of household i characteristics at time t; and M_{ibt} is a matrix of household-bank ib pair characteristics. Importantly, Equation (1) is saturated with vectors of time (d_t), geographical (d_g) and bank (d_b) fixed effects; ϵ_{ibt} is a vector of identically and independently distributed errors.

The matrix of bank characteristics B_{bt} contains our variables of interest, to respond to whether and how features of banks are associated with the financial decisions of their household customers. Among bank characteristics, we consider the main bank traits which are commonly used in the literature on bank lending to capture banks' business models and characteristics. First, B_{bt} contains the variable Bank Size (natural logarithm of total assets), which is important as banks of different size have often different conducts on several profiles. Second, two variables measure bank health and soundness: Bank Capital measures the available capital as a share of risk weighted assets; Bank NPLs measure the burden of non-performing loans to total loans. Third, three variables measure the relevance of each source of funding on the liability-side of bank balance-sheets: Funding via Deposits, Funding via Bonds, and Liquidity from Central Bank (all scaled to total assets). Fourth, two variables, Bank Loans to private sector and Bank Holdings of Gov't Bonds (as ratios to total assets), capture alternative specializations of different lines of business. Finally, variables drawn from income statements measure bank performance, either through Bank Return on Assets (ROA), that is, a brief indicator of banks' total profits, or through three relevant components of banks' income statements: net interest income, fees and commissions, and operating expenses (e.g., Affinito et al., 2023).[4]

The matrix of household-bank pair characteristics M_{ibt} contains two regressors, which are specific of each bank-household relationship. The former is a dummy capturing the use of a remote connection and it seizes the existence of a digital customer relationship between the household and the bank (Agarwal & Chua, 2020, Frost et al., 2022). The latter covariate computes whether the pair household-bank have an exclusive relationship or whether the household relies on more than one bank (e.g., Angelini et al., 1998; Morris & Shin, 2004; Gobbi & Sette, 2014; Affinito & Meucci, 2023).

The matrix of household characteristics H_{it} is also crucial for our identification strategy as it includes all the main investor-side features that the literature shows that may be relevant in explaining the financial decisions of individual investors. The inclusion of these variables is important because, while we are interested in the regressors in B_{bt} and M_{ibt}, the presence of regressors in H_{it} limits omission biases, allows for confounding effects and verifies the consistency of our results to the previous literature, which enables to contextualize our contribution. Moreover, estimating the household characteristics H_{it} helps shed light on what kind of customers the connections we (possibly) find for B_{bt} and M_{ibt} work and are effective. For example, we can verify whether certain links apply on average to poorer or richer households, and to more or less educated households. The variables in the matrix H_{it} refer to phenomena such as age, income, wealth, work status, household type, risk aversion, financial literacy (e.g., Guiso et al., 2002; Cardak & Wilkins, 2009; Lusardi, 2012; Hastings et al., 2012; Guiso & Sodini, 2013; Ampudia Fraile et al., 2016; Guiso & Viviano, 2015), and are detailed in the next Section.

The time d_t, geographical d_g and bank d_b fixed effects (henceforth, FEs) also contribute to our identification strategy by absorbing residual factors affecting financial investment decisions that are not explicitly controlled for by the explanatory variables, and removing all observable and unobservable differences in the cross-section, at geographical and bank level, and over time (Ludvigson, 1999; Nakajima & Ríos-Rull, 2014). In particular, geographical fixed effects d_g control for specific institutional settings and local economic and financial features that can influence holding decisions; bank fixed effects d_b control for the time-invariant components of bank characteristics, limit concerns about omitted variables, and are a first means to address self-selection biases; time fixed effects d_t control for changes in the business cycle and macro policies. The latter also is important because

[4] To avoid problems of multicollinearity, and assure parsimony, related variables (i.e., Bank Loans versus Bank Holdings of Gov't Bonds; Deposits versus Bonds; ROA versus income statement components) are used alternatively in the estimations.

changes in the business cycle, monetary policy stance, or government spending can also affect financial decisions of individual households.

We estimate Equation 1 using several regression models.[5] Our preferred estimation choice are the sample selection bias correction models (i.e., the Dubin & McFadden (1984) and the Dahl (2002) model), which allow us to control for sorting and sample selection effects. The selection bias correction models are made of two steps, and are described in detail in Section 5. However, we start with a linear fixed effects estimator (FE-OLS), which serves as a necessary benchmark for the second step of the sample selection bias correction models, which uses a linear estimation. To manage the heterogeneity in variability all our estimations are clustered at year, bank and geographical level (MacKinnon et al., 2023).

3 Data

We rely on two main data sources: the Bank of Italy's Survey on Household Income and Wealth (SHIW), and the supervisory reports that all banks operating in Italy have to transmit to the Bank of Italy. We merge these two sources by exploiting the information on the identity of the primary bank households rely upon. The final database spans the period 2004-2020, thus covering phases characterised by very different macroeconomic and financial conditions, for a total of about 60,000 observations. In the rest of the section, we detail the two data sources and the construction of our dataset. Tables 1-4 report the composition of our sample, the list of the variables, and the summary statistics.

Survey on Household Income and Wealth

The SHIW contains data on many socio-demographic characteristics of household (henceforth, HH) members and their sources of income and accumulated wealth.[6] We use the SHIW: (i) to collect information on financial asset investments of households (that is, the information underlying the dependent variable y_{ibt} of Equation 1); (ii) to compute the regressors in the matrices of household H_{it} and pair M_{ibt} characteristics; and (iii) to match each household with the reference financial institution. More in detail.

First, we take the data on household financial assets from the SHIW. Survey data on financial assets are generally affected by non-sampling errors. In the SHIW, these errors

[5] See robustness described in Section 7.

[6] The survey is conducted every two years on a sample of about 8,000 households representative of the Italian population, and about half of the households have already participated in previous surveys (panel households). The sample is obtained using a two-stage stratified sample design. The information about financial endowments is collected with great detail using specific techniques to reduce under-reporting (i.e., unfolding bracket technique). For more details, see Bank of Italy (2022).

mainly relate to the under-reporting of the value of assets (Neri & Rannalli, 2011), while the data on the holdings of financial assets are more accurate (D'Aurizio et al., 2006; Baffigi et al., 2016). For this reason, we measure households' portfolio decisions mainly by whether or not they hold certain categories of financial assets, rather than by the reported value of the amounts held. However, in Section 7 we also describe some exercises we performed with the amounts. We focus on financial assets *other than safe assets* (i.e. we exclude deposits, savings accounts and other smaller, highly liquid safe assets), as these are held by almost all households that report having a banking relationship and by many of those that use other reference institutions. Then, we group financial assets *other than safe assets* according to their underlying riskiness into two classes: (i) *medium-risk assets* (which include Italian government securities, bonds, mutual funds' units in bonds, money market and liquidity in euros); and *riskier assets* (which include shares and other equity, managed portfolio, foreign securities, mutual funds' units in equities, mixed or in foreign currencies, and complex financial assets such as option, futures, etc.). We also analyze in isolation the holdings of bank bonds (only collected since 2010), which are medium-risk assets.

Second, we extract from the SHIW the matrix of household characteristics H_{it}, that is, a rich amount of information about household characteristics that may affect financial decisions and the effectiveness of the influence exercised by banks. Many variables are typically used in the related literature, such as age class, equivalent income and wealth quintile, household type (e.g., single, couple, with or without children, etc.), work status (e.g., blue-collar, office worker, manager, sole proprietor, etc.), working in the financial sector, risk aversion, financial literacy.[7]

Third, we use the SHIW to match the identity of households and banks. We group all households in our dataset into three types (Table 1). The first type consists of "households with a bank" (or "banked HHs"), of which we know the exact match between the household and the primary bank.[8] Banked HHs include approximately 42,000 observations, due

[7]Individual characteristics refer to the head of the household, defined as the main income earner with the exception of risk aversion and financial literacy which is only collected for survey respondents. However, the two individuals coincide in most cases. To purge the effect of the family composition we consider quintiles of both equivalent income and equivalent wealth, obtained using the OECD modified equivalence scale (Hagenaars et al., 1994).

[8]SHIW collects the names of the banks used by the households and the indication of which is the main one. We focus on the bank indicated by the family as the primary one. In the analysed period about 80 per cent of the "banked" households have just one bank. In several cases, the names of the banks had not been coded, so it was not possible to match households' data with their banks. In these cases, we reconstructed backwards a - as much as possible - complete dataset on households' banks for the period 2004-2020. Information on foreign banks operating in Italy is not always fully available. In these cases, when data are available, we replace the main bank indicated by the household with another Italian bank

to almost 24,000 households who use 249 banks. The second type consists of "households with Poste". Poste Italiane is the national postal operator, which since several years is a major competitor of banks as it provides a bundle of financial services (e.g., overnight and saving deposits, payments, debt cards, etc.), and it is preferred by many households, although it cannot grant loans and it is not legally a bank.[9] Households relying only on Poste Italiane include around 6,000 families and about 9,000 observations. The last type consists of households who do not reveal their reference institution or declare not to have any reference financial institution ("non-banked HHs", around 7,000 families and 9,000 observations). We retain information on "households with Poste" and "non-banked households" to conduct our tests on sample selection bias (Section 5). We exclude from our final dataset some observations (approximately 800) referring to households that are clients exclusively of foreign banks, or that use financial institutions other than banks on which we do not have sufficient information, or declare banks that cannot be identified.[10]

Overall, as shown in Table 1, around 68 per cent of observations refer to households that declare to have at least one bank relationship; about 15 per cent to households using Poste Italiane; 15 per cent (with a share decreasing over time) to households that report having no financial institution or do not reveal the financial institution they use. A residual of less than 2 per cent (decreasing over time) are the observations we drop.

Banking data

As explained, the matrix B_{bt} includes a wide set of bank level data. We draw these information from the supervisory reports of the Bank of Italy.[11]

We aggregate all the information at the banking group level to which the bank belongs to instead of using individual bank level data, based on the fact that banking groups internally implement policies for the redistribution of resources.[12] To avoid data breaks following mergers and acquisitions, we apply the standard technique (the so called *pro-*

used by the family.

[9]A part of households that uses Poste has at least another banking account; in these cases, we performed exercises considering or not this other bank as the reference institution (and including or not the households between the households with a bank).

[10]The households we remove are too few for considering a separate group, and too different for including them with those who use Poste or no institution.

[11]We winsorize cases with outliers, first and last two percentiles of the distribution of the variables.

[12]In recent years, the Italian banking system has experienced a strong reform process and many cooperative banks aggregated into larger groups. During 2018 and 2019 more than 200 cooperative credit banks merged in two big groups (ICCREA Group and Cassa Centrale Group). As the integration process takes time, for these banks individual indicators provide a better assessment than the consolidated data for the year 2020; so we used individual data instead of a single value for all the banks belonging to the same group.

forma approach) of simulating that all M&As occurred at the beginning of the sample period. The banks in the sample account for about 80 per cent of the total assets of banks operating in Italy.

4 Baseline results

As mentioned, we start with FE-OLS estimations of Equation 1. Table 5 reports the results for the three alternative dependent variables (medium-risk assets, riskier assets, and bank bonds), and for two specifications for each dependent variable, which alternate as geographical fixed effects d_g either the macro-areas of Italy or the much more numerous administrative provinces.[13]

First of all, it is worthwhile stressing that the coefficients of our control regressors in the matrix H_{it} are all in line with what is generally found in the literature. As mentioned, this is important because, while we are interested in the links with bank-side characteristics, it is crucial also to take into account the other factors in order to avoid biased results on the variables of interest, and verify what clients concern. In this respect, our results confirm that the holding of medium-risk and riskier financial asset is positively related to income, wealth and financial literacy (Abreu & Mendes, 2010; Gaudecker, 2015).[14]

Regarding bank characteristics, our results indicate that certain features of banks are significantly related to the financial holdings of their customer households.[15] The links exist only for certain features, vary depending on the type of financial asset, and may reflect both the asset-side and liability-side conditions of bank balance sheets. More in detail.

First, the results on the soundness of banks in shaping the financial decisions of customer investors are mixed. Bank Capital is never significantly associated with the probability for households of investing in any kind of financial asset. Nor do Bank NPLs correlate with households' decision to buy medium-risk assets or bank bonds. However, Bank NPLs are the only bank trait that is statistically and positively associated with household investments in riskier assets.

[13]Provinces are Italian administrative "counties", that is, administrative divisions of intermediate level between municipalities and Regions. In Italy there are three geographical macro-areas (North, Center, and South and Islands), and about (depending on the year) 105 provinces.

[14]In addition, we find that the holding of any kind of financial asset is significantly positive when the household head works in the financial sector. As risk aversion grows, households have a lower probability to hold riskier assets; moderately risk averse individuals have a larger probability to hold medium-risk assets and bank bonds. The results on control regressors are not reported; they are available upon request.

[15]The log-likelihood ratio test rejects the hypothesis that the restricted model in Equation 1 (without banking variables in B_{bt} and pair variables in M_{ibt}) fits the data as well as the full model.

Second, with regard to sources of funding, customers of banks that rely more heavily on Funding via Bonds are more likely to hold medium-risk assets and, specifically, bank bonds. Conversely, customers of banks that receive more Liquidity from Central Bank tend to invest less in bank bonds. The positive correlation between banks' decision to use more bonds as a source of funding and households' decision to hold more bank bonds as a financial asset does not untie the knot of causality. It could be that households demand more bank bonds and thus induce their banks to issue more, or it could even be that households choose those highly issuing banks exactly because are more willing to invest in bonds (see below, Section 5), or banks could issue more bonds and then encourage their customers to purchase them.[16]

In the latter case, the result would be consistent with the hypothesis of a conflict of interests of banks when act at once as issuers and sellers of securities (e.g., Cabral & Santos, 2001; Focarelli & Pozzolo, 2005; Drucker & Puri, 2006; Mehran & Stulz, 2007; Albareto, 2019). However, even if it were the banks pursuing the purchase of the bonds, our result would not demonstrate any adverse impact on investors. This limitation is common in the literature on conflicts of interest (e.g., Cabral & Santos, 2001; Drucker & Puri, 2006; Mehran & Stulz, 2007; Bolton et al., 2007).

In order to try to shed more light on households' decision to invest in bank bonds, we also exploit another piece of information in the SHIW, which measures for each household the trust toward the main bank (e.g., Guiso et al., 2004; Guiso et al., 2008; Georgarakos & Inderst, 2011; Gennaioli et al., 2015).[17] Results show that trust in the main bank is not related to the holding of riskier assets, while it is associated with households' investments in medium-risk assets, and specifically in bank bonds. However, this information was only collected in the 2010 edition of the SHIW, and its use significantly limits the available observations and the ability to run robust estimations.

Third, regarding the characteristics of bank business, customers of banks that have higher Holdings of Gov't Bonds hold on average more medium-risk assets and bank bonds. This suggests that households' decision to hold medium-risk financial assets is more likely when they are customers of banks that specialize in proprietary financial investments

[16] It is important to highlight that we do not have information on the identity of the issuing bank of the bond held by each household. In this respect, the aggregated supervisory report data show that in Italy more than 70 per cent of the bank bonds held by households in their portfolios are issued by a credit institution belonging to the same group as the household's bank (Bank of Italy, 2011; Bank of Italy, 2015; Coletta & Santioni, 2016). Therefore, while there is a high probability that the bond is issued by a bank of the group, the information in the SHIW is lacking.

[17] We refer to the following question: "Do you trust your principal bank? Please assign a score of 1 to 10, where 1 means 'I do not trust it at all', and 10 means 'I trust it completely' and the intermediate scores serve to graduate your response."

rather than traditional lending. Conversely, however, this does not apply to the decision to hold riskier assets. Customers' investments in medium-risk assets are also associated to the evolution of their bank's profitability as households that are clients of banks with a lower ROA hold on average more medium-risk assets. This result could signal that when bank business is less profitable, banks are more willing to ease the financial participation of their clients as this entails brokerage fees (Campbell, 2006; Calvet et al., 2007; Bolton et al., 2007). The interpretation is similar when we decompose total bank profits by the relevant components of banks' income statements (net interest income, fees and commissions, and operating expenses), and find that the coefficient of fees is statistically significant and positive in the regressions for medium-risk assets and bank bonds. In contrast, however, we find no other statistically systematic correlation of ROA and its components with households' portfolio decisions in favour of riskier assets and bank bonds.

Fourth, for the pairwise variables, households that have remote access to banking services are more likely to invest in financial assets, indicating that the use of technology and a digital customer relationship can facilitate financial participation among bank customers (Agarwal & Chua, 2020, Michelangeli & Viviano, 2021; Frost et al., 2022).

5 Sample selection correction

The results of the previous Section serve as a benchmark for the estimation we run to control for the sample selection and sorting effects. As mentioned in the Introduction, in our framework there are potentially two sources of sample selection bias, both stemming from self-selection (or sorting) of households (HHs). First, households could report in the survey (the SHIW) in a non-random manner whether they have a bank, and the identity of their bank, for example, if they have specific investments, conditions, or preferences. Second, regardless of the survey, households' choice of a bank (or the matching with a particular bank) could be non-random as households with certain characteristics may be more likely to choose banks with specific attributes. For instance, risk-averse households may be more inclined to select sounder banks, as these banks would better align with their preferences, tastes, and concerns. Moreover, customers of certain types of banks may be more responsive to specific bank characteristics precisely because they chose those banks for those characteristics. In all these cases, the self-selection of households into a particular state (having or not a reference institution, and disclosing or not its identity) or into a specific alternative (choosing a particular bank), would cause a significant correlation between bank characteristics and the error term in Equation 1, and our FE-OLS estimates would be biased and inconsistent. The literature typically identifies the two cases of self-selection as *participation-selection* and *sector-selection*, and proposes several selectivity bias correction methods to control for them.[18]

We perform two models of selectivity bias correction: the model proposed by Dubin & McFadden (1984), and the semi-parametric approach developed by Dahl (2002). Similar to the seminal correction method introduced by Heckman (1979), both the Dubin & McFadden (DMF) and Dahl (DHL) models employ a two step procedure. In the first step, the *selection equation* estimates the matching or selection process between HHs and the reference institution (or the lack of a reference institution); in the second step, the *outcome equation*, which is the actual equation of interest, estimates the HH financial asset holdings.[19] The two-step procedure allows to enrich the *outcome equation* with the inclusion of correction terms that take into account the selection process estimated in the first step. While the Heckman (1979) model considered a *binary* choice in the selection equation of the first step, typically estimated through a probit model, the DMF and DHL models account for a more complex structure of selection, which extends to the case where selectivity concerns a larger number of exclusive choices, and is modeled as a *multinomial logit*.[20]

In comparison to other methods of correction (for example, the switching regression model or the matching models), these multinomial logit correction models are particularly suitable for our identification issue since they allow to consider a range of alternatives/states and to control for both the *participation-selection* (in our case, having and revealing a reference institution) and the *sector-selection* (in our case, choosing a particular bank).[21] In the next sub-sections we describe more formally our DHL and DMF

[18]Many of these issues, and the resulting selectivity bias correction methods, were developed in labor economics, where typical research question often involves the estimation of wages, a process that is particularly prone to selection bias. In fact, the estimation of wages generally entails a high correlation between unobservable factors affecting wages and those determining the decision to participate in the labor market (*participation-selection*), or the regions and sectors in which individuals live or work (*sector-selection*).

[19]The DHL model in particular stems from a Roy model, in which therefore we can assume that in the first step households select their main bank to maximize their welfare under perfect knowledge of the bank's relevant characteristics.

[20]Lee (1983) was the first to introduce a generalization of the two-step selection bias correction method with multiple selection options. Compared to Lee (1983), the DMF and DHL models allow for any sign structure of covariances between the outcome equation residual and the various error terms from the selection model. For the Dubin & McFadden (1984) correction method, we use the variant proposed by Bourguignon et al. (2007), which allows for normally distributed residuals in the main equation, showing that waiving the restrictions of the original model allows to achieve more robust estimators.

[21]See Nakosteen & Zimmer, 1980; Lokshin & Sajaia, 2006; Bourguignon et al., 2007; Kai & Prabhala, 2007; Jung, 2017. In particular, Bourguignon et al. (2007) present a survey of these models, and show that the DMF and DHL models provide fairly good correction for the outcome equation, even when the independence of irrelevant alternatives (IIA) hypothesis is violated. The IIA is a drawback of the

estimation equations, and the main features of the four exercises we conduct using these models.

A system of two equations

In formal terms, the DHL and DMF models can be both specified through a system of two equations. The first equation is the *selection equation*, which estimates through a multinomial logit whether the household i belongs to the state j, and may be described as follows:

$$x_{ijt} = \alpha_2 + \eta_2^\tau H_{it} + \kappa_2^\tau K_{jt} + \tau_2^\tau d_t + \gamma_2^\tau d_g + \phi_2^\tau d_j + \eta_{ijt} \quad (2)$$

where x_{ijt} indicates whether at time t household i belongs to the state j, where $j = 1, 2, ..., n$, and n is the total number of states. In the various exercises we perform, the state j indicates, for each household i, either the type of household (i.e., whether the household i is a "banked household", a "household with Poste", or a "non-banked household"), or also the bank that each household i has chosen.

It is to clarify that, in the first step, HHs do not select a specific bank, but a type of institution (bank, Poste, or nothing) or also a category of bank (for example, large or cooperative banks). One could think that the selection equation should ideally control for the HHs' choice of each specific bank, and then the second step should include a correction term for each chosen bank. This cannot be the case. If in the first step (i.e., in the multinomial logit selection equation) each alternative j corresponded to a bank, we would have a very large number n of alternatives/banks (the number of banks in our dataset is about 200), and the number of required parameters would not allow the estimation of the *selection equation* of each bank. Even more importantly, as we discuss more in detail soon, if one were to estimate the *selection equation* for each bank in the first step, then one would also have to estimate an *outcome equation* for each bank (i.e., for the households customers of each bank) in the second step.[22]

Therefore, in more general terms, we have to balance a trade-off: the number n of alternatives (in our case the number of bank categories) must be large enough so that we

multinomial logit model (McFadden, 1974), which is due to the assumption that all errors are independent and implies that (conditional upon observed characteristics) utility levels of any two alternatives are independent (even if very similar). Bourguignon et al. (2007) show that, when the aim of the multinomial logit is the correction of the selection bias in the outcome equation (rather than the estimation of the selection process itself) multinomial logit estimates remain valid even if the IIA is severely violated.

[22] The estimation would be unfeasible (that is, it would suffer from missing data as it would be a regression for a single bank with only one observation for each bank-time characteristic). Indeed, even if it were feasible, it would be uninformative.

can study the selection process of HHs, but at the same time this number must not be too large and form coherent groups of banks so that we can estimate the portfolio holdings across the possible alternatives in the second step. The four exercises we run differ exactly in that the j alternatives (that is, the possible bank categories) and the total number n vary.

The matrix of household characteristics H_{it} in Equation 2 serves to control for all customer characteristics that can influence the choice or matching with a bank. The literature has not explicitly addressed the question of which factors influence the match between banks and their customers. However, there is a stream of research, albeit limited, that has analysed households' decision to switch banks and examined the HH characteristics that are likely to be associated with the switch (e.g., Kiser, 2002; Brown et al., 2013; Brunetti et al., 2016; Brown & Hoffmann, 2016). Our list of explanatory variables in the matrix H_{it} is consistent with these studies. Moreover, Equation 2 includes time d_t and geographical dummies d_g to control for economic and financial developments over time and local differences in economic and financial conditions, which too are factors that can contribute to explain the choice of a bank.

Furthermore, Equation 2 includes the matrix K_{jt}, which contains bank-side exogenous covariates that can influence the choice of a bank, and are grouped by the same categories j as the dependent variable x_{ijt}. We use to the purpose the number of bank branches in the province and the average bank interest rates (on loans and deposits).[23] The number of bank branches in the province may be a relevant factor of choice of the own bank. The relevance of geographical proximity has reduced over time due the effect of the progress of information and communication technologies. However, distance has been until a recent past a decisive variable of choice for both lenders and borrowers (Petersen & Rajan, 2002; Bonaccorsi & Gobbi, 2007; Alessandrini et al., 2009), and in Italy the correlation between the share of bank branches in a province and the share of loans and deposits of retail clients located in the same province remains close to 90 per cent. Also interest rates can be an important criterion when choosing a bank, as they can be a clear and immediate indicator of the bank's prices (Affinito & Farabullini, 2009).

[23]Although it is not necessary, the inclusion of these additional variables in the selection model is relevant, as the identification in the second step could otherwise rely only on the distributional assumption of the error terms. In any case, our regressors in the matrix K_{jt} easily overcome the (potential) exclusion restriction requirements, since it suffices to assume (very plausibly) that the number of branches in a province, or the average interest rates, can affect investment decisions of households only through the choice of the bank. Instead, Equation 2 cannot include the matrices of individual bank characteristics B_{bt} and household-bank pair characteristics M_{ibt} given that in this first step banks are grouped into categories. We carried out on the matrix K_{jt} several robustness checks, including the use of average computations of variables in B_{bt} and M_{ibt}. See below, in Section 6.

The second equation, the *outcome equation* that estimates the holding of different asset classes, can be described as follows:

$$
\begin{aligned}
y_{ibt_{(j=1)}} = \alpha_{3_{(j=1)}} + \beta^{\tau}_{3_{(j=1)}} B_{bt} + \eta^{\tau}_{3_{(j=1)}} H_{it} + \mu^{\tau}_{3_{(j=1)}} M_{ibt} + \\
+ \tau^{\tau}_{3_{(j=1)}} d_t + \gamma^{\tau}_{3_{(j=1)}} d_g + \phi^{\tau}_{3_{(j=1)}} d_b + \\
+ \sigma_{(j=1)} \sum_{i=2}^{n} \rho_{(j=1)} \left[\frac{P_j ln(P_j)}{1 - P_j} + ln(P_1) \right] + \chi_{ibt} \quad (3)
\end{aligned}
$$

where $y_{ibt_{(j=1)}}$ is the sample bias corrected probability that household i customer of bank b at time t holds a specific class of assets (like in Equation 1, financial assets are, alternatively, medium-risk assets, riskier assets, and bank bonds). The matrices of regressors and coefficients of Equation 3 are analogous to those of Equation 1 and are defined at bank (B_{bt}), household (H_{it}), and pair (M_{ibt}) level. The crucial difference between Equation 1 and Equation 3 is that the latter includes the correction factor $\sigma_{(j=1)} \sum_{i=2}^{n} \rho_{(j=1)} \left[\frac{P_j ln(P_j)}{1-P_j} + ln(P_1) \right]$, where P_j is the probability to choose j and $\sigma\rho$ is the coefficient term for the (polychotomous) correction of the selectivity bias. Importantly, $y_{ibt_{(j=1)}}$ is specific to the alternative state $j = 1$. This means that the Equation 3 is estimated as many times as the the number n of alternatives j, and therefore the estimated coefficients of interest referred to the matrices B_{bt} and M_{ibt} (as well as all the other coefficients of control variables) are obtained for each j. Therefore, Equation 3 is estimated for a group of HHs at a time: first the HHs who have chosen the alternative $j = 1$, then HHs of the alternative $j = 2$, and so on. As mentioned, our exercises differ exactly in the number and content of alternatives j.

Our exercises on the HHs selection process (with varying j and n)

We estimate four exercises of selectivity, which are summarized in Tav 6.

In the first exercise, the total number of alternatives j is $n = 3$. The three alternatives correspond to the three types of households in our dataset, described since Section 3. Therefore, $j = 1$ includes all HHs that disclose the identity of their bank (HHs with a bank or "banked HHs", around 42,000 observations); $j = 2$ includes the HHs who use Poste Italiane for their financial services ("HHs with Poste", around 9,000 observations); $j = 3$ includes HHs who do not disclose their reference financial institution or declare not to have a reference institution, neither a bank nor the post office (HHs without a financial institution or "non-banked HHs", around 9,000 observations).[24] This first

[24]In terms of the multinomial logit model, the dependent variable assumes the same values of j. This holds in all four exercises.

exercise therefore allows us to control for the selection process of HHs who self-select in favour of a bank, compared to those HHs who do not have a bank (or prefer not to indicate this) and those who choose another relevant financial player.

In the second exercise, the total number of alternatives is $n = 4$. The first group of the previous exercise ("banked HHs") is split into two groups: customers of banks belonging to the Top 5 Italian banking groups ($j = 1$; around 25,000 observations), and customers of the rest of the Italian banking system ($j = 2$; around 17,000 observations). The other two groups are the same as before: HHs customers of Poste ($j = 3$), and HHs without a financial institution ($j = 4$). The aim of the second exercise is to take into account (also) the effect of the selection process of HHs in different kinds of banks. Here, the self-selection of HHs between kinds of banks is hypothesized very essential. HHs are expected to decide (only) whether to be customers of very large banks, or of banks other than very large ones. However, though apparently basic, this decision can be already very pervasive. For example, HHs could choose very large banks because argue that these banks have a better reputation (Ballester & Munuera-Alemán, 2001; Lewis & Soureli, 2006; Haan & Moraga-González, 2011; Gurun et al., 2016), or because associate a larger size to a better standing and solidity (e.g., Boyd & Runkle, 1993; Cyree, 2016), or even because are enough sophisticated to trust in a “too big to fail” market equilibrium (e.g., Giese & Haldane, 2020), or because larger banks offer a broader bundle of products, services and opportunities (e.g., Affinito et al., 2025; Affinito et al., 2024b). Conversely, HHs could choose smaller banks, for example, because they believe these banks are more customer-oriented, or specifically more HH-oriented, and therefore closer to their needs, or to the needs of their territory, or in any case somehow easier to reach and trustworthier (e.g., Berger & Udell, 2002). Therefore, the self-selection of HHs between larger and smaller banks, while essential, is already able to capture many possible sorting effects.[25]

In the third exercise, the total number of alternatives is $n = 6$, as we split the category of banked HHs into four groups: banked HHs customers of Top 5 banking groups ($j = 1$, as before), customers of medium-sized banks ($j = 2$; around 11,000 observation), very small banks ($j = 3$; around 3,400), and cooperative banks ($j = 4$; around 2,400).[26] Moreover, we continue to control for the other two groups: HHs customers of Poste ($j = 5$), and non-

[25] At the same time, we also continue to control for the effect of selection with respect to HHs who do not have a bank and those who prefer Poste. As clarified, in fact, the DMF and DHL models allow us to take into account the *participation-selection* even if we control for the *sector-selection*.

[26] Medium and small banks follow the traditional definition used in Bank of Italy for the analysis of the credit system: specifically, medium-sized banks are those with average total assets ranging from €9 billion to €26 billion; small banks are those with average total assets less than €9 billion (Bank of Italy, 2020). Cooperative banks are treated as a separate group and then are excluded from the size-based groups.

banked HHs ($j = 6$). Here, the self-selection of HHs between different types of banks is assumed more complex and potentially able to capture more sorting effects. For example, HHs could choose medium-sized banks because these are seen as different in terms of product and service offering, customer relationships and geographical network, possibly precisely because they are neither too big nor too small. Or HHs could sort very small banks because they prefer local banks, or because HHs are more likely to know directly the employees or even the managers, especially in small towns, and have more confidence based on personal relationships (e.g., Guiso et al., 2004; Guiso et al., 2008; Georgarakos & Inderst, 2011; Gennaioli et al., 2015). Likewise, HHs could match with cooperative banks as these banks are often regarded as peculiar institutions according to many profiles, including from a regulatory perspective, and some HHs could decide to turn to this kind of banks for some of these peculiarities. For example, cooperative banks are characterized by "per capita" voting rights and mutual assistance, and some HHs could decide to be their customers precisely in view of these features.

Finally, in the fourth exercise we again consider six alternative states, but now we hypothesize that banked HHs rely on a different, even more complex, perspective. In this exercise, banked HHs are grouped into four states depending on both bank size and capital: two groups refer, respectively, to *more capitalized* and *less capitalized* Top 5 banking groups (respectively, $j = 1$ and $j = 2$); and the other two groups to *more* and *less capitalized* banks of the rest of the system (respectively, $j = 3$ and $j = 4$). The last two groups are again HHs customers of Poste ($j = 5$), and non-banked HHs ($j = 6$). We identify *more* and *less capitalized* banks according to whether their Tier 1 capital ratio is above or below the median of the system. Obviously, this is not a criterion of bank soundness. The capital ratio cannot be the only measure of a bank's health, nor can the median be a cut-off point. The soundness of banks is assessed (by the public authority supervisors, but also by the market) on the basis of a variety of indicators and factors. In addition, values above or below the median (possibly only just above or below) cannot be viewed as a warning signal. However, the aim of this exercise is not to make an assessment of the banks, but to attempt to allow for a more complex selection process of the HHs in the banks, assuming that they even take into account some aspects of banks' health. In other words, this fourth exercise assumes that HHs self-select for different types of banks and that they are sophisticated enough to gather and compare some information about banks' capital. If the HHs were indeed experienced enough to choose their bank (also) based on their capital, they could do this, given that the bank's capital is public information. Furthermore, using the median value allows us both to hypothesise that HHs make a comparison between banks and to keep all banks in the dataset.[27] All in all,

[27]It would not be possible, say, if we kept only the banks with the highest and lowest levels of capital

the distinction between more and less capitalized banks may be relevant to the purposes of our self-selection correction as HHs could purposely sort into more capitalized banks (for example, if they are risk-averse and prefer a very sound bank), but they could also purposely sort into less capitalized banks (for example, if they are more prone to risk, and might think that less sound banks can better suit their risky preferences).

Splits of our sample

In order to have a benchmark, we compare the results of the DMF and DHL models of Equations (2)-(3) with our FE-OLS results of Equation (1). This comparison is directly feasible in the first exercise, since in this case the results of both Equation (1) and the system of Equations (2)-(3) refer indistinctly to all HHs with a bank.[28] Instead, in the other three exercises, in order to have a consistent comparison we re-estimate Equation (1) for sub-samples of our dataset. More specifically, the last three exercises make as many subdivisions of our sample as the number of groups n (i.e., banked HHs are split into two groups in the second exercise; four groups in the third exercise; and other four groups in the fourth exercise). Therefore, we have the results of as many outcome equations as the number of these groups. Symmetrically, to enable a proper comparison, we re-estimate Equation (1) using FE-OLS, but allow the coefficients to vary across the same groups. It is important to emphasize that these FE-OLS estimates allocate the results to customers of the various groups of banks, similar to the sample correction models, but without accounting for the selection process of the HHs. Therefore, while these results are splits, they are not corrected. They serve as a useful benchmark for comparing the corrected estimation results.[29]

6 Results with sample selection correction

HHs with a bank versus the other HHs

Table 7 and Tables 8-10 report the results of the first exercise of sample selection correction described above, and for comparison the corresponding results of the FE-OLS regression

(for example, below 25 and above 75 per cent, respectively).

[28]In the DMF and DHL models, we also have the results of the outcome equation for HHs using Poste and for HHs without a financial institution. However, for our purposes, these groups of HHs serve (only) to correct the results of the second step; in particular, they cannot contain bank-side covariates (in which instead we are interested) and are therefore not reported, but available upon request.

[29]In operational terms, to perform the FE-OLS regressions for subgroups of households, we do not merely split the dataset; instead, we re-estimate Equation (1) by interacting each regressor with the group dummy. This method ensures greater efficiency and allows for a direct comparison of the coefficients.

of Equation (1).[30] Table 7 reports the results of the first step (*selection equation*), Tables 8-10 those of the second step (*outcome equation*).

The results of both steps confirm that it is appropriate to apply a method that takes into account the sorting of HHs into different groups. Table 7 shows that the coefficients of the first step are statistically significant, which confirms that the selection into the three states (HHs with a bank, HHs with Poste Italiane, and HHs without a financial institution) is not random, and that the selection variables have an influence on the probability of choosing one's own alternative state. Specifically, the probability of being a "banked HH" increases with financial literacy, income and wealth, varies depending on the age, is stronger for families with children and self-employees. In addition, Tables 8-10 show that all the correction terms of the second step are statistically significant, which corroborates once more the finding that there is a correlation between the two steps, and therefore that the selection is not random. Notably, the positive sign of the covariance of the correction terms indicates that there is a positive correlation between the decision to invest in financial assets and the decision to use a bank. In other words, households that are "banked" are also more likely to invest in financial assets. On the one hand, this result was predictable based on aggregated or anecdotal evidence, as banks serve as a crucial access point to financial markets. On the other hand, it constitutes a micro-based econometric confirmation that banks (even taking into account the individual characteristics of households and the selection process) play a key role, due to their position, in reducing households' non-participation in financial markets (Haliassos & Bertaut, 1995; Guiso & Jappelli, 2005; Campbell, 2006; Bolton et al., 2007).

Importantly, despite the statistical significance of the selection process of HHs across the three j states, the corrected results of the system of Equations (2)-(3) coincide with those of Equation (1). Therefore, even once controlled for the sorting effects, we find a statistically significant relationship between HHs investment decisions and bank characteristics. The comparison also confirms the specific bank characteristics that are most significantly associated with the portfolio choices of HHs. First, Holdings of Gov't Bonds and Funding via Bonds remain significantly and positively associated to the HH financial decision to invest in medium-risk assets, while Bank ROA remains significantly and negatively associated with this decision (Table 8). Second, NPLs remain significantly and positively associated to the HH financial decision to invest in riskier assets (Table 9). Third, Holdings of Gov't Bonds and Funding via Bonds (positively) and Liquidity from Central Bank (negatively) remain significantly associated to the HH decision to invest in bank bonds (Table 10). Fourth, the availability of a digital remote connection remains positively linked to the decision of holding all classes of financial assets (Tables 8-10).

[30]The FE-OLS results of Equation (1) correspond to those of specification 1 in Table 5.

Also the economic impact of these results is not negligible. Being a HH with a bank increases the probability of holding medium-risk assets from less than 6 per cent to around 20 per cent, the probability of holding higher-risk assets from less than 4 per cent to about 14 per cent, and the probability of holding bank bonds from 0.1 per cent to about 8 per cent. The effect is also relevant for the specific bank characteristics. Using the DHL specifications of Tables 8-10 as benchmarks and looking only at the statistically significant regressors, the probability of holding medium-risk assets increases by more than one-fifth when moving from clients of banks at the 25th to clients of banks at the 75th percentile of the variable Holdings of Gov't Bonds, and by more than one-third when moving from the 25th to the 75th percentile of the variable Funding via Bonds, while the decreasing effect of the variable ROA is less strong (it decreases by around one-tenth). The probability of holding riskier assets increases by around 16 per cent moving from the 25th to the 75th percentile of the variable NPLs, while the probability of holding bank bonds grows very strongly when moving from the 25th to the 75th percentile of all statistically significant bank variables (the probability more than doubles for Holdings of Gov't Bonds and Liquidity from Central Bank; it is even stronger for Funding via Bonds).

HHs customers of different categories of bank

To start the analysis of HHs customers of different bank categories, we show in Table 11 the results of nine regressions where the dependent variables are defined as in Equations (1) and (3) (that is, they are alternatively equal to 1 when households hold the different classes of financial assets) and the list of covariates includes the dummies capturing the categories of bank defined above. These regressions cannot contain bank fixed effects; however, they still contain time and geographical fixed effects, and all regressors used so far, defined at household, bank and pairwise level, and moreover they are clustered by time and geographical territory. These regressions are useful to introduce the results of the bank characteristics across specific categories. In fact, they indicate that: (i) HHs customers of the Top 5 banks are on average less likely to participate in financial markets than customers of banks of the rest of the system (Table 11, columns 1-3); (ii) customers of medium-sized banks are systematically more likely, while those of cooperative banks are systematically less likely to invest in financial assets than Top 5 bank customers (columns 4-6); (iii) customers of *less capitalized* banks are more likely to invest in riskier financial assets than customers of *more capitalized* banks, while customers of *less capitalized* Top 5 banks are less likely to invest in bank bonds (columns 7-9).

Tables 12-18 report the results of the last three exercises of sample selection correction described above, for HHs customers of each bank category, and for comparison the corresponding FE-OLS results for each split.[31] These comparisons allow us to point out four main results.

First, also in the sub-samples of HHs who are customers of different types of bank, the FE-OLS results largely tally with those of the sample bias correction models. Indeed, the number of statistically significant coefficients among bank regressors grows in the correction methods. Therefore, once taken into account the selection process of HHs and the match with the bank category, the links between bank features and HHs financial decisions appear even clearer. These results thus amplify the findings that bank characteristics are associated with financial decisions of customer HHs.

Second, the results of the specific bank categories (both FE-OLS splits and bias-corrected estimates) show that the relationships between bank characteristics and HH portfolio decisions vary across bank types. In a few cases, the relationship between financial decisions and bank characteristic is even opposite in different categories of bank. These differences across bank categories are a typical result of the heterogeneity analysis, and reveal that each bank type can impact customers' decisions on different profiles. To our analysis, it is therefore a further confirmation that bank traits matter in their clients' financial decisions.

Third, looking in detail to the results of bank categories, and focusing only on regressors that are systematically and consistently significant in both FE-OLS and selectivity estimations, the main results can be summarized as follows.[32] (i) Top 5 banks are related to financial medium-risk investments of their customers through larger Liquidity from Central Bank, and a lower level of Capital (Table 12, columns 1-3). (ii) In contrast,

[31] For reasons of space, the results of the first step are not reported in the case of these exercises. In general, the results of the selection equation and the correction terms in the outcome equation confirm that selection into the different bank types is not random, and therefore that HHs self-select across banks, and the probability of holding financial assets changes in different bank categories. Some relevant results of the first step are reported in this Section 6; all the details are available on request.

[32] As mentioned above, to test the robustness of the results, we performed the selectivity bias correction models by modifying the use of the regressors in the matrix K_{jt} of Equation 2. In particular: (i) we tested the separate or combined use of the two exogenous bank-side variables (number of bank branches in the province by category and average bank interest rates by province-category); (ii) we added the average value (by category) of the explanatory variables in B_{bt}; (iii) we also tried specifications that exclude bank-side variables in the first step, and treat two HH variables as the exogenous variables (financial education and job in the financial sector). The results have always remained the same. To show some of these different specifications (the others are available upon request), Tables 12-14 include two different DHL specifications instead of the same specifications with different models (DFM and DHL) as in the other tables. Specifically, in Tables 12-14, the selection equation of DHL-1 contains only HH-side covariates; DHL-2 contains both branches and bank rates. It is to notice that the inclusion of bank-side covariates in the selection equation becomes important precisely when there is a choice between different bank types, such as the case in our last three exercises.

banks of the rest of the system are linked to HHs' medium-risk investments through larger Holdings of Gov't Bonds and a lower ROA (Table 12, columns 4-6); are linked to riskier investments through higher NPLs (Table 13, columns 4-6), and to investments in bonds through more Holdings of Gov't Bonds and Funding via Bonds, and lower Liquidity from Central Bank (Table 14, columns 4-6). (iii) Among the banks of the rest of the system (focusing again only on regressors that are consistently significant in both FE-OLS splits and corrected estimations), the medium-sized banks present the highest number of statistically significant characteristics associated with HH decisions. In particular, the results of the medium-sized banks confirm and explain the positive sign of Funding via Bonds and the negative sign of ROA in medium-risk investments (Table 15, columns 3-4), the positive sign of NPLs in high-risk investments (Table 16, columns 3-4), and the positive signs of Gov't Bonds and Funding via Bonds in investments in bank bonds (Table 17, columns 3-4). (iv) However, also the other categories of banks present numerous significant characteristics. Very small banks are more likely to be associated to medium-risk investments of their customers when they have higher NPLs (Table 15, columns 5-6), and to riskier investments of their customers when they issue more bonds (Table 16, columns 5-6). Cooperative banks are linked to HHs purchasing bank bonds when their ROA and NPLs are lower (Table 17, columns 7-8). (v) The effect of digital connections remains generally significant and positive, except in small and cooperative banks (Tables 12-17).

Fourth, when we consider bank categories with different size and different levels of capital in the last exercise, the number of statistically significant regressors decreases, both in the FE-OLS splits and in the bias correction estimates (Table 18). This reduction could signal that, when the sorting accounts for several and precise bank features, the categories of bank absorb by themselves a large part of bank heterogeneity, and therefore partially embody the effect of bank characteristics. Or, put more simply, the results could indicate that our sample data are not suitable for considering too sharp and restrictive bank categories as this leads to a loss of representativeness. Most importantly, several coefficients still remain in any case statistically significant. The highest number of significant regressors lies in the estimates of *less capitalized* banks of the rest of the system. Looking again only at the systematic results, customers of *less capitalized* banks of the rest of the system are more likely to buy medium-risk assets when their banks issue more bank bonds (Table 18, column 4); are more likely to purchase riskier assets when their banks have higher NPLs (column 8); and finally are more likely to invest in bank bonds when their banks are less profitable and receive less Liquidity from Central Bank (column 12).

Again also the economic impact of these results is significant, both by bank category and bank characteristics. For example, compared to the other categories, being a HH client of a medium-sized bank increases on average the probability of holding medium-risk assets by about 20 per cent, the probability of holding riskier assets by more than 10 per cent, and the probability of holding bank bonds by more than 25 per cent.

7 Other robustness checks

We cite five further checks. First, as noticed in Section 2, despite the use of time, bank and geographical fixed effects, we also included in the matrices $B_{b,t}$ and M_{ibt} other bank- and pairwise- level regressors to avoid omitted variable biases and control for other confounding effects. In particular, we experimented with alternative measures of funding on the liability-side (Funding via Deposits), and alternative uses of resources on the asset-side (Bank Loans to households and non-financial firms). We also used variables measuring bank profits through relevant components of banks' income statements instead of ROA (net interest income, fees and commissions, and operating expenses). The variables measuring the specific components of income were hardly significant. Only the coefficient of fees was statistically significant and positive in the regressions for medium-risk assets and bank bonds, signalling that clients of banks that rely more on fees and commissions as a source of income tend to invest more in medium-risk assets. At the same time, we experimented with the variable capturing whether the household has more than one bank relationship. The coefficient of this variable tends to be positive suggesting that households that rely on more than one bank increases the holding of medium-risk and riskier assets. In any case, including or not these variables, all other results remained stable. In addition, while our baseline estimations include bank fixed effects – in order to take into account (as argued in Section 2) time-invariant components of bank characteristics, mitigate concerns about omitted variable bias and as an initial means of addressing self-selection bias – we also estimated our models without these fixed effects exactly to verify whether our results depend solely on the time-varying component of bank characteristics. The results remained confirmed.

Second, to estimate Equation 1 and predict the binary variables associated with holdings of each specific class of assets, we also used quasibinomial generalized linear models with the logit link function.[33] The results were unchanged. Nonetheless, in our baseline

[33]Moreover, the use of the logit link function is useful to account for data over-dispersion due to the high concentration of financial assets among Italian households (McCullagh & Nelder, 2019; Demétrio et al., 2014). It is to highlight that, even when we ran the logit version of Equation 1 with fixed effects, ours was not a fixed effects logit model (nor therefore did it suffer from the related limitations). In fact, in fixed effects logit models, only individuals who change their status at least once are relevant for the estimation of the coefficients. However, this applies to fixed effects that refer to individuals who form the reference unit of analysis (in our case households i). Instead, our fixed effects refer to t, g and b, but not

estimations, we presented the results obtained by estimating Equation 1 through a linear fixed effects estimator, since this served as a necessary benchmark for the estimation of the sample selection bias correction models.

Third, another test concerned the stability over time of the effects. Our estimation model contain time-fixed effects to account for general market movements that may affect household investment decisions in each period. However, we also inspected how the results regarding bank characteristics have changed over time. Banking variables significantly associated to household portfolios appear to be relevant in most of the periods.

Fourth, we estimated all our models with one-year lagged bank characteristics, in order to further investigate whether there are demand-side effects on household investment choices related to banks' characteristics, that is, whether households decide to invest in certain asset classes after checking their bank's financial conditions, as one-year lagged bank conditions are more easily consulted by investors. All results were confirmed.

Fifth, we also ran analogous specifications to those of Equation (1) and Equation (2)-(3) for the amounts, and the results were very similar. However, as mentioned in Section 3, we chose to measure households' portfolio decisions based on whether or not they hold certain classes of financial assets rather than on the reported value of the amounts held, because our data on financial assets are typically affected by under-reporting of the value of assets and are less stable, while data on holdings of financial assets are more accurate (Neri & Rannalli, 2011; D'Aurizio et al., 2006; Baffigi et al., 2016).

8 Conclusions

We analyse the relationship between households' participation in financial markets and the characteristics of the banks they are customers of. Our analysis uses a unique dataset that matches household and bank information, takes into account all individual household characteristics typically considered in the literature on household investment decisions, and controls for the selection process of households across banks. Our main findings can be summarized in four key conclusions.

First, we show that banks support household participation in financial markets. This could represent an important contribution to the economic welfare of their customers and to overall growth. Non-participation in risky asset markets and insufficient diversification of risky portfolios are widely recognized in the literature as major investment failures among many households. Our results show that participation can be increased when households choose banks, as these facilitate access to financial markets and offer guidance. Without such support, households may lack the time, expertise or confidence,

to i.

and ultimately the gateway, to invest in the financial markets.

Second, the process by which households select and match with banks is not random, in the sense that households choose the characteristics of their reference financial institution. Nevertheless, even once this selection process is accounted for, bank characteristics remain statistically associated to household decisions to participate in financial markets and invest in different asset classes.

Third, the links between households' financial investment decisions and their bank characteristics vary depending on the class of financial assets, and are heterogeneous across the categories of banks. Households are more likely to invest in medium-risk assets when they are clients of banks more specialized in proprietary financial investments rather than traditional lending, that rely more on bond issuance for financing, and have lower profit levels. Conversely, the decision to invest in riskier assets is more likely for clients of banks with higher levels of non-performing loans. Households are also more likely to invest in bank bonds when their bank finances itself more through bonds and less through monetary transactions with the central bank. Additionally, households with remote access to their bank are more likely to hold financial assets.

Fourth, even once taken into account the sorting effects, our findings cannot unknot the causal channels through which these links arise. On the household-side, our findings confirm that holdings of medium-risk and riskier financial assets are positively correlated with income, wealth, and financial literacy, which suggests that the relationships we observe between households' financial decisions and bank characteristics apply even to wealthier and more educated households, who are generally considered less susceptible to banks' steering. Uncovering potential causal links would require for example the use of exogenous shocks, a task we leave for future research.

References

Abreu, M., & Mendes, V. (2010). Financial literacy and portfolio diversification. *Quantitative finance*, *10*, 515–528.

Acharya, V., & Johnson, T. C. (2007). Insider trading in credit derivatives. *Journal of Financial Economics*, *84*, 110–141.

Affinito, M., D'Amato, M., & Santioni, R. (2023). *The evolution of bank fees as a source of income: trends and new business models - evidence from Italy*. Questioni di Economia e Finanza (Occasional Papers) 777 Bank of Italy.

Affinito, M., & Farabullini, F. (2009). Does the law of one price hold in euro-area retail banking? an empirical analysis of interest rate differentials across the monetary union. *Sixteenth issue (March 2009) of the International Journal of Central Banking*, .

Affinito, M., Galotto, L., & Privitera, F. (2024a). *The case for mindful customer protection: a review and some thoughts on neuroeconomics and neurofinance*. Questioni di Economia e Finanza (Occasional Papers) 888 Bank of Italy.

Affinito, M., & Meucci, G. (2023). Return of the NPLs to the Bright Side: Which Unlikely to Pay Firms are More Likely to Pay? *Quarterly Journal of Finance (QJF)*, *13*, 1–48.

Affinito, M., Sabbi, F., Santioni, R., & Santorelli, F. (2024b). *Types of consumer credit and bank distribution channels: insights from Italy*. Questioni di Economia e Finanza (Occasional Papers) 879 Bank of Italy.

Affinito, M., Santioni, R., & Tomassetti, L. (2025). Inside household debt: disentangling mortgages and consumer credit, and household and bank factors. Evidence from Italy. *Italian Economic Journal*, *on line*.

Agarwal, S., & Chua, Y. H. (2020). FinTech and household finance: a review of the empirical literature. *China Finance Review International*, *10*, 361–376.

Albareto, G. (2019). *Banks and retail financial investors in Italy: main evidence and open issues*. Technical Report.

Albareto, G., Cardillo, A., Hamaui, A., & Marinelli, G. (2020). *Mutual funds' performance: the role of distribution networks and bank affiliation*. Bank of Italy Economic working papers 1272.

Alessandrini, P., Presbitero, A., & Zazzaro, A. (2009). Banks, distances and firms' financing constraints. *Review of Finance*, *13*, 261–307.

Ampudia Fraile, M., van Vlokhoven, H., & Żochowski, D. (2016). Financial fragility of euro area households. *Journal of Financial Stability*, *27*, 250–262.

Angelini, P., Di Salvo, R., & Ferri, G. (1998). Availability and cost of credit for small businesses: Customer relationships and credit cooperatives. *Journal of Banking amp; Finance*, *22*, 925–954.

Baffigi, A., Cannari, L., & D'Alessio, G. (2016). Cinquant'anni di indagini sui bilanci delle famiglie italiane: Storia, metodi, prospettive (fifty years of household income and wealth surveys: History, methods and future prospects). *Bank of Italy Occasional Paper*, *368*.

Ballester, E., & Munuera-Alemán, J.-L. (2001). Brand trust in the context of consumer loyalty. *European Journal of Marketing*, *35*, 1238–1258.

Bank of Italy (2011). *Annual Report for 2010*. Rome, May.

Bank of Italy (2015). *Annual Report for 2014*. Rome, May.

Bank of Italy (2020). *Banks and Financial Institutions: Branch Network*. Methods and Sources: Methodological notes.

Bank of Italy (2022). *The Survey on Household Income and Wealth*. Methods and Sources: Methodological notes.

Berger, A., & Udell, G. (2002). Small business credit availability and relationship lending: The importance of bank organisational structure. *Economic Journal*, *112*, F32–F53.

Bluethgen, R., Gintschel, A., Hackethal, A., & Mueller, A. (2008). *Financial advice and individual investors' portfolios*. Working Paper European Business School.

Bolton, P., Freixas, X., & Shapiro, J. (2007). Conflicts of interest, information provision, and competition in the financial services industry. *Journal of Financial Economics*, *85*, 297–330.

Bonaccorsi, E., & Gobbi, G. (2007). Winners or Losers? The Effects of Banking Consolidation on Corporate Borrowers. *Journal of Finance*, *62*, 669–695.

Bourguignon, F., Fournier, M., & Gurgand, M. (2007). Selection bias corrections based on the multinomial logit model: Monte carlo comparisons. *Journal of Economic Surveys*, *21*, 174–205.

Boyd, J. H., & Runkle, D. E. (1993). Size and performance of banking firms : Testing the predictions of theory. *Journal of Monetary Economics*, *31*, 47–67.

Brown, M., Guin, B., & Morkoetter, S. (2013). Switching costs, deposit insurance and deposit withdrawals from distressed banks. *Swiss Institute of Banking and Finance, Working Paper on Finance*, *2013/19*.

Brown, M., & Hoffmann, M. (2016). Relationship Banking in the Residential Mortgage Market? Evidence from Switzerland. *Swiss Journal of Economics and Statistics (SJES)*, *152*, 23–48.

Brunetti, M., Ciciretti, R., & Djordjevic, L. (2016). The determinants of household's bank switching. *Journal of Financial Stability*, *26*, 175–189.

Cabral, L., & Santos, J. (2001). Cross selling and banking efficiency. *SSRN Electronic Journal*, .

Calvet, L., Campbell, J., & Sodini, P. (2007). Down or out: Assessing the welfare costs of household investment mistakes. *Journal of Political Economy*, *115*, 707–747.

Campbell, J. Y. (2006). Household finance. *The Journal of Finance*, *61*, 1553–1604.

Cardak, B., & Wilkins, R. (2009). The determinants of household risky asset holdings: Australian evidence on background risk and other factors. *Journal of Banking Finance*, *33*, 850–860.

Coletta, M., & Santioni, R. (2016). Le obbligazioni bancarie nel portafoglio delle famiglie italiane (bank bonds in italian householdss portfolios). *SSRN Electronic Journal*, .

Crockett, A. D., Harris, T. S., Mishkin, F. S., & White, E. N. (2003). Conflicts of interest in the financial services industry: What should we do about them?

Cyree, K. B. (2016). The Effects Of Regulatory Compliance For Small Banks Around Crisis-Based Regulation. *Journal of Financial Research*, *39*, 215–246.

Dahl, G. B. (2002). Mobility and the Return to Education: Testing a Roy Model with Multiple Markets. *Econometrica*, *70*, 2367–2420.

Damar, H. E., Gropp, R., & Mordel, A. (2020). Banks' Funding Stress, Lending Supply, and Consumption Expenditure. *Journal of Money, Credit and Banking*, *52*, 685–720.

D'Aurizio, L., Faiella, I., Iezzi, S., & Neri, A. (2006). The under-reporting of financial wealth in the survey on household income and wealth. *Bank of Italy Economic Research Paper*, *610*.

Demétrio, C. G., Hinde, J., & Moral, R. A. (2014). Models for overdispersed data in entomology. In *Ecological modelling applied to entomology* (pp. 219–259). Springer.

Drucker, S., & Puri, M. (2006). Banks in capital markets: A survey. In B. Eckbo (Ed.), *Handbook in Corporate Finance: Empirical Corporate Finance* Handbooks in Finance. San Diego: Elsevier.

Dubin, J. A., & McFadden, D. (1984). An econometric analysis of residential electric appliance holdings and consumption. *Econometrica*, *52*, 345–62.

Egan, M. (2019). Brokers versus retail investors: Conflicting interests and dominated products. *Journal of Finance*, *74*, 1217–1260.

Fecht, F., Hackethal, A., & Karabulut, Y. (2018). Is proprietary trading detrimental to retail investors? *Journal of Finance*, *73*, 1323–1361.

Focarelli, D., & Pozzolo, A. F. (2005). *Conflicts of Interest in Financial Markets - Evidence from Bond Underwriting in the Nineties*. Economics & Statistics Discussion Papers esdp05023 University of Molise, Department of Economics.

Foà, G., Gambacorta, L., Guiso, L., & Mistrulli, P. E. (2019). The supply side of household finance. *The Review of Financial Studies*, *32*, 3762–3798.

Frost, J., Gambacorta, L., & Gambacorta, R. (2022). On the nexus between wealth inequality, financial development and financial technology. *Journal of Economic Behavior & Organization*, *202*, 429–451.

Gaudecker, H.-M. V. (2015). How does household portfolio diversification vary with financial literacy and financial advice? *The Journal of Finance*, *70*, 489–507.

Gennaioli, N., Shleifer, A., & Vishny, R. (2015). Money doctors. *The Journal of Finance*, *70*, 91–114.

Georgarakos, D., & Inderst, R. (2011). *Financial advice and stock market participation*. Working Paper Series 1296 European Central Bank.

Giese, J., & Haldane, A. (2020). Covid-19 and the financial system: a tale of two crises. *Oxford Review of Economic Policy*, *36*, S200–S214.

Gobbi, G., & Sette, E. (2014). Do Firms Benefit from Concentrating their Borrowing? Evidence from the Great Recession. *Review of Finance*, *18*, 527–560.

Guiso, L., Haliassos, M., Jappelli, T. et al. (2002). *Household portfolios*. MIT press.

Guiso, L., & Jappelli, T. (2005). Awareness and stock market participation. *Review of Finance*, *9*, 537–567.

Guiso, L., Pozzi, A., Tsoy, A., Gambacorta, L., & Mistrulli, P. E. (2022). The cost of steering in financial markets: Evidence from the mortgage market. *Journal of Financial Economics*, *143*, 1209–1226.

Guiso, L., Sapienza, P., & Zingales, L. (2004). The Role of Social Capital in Financial Development. *American Economic Review*, *94*, 526–556.

Guiso, L., Sapienza, P., & Zingales, L. (2008). Trusting the stock market. *the Journal of Finance*, *63*, 2557–2600.

Guiso, L., & Sodini, P. (2013). Chapter 21 - household finance: An emerging field. (pp. 1397–1532). Elsevier volume 2 of *Handbook of the Economics of Finance*.

Guiso, L., & Viviano, E. (2015). How much can financial literacy help? *Review of Finance*, *19*, 1347–1382.

Gurun, U. G., Matvos, G., & Seru, A. (2016). Advertising Expensive Mortgages. *Journal of Finance*, *71*, 2371–2416.

Haan, M. A., & Moraga-González, J. L. (2011). Advertising for attention in a consumer search model. *The Economic Journal*, *121*, 552–579.

Hackethal, A., Haliassos, M., & Jappelli, T. (2012). Financial advisors: A case of babysitters? *Journal of Banking & Finance*, *36*, 509–524.

Hagenaars, A. J., De Vos, K., & Asghar Zaidi, M. (1994). *Poverty statistics in the late 1980s: Research based on micro-data*. Luxembourg (Luxembourg) Office for Official Publications of the European Communities.

Haliassos, M., & Bertaut, C. C. (1995). Why Do So Few Hold Stocks? *Economic Journal*, *105*, 1110–1129.

Hastings, J. S., Madrian, B. C., & Skimmyhorn, W. L. (2012). *Financial Literacy, Financial Education and Economic Outcomes*. NBER Working Papers 18412 National Bureau of Economic Research, Inc.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*, 153–161.

Inderst, R., & Ottaviani, M. (2009). Misselling through agents. *The American Economic Review*, *99*, 883–908.

Ivashina, V., & Sun, Z. (2011). Institutional demand pressure and the cost of corporate loans. *Journal of Financial Economics*, *99*, 500–522.

Jung, S. (2017). The gender wage gap and sample selection via risk attitudes. *International Journal of Manpower*, *38*, 318–335.

Kai, L., & Prabhala, N. R. (2007). Chapter 2 - self-selection models in corporate finance. In B. E. Eckbo (Ed.), *Handbook of Empirical Corporate Finance* Handbooks in Finance (pp. 37–86). Elsevier.

Kiser, E. (2002). Predicting household switching behavior and switching costs at depository institutions. *Review of Industrial Organization*, *20*, 349–365.

Lee, L.-F. (1983). Generalized econometric models with selectivity. *Econometrica*, *51*, 507–512.

Lewis, B., & Soureli, M. (2006). The antecedents of consumer loyalty in retail banking. *Journal of Consumer Behaviour*, *5*, 15 – 31.

Lokshin, M., & Sajaia, Z. (2006). MOVESTAY: Stata module for maximum likelihood estimation of endogenous regression switching models. Statistical Software Components, Boston College Department of Economics.

Ludvigson, S. (1999). Consumption and credit: a model of time-varying liquidity constraints. *Review of economics and statistics*, *81*, 434–447.

Lusardi, A. (2012). *Numeracy, financial literacy, and financial decision-making*. NBER Working Papers 17821 National Bureau of Economic Research, Inc.

Lusardi, A., & Mitchell, O. (2007). Financial literacy and retirement preparedness: Evidence and implications for financial education. *Business Economics*, *42*, 35–44.

MacKinnon, J. G., Ørregaard Nielsen, M., & Webb, M. D. (2023). Cluster-robust inference: A guide to empirical practice. *Journal of Econometrics*, *232*, 272–299.

Massa, M., & Rehman, Z. (2008). Information flows within financial conglomerates: Evidence from the banks-mutual funds relation. *Journal of Financial Economics*, *89*, 288–306.

McCullagh, P., & Nelder, J. A. (2019). *Generalized linear models*. Routledge.

McFadden, D. (1974). *Conditional Logit Analysis of Qualitative Choice Behavior" in P. Zarembka Eds. Frontiers in Econometrics, New York Acedemic press*.

Mehran, H., & Stulz, R. M. (2007). The economics of conflicts of interest in financial institutions. *Journal of Financial Economics*, *85*, 267–296.

Michelangeli, V., & Viviano, E. (2021). *Can internet banking affect households' participation in financial markets and financial awareness?*. Temi di discussione (Economic working papers) 1329 Bank of Italy, Economic Research and International Relations Area.

Morris, S., & Shin, H. S. (2004). Coordination risk and the price of debt. *European Economic Review*, *48*, 133–153.

Nakajima, M., & Ríos-Rull, J.-V. (2014). *Credit, bankruptcy, and aggregate fluctuations*. Technical Report National Bureau of Economic Research.

Nakosteen, R. A., & Zimmer, M. (1980). Migration and income: The question of self-selection. *Southern Economic Journal*, *46*, 840–851.

Neri, A., & Rannalli, M. G. (2011). To misreport or not to report? the case of the italian survey on household income and wealth. *Statistics in Transition new series*, *12*, 281–300.

Petersen, M. A., & Rajan, R. (2002). Does distance still matter? the information revolution in small business lending. *Journal of Finance*, *57*, 2533–2570.

van Rooij, M., Lusardi, A., & Alessie, R. (2011). Financial literacy and stock market participation. *Journal of Financial Economics*, *101*, 449–472.

Uppal, R., & Bhamra, H. S. (2017). *Does Household Finance Matter? Small Financial Errors with Large Social Costs*. CEPR Discussion Papers 12414 C.E.P.R. Discussion Papers.

Tables

Table 1: Composition of our sample

The table reports the type of households (HHs) in our sample based on the category of financial institution they use, and the corresponding number of observations.

Type	Definition	Obs	Freq	#HHs	#banks
HHs with a bank	have and reveal the identity of their bank	42,400	68.75	23,940	249
HHs with Poste	use Poste Italiane for their financial services	9,327	15.12	6,131	-
HHs with no financial institution	do not have or disclose any reference financial institution	9,125	14.8	7,000	-
Excluded HHs	use only foreign banks	258	0.42	-	-
Excluded HHs	use other institutions or declare banks that is not possible to identify	561	0.91	-	-
Total		61,671	100	37,071	249

Table 2: HHs descriptive statistics - Frequencies

	Frequency	Percent
Age		
Age group less then 30	2552	4.14
Age group 31-40	6830	11.07
Age group 41-50	11287	18.30
Age group 51-65	18204	29.52
Age group 65+	22798	36.97
Equivalent income quintiles		
1st eq. income quintile	11036	17.89
2nd eq. income quintile	11736	19.03
3rd eq. income quintile	12305	19.95
4th eq. income quintile	12528	20.31
5th eq. income quintile	14066	22.81
Equivalent wealth quintiles		
1st eq. wealth quintile	10735	17.41
2nd eq. wealth quintile	10457	16.96
3rd eq. wealth quintile	12312	19.96
4th eq. wealth quintile	13586	22.03
5th eq. wealth quintile	14581	23.64
Risk aversion		
Low risk aversion	527	0.85
Low-medium risk aversion	7389	11.98
Medium-high risk aversion	17458	28.31
High risk aversion	36297	58.86
Household type		
Singles older than 65	9873	16.01
Singles younger than 65	6563	10.64
Couple no children	15801	25.62
Couple with children	22322	36.20
Single parent with children	4725	7.66
Other households	2387	3.87
Work status		
Blue-collar worker	11537	18.71
Office worker or teacher	10034	16.27
Junior manager	1808	2.93
Manager	1088	1.76
Member arts or professions	2046	3.32
Sole proprietor or freelance	5134	8.32
Not employed	30024	48.68
Geographical area		
North	27708	44.93
Center	12964	21.02
South and islands	20999	34.05
Job in financial sector		
No	60303	97.78
Yes	1368	2.22

Table 3: **HHs descriptive statistics - financial literacy**

	count	mean	sd
Financial Literacy	61671	0.545	0.306

Table 4: **Bank variables' descriptive statistics**

Variable	Description	count	mean	sd
Size	Natural logarithm of total assets	42,489	11.664	1.869
Holdings of Gov't Bonds	Portfolio of Government securities / Total assets	42,489	0.082	0.079
Liquidity from Central Bank	Total refinancing / Total assets	42,489	0.063	0.067
Funding via Bonds	Total Bonds issued / Total assets	42,489	0.143	0.083
Capital	Tier 1 / Risk weighted assets	42,356	0.104	0.039
NPLs	Total non-performing loans / Total loans	42,489	0.032	0.024
ROA	Return on Assets	42,489	0.002	0.008
Bank loans to HHs and NFCs	Loans to private sector / Total assets	42,489	0.458	0.121
Funding via Deposits	Total deposits / Total assets	42,489	0.511	0.142
Net interest income	Net interest income / Total assets	42,489	0.013	0.005
Fees and commissions	Fees and commissions / Total assets	42,489	0.008	0.002
Operating expenses	Operating expenses / Total assets	42,489	0.017	0.005
Remote connection to bank	Dummy 0-1	42,489	0.260	0.439
More bank relationships	Dummy 0-1	42,489	0.251	0.434

Table 5: Financial assets holdings: baseline results, fixed effect OLS estimations

The table reports OLS regression coefficients and associated robust standard errors in parentheses of Equation (1). The dependent variables are equal to 1 if household i customer of bank b at time t holds a specific class of assets (alternatively, medium-risk assets, riskier assets, and bank bonds). Standard errors are White-corrected for heteroskedasticity and clustered at the bank, time and geographical level. ***, **, and * indicate that the coefficient estimate is significantly different from zero at 1%, 5%, and 10%, respectively. Control regressors at HH level include: age class, equivalent income and wealth quintile, household type (e.g., single, couple, with or without children, etc.), work status (e.g., blue-collar, office worker, manager, sole proprietor, etc.), working in the financial sector, risk aversion, financial literacy.

	(1) Medium-risk	(2) Medium-risk	(3) Riskier	(4) Riskier	(5) Bank bonds	(6) Bank bonds
Size	-0.01	-0.01	0.01	0.01*	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Holdings of Gov't Bonds	0.26***	0.27***	-0.01	-0.01	0.17***	0.19***
	(0.08)	(0.08)	(0.06)	(0.06)	(0.06)	(0.06)
Liquidity from Central Bank	-0.01	-0.02	-0.04	-0.03	-0.14**	-0.14**
	(0.07)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)
Funding via Bonds	0.16**	0.15*	0.09	0.08	0.28***	0.28***
	(0.08)	(0.08)	(0.07)	(0.06)	(0.07)	(0.07)
Capital	-0.07	-0.07	0.09	0.13	0.04	0.05
	(0.16)	(0.16)	(0.14)	(0.14)	(0.12)	(0.12)
NPLs	-0.34	-0.36	0.54***	0.45**	-0.26	-0.26
	(0.24)	(0.24)	(0.19)	(0.19)	(0.19)	(0.19)
ROA	-1.42***	-1.38***	-0.25	-0.25	-0.30	-0.42
	(0.47)	(0.47)	(0.43)	(0.42)	(0.33)	(0.33)
Remote connection to bank	0.04***	0.04***	0.09***	0.08***	0.03***	0.02***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Regressors at HH level	YES	YES	YES	YES	YES	YES
Time FEs	AREA	PROV	AREA	PROV	AREA	PROV
Geographical FEs	YES	YES	YES	YES	YES	YES
Bank FEs	YES	YES	YES	YES	YES	YES
Observations	42,285	42,285	42,285	42,285	27,420	27,420

Table 6: Our four exercises on households' selection process and selectivity bias correction methods

Observations concerning HHs that declare to:	Number of observations	Freq.	First exercise			Second exercise			Third exercise			Fourth exercise		
			Alternative states	**Mult. logit alternatives**	**Obs. per alternative**	**Alternative states**	**Mult. logit alternatives**	**Obs. per alternative**	**Alternative states**	**Mult. logit alternatives**	**Obs. per alternative**	**Alternative states**	**Mult. logit alternatives**	**Obs. per alternative**
Have and reveal the identity of their bank	42,400	68.75	Have and reveal the identity of their bank	1	42,400	Top 5 banking groups	1	25,023	Top 5	1	25,023	More cap. top 5	1	14,804
						Rest of the system	2	17,377	Medium	2	11,473	Less cap. top 5	2	10,219
									Small	3	3,441	More cap. rest of the system	3	8,948
									Cooperatives	4	2,463	Less cap. rest of the system	4	8,429
Use Poste Italiane for their financial services	9,327	15.12	Poste Italiane	2	9,327	Poste Italiane	3	9,327	Poste Italiane	5	9,327	Poste Italiane	5	9,327
Not to have any reference financial institution	9,125	14.8	No financial institution	3	9,125	No financial institution	4	9,125	No financial institution	6	9,125	No financial institution	6	9,125
Use foreign banks	258	0.42	-	-	-	-	-	-	-	-	-	-	-	-
Use financial institutions other than banks or declare banks that it is not possible to identify	561	0.91	-	-	-	-	-	-	-	-	-	-	-	-
Total	**61,671**	**100**			**60,852**			**60,852**			**60,852**			**60,852**

Table 7: Selection Equation between "banked HHs", "HHs with Poste" and "non-banked HHs"

The table reports multinomial logit regression coefficients and associated robust standard errors in parentheses of Equation (2), i.e., the first step of the first exercise described in Table 6. The reference case are HHs with a bank (banked HHs). ***, **, and * indicate that the coefficient estimate is significantly different from zero at 1%, 5%, and 10%, respectively.

	(1)			
	HHs with Poste		Unbanked HHs	
Financial literacy	-1.02***	(0.05)	-1.80***	(0.06)
Job in financial sector	-0.30*	(0.17)	0.31***	(0.12)
Low-medium risk aversion	-0.04	(0.17)	-0.20	(0.15)
High-medium risk aversion	0.06	(0.17)	-0.24	(0.15)
High risk aversion	0.28*	(0.17)	-0.28*	(0.15)
2nd eq. wealth quintile	-0.22***	(0.04)	-0.74***	(0.04)
3rd eq. wealth quintile	-0.28***	(0.04)	-0.59***	(0.04)
4th eq. wealth quintile	-0.48***	(0.04)	-0.57***	(0.05)
5th eq. wealth quintile	-0.83***	(0.05)	-0.61***	(0.05)
2nd eq. income quintile	-0.26***	(0.04)	-0.76***	(0.04)
3rd eq. income quintile	-0.52***	(0.04)	-1.17***	(0.05)
4th eq. income quintile	-0.84***	(0.05)	-1.27***	(0.05)
5th eq. income quintile	-1.51***	(0.07)	-1.43***	(0.06)
Age group 31-40	-0.14*	(0.08)	-0.18**	(0.07)
Age group 41-50	-0.19***	(0.07)	-0.20***	(0.07)
Age group 51-65	-0.10	(0.07)	-0.35***	(0.07)
Age group 65+	0.20**	(0.08)	-0.64***	(0.08)
Singles younger than 65	-0.05	(0.06)	-0.00	(0.06)
Couple no children	-0.66***	(0.04)	-0.46***	(0.05)
Couple with children	-1.07***	(0.05)	-0.63***	(0.06)
Single parent with children	-0.63***	(0.06)	-0.34***	(0.06)
Other households	-0.45***	(0.07)	-0.30***	(0.08)
Office worker or teacher	-0.24***	(0.05)	-0.35***	(0.05)
Junior manager	-0.21*	(0.12)	-0.25**	(0.11)
Manager	-1.09***	(0.27)	0.11	(0.14)
Member arts or professions	-1.07***	(0.16)	-0.13	(0.11)
Sole proprietor or freelance	-0.74***	(0.07)	0.04	(0.06)
Not employed	0.06	(0.05)	0.49***	(0.05)
Center	0.78***	(0.04)	0.77***	(0.04)
South and islands	1.18***	(0.03)	1.05***	(0.03)
Constant	-0.43**	(0.19)	2.18***	(0.17)
Regressors at HH level	YES			
Time FEs	YES			
Geographic FEs	YES			
Observations	60,852			

Table 8: Outcome equation of medium-risk assets holdings

First exercise described in Table 6. The table reports the results of Equation (3) (i.e., the results of the second step or outcome equation) for two different selectivity bias correction methods (DMF stands for Dubin & McFadden (1984), and DHL for Dahl (2002)), and the FE OLS results of Equation (1) for comparison. The table shows regression coefficients and associated robust standard errors in parentheses. The dependent variables are equal to 1 if household i customer of bank b at time t holds medium-risk financial assets. Standard errors are White-corrected for heteroskedasticity and clustered at the bank, time and geographical level. ***, **, and * indicate that the coefficient estimate is significantly different from zero at 1%, 5%, and 10%, respectively. For the list of control regressors at HH level, see Table 5.

	(1) OLS		(2) DMF		(3) DHL	
Size	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)
Holdings of Gov't Bonds	0.26***	(0.08)	0.20***	(0.05)	0.22***	(0.06)
Liquidity from Central Bank	-0.01	(0.07)	0.03	(0.05)	-0.01	(0.06)
Funding via Bonds	0.16**	(0.08)	0.11*	(0.06)	0.11*	(0.06)
Capital	-0.07	(0.16)	-0.10	(0.12)	-0.10	(0.13)
NPLs	-0.34	(0.24)	0.02	(0.18)	0.08	(0.19)
ROA	-1.42***	(0.47)	-0.59*	(0.35)	-0.66*	(0.38)
Remote connection to bank	0.04***	(0.01)	0.05***	(0.01)	0.05***	(0.01)
_m1			0.28***	(0.02)		
_m2			0.36***	(0.04)		
_m3			0.70***	(0.04)		
Constant	0.02	(0.12)	0.75	(0.47)	-1.23***	(0.33)
Anciliary						
Sigma2			0.27***	(0.02)		
rho1			0.54***	(0.02)		
rho2			0.69***	(0.07)		
rho3			1.34***	(0.05)		
Regressors at HH level	YES		YES		YES	
Time FEs	YES		YES		YES	
Geographic FEs	YES		YES		YES	
Bank FEs	YES		YES		YES	
Observations	42,285					

Table 9: Outcome equation of riskier assets holdings

First exercise described in Table 6. The table reports the results of Equation (3) (i.e., the results of the second step or outcome equation) for two different selectivity bias correction methods (DMF stands for Dubin & McFadden (1984), and DHL for Dahl (2002)), and the FE OLS results of Equation (1) for comparison. The table shows regression coefficients and associated robust standard errors in parentheses. The dependent variables are equal to 1 if household i customer of bank b at time t holds riskier financial assets. Standard errors are White-corrected for heteroskedasticity and clustered at the bank, time and geographical level. ***, **, and * indicate that the coefficient estimate is significantly different from zero at 1%, 5%, and 10%, respectively. For the list of control regressors at HH level, see Table 5.

	(1) OLS		(2) DMF		(3) DHL	
Size	0.01	(0.01)	0.01	(0.01)	-0.00	(0.01)
Holdings of Gov't Bonds	-0.01	(0.06)	0.07	(0.06)	0.08	(0.05)
Liquidity from Central Bank	-0.04	(0.06)	0.03	(0.07)	-0.00	(0.05)
Funding via Bonds	0.09	(0.07)	0.03	(0.07)	-0.00	(0.05)
Capital	0.09	(0.14)	0.23	(0.16)	0.10	(0.10)
NPLs	0.54***	(0.19)	0.24**	(0.11)	0.27*	(0.15)
ROA	-0.25	(0.43)	-0.28	(0.45)	0.09	(0.29)
Remote connection to bank	0.09***	(0.01)	0.08***	(0.01)	0.09***	(0.01)
_m1			0.47***	(0.07)		
_m2			0.00	(0.07)		
_m3			1.11***	(0.17)		
Constant	0.01	(0.10)	0.97**	(0.44)	-1.28***	(0.30)
Anciliary						
Sigma2			0.26***	(0.02)		
rho1			0.91***	(0.12)		
rho2			0.00	(0.14)		
rho3			2.18***	(0.28)		
Regressors at HH level	YES		YES		YES	
Time FEs	YES		YES		YES	
Geographic FEs	YES		YES		YES	
Bank FEs	YES		YES		YES	
Observations	42,285					

Table 10: Outcome equation of holdings of bank bonds

First exercise described in Table 6. The table reports the results of Equation (3) (i.e., the results of the second step or outcome equation) for two different selectivity bias correction methods (DMF stands for Dubin & McFadden (1984), and DHL for Dahl (2002)), and the FE OLS results of Equation (1) for comparison. The table shows regression coefficients and associated robust standard errors in parentheses. The dependent variables are equal to 1 if household i customer of bank b at time t holds bank bonds. Standard errors are White-corrected for heteroskedasticity and clustered at the bank, time and geographical level. ***, **, and * indicate that the coefficient estimate is significantly different from zero at 1%, 5%, and 10%, respectively. For the list of control regressors at HH level, see Table 5.

	(1) OLS		(2) DMF		(3) DHL	
Size	-0.01	(0.01)	0.01	(0.01)	0.01	(0.01)
Holdings of Gov't Bonds	0.17***	(0.06)	0.14***	(0.05)	0.16***	(0.05)
Liquidity from Central Bank	-0.14**	(0.06)	-0.08*	(0.05)	-0.10**	(0.05)
Funding via Bonds	0.28***	(0.07)	0.22***	(0.06)	0.24***	(0.06)
Capital	0.04	(0.12)	0.11	(0.10)	0.10	(0.10)
NPLs	-0.26	(0.19)	-0.06	(0.13)	-0.06	(0.15)
ROA	-0.30	(0.33)	-0.33	(0.28)	-0.37	(0.31)
Remote connection to bank	0.03***	(0.01)	0.03***	(0.00)	0.03***	(0.00)
_m1			0.19***	(0.02)		
_m2			0.10**	(0.05)		
_m3			0.45***	(0.03)		
Constant	0.05	(0.13)	-0.33	(0.46)	-0.96***	(0.37)
Anciliary						
Sigma2			0.11***	(0.01)		
rho1			0.55***	(0.05)		
rho2			0.31**	(0.14)		
rho3			1.34***	(0.06)		
Regressors at HH level	YES		YES		YES	
Time FEs	YES		YES		YES	
Geographic FEs	YES		YES		YES	
Bank FEs	YES		YES		YES	
Observations	27,420					

Table 11: Regressions including dummies for the different bank categories

Table reports the results of nine regressions where the dependent variables are alternatively medium-risk assets (m-r); riskier assets; or bank bonds (b-b), and the list of covariates includes the dummies capturing the categories of bank (Top 5, medium-sized, very small, cooperative; *more capitalized* and *less capitalized*). In columns 1-3, the reference case are the banks of the rest of the system (i.e., those different from banks belonging to the Top 5 banking groups); in columns 4-6, the reference case are the banks belonging to the Top 5 banking groups; in columns 7-9, the reference case are *less capitalized* Top 5 banks. See description in Table 8.

	Ref. case: Other banks			Ref. case: Top 5			Ref. case: Less cap. Top 5		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	m-r	riskier	b-b	m-r	riskier	b-b	m-r	riskier	b-b
Top 5	-0.03***	-0.03***	-0.02***						
	(0.01)	(0.01)	(0.01)						
Medium size				0.02***	0.02***	0.01**			
				(0.01)	(0.01)	(0.01)			
Very small				-0.00	-0.03***	-0.01			
				(0.01)	(0.01)	(0.01)			
Cooperative				-0.06***	-0.07***	-0.06***			
				(0.02)	(0.02)	(0.01)			
More cap. Top 5							0.00	-0.03***	0.02*
							(0.01)	(0.01)	(0.01)
Less cap. Other							0.01	0.02***	0.01*
							(0.01)	(0.01)	(0.01)
More cap. Other							0.04***	0.00	0.04***
							(0.01)	(0.01)	(0.01)
Regressors at bank level	YES	YES	YES	YES	YES	YES	YES	YES	YES
Regressors at HH level	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Geographic FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	42,349	42,349	27,491	42,349	42,349	27,491	42,349	42,349	27,491

Table 12: Outcome equation of holdings of medium-risk assets

Second exercise described in Table 6: households who are customers of Top 5 bank groups (Top 5) versus households who are customers of the banks of the rest of the system (ROS). In DHL1 the selection equation in the first step contains only HH-side covariates; in DHL2 the selection equation contains both branches and bank rates. See description in Table 8.

		Top 5			ROS	
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	DHL1	DHL2	OLS	DHL1	DHL2
Size	-0.02	0.01	0.01	-0.01	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Holdings of Gov't Bonds	0.23	-0.08	-0.11	0.29***	0.22***	0.23***
	(0.20)	(0.16)	(0.16)	(0.09)	(0.08)	(0.08)
Liquidity from Central Bank	0.53***	0.34**	0.36**	-0.10	-0.12	-0.10
	(0.17)	(0.15)	(0.16)	(0.08)	(0.07)	(0.07)
Funding via Bonds	-0.00	0.35***	0.33**	0.15	0.08	0.10
	(0.12)	(0.13)	(0.13)	(0.09)	(0.08)	(0.08)
Capital	-0.49*	-1.09***	-1.19***	-0.06	-0.00	-0.02
	(0.28)	(0.34)	(0.33)	(0.19)	(0.15)	(0.16)
NPLs	0.21	-0.63	-0.50	-0.19	0.78***	0.76***
	(0.33)	(0.39)	(0.41)	(0.28)	(0.28)	(0.25)
ROA	-0.16	0.71	0.59	-1.94***	-1.06*	-1.11**
	(0.71)	(0.67)	(0.61)	(0.63)	(0.56)	(0.50)
Remote connection to bank	0.02***	0.04***	0.06***	0.03***	0.05***	0.06***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	0.04	-0.57	0.73*	0.04	-1.15*	0.15
	(0.12)	(0.66)	(0.41)	(0.12)	(0.68)	(0.33)
Regressors at HH level	YES	YES	YES	YES	YES	YES
Time FEs	YES	YES	YES	YES	YES	YES
Geographical FEs	YES	YES	YES	YES	YES	YES
Bank FEs	YES	YES	YES	YES	YES	YES
Observations	42,285			42,285		

Table 13: Outcome equation of holdings of riskier assets

Second exercise described in Table 6: households who are customers of Top 5 bank groups (Top 5) versus households who are customers of the banks of the rest of the system (ROS). In DHL1 the selection equation in the first step contains only HH-side covariates; in DHL2 the selection equation contains both branches and bank rates. See description in Table 8.

		Top 5			ROS	
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	DHL1	DHL2	OLS	DHL1	DHL2
Size	0.00	-0.01	-0.01	0.02	-0.01	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Holdings of Gov't Bonds	0.06	-0.01	-0.08	0.02	0.12*	0.13**
	(0.14)	(0.12)	(0.12)	(0.07)	(0.07)	(0.06)
Liquidity from Central Bank	-0.18	0.01	0.04	-0.04	0.02	0.03
	(0.12)	(0.11)	(0.12)	(0.07)	(0.06)	(0.06)
Funding via Bonds	-0.01	0.04	0.06	0.10	-0.02	-0.01
	(0.10)	(0.11)	(0.11)	(0.08)	(0.07)	(0.07)
Capital	0.35	0.15	-0.01	0.08	-0.05	-0.01
	(0.22)	(0.25)	(0.27)	(0.17)	(0.14)	(0.14)
NPLs	0.46*	0.09	0.16	0.55**	0.49**	0.59***
	(0.27)	(0.31)	(0.30)	(0.22)	(0.21)	(0.22)
ROA	0.11	0.54	0.51	-0.57	0.12	0.08
	(0.54)	(0.47)	(0.50)	(0.63)	(0.44)	(0.45)
Remote connection to bank	0.08***	0.09***	0.10***	0.07***	0.07***	0.09***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	-0.02	-1.54***	0.40	-0.02	-2.25***	0.48*
	(0.10)	(0.59)	(0.29)	(0.10)	(0.66)	(0.26)
Regressors at HH level	YES	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES	YES
Geographic FE	YES	YES	YES	YES	YES	YES
Bank FE	YES	YES	YES	YES	YES	YES
Observations	42,285			42,285		

Table 14: Outcome equation of holdings of bank bonds

Second exercise described in Table 6: households who are customers of Top 5 bank groups (Top 5) versus households who are customers of the banks of the rest of the system (ROS). In DHL1 the selection equation in the first step contains only HH-side covariates; in DHL2 the selection equation contains both branches and bank rates. See description in Table 8.

		Top 5			ROS	
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	DHL1	DHL2	OLS	DHL1	DHL2
Size	-0.00	0.06**	0.06**	-0.02	-0.01	-0.01
	(0.03)	(0.02)	(0.03)	(0.01)	(0.01)	(0.01)
Holdings of Gov't Bonds	0.09	0.16	0.17	0.25***	0.16**	0.19***
	(0.15)	(0.19)	(0.17)	(0.07)	(0.07)	(0.07)
Liquidity from Central Bank	0.04	-0.16	-0.17	-0.17***	-0.14***	-0.15***
	(0.13)	(0.16)	(0.17)	(0.06)	(0.06)	(0.06)
Funding via Bonds	0.02	0.47***	0.46***	0.32***	0.20***	0.19***
	(0.11)	(0.15)	(0.15)	(0.08)	(0.07)	(0.07)
Capital	-0.33	0.58	0.54	-0.04	0.01	-0.03
	(0.28)	(0.41)	(0.41)	(0.14)	(0.13)	(0.13)
NPLs	-0.23	0.18	0.19	-0.13	0.23	0.26
	(0.29)	(0.55)	(0.57)	(0.20)	(0.19)	(0.20)
ROA	0.50	-0.17	-0.03	-0.49	-0.19	-0.14
	(0.53)	(0.53)	(0.55)	(0.44)	(0.42)	(0.39)
Remote connection to bank	0.02***	0.03***	0.04***	0.02**	0.02***	0.03***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	0.06	-1.21	-0.81*	0.06	-1.66**	-0.21
	(0.20)	(0.88)	(0.47)	(0.20)	(0.80)	(0.26)
Regressors at HH level	YES	YES	YES	YES	YES	YES
Time FEs	YES	YES	YES	YES	YES	YES
Geographical FEs	YES	YES	YES	YES	YES	YES
Bank FEs	YES	YES	YES	YES	YES	YES
Observations	27,420			27,420		

Table 15: Outcome equation of holdings of medium-risk assets

Third exercise of Table 6: Top 5, medium-sized, very small, and cooperative banks. See description in Table 8.

	Top 5		Medium size		Very small		Cooperative	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	DHL	OLS	DHL	OLS	DHL	OLS	DHL
Size	-0.01	0.01	-0.01	-0.01	0.00	-0.00	-0.07	-0.09
	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.05)	(0.06)
Holdings of Gov't Bonds	0.21	-0.11	0.36***	0.14	-0.08	0.03	0.53***	0.30
	(0.20)	(0.16)	(0.14)	(0.11)	(0.21)	(0.19)	(0.19)	(0.25)
Liquidity from Central Bank	0.53***	0.35**	-0.05	-0.15	-0.08	-0.25	0.06	0.21
	(0.17)	(0.15)	(0.11)	(0.10)	(0.18)	(0.22)	(0.18)	(0.17)
Funding via Bonds	0.01	0.36***	0.28**	0.21*	0.06	0.12	0.06	-0.18
	(0.12)	(0.13)	(0.14)	(0.12)	(0.15)	(0.21)	(0.14)	(0.23)
Capital	-0.54*	-1.06***	-0.17	0.05	0.31	0.17	-0.69*	-0.34
	(0.29)	(0.32)	(0.27)	(0.21)	(0.44)	(0.48)	(0.41)	(0.48)
NPLs	0.20	-0.65*	-0.40	0.40	0.77*	2.52***	-1.51***	-1.26*
	(0.33)	(0.37)	(0.36)	(0.37)	(0.43)	(0.56)	(0.55)	(0.71)
ROA	-0.15	0.61	-2.30**	-1.58**	-1.40	0.54	-2.59	-3.42
	(0.71)	(0.62)	(0.95)	(0.80)	(0.90)	(0.87)	(2.18)	(2.49)
Remote connection to bank	0.03***	0.05***	0.05***	0.06***	0.05**	0.03	-0.04**	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Constant	0.07	-0.71	0.07	-0.92	0.07	-1.06	0.07	-0.81
	(0.12)	(0.65)	(0.12)	(0.89)	(0.12)	(1.66)	(0.12)	(3.36)
Regressors at HH level	YES	YES	YES	YES	YES	YES	YES	YES
Time FEs	YES	YES	YES	YES	YES	YES	YES	YES
Geographical FEs	YES	YES	YES	YES	YES	YES	YES	YES
Bank FEs	YES	YES	YES	YES	YES	YES	YES	YES
Observations	42,285		42,285		42,285		42,285	

Table 16: Outcome equation of holdings of riskier assets

Third exercise of Table 6: Top 5, medium-sized, very small, and cooperative banks. See description in Table 8.

	Top 5		Medium size		Very small		Cooperative	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	DHL	OLS	DHL	OLS	DHL	OLS	DHL
Size	0.00	-0.00	0.02	-0.00	0.01	-0.01	0.00	-0.03
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.04)	(0.05)
Holdings of Gov't Bonds	0.07	-0.05	-0.12	0.00	0.12	0.28*	0.08	0.11
	(0.14)	(0.12)	(0.12)	(0.10)	(0.15)	(0.16)	(0.16)	(0.20)
Liquidity from Central Bank	-0.18	0.03	-0.07	-0.05	0.09	0.05	0.00	0.17
	(0.13)	(0.11)	(0.10)	(0.08)	(0.17)	(0.18)	(0.14)	(0.13)
Funding via Bonds	-0.03	0.05	0.17	-0.09	0.24*	0.40**	-0.02	-0.17
	(0.10)	(0.11)	(0.12)	(0.10)	(0.13)	(0.17)	(0.13)	(0.18)
Capital	0.38*	0.12	0.27	0.11	0.92**	0.32	-0.40	-0.42
	(0.22)	(0.27)	(0.24)	(0.19)	(0.37)	(0.40)	(0.41)	(0.40)
NPLs	0.49*	0.08	0.83***	0.80**	0.14	0.18	0.31	0.48
	(0.28)	(0.32)	(0.28)	(0.32)	(0.31)	(0.49)	(0.37)	(0.49)
ROA	0.11	0.47	-0.04	0.66	-0.97	0.24	-1.33	1.32
	(0.54)	(0.49)	(0.87)	(0.63)	(1.12)	(0.77)	(1.58)	(1.77)
Remote connection to bank	0.08***	0.09***	0.08***	0.07***	0.05***	0.07***	0.05**	0.09***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Constant	-0.01	-1.98***	-0.01	-2.06***	-0.01	-2.16	-0.01	-2.40
	(0.10)	(0.63)	(0.10)	(0.80)	(0.10)	(1.42)	(0.10)	(2.72)
Regressors at HH level	YES	YES	YES	YES	YES	YES	YES	YES
Time FEs	YES	YES	YES	YES	YES	YES	YES	YES
Geographic FEs	YES	YES	YES	YES	YES	YES	YES	YES
Bank FEs	YES	YES	YES	YES	YES	YES	YES	YES
Observations	42,285		42,285		42,285		42,285	

Table 17: Outcome equation of holdings of bank bonds

Third exercise of Table 6: Top 5, medium-sized, very small, and cooperative banks. See description in Table 8.

	Top 5		Medium size		Very small		Cooperative	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	DHL	OLS	DHL	OLS	DHL	OLS	DHL
Size	0.00	0.06^{**}	-0.03^{*}	-0.02	-0.01	-0.03	-0.05	-0.05
	(0.03)	(0.03)	(0.02)	(0.01)	(0.02)	(0.02)	(0.04)	(0.05)
Holdings of Gov't Bonds	0.06	0.13	0.36^{***}	0.22^{*}	-0.02	0.08	0.25^{*}	-0.00
	(0.16)	(0.17)	(0.13)	(0.12)	(0.15)	(0.16)	(0.15)	(0.19)
Liquidity from Central Bank	0.06	-0.18	0.01	0.02	0.00	-0.11	-0.21	-0.14
	(0.13)	(0.17)	(0.08)	(0.08)	(0.13)	(0.18)	(0.15)	(0.15)
Funding via Bonds	0.13	0.47^{***}	0.61^{***}	0.59^{***}	0.45^{***}	0.16	0.20^{**}	-0.08
	(0.11)	(0.15)	(0.13)	(0.13)	(0.12)	(0.17)	(0.10)	(0.17)
Capital	-0.32	0.58	-0.25	-0.15	0.62^{*}	0.44	-0.21	0.57
	(0.29)	(0.42)	(0.20)	(0.20)	(0.36)	(0.42)	(0.33)	(0.42)
NPLs	-0.27	0.06	-0.19	0.23	0.23	0.31	-1.29^{***}	-1.14^{**}
	(0.29)	(0.55)	(0.25)	(0.28)	(0.28)	(0.45)	(0.44)	(0.57)
ROA	0.56	-0.33	-0.22	0.09	0.23	0.69	-3.69^{**}	-4.74^{***}
	(0.53)	(0.57)	(0.79)	(0.75)	(0.54)	(0.60)	(1.54)	(1.83)
Remote connection to bank	0.03^{***}	0.03^{***}	0.02^{*}	0.01^{*}	0.05^{***}	0.05^{***}	-0.00	0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)
Constant	0.04	-1.21	0.04	-1.23	0.04	-0.81	0.04	0.59
	(0.20)	(0.93)	(0.20)	(1.10)	(0.20)	(2.05)	(0.20)	(2.66)
Regressors at HH level	YES	YES	YES	YES	YES	YES	YES	YES
Time FEs	YES	YES	YES	YES	YES	YES	YES	YES
Geographical FEs	YES	YES	YES	YES	YES	YES	YES	YES
Bank FEs	YES	YES	YES	YES	YES	YES	YES	YES
Observations	27,420		27,420		27,420		27,420	

Table 18: Outcome equation of holdings of financial assets

Fourth exercise described in Table 6: households who are customers of *more capitalized* Top 5 bank groups (Top 5 S) versus households who are customers of *less capitalized* Top 5 bank groups (Top 5 W), customers of *more capitalized* banks of the rest of the system (ROS S), and customers of *less capitalized* banks of the rest of the system (ROS W). See description in Table 8.

	Medium risk				Riskier				Bank bonds			
	Top 5 S	Top 5 W	ROS S	ROS W	Top 5 S	Top 5 W	ROS S	ROS W	Top 5 S	Top 5 W	ROS S	ROS W
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Size	-0.01	0.03	0.02	0.03*	-0.02	0.02	0.00	-0.02	0.10	-0.00	-0.02	0.01
	(0.02)	(0.05)	(0.02)	(0.02)	(0.01)	(0.04)	(0.02)	(0.02)	(0.28)	(0.03)	(0.04)	(0.01)
Holdings of Gov't Bonds	-0.26	0.33	0.29	0.22	-0.20	0.13	0.02	0.19	0.50	-0.56	0.65***	0.15
	(0.27)	(0.51)	(0.19)	(0.14)	(0.22)	(0.37)	(0.14)	(0.12)	(0.64)	(0.35)	(0.23)	(0.11)
Liquidity from Central Bank	0.36	0.08	-0.10	-0.13	-0.03	0.47	0.07	0.02	0.20	-0.08	0.20	-0.27***
	(0.24)	(0.40)	(0.16)	(0.10)	(0.17)	(0.35)	(0.12)	(0.08)	(0.35)	(0.27)	(0.14)	(0.07)
Funding via Bonds	0.33	0.16	0.27	0.22*	-0.16	0.07	-0.00	-0.03	-0.90	0.73**	0.48	0.12
	(0.23)	(0.41)	(0.17)	(0.13)	(0.19)	(0.32)	(0.14)	(0.11)	(0.64)	(0.31)	(0.33)	(0.09)
Capital	-1.48**	1.32	-0.29	0.13	0.25	0.63	-0.08	-0.26	1.48	0.81	-0.44	0.18
	(0.58)	(1.05)	(0.63)	(0.24)	(0.48)	(0.82)	(0.51)	(0.21)	(1.85)	(0.71)	(0.93)	(0.19)
NPLs	-1.01*	2.24	0.88	0.23	0.02	0.91	1.03*	0.57*	-0.58	0.66	0.21	0.16
	(0.56)	(1.37)	(0.75)	(0.38)	(0.45)	(1.14)	(0.63)	(0.30)	(2.67)	(0.90)	(0.80)	(0.26)
ROA	2.63	1.77	1.02	0.06	2.32*	1.04	0.14	-1.18	6.30**	2.88	2.01*	-1.18*
	(1.62)	(2.79)	(1.25)	(1.03)	(1.38)	(2.21)	(0.95)	(0.84)	(2.59)	(2.06)	(1.06)	(0.62)
Remote connection to bank	0.04***	0.06***	0.07***	0.04***	0.10***	0.08***	0.10***	0.06***	0.04***	0.03***	0.03**	0.02***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	0.43	5.03**	-0.48	-0.04	-0.78	-2.35	0.68	-1.03	-1.03	0.52	-0.68	-0.75
	(0.74)	(2.25)	(0.89)	(0.89)	(0.69)	(1.77)	(0.77)	(0.79)	(3.02)	(1.13)	(0.96)	(0.64)
Regressors at HH level	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Geographical FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Bank FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations		42,285				42,285				27,420		