Politecnico di Torino

Master's Degree Engineering and Management (LM-31)

Academic Year 2021-2022

Graduation Session December 2022

Determinants of household stock holdings: Evidence from the Survey of Consumer Finances

Supervisor:

Prof. Riccardo CALCAGNO

Candidate:

Lorenzo FOGLIA PARRUCIN

Abstract

The purpose of this study is to understand the factors that influence households' stock investment choices. Using the cross-sectional surveys of the Survey of Consumer Finances in the period 1992-2019 two types of empirical models were studied. In the first analysis, with probit models that capture the probability of participation in the stock market, the characteristics associated with the choice to own shares were analysed. Subsequently, conditional on shareholding, linear regression models were used to estimate the fraction of financial assets invested in shares. The data show that an important, albeit decreasing, portion of households do not invest in equities; through the probit models developed it is possible to estimate with an acceptable degree of accuracy whether a household has made the decision to enter the equity markets. It has been shown how net worth, the possession of a college degree and the intention to save for retirement are positively related with the ownership of shares; on the contrary investments in real estate or private business, risk aversion and having more than two children negatively affect shareholding. It was also proved that the use of the internet was significantly positively related with the holding of shares in the period 1998-2010. The linear regression models, that were used to study the choice of the fraction of financial assets invested in stocks, show that the allocation decisions are highly variable among households. Even by controlling for many characteristics the predictive power of the models is limited. Nevertheless, it is possible to affirm that: aversion to risk, the possession of investments in real estate or business are negatively related with the fraction of risky assets held. On the other hand, a college degree, the net worth and a high propensity risk are associable with a higher equity component in the linear regression models. Finally, the use of the internet to make financial choices has lost the significance that it had shown in the probit models.

Summary

1. Introduction

Participation of households in equity markets is a topic of remarkable interest due to the multiple financial and social implications it entails. The behaviour of families in this context is particularly complex to predict because empirically, it manifests a degree of heterogenicity not foreseen by theoretical models (Ameriks and Zeldes, 2004). Indeed, the Merton model (1969) or the CCAPM prescribes that stock market participation should be considerably higher than the one observed. In the simplest specification of Merton portfolio choice model, the optimal investment share in risky assets of household i should be ω_i:

$$\omega_i = \frac{\mathrm{E}[\mathrm{r}_i^e]}{\gamma_i \sigma_i^2} \qquad (1.1)$$

Where $\mathrm{E}[\mathrm{r}_i^e]$ is the expected risk premium, γ_i is the Arrow-Pratt measure of relative risk aversion and σ_i^2 the volatility of risky asset. Consequently, every positive net wealth household should hold a stake in the market portfolio assuming reasonable values of γ_i. However, empirical data do not confirm these predictions, neither regarding the share of risky assets nor for the decision to invest in the market portfolio.

The discrepancy between model predictions and actual manifestations is known to insiders as stock market participation puzzle. Various reasons have been proposed in the literature to explain the phenomenon such as: transaction costs, information costs, limited access, non-standard preferences and beliefs.

The extensive development and diffusion of internet technologies that characterized the 2000s has reduced the impact of some of these factors: transaction costs have largely decreased due to the massive introduction of telematic services and the entry into the sector of new players with low commissions; instruments that have reduced management costs such as ETFs, launched in 1993, and that allow significant diversification of investments have become widespread. Information costs have also been reduced because of the wide dissemination of public information by companies, regulatory interventions aimed at regulating disclosure of financial products and the large amount of freely accessible material on internet. All this publics available information is useful, at least in

principle, to make informed financial choices. For an introduction to the topic and other suggested readings see Guiso and Sodini (2013) chapter 4.1.

Furthermore, thanks to specific legislative interventions and automatisms offered by companies, the financial situation of families has also been subject, especially in USA, to a progressive increase of investments in retirement accounts (for example 401(k), IRAs). As can be seen in table 1.1, there has been a positive trend in households who hold retirement accounts and, conditionally on ownership, also the average amount invested raised.

Table 1.1: percentage of US households' participation in quasi-liquid retirement accounts and average investment value, evolution over time.[1]

Retirement accounts	1989	1992	1995	1998	2001	2004	2007	2010	2013	2016	2019
Not participant	62.8	59.9	54.7	51.1	47.2	50.1	47	49.6	50.8	47.9	49.5
Participant	37.2	40.1	45.3	48.9	52.8	49.9	53	50.4	49.2	52.1	50.5
Average value (k$)	76	80	96	120	151	167	182	201	221	243	255

Therefore, considering this quick summary of changes that have occurred over the last two decades, an evolution of households' participation in financial markets is expected. Based on theoretical models, a costs reduction associated with investments in financial markets should reduce barriers to entry. As a result, over time one could expect an increasing trend of households that join equity stakes, at least in the deciles with more net worth.

This empirical research aims to contribute to the existing literature by analysing recent data and explaining the characteristics of households that determine participation in stock markets and, conditional on participation, the financial amount allocated in equities. The choice of the Survey of Consumer Finances as the dataset to be analysed is motivated by the fact that it is representative of American households, which is the population that originated the most developed financial market in the world. In addition, the large sample under investigation, combined with great meticulousness used in the data collection phase, make it an ideal study base. Furthermore, the conspicuous amount of information collected makes it possible to estimate the effects of explanatory variables that can hardly be studied with other datasets.

[1] SCF cross sectional surveys, population weighted, inflation-adjusted values to 2019 dollars.

The following sections of the study are organized as follows: first, previous knowledge will be reviewed giving references to studies carried out on the subject. Section 3 will discuss the issues related to the use of the SCF surveys and the choice of variables and population under analysis. Descriptive statistics on the data used for the estimation of multivariate models will provided in chapters four. Probit regressions regarding the decision to enter the equity markets and linear regressions aimed at estimating the share of the equity participation relative to the overall financial portfolio of households will be shown in section 5. Finally, conclusions and implications of the analyses will be discussed.

2. Literature review

The literature regarding the determinants of household equity participation is relatively broad, but not always very up-to date despite the fact that there is reason to believe socio-economic events, such as those mentioned above, may have induced changes in the behaviour of households.

With regard to demographic variables Ameriks and Zeldes (2004) focused on the relationship between age and amount of wealth invested in stock market, using pooled cross sectional data form SCF and panel data form TIAA-CREF, they found no evidence that the fraction invested in shares tends to decrease with age. However, the financial planner's suggestion is generally to progressively decrease equity exposure after a certain age; while the aforementioned study found no support for this disposition, others such as those of Bodie and Crane (1997), Agnew et al. (2003) documented a decrease in equity exposure. Empirical findings of Aizcorbe et al. (2003) and Shum and Faig (2006) suggest a hump-shape age effect.

Studying the influence of sex and marital status, papers documented that risky asset allocation are higher for males (Coleman 2003, which result was in line with previous findings of Jianakoplos and Bernasek, 1998) and married investors (Agnew et al., 2003). Barber and Odean (2001) analyse the common stock investments of men and women documenting men trade 45% more than women, negatively affecting their returns.

Another often considered explanatory variable is the level of education which was found statistically significant in explaining stock holdings: Bertaut (1998), Bogan (2008), Hanna et al. (2008).

Regarding the influence of investments in other assets, such as primary residence and other real estate ownerships, the results in literature are not always consistent. Kullmann and Siegel (2003) finds that a larger real estate exposure and variability of homeowners' house values is negatively correlated with likelihood of stock market participation, the first finding is also confirmed by Cocco (2005). On the contrary, in the logistic regression of Hanna et al. (2008), ownership of investment real estate had a positive effect on stock holdings.

A widely accepted negative relationship is present between private businesses and equity investment in financial markets (see for example Shum and Faig, 2006 and Hanna et al.,

2008). The reason could be due to the fact that private businesses are perceived as a substitute of stock assets (individuals who own risky assets related to their business may be reluctant to invest additional capital in financial market equities). This latter interpretation can contribute, albeit only partially, to explaining the age-old question of the non-participation of some HNWIs[2] in financial markets, Campbell (2006).

Further well-known positively related determinants of investment choices in stocks are income uncertainty and total net worth, both were widely analysed in the literature mentioned above. Uncertainty of future income could play a role in financial decision making, in fact Hyun and Tae (2018) using the 2007–2009 SCF panel dataset and control for robustness check found that as income uncertainty increases, the amount allocated to risky financial markets tends to decrease. This evidence confirms the importance of human capital in financial decisions. Regarding net worth, although correlated both to participation and allocations choices, it is impossible using cross-sectional survey to assess whether those who invest more in risky financial markets are richer due to these investments or whether richer households tend to invest more (Guiso and Sodini, 2013). Calvet et al. (2009), after correcting for endogeneity, found that households as they become richer rebalance their portfolios towards stocks.

Probably one of the most important characteristics that determine the choice of capital allocation is risk aversion, for an introduction to the topic see Guiso and Sodini (2013) chapter 3.

Many other variables have been studied such as: the intention to incur expenses in the future Shum and Faig (2006); the effect of financial advice Georgarakos and Inderst (2014); community effects Brown et al. (2008). However, perhaps one of the most interesting features to study, is the use of the internet for making investment decisions. A remarkable study based on a very detailed microdata database, provided by the Norwegian Central Securities Depository, was carried out by Hvide et al. (2022). They show that internet use

[2] High-net-worth individual (HNWI) indicate persons whose investible wealth exceeds a given amount. Usually, these individuals are defined as holding financial assets (excluding their primary residence) with a value greater than US$1 million.

causes a substantial increase in stock market participation, as well as improving investment choices by spreading diversification and favouring behaviours more in line with portfolio theory. Comparable results, confirming the external validity of these conclusions, have recently been found by Ye et al. (2022) using data from China.

Despite all the literature briefly exposed in this paragraph, even if all the characteristics previously analysed are considered together they are not able to explain the stock market participation puzzle. Other factors, that are difficult to consider in mathematical models due to the non rational framework that they entail, could be involved in the financial decision households: familiarity and home bias, overconfidence, past experiences, cognitive limits, information asymmetry.

3. Dataset: Survey of Consumer Finances

The Survey of Consumer Finances (SCF), sponsored by the Board of Governors of the Federal Reserve System is one of the most valuable sources of information about balance sheets, financial markets participation, income and demographic characteristics of American households.

The survey has been carried out every three years since 1983, and therefore allows, thanks to thirteen cross-sectional surveys, to have an overall view of the evolution of the assets of American families. Over time, the scope of the survey has been expanded and additional questions have been created to amplify and extend the information collected.

The most reliable source from which to derive the information useful for interpreting the data of the various annual datasets is the respective "codebook" published by the FED.[3]

3.1. Primary economic unit, respondent and reference person

The survey unit, which for the sake of simplicity will henceforth be defined “household”, is a subset of the household unit called the "primary economic unit" (PEU). This aspect should be considered in order to understand the analyses carried out starting from the data. In fact, in most cases the measured variables refer to the PEU, which consists of an economically dominant individual or couple (married or partners) and all other individuals who are financially dependent from that individual or couple.

Another important distinction worthy of attention is that between the respondent and the PEU reference person. The first is the one who is interviewed (interviewers try to identify him in such a way that he is the most financially informed of the family). The second is conventionally identified in the man in heterosexual couples and in the elder in homosexual couples. Demographic data on both figures are not always collected if they do not coincide.

For the aspects listed above it follows that particular attention should be paid in the interpretation of the statistical analysis, especially when analysing relationships between variables of the individual (age, sex, race, education) and those aggregated at the family

[3] Codebooks refers to the variables in the format of a number prefixed by an "X". For example, the variable X3915 reflect the response give to the question: “What is the total market value of the publicly traded stock that your household held?”. The interested reader is therefore advised to refer to the codebook published for each survey year for any clarification regarding the meaning of the variables defined in the X format.
Link to the 2019 codebook: https://www.federalreserve.gov/econres/files/codebk2019.txt

level (such as shareholding). For example, the case could arise in which two spouses have very different ages: the investment choices are mainly decided by the young spouse, but the age considered in a regression is that of the older spouse.

Even when the variables are required at the family level, but are inherently subjective, the respondent may inadvertently provide an answer that reflects his thinking, which does not necessarily coincide with that of other family members.[4]

3.2. Sample design

The SCF is based on a dual-frame sample design. The total observations in each survey can be traced back into two different subsets: one was selected from a standard multi-stage area-probability design, the other was selected as a list sample from statistical tax records of the Statistics of Income Division of the USA Internal Revenue Service (SCF codebook, 2019). The rationale behind this choice is to be able to collect data, while maintaining a reasonable sample size, even from subjects who, following a simple random sampling process, would not be adequately represented (especially HNWI[2] individuals). The drawback is that the sample is consequently not representative of the US population if analysed as coming from a Simple random sampling process. Because of the previously mentioned complex sample design, weights play a critical role in statistical analysis on SCF records, as will be addressed later.

3.3. Imputation and implicates

The survey of consumers finances makes public for each survey year a database consisting of five implicates calculated with an imputation system. In order to carry out correct statistical analyses on the SCF data, it is necessary to consider this peculiarity of the databases. All the analysis in this paper uses repeated-imputation inference RII techniques if not differently specified. **The Stata ado "micombine" was used to consider the implicates.** For further information on the topic, please refer to **Appendix 1**.

[4] One case could be the attitude to risk aversion linked to the financial markets. The question is asked at family level but the possibility that it represents the respondent's attitude cannot be excluded. The asked question is: which of the following statements comes closest to describing the amount of financial risk that you and your husband/wife/partner are willing to take when you save or make investments?

3.4. Sampling and bootstrap weights

The staff of Survey of Consumer Finances provide two different types of weights: sampling weights and bootstrap weights.

An in-depth analysis is available in **Appendix 2**, aimed at explaining the weights system, the reasons for their introduction and the statistical techniques that can be used to make appropriate statistical inferences on data from SCF.

The Stata ado file used to consider bootstrap weights was "scfcombo".

3.5. Criteria for estimating descriptive statistics

Being based on a non-SRS sample, consumer finances surveys are not representative of the US population if the data are treated as if they were. Therefore, even for simple descriptive statistics such as means and medians, it is necessary to use sample weights whenever there is an association between the variable of interest and the sampling probabilities. As pointed out by Lindamood and Hanna (2007) in an example, previously mentioned differences in simple statistics can be broad, even by more than one order of magnitude (table 3.1). For further theoretical insight see Angus Deaton (The Analysis of Household Survey, chapter 1, 2019).

All the descriptive statistics in this paper are weighted if no differently mentioned.

Table 3.1: Weighted and unweighted mean for net worth percentile groups, 2019 SCF data.

Mean ($) of net worth for percentile groups	Unweighted	Weighted[5]
0-24.9	-14,588.10	-13,634.12
25-49.9	56,000.21	58,182.10
50-74.9	239,014.76	236,279.91
75-89.9	721,682.50	703,586.03
90-100	48,268,479.73	5,710,344.67
Total population	13,458,400.40	746,821.05

[5] The mean estimation considering weights is obtained by applying the following formula: $\overline{x_w} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$

3.6. Criteria for estimating multivariate analyses

In regressions analysis the decision to use sampling weights is controversial. Perhaps also for this reason the researchers who have analysed the SCF do not always clarify in their studies the procedure they adopted for the estimation of the coefficients. This has the clear implication of making it difficult to replicate results. Winship and Radbill (1994) suggest comparing weighted and unweighted analysis to check if the parameters differ between OLS and WOLS (weighted ordinary least square) due to sample selection bias (which is expected). They also point out that weighting will yield to consistent estimates of parameters but wrong variances. To analyse this problem Shin and Hanna (2017) compared unweighted, sample weighted and sample and bootstrap weighted analysis; although in their application bootstrap weights make little differences, they suggest using both weights to account for SCF complex sample design. Deaton (2019) emphasizes that the decision about the use of weights depends on the purpose of the research, whether econometric or statistical; given that there is interest in observing, for some variables, the magnitude of the effects on shareholding and for others their significance, it was decided to carry out both analyses, unweighted and sample and bootstrap weighted. However, considering the ambiguity on the proper methodology needed for integrating the bootstrap weights of several years together, in the latter case the analyses were conducted only for single survey year and not for pooled data.

In the regressions the heteroskedasticity of the random variables was considered by calculating the robust standard errors.

3.7. Variables

Based on the literature presented in section 2, the independent variables identified for the multivariate models are the following:

I. Age, Age^2: since as mentioned before previous studies found a hump-shape effect, the quadratic term is included to account for possible non linearities. An increase of the share of equity holdings is expected in the overall financial households' portfolio until an age near retirement, followed by a possible decrease in the share of stocks.

II. College: dummy variable whose value is equal to one if the reference person obtained a college degree, zero in all other cases. Everything else let equal, both in probit and conditional linear regression it is expected that the effect of a college degree is positively related with dependent variables.

III. Three dummy variables that measure the number of children in the PEU: one two or more than two, the reference category is without children.

IV. Married or living with partner: based on Bertaut (1998) or Guiso et al. (2003) it is expected that marriage increases stock market participation.

V. Log of net worth: the logarithm is considered to avoid skewness. In the analysis only households with a positive net worth are considered. It is expected that this variable is statistically significant both in explaining participation and share of equity over the total financial portfolio.

VI. Human capital: following Merton (1971) as Human capital over financial wealth increases, the optimal share of wealth invested in risky assets increases too.[6] Following Shum and Faig (2006), the logarithm of income was used as a measure of human capital. Moreover, the level of education (which is accounted for in the model with the degree dummy) is often considered as a proxy for human capital.

VII. Time dummies to account for time effects.

VIII. Share of primary residence value over total net worth.

[6] The relationship, generalization of (1.1) is the following:

$$\omega_i = \frac{\mathrm{E}[\mathrm{r}_i^e]}{\gamma_i \sigma_i^2}\left[1 + \frac{HC(a,T)}{W_{i,a}}\right] \quad (3.1)$$

$W_{i,a}$ is the financial wealth and $HC(a,T)$ the human capital obtained discounting future income cash flows of a fixed income y is given by:

$$HC(a,T) = \frac{y\left(1 + e^{-r_f(T-a)}\right)}{r_f} \quad (3.2)$$

Where: a is the actual age of the investor with horizon T, r_f is the risk-free rate. Since an estimate of human capital based on the data available following formula (3.2) would imply the need to make many assumptions, it was considered better to avoid them.

IX. Share of other real estate investment value over total net worth.

X. Share of private business value over total net wort: it is predicted that the variable is statistically significant both in explaining participation and share of equity over the total financial portfolio.

XI. Leverage ratio: this variable is taken into consideration also to avoid possible bias in the three previously mentioned variables.

XII. Attitude toward risk: it is supposed that risk attitude is statistically significant in explaining both equity participation and the fraction of stocks in the financial portfolio.

XIII. Saving reasons: in the SCF there is a question concerning reason for saving. Specifically, it is asked: "What are your most important reasons for saving?" One of the possible responses is "retirement". It is predicted that, inserting a dummy accounting for this saving reason, those who are aware of the need to save for retirement are more likely to participate in stock markets. In the models are also included saving reasons such as education (children's education, education of grandchildren, own education, spouse's education) and buying a primary residence.

XIV. Internet dummy: internet usage of households is accounted for using the variable INTERNET as defined in the FED bulletin. The variable contains information on the use of online banking and internet for financial decisions. Unfortunately, there are no more general questions in the survey such as stable access to broadband. For this reason, it is necessary to interpret the regression coefficients with caution due to possible endogeneity. It is expected that the use of internet generates a positive effect on participation.

XV. Gender of the reference person: if the reference person is a woman the variable is equal to one. Many studies confirm lower risk exposure in women. For this reason, capital allocation is supposed less favourable to equities in women.

XVI. Ethnicity: although this information is only available for the reference person, it is expected that native African or Hispanic people are less likely to participate in stock markets and if they participate, they are on average

more risk averse. However, the ethnic effects calculated in this study should be considered with caution as detailed information on the sex of each member of the PEU is not available.

For precise definitions of the variables used to estimate the models, refer to Annex 1.

3.8. Sample selection

The broad purpose of the analysis is to depict the behaviour of households that are in the condition to construct a stocks portfolio even if modest in value. Therefore, the choice was made to consider households that have a positive total net worth and total financial assets grater than 0$.[7] These conditions are also required because financial asset and net worth appear respectively in the denominator of a dependent variable and of the explanatory variables.

In order to reduce standard errors of estimates or consider the evolution of shareholding over time, as many surveys as possible were used (data from 1992 has generally been used). The survey years considered, however, could vary in the sections of the study as a result of the availability of data of interest in the older SCF surveys. For example, the questions regarding the use of the internet to make decisions about saving and investments are only available starting from the year 1998. If it is not evident in the representations provided, the investigation time span will be clarified in the margins of the analysis. Especially for multivariate analyses further filters could be applied to avoid possible outliers as will be specified later.

All monetary values used for the analyses were adjusted for inflation to 2019.

[7] More restrictive rules on net worth or assets are possible, but they have the drawbacks to exclude a conspicuous number of families in which the components are youth. On the contrary not consider a lower bound for total net worth is a possibility, but since there are households with negative value of this variable it cannot be excluded that, in this case, a large allocation of capital in the financial sector is due to unusual risk incentives.

4. Descriptive statistics

In this chapter the descriptive statistics that are of interest for this research are analysed.

4.1. Stock holdings

The section 4.1 shows the households' assets allocation choices in equity markets relative to total financial net worth. The focus is on direct stocks holding, stocks in funds (outside pension funds), equity in retirement accounts and the overall portfolio share considering all the categories. Only households with positive net worth and financial assets are considered. The format of the tables follows the one proposed by Shum and Faig (2006).

A detailed definition of all the variables in capital letters mentioned below are available in the annex 1 or in the SCF bulletin.[8]

4.1.1. Direct stocks holdings:

As can be seen in table 4.1, many families do not own shares directly, this attitude has remained almost constant in the period analysed. Moreover, less than 11% of the PEUs held a fraction of capital greater than 20% in directly owned shares in all the years considered.

Table 4.1: Evolution over time of percentages class of direct stocks holdings (STOCKS) over total financial wealth. The variable used to compute the value of stocks is X3915. Implicate 1, N=44736.

Direct shareholding	1992	1995	1998	2001	2004	2007	2010	2013	2016	2019
STOCKS=0	80.3[9]	82.6	78.5	76.4	77.0	79.9	82.6	84.0	84.4	83.5
0<STOCKS<=0.2	11.5	10.3	11.2	12.9	14.2	12.5	11.2	9.3	10.2	10.1
0.2<STOCKS<=0.4	3.9	3.2	4.6	5.4	4.2	3.6	3.3	2.9	2.8	3.4
0.4<STOCKS<=0.6	2.1	1.9	2.8	2.6	2.3	1.8	1.3	1.7	1.3	1.3
0.6<STOCKS<=0.8	1.3	1.0	1.8	1.3	1.4	1.3	0.8	1.1	0.8	1.0
0.8<STOCKS<1	0.7	1.1	1.0	1.2	0.8	0.9	0.6	0.8	0.6	0.8
STOCKS=1	0.1	0	0.1	0.1	0	0	0.1	0	0	0

4.1.2. Stocks in mutual funds (outside pension funds)

The variables used to compute the total values of stocks held in mutual funds outside pension funds were: STMUTF (X3822), COMUTF (X3830), OMUTF (X7787). Applying the following formula is possible to approximate the amount of equity held in funds:

[8] SCF Bulletin macro is available al this link: https://www.federalreserve.gov/econres/files/bulletin.macro.txt

[9] For example, this means that in 1992: 80.3% of American households did not hold shares directly; 11.5% of American families had a fraction of directly held shares over financial wealth between (0; 0.2].

$$FUND = STMUTF + 0.5 * COMUTF + OMUTF \quad (4.1)$$

No significant increase in participation in equity funds was detected between 1992 and 2019. In general, over the considered timeframe, the fraction of families in the sample that relies on these instruments to invest in equity was always lower than 20% and often half of this subset allocated a fraction lower than 1/5 of their financial endowment (table 4.2).

Table 4.2: Evolution over time of percentages class of stocks holdings in funds over total financial wealth (FIN). Implicate 1, N=44736.

Stocks in funds	1992	1995	1998	2001	2004	2007	2010	2013	2016	2019
STOCKS=0	90.2	87.0	83.0	81.4	83.8	87.8	90.2	90.9	89.0	90.2
0<STOCKS<=0.2	6.7	7.4	8.4	10.2	7.9	6.2	5.4	4.7	4.5	4.5
0.2<STOCKS<=0.4	2.3	2.9	4.7	4.3	4.5	3.0	2.2	2.2	3.0	2.5
0.4<STOCKS<=0.6	0.7	1.5	2.3	2.5	2.3	1.8	1.2	1.3	2.0	1.9
0.6<STOCKS<=0.8	0.1	0.4	0.9	1.1	1.0	0.8	0.6	0.5	1.0	0.6
0.8<STOCKS<1	0	0.7	0.6	0.5	0.4	0.4	0.4	0.4	0.5	0.3
STOCKS=1	0	0	0	0	0	0	0	0	0	0

4.1.3. Stocks in quasi-liquid retirement accounts

As has already been mentioned in the introduction to this study, in the last thirty years in the USA there has been an increase in families who own retirement accounts. An increasing trend is also present in the equity component of retirement accounts. Participation, in the sample analysed, increased from 28.9% to 48.2% (table 4.3); the difference is statistically significant at the one percent level.

Table 4.3: Retirement account category considers: Individual retirement accounts/Keoghs (IRAKH), Account-type pensions on current job (THRIFT) and other minor items; for a precise definition of this variable consider the RETEQ variable. Implicate 1, N=44736.

Retirement account	1992	1995	1998	2001	2004	2007	2010	2013	2016	2019
STOCKS=0	71.1	66.3	58.7	52.7	54.6	50.1	51.9	51.8	50.0	51.8
0<STOCKS<=0.2	12.3	11.6	13.6	13.8	15.2	16.9	18	17.8	19.2	18.5
0.2<STOCKS<=0.4	8.0	8.2	10.2	11.7	12.6	13.3	12.5	12.3	12.4	11.3
0.4<STOCKS<=0.6	4.8	6.9	8.3	10.1	8.7	9.1	8.5	8.8	8.8	8.8
0.6<STOCKS<=0.8	1.7	3.8	4.5	5.1	5.2	6.0	5.1	5.0	5.1	5.3
0.8<STOCKS<1	2.0	3.1	4.6	6.3	3.5	4.4	4	4.2	4.5	4.4
STOCKS=1	0.2	0.1	0.1	0.2	0.1	0.2	0.1	0.1	0.1	0

4.1.4. Total stocks holdings

Looking at the overall allocation of households' financial resources in equity (table 4.4), it is possible to note the trend already highlighted in the previous table. Indeed, participation in the financial markets has grown over the years mainly due to the spread of retirement

accounts. Households without stocks decreased from 57.5% to 44.7% between 1992 and 2019. The difference is statistically significant at the one percent level. The increase of households that allocate fractions greater than 3/5 of their financial capital in shares should also be noted.

Table 4.4: Total stocks holdings consider the previously mentioned categories plus others like managed assets with equity interest (annuities, trusts, etc.), a full definition is available under the variable EQUITY in FED bulletin. Implicate 1, N=44736.

Total shareholding	1992	1995	1998	2001	2004	2007	2010	2013	2016	2019
STOCKS=0	57.5	53.8	45.9	41.7	43.8	41.4	44.7	45.5	44.0	44.7
0<STOCKS<=0.2	14.9	12.6	10.5	9.4	10.7	13.0	14.6	13.3	14.5	13.9
0.2<STOCKS<=0.4	10.8	9.9	10.8	11.4	12.6	13.1	13.3	12.3	11.8	12.0
0.4<STOCKS<=0.6	7.6	9.3	12.5	12.6	13.7	12.8	11.4	12.1	11.9	11.8
0.6<STOCKS<=0.8	4.5	7.1	9.3	10.0	10.6	10.2	8.8	8.8	9.1	9.5
0.8<STOCKS<1	4.5	7.1	10.8	14.6	8.5	9.2	7.0	7.7	8.6	8.2
STOCKS=1	0.3	0.2	0.2	0.3	0.1	0.2	0.2	0.2	0.1	0

Although it is clear that households' portfolios have migrated towards equity in the last thirty years, the reasons for this phenomenon are many. Not just the already mentioned reasons, such as the development of telematic markets and the emergence of online platforms with low commissions and ETFs, but also many other causes could have favoured the phenomenon. For example, the legislation introduced regarding pension funds and the reduction over time of the interest rates induced by central banks including the FED could have affected equity investments. Establishing the quantitative effect of these changes on shareholding is complex; however, considering the data available, this research will try to provide an answer with regard to the spread of the internet and the evolutions that have taken place in the pension sector.

4.2. Explanatory variables

These paragraphs present the trend over time of explanatory variables weighted averages. All the statistics consider only households with strictly positive net worth and financial assets. Further constraints imposed for specific predictors are clarified in the relative sections. For reasons of space, the following abbreviations will be used: NEH and EH refer to not equity holders and equity holders respectively. The statistical hypothesis tests refer to an $\alpha = 0.05$.

4.2.1. Age

Regarding the mean age of the reference person (Table 4.5), it can be noted that between 1992 and 2019 there was an increase in the average age. The increase is statistically significant at 95%. Mean age of NEH and EH were similar in 2019.

Table 4.5: Average age of household reference person. N=44723.

Mean age	1992	1995	1998	2001	2004	2007	2010	2013	2016	2019
Equity holders	48.8	48.0	48.5	48.2	50.1	51.0	52.1	52.3	53.5	53.0
Not equity holders	50.4	51.0	52.3	53.0	52.0	52.3	52.4	53.6	53.0	53.6
Total	49.7	49.6	50.2	50.2	51.0	51.6	52.2	52.8	53.3	53.3

4.2.2. College degree

In table 4.6 it is possible to see how the percentage of graduates differs substantially between EH and NEH. Although graduates are around 36% in 2019, they are more than 50% in the class of equity holders. This suggests that graduation leads to benefit from the equity risk premium more than people with lower educational qualifications.

Table 4.6: Percentage of reference people that were graduated. N=44723.

Graduated percentage	1992	1995	1998	2001	2004	2007	2010	2013	2016	2019
Not equity holder	19	16	16	14	16	15	17	15	16	18
Equity holder	45	41	39	43	45	43	46	49	50	51
Total	30	27	28	31	32	31	33	34	35	36

4.2.3. Children

The number of dependent children seems to be related to the probability of holding shares; in particular the highest participation probability is found among families with 2 children (Table 4.7). The lower participation associated with not having children is reasonably linked to age; in fact, the elderly and young people should tend to participate less in the stock markets. On the contrary, the reduction associated with having more than two children could be due to ethnicity or the important financial commitment that requires their maintenance.

Table 4.7: Percentage of shareholder by number of children. N=44723.

Equity holders	1992	1995	1998	2001	2004	2007	2010	2013	2016	2019
No kids	41	44	51	54	55	58	55	54	55	54
1 kid	43	44	57	60	55	60	56	51	55	57
2 kids	52	54	65	67	62	62	61	61	64	65
From 3 to 10 kids	38	46	54	65	55	55	50	54	51	51

4.2.4. Married or living with a partner

Table 4.8 shows how the EHs, on average, cohabite more with a spouse or a partner compared to the NEHs; the phenomenon could be related to age. it is reasonable to assume that young single men may have a different investment profile compared to married people, single women or widowed retired people.

Table 4.8: Percentage of reference people that are married or living with partner. N=44723.

Married/partner	1992	1995	1998	2001	2004	2007	2010	2013	2016	2019
Not equity holder	53	55	51	52	49	50	51	50	47	46
Equity holder	71	70	70	72	70	69	68	69	68	68
Total	61	62	61	63	61	61	61	60	59	58

4.2.5. Net worth

In table 4.9 it is possible to observe the average net worth over time of households by percentile groupings and, in the last three columns, the same variable assumed to be EH or NEH. Especially in the first group 0-24.9 average wealth seems high, however it should be remembered that all households with negative net worth have been excluded from the calculation of values. In fact, considering all respondents to the survey and weighting the observations, net worth would assume negative values in the first group (about -13,800 $ in 2019).

Observing the evolution of wealth over time, an increase in inequality among American families is evident, as confirmed by the data published by the World Bank which attests that the Gini coefficient in the USA went from 38.4 to 41.5 between 1992 and 2019. Considering the analysed data, between 1992 and 2019 there was a reduction, albeit not statistically significant, in the average net wealth of the poorest households and a significant increase in the economic endowment of 50% of the richest households (the last decile has seen its wealth more than double).

Dividing the population analysed between equity holders and non-equity holders, the data shows that the average net worth between the two categories is statistically different for all the years of survey. Shareholders' wealth was around three times that of NEHs in 1992. The gap has further grown over the years to reach a ratio of close to seven in 2019.

Table 4.10 shows the participation percentage among the different percentile groups. It is evident that the major change in participation occurs in the two last percentile groups.

Despite an already high participation rate, the fourth quartile shows the highest variation in participation.

Looking at the data presented and the evolution over time of the American stock indices, there are reasonable elements to suppose that there is a two-way phenomenon of causality between wealth and equity participation and wealth and capital allocation choices in shares. For these reasons, the models analysed could be affected by endogeneity.

Table 4.9: Average net worth ($) by percentile groups, average wealth of stockholders. N=44723.

Net worth ($)	1992	1995	1998	2001	2004
0-24.9 percentile	5,199	6,829	5,845	7,423	6,787
25-49.9 percentile	45,504	51,538	56,884	64,159	64,130
50-74.9 percentile	162,546	166,702	202,645	242,631	252,505
75-89.9 percentile	391,935	398,501	507,610	652,421	717,016
90-100 percentile	2,240,301	2,416,063	3,058,245	3,993,964	4,227,998
Not equity holders	200,710	201,285	178,220	196,457	204,065
Equity holders	649,509	663,149	798,566	982,957	1,085,092
Total	392,587	412,973	517,468	654,732	697,151
Net worth	2007	2010	2013	2016	2019
0-24.9 percentile	7,061	4,311	4,035	3,877	4,785
25-49.9 percentile	72,600	42,272	39,512	47,623	58,182
50-74.9 percentile	281,639	198,080	195,841	217,256	236,141
75-89.9 percentile	725,859	619,245	600,797	701,164	704,055
90-100 percentile	4,905,889	4,338,171	4,357,119	5,648,505	5,710,345
Not equity holders	237,911	185,756	184,643	187,488	199,998
Equity holders	1,174,634	1,106,508	1,116,664	1,356,384	1,366,049
Total	788,327	696,369	691,904	839,709	848,806

Table 4.10: participation percentage in equity market by percentile intervals.

Equity participant	1992	1995	1998	2001	2004	2007	2010	2013	2016	2019
0-24.9 percentile	11.4	16.2	17.5	21.7	14.3	20.7	14.8	8.5	11.2	11.3
25-49.9 percentile	30.8	38.3	45.6	45.8	44.9	48.1	41.8	39.5	41.3	42.2
50-74.9 percentile	45.7	45.7	55.2	63.9	59.4	60.8	57.8	58.2	63.4	61.7
75-89.9 percentile	58.7	61.3	77.5	79.8	82.8	83.4	77.5	79.7	82.4	80.9
90-100 percentile	75.7	80.9	87.8	92.1	92.0	89.8	91.4	92.1	93.8	94.3
TOTAL	42.5	46.2	54.1	58.3	56.2	58.6	55.3	54.5	56.0	55.3

4.2.6. Income

Similar considerations to net worth can also be made for income (table 4.11):

- The average is statistically different between the two groups for all surveys.

- Between 1992 and 2019 the hypothesis that there were no income increases among NEHs cannot be rejected.
- The average income between 1992 and 2019 or 1992 and 2016 statistically increased for EHs.

Again, the likelihood of possible endogeneity between income and dependent variables studied in the models is emphasized, although the severity of the problem should be lower than that which could occur with net worth.

Table 4.11: Average income ($) of shareholder and not shareholder households.

Mean income ($)	1992	1995	1998	2001	2004
Not equity holder	52,508	52,294	48,458	51,893	51,212
Equity holder	115,075	117,341	127,891	150,979	146,957
Total	79,258	82,107	91,898	109,628	104,798
Mean income	2007	2010	2013	2016	2019
Not equity holder	52,248	50,903	47,382	46,722	52,577
Equity holder	157,308	142,984	155,007	174,371	163,826
Total	113,981	101,967	105,958	117,948	114,478

4.2.7. Ratio of primary residence, real estate and business over net worth

For the tables in this paragraph, further constraints were placed on the sample studied. It seemed reasonable to consider only the families that owned the assets analysed from time to time. Therefore, for example, for table 4.12 only the families that owned a primary residence entered in the computed values. Furthermore, outliers were eliminated because they would have deeply affected the interpretation of the results, so only households with a ratio of primary residence/net worth<4 were considered in table 4.12. Analogous considerations hold for table 4.13 and 4.14.

The ratios in the table below show that the primary residence, if purchased, is one of the main investments for NEHs. Moreover, the ratios constantly higher than 1, since the 2000s, indicate the large recourse to mortgages by Americans for the purchase of their primary residence. On the other hand, other real estate investments (table 4.13) do not show such high ratios as those of table 4.12.

Finally, the relationship studied in table 4.14 is particularly interesting because it highlights that among entrepreneurs who do not hold shares, the net worth is largely constituted by the value of their entrepreneurial activities. In other words, the data seems to confirm the

hypothesis that those with wealth heavily dependent on entrepreneurial activities tend to invest less in financial markets.

Table 4.12: Average of the ratio between the value of primary residence and net worth. N=32723.

Houses/net worth	1992	1995	1998	2001	2004	2007	2010	2013	2016	2019
Not equity holder	0.98	0.98	1	1.03	1.12	1.14	1.19	1.09	1.18	1.13
Equity holder	0.84	0.87	0.8	0.77	0.87	0.87	0.86	0.77	0.77	0.81
Total	0.91	0.92	0.88	0.86	0.96	0.96	0.98	0.89	0.91	0.92

Table 4.13: Average of the ratio between the value of real estate investment (excluding primary residence) and net worth. N=35136.

Real estate/net worth	1992	1995	1998	2001	2004	2007	2010	2013	2016	2019
Not equity holder	0.09	0.09	0.09	0.08	0.07	0.09	0.09	0.08	0.08	0.09
Equity holder	0.11	0.1	0.09	0.08	0.1	0.11	0.11	0.11	0.09	0.08
Total	0.1	0.1	0.09	0.08	0.09	0.1	0.1	0.1	0.09	0.09

Table 4.14: Average of the ratio between the value of private business and net worth. N=13020.

business/net worth	1992	1995	1998	2001	2004	2007	2010	2013	2016	2019
Not equity holder	0.48	0.47	0.47	0.47	0.51	0.51	0.54	0.57	0.48	0.47
Equity holder	0.37	0.28	0.33	0.33	0.34	0.33	0.33	0.33	0.31	0.31
Total	0.42	0.37	0.37	0.37	0.39	0.39	0.39	0.4	0.36	0.35

4.2.8. Financial leverage

Table 4.15 shows the average leverage by investment profile. Only households that have debt and whose asset value exceeds the debt itself were considered. On average, those who do not hold shares have a higher leverage ratio.

Table 4.15: Average of the ratio of debt over asset (leverage). N=32682.

Debt/asset ratio	1992	1995	1998	2001	2004	2007	2010	2013	2016	2019
Non equity holder	0.32	0.32	0.35	0.35	0.38	0.38	0.41	0.4	0.39	0.36
Equity holder	0.3	0.32	0.31	0.3	0.32	0.33	0.35	0.33	0.31	0.31
Mean	0.31	0.32	0.32	0.32	0.35	0.35	0.37	0.35	0.34	0.33

4.2.9. Financial risk attitude

The average trend of the variable that is used to measure the degree of household risk aversion is summarized in table 4.16. More than half of interviewees state that they adopt a moderate risk profile (obtain returns in line with the average by assuming average risks); while less than 5% declare that they want to take substantial risks for obtaining above-average earnings. In absolute terms, the percentages of those willing to take considerable

risks do not vary much between NEH and EH. Finally, the desire not to take risks is widespread among NEHs (60% of respondents in 2019).

Table 4.16: Attitude of households regarding investment in financial asset, percentage of families that report the respective attitude. N=44736.

Attitude towards risk	1992	1995	1998	2001	2004	2007	2010	2013	2016	2019
Dislike fin. risk, NEH	60	58	58	63	64	62	67	66	61	60
Dislike fin. risk, EH	29	25	18	18	20	23	27	25	23	22
Dislike fin. risk	47	43	36	37	39	39	45	44	40	39
Medium fin. risk, NEH	37	39	39	35	34	36	30	32	35	36
Medium fin. risk, EH	68	71	76	76	76	73	69	72	73	74
Medium fin. risk	50	54	59	58	58	58	52	53	56	57
Substantial fin. risk NEH	3	3	3	2	2	2	3	2	4	4
Substantial fin. risk, EH	3	4	6	6	4	4	4	3	4	4
Substantial fin. risk	3	3	5	5	3	3	3	3	4	4

4.2.10. Reasons for savings: retirement, education and home purchase

Commitment to purpose can be a strong incentive to save and financial planners often ask for this kind of information to assess the expectations and needs of a client. Among the savings reasons presented, retirement is undoubtedly the most widespread. EHs declare retirement as a reason for saving almost twice as often as NEHs (lines 2 and 3 of table 4.17). Education in the United States is often much more expensive than in other countries such as Europe, so it is reasonable to assume that, in order to ensure a satisfactory education for children, grandchildren, yourself or a partner, it is appropriate to dedicate planning to this cause. Between 1992 and 2004 more than 10% of the EHs interviewed declared education as a reason for saving, but this percentage has decreased over time.

Table 4.17: Households percentage that save for Retirement, education, primary home purchase. N=44736.

Saving reasons	1992	1995	1998	2001	2004	2007	2010	2013	2016	2019
Retirement, NEH	16	17	24	23	26	22	20	21	19	18
Retirement, EH	29	37	46	43	46	47	42	41	42	40
Retirement	22	26	36	35	37	37	32	32	32	30
Education, NEH	8	10	10	10	11	9	8	9	8	6
Education, EH	10	12	12	11	12	7	7	8	7	5
Education	9	11	11	11	11	8	8	8	7	5
Home, NEH	4	5	4	4	5	4	4	3	5	6
Home, EH	3	4	3	3	3	2	2	2	3	3
Home	3	5	3	3	4	3	3	2	4	4

4.2.11. Sex of the reference person

Analysing the percentage of women in the variable "sex of the family referent", it is possible to see that females are a minority compared to males (table 4.18). The reason lies in the conventions used by the SCF presented in 3.1. However, assuming one considers only the families that invest in shares, the prevalence of women is further reduced, so much so that they never represented more than 20% of respondents.

Table 4.18: Percentage of woman that are the reference person of the household. N=44736.

Woman	1992	1995	1998	2001	2004	2007	2010	2013	2016	2019
Non equity holders	30	31	34	35	35	35	32	33	34	32
Equity holders	18	20	19	17	18	20	20	20	19	19
Total	25	26	26	25	26	26	25	26	26	25

4.2.12. Ethnicity

Ethnicity has often been studied as an element that could influence investments in stocks. In fact, in table 4.19 it is possible to observe how African or Hispanic people invest less in equity than white people. For example, in 2019 only 29% of the Hispanics interviewed declared they had equity against 63% of whites. The reasons for this trend could be multiple and are potentially attributable to socio-cultural, economic and historical differences.

Table 4.19: Percentage of equity investors among the different ethnicities. N=44736.

Equity investors	1992	1995	1998	2001	2004	2007	2010	2013	2016	2019
White non-Hispanic	47	49	58	62	62	63	62	61	63	63
Black / African American	24	27	39	43	35	43	38	38	35	36
Hispanic	21	34	29	40	27	37	30	27	33	29
Other	31	44	55	61	54	67	54	54	58	66
Total	43	46	55	58	56	59	55	54	56	56

4.2.13. Internet usage

Based on the data in table 4.20, it is clear that EHs have a much higher probability of using the internet than NEHs. However, a causal relationship between internet use and shareholding is difficult not only to estimate but also to prove. Possible biases associated with the lack of control of characteristics related to the use of internet must be considered; for example, having a STEM degree could increase the chances of having to know how to use a computer. Through multivariate analysis it is possible to reduce these biases, with the awareness that the endogeneity problems already mentioned can affect the validity of the results.

Table 4.20: Percentage Households that use internet to make financial decision. N=41248.

Internet	1998	2001	2004	2007	2010	2013	2016	2019
Non equity holders	7	18	30	44	53	60	67	75
Equity holders	24	46	62	74	81	84	88	91
Total	17	34	48	61	68	73	79	84

4.3. Correlation matrices

In order to investigate possible linear relationships between dependent and independent variables, two correlation matrices were created before estimating the models in the next chapter. The first concerns shareholding, while the second concerns the fraction of financial capital invested directly or indirectly in shares. Correlation tables are presented in annex 3.

As one could expect, participation in equity investments and net worth are highly correlated. Other variables positively correlated to the former, with ρ coefficients between [0.2; 0.5), are income and the dummy variables that measure the fact of being graduated, living in a couple, saving for retirement and using internet. Conversely, high risk aversion and being a woman are negatively correlated with equity participation, ρ=-0.45 and ρ=-0.22 respectively. The second correlation matrix inherent to the fraction of financial capital invested in equity shows much lower correlation forces. In fact, all predictors have a correlation coefficient in absolute value lower than 0.2 with the dependent variable. This is symptom of the fact that the predictive capacity of the regression models will probably be lower than that of the probit models. The higher ρ between dependent and independent variables is found with net wealth followed by income and the degree dummy, while the lower one is found with risk aversion in financial investments. Note the inversion of the sign related to the coefficient of the real estate investment over net worth and dependent variables between the two tables.

A further useful fact of the correlation matrix is that it allows in first analysis to exclude the possibility that the models are affected by multicollinearity problems. Excluding the high ρ between the two age-related variables, which is perfectly predictable, there are no other correlation values which could suggest problems of multicollinearity between the independent variables. The highest positive correlations between independent variables

are found between net wealth and income, leverage ratio and value of the primary residence on net wealth. The lowest ρ occurs between the dummies married and woman. This value, which would normally seem strange, is due to the reasons given in paragraph 3.1.

5. Multivariate analyses

For both multivariate analyses outliers were eliminated because they could have affected the goodness of results. In addition to the criteria set out in paragraph 3.8, the following constraints were used:

- HOUSES/NET WORTH<4: this ratio can assume very high values in subjects who own a primary residence, but who find themselves in financial difficulties. The outliers must be removed because, theoretically, the investment choices of these individuals do not fall within the scope of this analysis; practically, statistical software would not be able to converge to a solution. However, the number of observations rejected through this filter are rather small (about 2.83% of the families). Furthermore, considering the ratio of the house net value (subtracting residential mortgages) and net worth, the percentage becomes negligible (0.20%).
- -4<BUSINESS/NET WORTH>4: Given that, especially in the initial phase, the net fair assets value of an entrepreneurial activity could be negative, it was considered appropriate not to exclude all negative values for this indicator A eventually negative value is due to a business figure less than zero since net worth is keep positive. The percentage of observations that do not satisfy this condition is about 0.06%.
- -4<REAL ESTATE INVESTMENT/ NET WORTH <4: analogous consideration made for the ratio previously mentioned are true also in this case. The percentage of observations that do not satisfy this condition is about 0.18%.
- DEBT/ASSET<1: an analysis on the investment choices of these individuals is outside the scope of this research.

The filters, if considered together, eliminate 13.8% of the observations in the period 1992-2019. 2010 was the year with the highest percentage of discarded observations (18.2%). Most of the observations are discarded by the positivity constraints of net worth and financial assets.

5.1. Probit models: decision to invest in equity, pooled data

This section will analyse the choice of families to participate in the equity markets whatever the methodology undertaken to invest in stock: direct stock holdings, mutual fund, pension fund. A PEU is assumed to be a participant if it holds an amount greater than zero in equity financial markets.

Considering the aspect raised in paragraph 3.1 (related to the use of individual demographic variables to predict family behaviours) in the model presented in table 5.1 it was decided to include only the variables that are probably less affected by drawbacks. The sex of reference person and his ethnicity were therefore excluded in the first analyses. Table 5.2 shows the results obtained including gender and ethnicity. The use of the internet as a predictor has been included in table 5.3. The reason for a separate analysis is due to the fact that the variable has been available since 1998. In general, the models predict in a satisfactory way who owns shares; in fact, in every probit the percentage of correct forecast is always higher than 80% (table 5.5).

5.1.1. Unweighted analyses

Following the suggestion of Hanna and Shin (2017) unweighted analysis should be used for hypothesis testing, such advice is followed in this research. The next comments refer to unweighted results of tables 5.1, 5.2 and 5.3. Hypothesis tests are referred at the five percent level.

In all models, the variables degree, net worth, income, debt/asset ratio, substantial financial risk and saving for retirement are statistically significant and increase the probability of holding shares. Conversely three or more children, no financial risk, the ratios of primary residence, real estate investments and business are statistically significant, but reduce the probability of being a shareholder. For a quick overview of the significant variables, see the table 5.4. Examining the other variables common to the three models it is possible to note that:

- the predictor “two children” is never significant at five per cent.
- Saving for a home is never significant, so the wish to buy a house does not seem to influence the shareholding.
- The coefficients of saving for education and “one child” have a P-value greater than 0.05 in the model with internet.

Particular attention should be paid to studying age. In every probit the quadratic term is significant, but the linear term is significant only in table 5.2**.** The reason is probably due to the imperfect, but strong, collinearity between age and its square. This may cause an inaccurate estimate of the coefficients and inflated standard errors. Following the estimates made, fixed the other factors, the age at which there is the greatest probability of holding shares is in the 19-22 years range. However, these age-related results should be further analysed for the reasons already set out and also because it would be more reasonable to use panel data to study how age affects the shareholding. In other words, to follow an individual's decisions over time would provide more robust results.

The dummy variables concerning ethnicity are all statistically significant (table 5.2) and have a negative coefficient with respect to the reference category “white not Hispanic”, so it can be considered that, on average, a non-white member in the household negatively affects equity participation.

By including the internet among the predictors (table 5.3) it is concluded that those who use it to entertain relationships with financial institutions (make payments, access their online banking, find out about investment choices or to search for a loan) are more likely to hold shares. A strong increase in stock market participation was also found also by Bogan (2008) but by analysing data on computer usage. Unfortunately, on the basis of the data, it was not possible to have a more plausible exogenous measure (such as the availability of broadband in the city of residence over time) to measure the effect of the diffusion of the internet on investments.

5.1.2. Weighted analyses

The following paragraph will focus more on the effects that explanatory variables have on participation; in fact, by weighting the observations it is possible to obtain an estimate that is representative of the population. Clearly, this does not mean that the value of the coefficients cannot show bias for omitted variables, or that, considering interaction terms, the effects are always equal between households with different characteristics.

A comparison between weighted and unweighted analysis shows how the variables that were statistically significant in each unweighted probit remain significant also in weighted

analysis. The only exception is age^2 in the specification with internet. The finding is interesting because it shows a certain stability in the result.

Looking at the coefficients of the different model specifications table 5.1, 5.2 and 5.3, it emerges in the weighted models that, for a fixed predictor, they always agree in sign (ex. coefficients of degree are all positive). The only exception is marital status. Furthermore, for the logarithm net worth and income, houses/net worth, real estate/net worth, business/net worth, debt/asset, no financial risk and saving for retirement, the estimated differences are always less than 6%. For the dummies “degree” and “three or more children” the maximum difference among the estimated effects is 20%.

Once a relative stability of the results is ascertained, it is possible to observe the marginal effect of a specific variable on shareholding. The model in table 5.3 is used to carry out some calculations. Naturally, in the probit model the effect generated by varying a predictor is a function of the values assumed by the other variables. It is therefore necessary to make assumptions regarding the characteristic of the household. For example, consider a 30-year-old single person with no children, internet user, without non-financial investments and debt, with net wealth of $ 100,000 and income of $ 40,000. The willingness to save for retirement causes an increase in the probability of holding shares by 7.35%. In a similar individual of 60 years who saves for retirement the use of internet increases the probability of holding stocks by 7.16%.

Table 5.1: probit models for equity participation, not including ethnicity and gender. Period: 1992-2019. Pooled data. One, two or three asterisks means that the predictor is significant at the ten, five or one percent level respectively. Models estimated considering data of all the five implicates through the Stata ado "micombine".

	Unweighted		Weighted	
	Coefficient	P-value	Coefficient	P-value
AGE	5.85×10^{-3}*	0.054	-1.99×10^{-3}	0.581
AGE^2	-1.38×10^{-4}***	<0.001	-8.24×10^{-5}**	0.015
Education (reference: not have a degree)				
COLLEGE DEGREE	0.2930***	<0.001	0.2298***	<0.001
Children (reference: without children)				
ONE CHILD	-0.0541**	0.022	-0.0669**	0.015
TWO CHILDREN	-0.0155	0.550	-0.0311	0.295
THREE OR MORE CHILDREN	-0.1607***	<0.001	-0.1897***	<0.001
Married (ref: not married/living with partner)				
MARRIED	0.0298	0.115	-0.0521**	0.025
ln(NET WORTH)	0.3836***	<0.001	0.4520***	<0.001
ln(INCOME)	0.1048***	<0.001	0.2094***	<0.001
HOUSES/NET WORTH	-0.3741***	<0.001	-0.4314***	<0.001
REAL ESTATE/NET WORTH	-0.6110***	<0.001	-0.5943***	<0.001
BUSINESS/NET WORTH	-1.4044***	<0.001	-1.3131***	<0.001
DEBT/ASSET	1.4344***	<0.001	1.6138***	<0.001
Risk attitude (reference: medium financial risk)				
NO FINANCIAL RISK	-0.6117***	<0.001	-0.5534***	<0.001
SUBSTANTIAL FINANCIAL RISK	-0.2441***	<0.001	-0.0725	0.140
Saving reasons				
SAVING FOR RETIREMENT	0.3010***	<0.001	0.2566***	<0.001
SAVING FOR EDUCATION	0.0596**	0.047	0.0031	0.927
SAVING FOR HOME	0.0277	0.539	0.0344	0.485
Survey year (reference: 1992)				
1995	0.0515	0.216	0.0010	0.984
1998	0.2222***	<0.001	0.1722***	<0.001
2001	0.3246***	<0.001	0.2170***	<0.001
2004	0.2340***	<0.001	0.1526***	0.001
2007	0.3259***	<0.001	0.2401***	<0.001
2010	0.3732***	<0.001	0.2930***	<0.001
2013	0.3343***	<0.001	0.2534***	<0.001
2016	0.3437***	<0.001	0.2481***	<0.001
2019	0.3004***	<0.001	0.2149***	<0.001
constant	-5.5796***	<0.001	-7.1328***	<0.001
Sample size	N=43396		N=43396	

Table 5.2: probit models for equity participation, including ethnicity and gender. Period: 1992-2019. Pooled data. One, two or three asterisks means that the predictor is significant at the ten, five or one percent level respectively. Models estimated considering data of all the five implicates through the Stata ado "micombine".

	Unweighted		Weighted	
	Coefficient	P-value	Coefficient	P-value
AGE	$6.84x10^{-3}$**	0.025	$-8.36x10^{-4}$	0.817
AGE^2	$-1.59x10^{-4}$***	<0.001	$-1.05x10^{-4}$***	0.002
Education (reference: not have a degree)				
COLLEGE DEGREE	0.2860***	<0.001	0.2211***	<0.001
Children (reference: without children)				
ONE CHILD	-0.0528**	0.026	-0.0694**	0.012
TWO CHILDREN	-0.0052	0.846	-0.0242	0.425
THREE OR MORE CHILDREN	-0.1372***	<0.001	-0.1686***	<0.001
Married (ref: not married/living with partner)				
MARRIED	0.1116***	<0.001	0.0503*	0.092
ln(NET WORTH)	0.3750***	<0.001	0.4429***	<0.001
ln(INCOME)	0.1073***	<0.001	0.2126***	<0.001
HOUSES/NET WORTH	-0.3752***	<0.001	-0.4329***	<0.001
REAL ESTATE/NET WORTH	-0.5931***	<0.001	-0.5768***	<0.001
BUSINESS/NET WORTH	-1.4035***	<0.001	-1.3180***	<0.001
DEBT/ASSET	1.4041***	<0.001	1.5831***	<0.001
Risk attitude (reference: medium financial risk)				
NO FINANCIAL RISK	-0.5967***	<0.001	-0.5423***	<0.001
SUBSTANTIAL FINANCIAL RISK	-0.2178***	<0.001	-0.0368	0.455
Saving reasons				
SAVING FOR RETIREMENT	0.2896***	<0.001	0.2451***	<0.001
SAVING FOR EDUCATION	0.0654**	0.030	0.0079	0.819
SAVING FOR HOME	0.0405	0.373	0.0492	0.320
Sex (reference: male)				
WOMAN	0.1359***	<0.001	0.1651***	<0.001
Ethnicity (reference: white not Hispanic)				
BLACK/AFRICAN AMERICAN	-0.1775***	<0.001	-0.1723***	<0.001
HISPANIC	-0.4088***	<0.001	-0.3854***	<0.001
OTHER ETHNICITY	-0.2273***	<0.001	-0.1802***	<0.001
Survey year (reference: 1992)				
1995	0.0490	0.238	-0.0012	0.981
1998	0.2290***	<0.001	0.1780***	<0.001
2001	0.3320***	<0.001	0.2269***	<0.001
2004	0.2553***	<0.001	0.1728***	<0.001
2007	0.3447***	<0.001	0.2609***	<0.001
2010	0.4033***	<0.001	0.3251***	<0.001
2013	0.3644***	<0.001	0.2840***	<0.001
2016	0.3850***	<0.001	0.2897***	<0.001
2019	0.3464***	<0.001	0.2585***	<0.001
constant	-5.5368***	<0.001	-7.1158***	<0.001
Sample size	N=43396		N=43396	

Table 5.3: probit model for equity participation, including internet variable available since 1998. Period: 1998-2019. Pooled data. One, two or three asterisks means that the predictor is significant at the ten, five or one percent level respectively. Models estimated considering data of all the five implicates through the Stata ado "micombine".

	Unweighted		Weighted	
	Coefficient	P-value	Coefficient	P-value
AGE	$4.22x10^{-3}$	0.205	$-3.47x10^{-3}$	0.373
AGE^2	$-1.12x10^{-4}$***	<0.001	$-5.92x10^{-5}$	0.107
Education (reference: not have a degree)				
COLLEGE DEGREE	0.2612***	<0.001	0.1923***	<0.001
Children (reference: without children)				
ONE CHILD	-0.0469*	0.071	-0.0592**	0.048
TWO CHILDREN	-0.0090	0.761	-0.0367	0.274
THREE OR MORE CHILDREN	-0.1565***	<0.001	-0.1924***	<0.001
Married (ref: not married/living with partner)				
MARRIED	0.0086	0.679	-0.0576**	0.027
ln(NET WORTH)	0.3941***	<0.001	0.4579***	<0.001
ln(INCOME)	0.1090***	<0.001	0.2104***	<0.001
HOUSES/NET WORTH	-0.3880***	<0.001	-0.4394***	<0.001
REAL ESTATE/NET WORTH	-0.5849***	<0.001	-0.6022***	<0.001
BUSINESS/NET WORTH	-1.4352***	<0.001	-1.3410***	<0.001
DEBT/ASSET	1.4014***	<0.001	1.5658***	<0.001
Risk attitude (reference: medium financial risk)				
NO FINANCIAL RISK	-0.5973***	<0.001	-0.5567***	<0.001
SUBSTANTIAL FINANCIAL RISK	-0.2407***	<0.001	-0.0880	0.112
Saving reasons				
SAVING FOR RETIREMENT	0.2866***	<0.001	0.2447***	<0.001
SAVING FOR EDUCATION	0.0557	0.103	0.0152	0.695
SAVING FOR HOME	0.0178	0.721	0.0279	0.622
Internet usage (reference: not used)				
INTERNET	0.2457***	<0.001	0.2045***	<0.001
Survey year (reference: 1998)				
2001	0.0599	0.119	0.0093	0.833
2004	-0.0655	0.108	-0.0821*	0.085
2007	-0.0107	0.791	-0.0228	0.621
2010	0.0215	0.572	0.0145	0.737
2013	-0.0360	0.352	-0.0403	0.377
2016	-0.0408	0.286	-0.0582	0.201
2019	-0.0972**	0.017	-0.1024**	0.027
constant	-5.5242***	<0.001	-7.0230***	<0.001
Sample size	N=36244		N=36244	

Table 5.4: coefficient significance at the five percent level, comparison among the different probit model specifications. Legend: "ns" means not significant (blue background), "s" means significant, "-" means coefficient not available.

	1992 Unweighted		1992 Weighted		1998 Unweighted	1998 Weighted
AGE	ns	s	ns	ns	ns	ns
AGE^2	s	s	s	s	s	ns
COLLEGE DEGREE	s	s	s	s	s	s
ONE CHILD	s	s	s	s	ns	s
TWO CHILDREN	ns	ns	ns	ns	ns	ns
THREE OR MORE CHILDREN	s	s	s	s	s	s
MARRIED	ns	s	s	ns	ns	s
ln(NET WORTH)	s	s	s	s	s	s
ln(INCOME)	s	s	s	s	s	s
HOUSES/NET WORTH	s	s	s	s	s	s
REAL ESTATE/NET WORTH	s	s	s	s	s	s
BUSINESS/NET WORTH	s	s	s	s	s	s
DEBT/ASSET	s	s	s	s	s	s
NO FINANCIAL RISK	s	s	s	s	s	s
SUBSTANTIAL FINANCIAL RISK	s	s	ns	ns	s	ns
SAVING FOR RETIREMENT	s	s	s	s	s	s
SAVING FOR EDUCATION	s	s	ns	ns	ns	ns
SAVING FOR HOME	ns	ns	ns	ns	ns	ns
WOMAN	-	s	-	s	-	-
BLACK/AFRICAN AMERICAN	-	s	-	s	-	-
HISPANIC	-	s	-	s	-	-
OTHER ETHNICITY	-	s	-	s	-	-
INTERNET	-	-	-	-	s	s

Table 5.5: Percentage of accordant predictions. A household is assumed to be an equity holder if the probability of owning stocks is grater than 50%.

	1992 Unweighted		1992 Weighted		1998 Unweighted	1998 Weighted
Accordant predictions	81.75	81.98	81.93	82.07	82.70	82.86

5.2. Linear regression: equity share in the financial portfolio, pooled data

Having analysed the choice to participate in the stock markets, now among the families that own shares, the allocation choices of financial capital in shares are studied. The comments provided in 5.2 for hypothesis tests refer to the five percent level.

5.2.1. Unweighted analyses

The unweighted analyses show that, similarly to what happens for participation, degree, net worth, saving for retirement and substantial financial risk attitude are significant in explaining the risky share allocation of financial resources and are positively related. Real estate investments, business and risk aversion are significant and have a negative effect on financial risk profile, tables 5.6, 5.7 and 5.8.

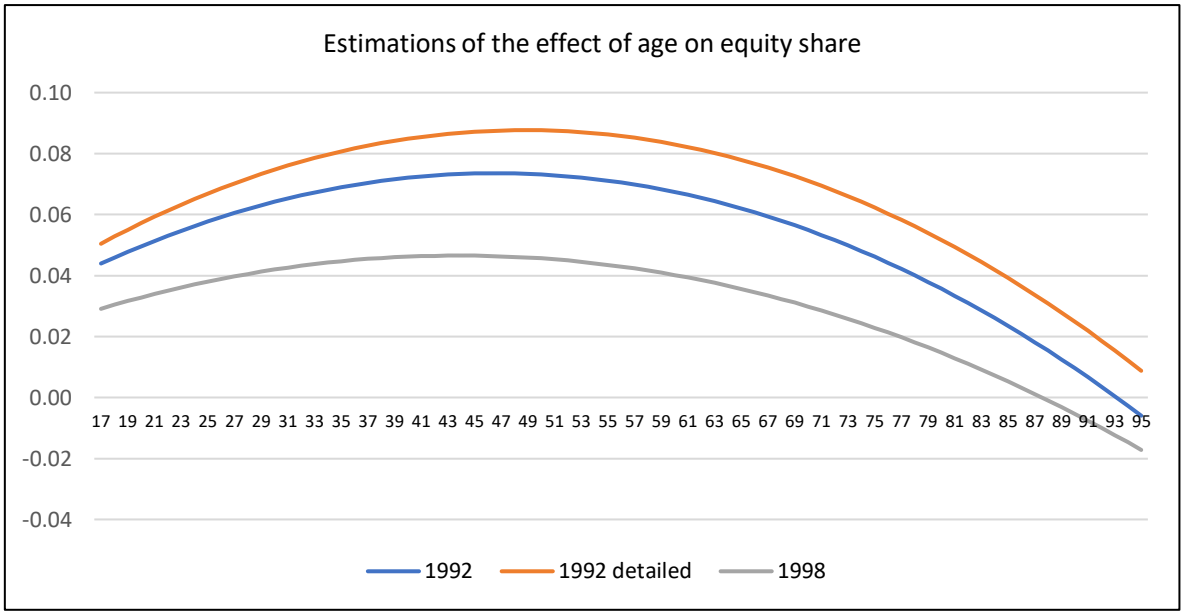
The presence of three or more children in the family is always significant in the models and has a positive effect. In other words, having many children instead of zero increases the investment in shares. The age coefficients individually do not seem to be significant, however by carrying out a Walt test in the regressions of tables 5.6, 5.7 the null hypothesis that they are simultaneously null is rejected. Regarding the other variables, the P-value for the dummies related to ethnicity are close to 0.05. Marital status and gender are significant, the latter only at the ten percent level.

5.2.2. Weighted analyses

The estimated effects will now be analysed from a quantitative point of view, refer to weighted analyses of tables 5.6, 5.7 and 5.8 for a quick view of the results.

Age

The effect of age on financial equity share shows a hump shape effect that, based on the estimations, peaks between 44 and 49 years old as shown in Graph 5.1. The result is consistent with the theory that human capital in youth should not be considered a risk-free asset, see for example Benzoni et al. (2007). Based on this consideration, young individuals would be already implicitly exposed to a risk comparable to the possession of shares. As a consequence of this theory, a hump-shape effect should be found in data. Minderhoud et al. (2011) give a theoretical framework that justifies the result obtained.



Graph 5.1: Estimate of the effect of age on the equity / net wealth ratio. for the three regressions in Tables 5.6, 5.7 and 5.8.

College degree

In the regression model a college degree generates an average increase of about 2.3 percentage points in stock. Although, due to different model specifications, it is difficult to make analogies between the results, Cupák et al. (2020) suggest that the effect varies with wealth.

Children

Based on the estimates made, having many children increases the share component compared to families without them. Even controlling for other factors, the percentage increase estimated with the coefficient is similar to that which occurs, on average, between the two groups in a descriptive analysis.

Marital status and female headed households

Being a single woman significantly decreases the average risky share component compared to single men by a value of approximately 4.15%. Similarly, Halko et al. (2012) find that, controlling for other variables, the residual effect of being male was 3%. A negative relationship exists also for married couples versus singles, even if the magnitude of the effect is smaller. Similar results were found empirically by Bertaut et al. (2000).

Ratio of business value, real estate investment and houses over net worth

Each of the ratios business/net worth, real-estate/net worth, houses/net worth reduces the risky assets share component. Approximately, it can be said that on average they cause a reduction of 10%, 5% and 3% respectively. The models control for the leverage ratio, as debt is deeply correlated with real estate investments and represents an important source of financial commitment, considering it could reduce a possible omitted variable bias. Other studies have shown that higher home value to wealth is associated with a lower exposition to stocks (Cho, 2014), or other real estate investment (Kullmann end Siegel, 2003).

Attitude toward risk

The dummies associated with the attitude toward risk are among the variables that have the greatest absolute effects on the independent variable. The result is widely predictable theoretically (see for example formula 1.1).

Internet

The use of the internet is not significantly linked to the financial risky share component held even in the weighted regression.

Retirement

The intention to save for retirement induces the growth of the equity component. The estimated effect is always greater than 1.5 percentage points. A positive relation was therefore found both in the probit models and in the conditional regression models; however, the estimation of these effects is approximate because it is reasonable to assume that they depend on specific characteristics of the individual such as age. In the paragraph 5.7 possible interaction terms will be considered.

Net worth

Consistent with the fact that holding equity investments involves fixed costs, net worth strongly influences shareholding. It is estimated that a doubling of net worth corresponds to a 2.5 percent point increase in equity share. A similar result was found by Bertaut (2000).

Table 5.6: linear regression models for equity shares, not including ethnicity and gender. Period: 1992-2019. Pooled data. One, two or three asterisks means that the predictor is significant at the ten, five or one percent level respectively. Models estimated considering data of all the five implicates through the Stata ado "micombine".

	Unweighted		Weighted	
	Coefficient	P-value	Coefficient	P-value
AGE	$1.09x10^{-3}$	0.228	$3.16x10^{-3}$***	0.003
AGE^2	$-1.29x10^{-5}$	0.102	$-3.39x10^{-5}$***	0.001
Education (reference: not have a degree)				
COLLEGE DEGREE	0.0304***	<0.001	0.0243***	<0.001
Children (reference: without children)				
ONE CHILD	-0.0031	0.622	-0.0044	0.603
TWO CHILDREN	0.0053	0.388	0.0003	0.973
THREE OR MORE CHILDREN	0.0197***	0.008	0.0164	0.115
Married (ref: not married/living with partner)				
MARRIED	-0.0139***	0.002	-0.0215***	<0.001
ln(NET WORTH)	0.0244***	<0.001	0.0241***	<0.001
ln(INCOME)	-0.0025	0.130	-0.0008	0.796
HOUSES/NET WORTH	-0.0387***	<0.001	-0.0308***	<0.001
REAL ESTATE/NET WORTH	-0.0772***	<0.001	-0.0528***	<0.001
BUSINESS/NET WORTH	-0.1260***	<0.001	-0.1009***	<0.001
DEBT/ASSET	0.1418***	<0.001	0.1197***	<0.001
Risk attitude (reference: medium financial risk)				
NO FINANCIAL RISK	-0.0721***	<0.001	-0.0494***	<0.001
SUBSTANTIAL FINANCIAL RISK	0.0580***	<0.001	0.0689***	<0.001
Saving reasons				
SAVING FOR RETIREMENT	0.0166***	<0.001	0.0172***	0.001
SAVING FOR EDUCATION	0.0062	0.385	0.0040	0.673
SAVING FOR HOME	-0.0316*	0.059	-0.0214	0.285
Survey year (reference: 1992)				
1995	0.0494***	<0.001	0.0658***	<0.001
1998	0.1255***	<0.001	0.1130***	<0.001
2001	0.1383***	<0.001	0.1462***	<0.001
2004	0.0832***	<0.001	0.0851***	<0.001
2007	0.0926***	<0.001	0.0801***	<0.001
2010	0.0543***	<0.001	0.0490***	<0.001
2013	0.0763***	<0.001	0.0656***	<0.001
2016	0.0699***	<0.001	0.0564***	<0.001
2019	0.0765***	<0.001	0.0558***	<0.001
constant	0.0892***	0.001	0.0334	0.395
R-squared (average over the five implicates)	0.07016		0.05332	
Sample size	N=28089		N=28089	

Table 5.7: linear regressions model for equity shares, including ethnicity and gender. Period: 1992-2019. Pooled data. One, two or three asterisks means that the predictor is significant at the ten, five or one percent level respectively. Models estimated considering data of all the five implicates through the Stata ado "micombine".

	Unweighted		Weighted	
	Coefficient	P-value	Coefficient	P-value
AGE	$1.38x10^{-3}$	0.133	$3.60x10^{-3}$***	0.001
AGE^2	$-1.49x10^{-5}$*	0.064	$-3.69\ x10^{-5}$***	<0.001
Education (reference: not have a degree)				
COLLEGE DEGREE	0.0305***	<0.001	0.0249***	<0.001
Children (reference: without children)				
ONE CHILD	-0.0002	0.974	-0.0002	0.977
TWO CHILDREN	0.0084	0.175	0.0046	0.564
THREE OR MORE CHILDREN	0.0228***	0.002	0.0208**	0.046
Married (ref: not married/living with partner)				
MARRIED	-0.0351***	<0.001	-0.0472***	<0.001
ln(NET WORTH)	0.0236***	<0.001	0.0229***	<0.001
ln(INCOME)	-0.0026	0.111	-0.0013	0.698
HOUSES/NET WORTH	-0.0390***	<0.001	-0.0309***	<0.001
REAL ESTATE/NET WORTH	-0.0773***	<0.001	-0.0525***	<0.001
BUSINESS/NET WORTH	-0.1265***	<0.001	-0.1020***	<0.001
DEBT/ASSET	0.1432***	<0.001	0.1210***	<0.001
Risk attitude (reference: medium financial risk)				
NO FINANCIAL RISK	-0.0691***	<0.001	-0.0469***	<0.001
SUBSTANTIAL FINANCIAL RISK	0.0573***	<0.001	0.0674***	<0.001
Saving reasons				
SAVING FOR RETIREMENT	0.0158***	<0.001	0.0164***	0.002
SAVING FOR EDUCATION	0.0064	0.372	0.0039	0.678
SAVING FOR HOME	-0.0311*	0.065	-0.0215	0.284
Sex (reference: male)				
WOMAN	-0.0381***	<0.001	-0.0415***	<0.001
Ethnicity (reference: white not Hispanic)				
BLACK/AFRICAN AMERICAN	-0.0179*	0.057	-0.0220**	0.029
HISPANIC	-0.0198**	0.047	-0.0111	0.323
OTHER ETHNICITY	-0.0175*	0.056	-0.0129	0.321
Survey year (reference: 1992)				
1995	0.0500***	<0.001	0.0668***	<0.001
1998	0.1261***	<0.001	0.1140***	<0.001
2001	0.1389***	<0.001	0.1473***	<0.001
2004	0.0840***	<0.001	0.0861***	<0.001
2007	0.0936***	<0.001	0.0815***	<0.001
2010	0.0550***	<0.001	0.0499***	<0.001
2013	0.0774***	<0.001	0.0670***	<0.001
2016	0.0715***	<0.001	0.0577***	<0.001
2019	0.0781***	<0.001	0.0572***	<0.001
constant	0.1131***	<0.001	0.0633	0.111
R-squared (average over the five implicates)	0.0717		0.05546	
Sample size	N=28089		N=28089	

Table 5.8: linear regression model for equity shares, including data on internet usage available since 1998. Period: 1998-2019. Pooled data. One, two or three asterisks means that the predictor is significant at the ten, five or one percent level respectively. Models estimated considering data of all the five implicates through the Stata ado "micombine".

	Unweighted		Weighted	
	Coefficient	P-value	Coefficient	P-value
AGE	$5.49x10^{-4}$	0.581	$2.13x10^{-3}$*	0.072
AGE^2	$-7.59x10^{-6}$	0.393	$-2.43x10^{-5}$**	0.026
Education (reference: not have a degree)				
COLLEGE DEGREE	0.0289***	<0.001	0.0221***	<0.001
Children (reference: without children)				
ONE CHILD	-0.0011	0.870	-0.0006	0.948
TWO CHILDREN	0.0074	0.243	0.0018	0.832
THREE OR MORE CHILDREN	0.0229***	0.003	0.0225**	0.033
Married (ref: not married/living with partner)				
MARRIED	-0.0135***	0.005	-0.0200***	0.001
ln(NET WORTH)	0.0270***	<0.001	0.0277***	<0.001
ln(INCOME)	-0.0043**	0.015	-0.0036	0.305
HOUSES/NET WORTH	-0.0417***	<0.001	-0.0331***	<0.001
REAL ESTATE/NET WORTH	-0.0727***	<0.001	-0.0533***	<0.001
BUSINESS/NET WORTH	-0.1307***	<0.001	-0.1068***	<0.001
DEBT/ASSET	0.1367***	<0.001	0.1201***	<0.001
Risk attitude (reference: medium financial risk)				
NO FINANCIAL RISK	-0.0715***	<0.001	-0.0500***	<0.001
SUBSTANTIAL FINANCIAL RISK	0.0615***	<0.001	0.0698***	<0.001
Saving reasons				
SAVING FOR RETIREMENT	0.0145***	<0.001	0.0160***	0.004
SAVING FOR EDUCATION	0.0016	0.844	0.0022	0.837
SAVING FOR HOME	-0.0381**	0.044	-0.0289	0.216
Internet usage (reference: not used)				
INTERNET	0.0094	0.119	0.0033	0.647
Survey year (reference: 1998)				
2001	0.0106	0.194	0.0321***	0.003
2004	-0.0456***	<0.001	-0.0294***	0.009
2007	-0.0382***	<0.001	-0.0350***	0.004
2010	-0.0763***	<0.001	-0.0661***	<0.001
2013	-0.0547***	<0.001	-0.0496***	<0.001
2016	-0.0616***	<0.001	-0.0591***	<0.001
2019	-0.0556***	<0.001	-0.0600***	<0.001
constant	0.2134***	<0.001	0.1583***	<0.001
R-squared (average over the five implicates)	0.06982		0.05066	
Sample size	N=23937		N=23937	

Table 5.9: coefficient significance at five percent level, comparison among the different regression model specifications. Legend: “ns” means not significant (blue background), “s” means significant, “-“ means coefficient not available.

	1992 Unweighted		1992 Weighted		1998 Unweighted	1998 Weighted
AGE	ns	ns	s	s	ns	ns
AGE^2	ns	ns	s	s	ns	s
COLLEGE DEGREE	s	s	s	s	s	s
ONE CHILD	ns	ns	ns	ns	ns	ns
TWO CHILDREN	ns	ns	ns	ns	ns	ns
THREE OR MORE CHILDREN	s	s	ns	s	s	s
MARRIED	s	s	s	s	s	s
ln(NET WORTH)	s	s	s	s	s	s
ln(INCOME)	ns	ns	ns	ns	s	ns
HOUSES/NET WORTH	s	s	s	s	s	s
REAL ESTATE/NET WORTH	s	s	s	s	s	s
BUSINESS/NET WORTH	s	s	s	s	s	s
DEBT/ASSET	s	s	s	s	s	s
NO FINANCIAL RISK	s	s	s	s	s	s
SUBSTANTIAL FINANCIAL RISK	s	s	s	s	s	s
SAVING FOR RETIREMENT	s	s	s	s	s	s
SAVING FOR EDUCATION	ns	ns	ns	ns	ns	ns
SAVING FOR HOME	ns	ns	ns	ns	s	ns
WOMAN	-	s	-	s	-	-
BLACK/AFRICAN AMERICAN	-	ns	-	s	-	-
HISPANIC	-	s	-	ns	-	-
OTHER ETHNICITY	-	ns	-	ns	-	-
INTERNET	-	-	-	-	ns	ns

5.3. Limitations of the analyses in 5.1 and 5.2

Chapters 5.1 and 5.2 present two types of models that consider the main predictors that have been associated with household equity choices. However, these models have limitations. In addition to the possible phenomena of endogeneity widely discussed, it should be noted that in order to carry out accurate analyses on the SCF surveys, it is necessary to consider the bootstrap weights to keep the sampling variability error in consideration.

The attentive reader will not have failed to note that the dummies inherent to the years of investigation are generally significant (tables 5.1, 5.2, 5.6, 5.7 and 5.8). With regard to the linear regressions, this fact can be closely linked to the performance of shares (a reduction in shares value causes a reduction in the equity / financial assets ratio if no rebalancing action is taken). For the probit models this argument is not so strong. it is therefore useful to compare the estimates of the coefficients of the variables, calculated on the single year, to observe if there has been an evolution of their effects over time.

To deal with the last two issues the multivariate models in 5.4 and 5.5 were developed.

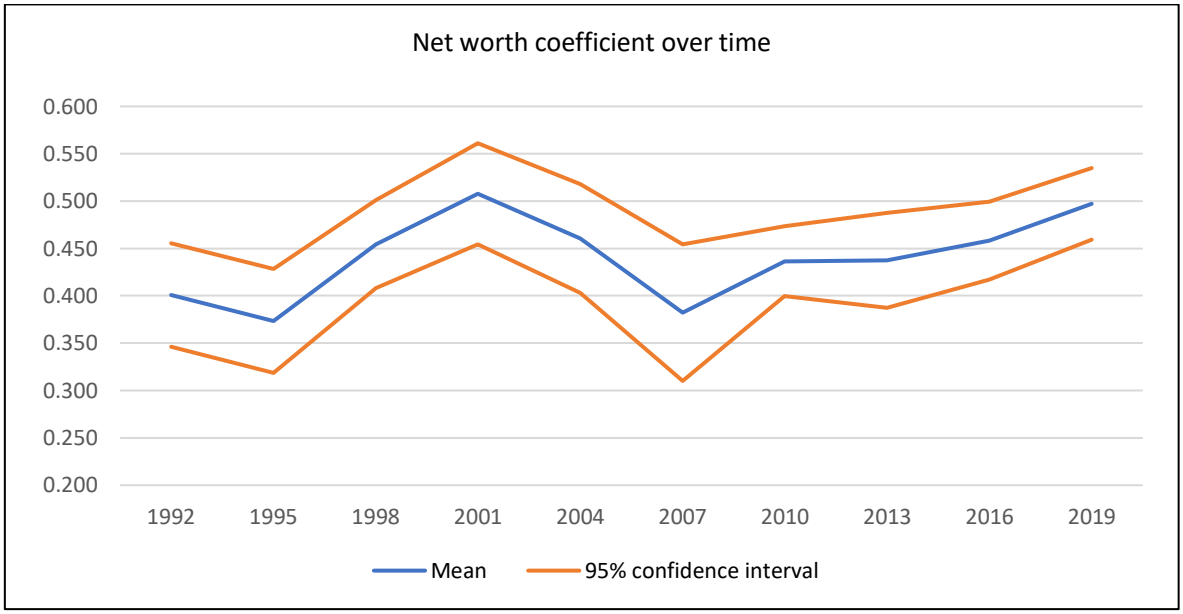
Another possible problem with the estimated coefficients is the absence of interaction terms in the analyses. For example, it is reasonable to assume that the effect on participation of “saving for retirement” changes as wealth or age varies. There is interest to further study the variable saving for retirement, so in 5.7 additional analyses will be shown.

5.4. Probit models: decision to invest in equity, bootstrap weights

This sub-chapter contains a brief analysis of the results that emerge from the probit regressions with bootstrap weights performed on the single years. The results are available in Annex 4.

The outputs in the annex do not show excessive differences with respect to what is claimed in 5.1. College degree, net worth, ratios, and risk aversion show coefficients that agree in sign with those previously estimated and are significant in the majority of cases. Other variables such as ethnicity, gender and number of children no longer have the constant significance that they had previously. The reason is due to the reduction in the sample size and the fact that sampling variability has now been taken into account.

Analysing the coefficient related to net worth, it emerges that, by comparing the years 1992-2016 and 1992-2019, it has increased. This evidence is also confirmed by a probit with pool data and terms of interaction between net wealth and year: i.e., the null hypothesis that there was no change in the effect of net worth is rejected in favour of an increase in the influence of net wealth for participation. This result is in some ways a puzzle because, on basis of a reduction in transaction and information costs, one would expect less influence of wealth under a framework of rational investors. In any case, these annual comparisons for net worth should be cautiously considered. A comparison between 2001 and 2019 does not allow the same conclusion. Analysing graph 5.2, it seems that the trend could be related to economic cycles.

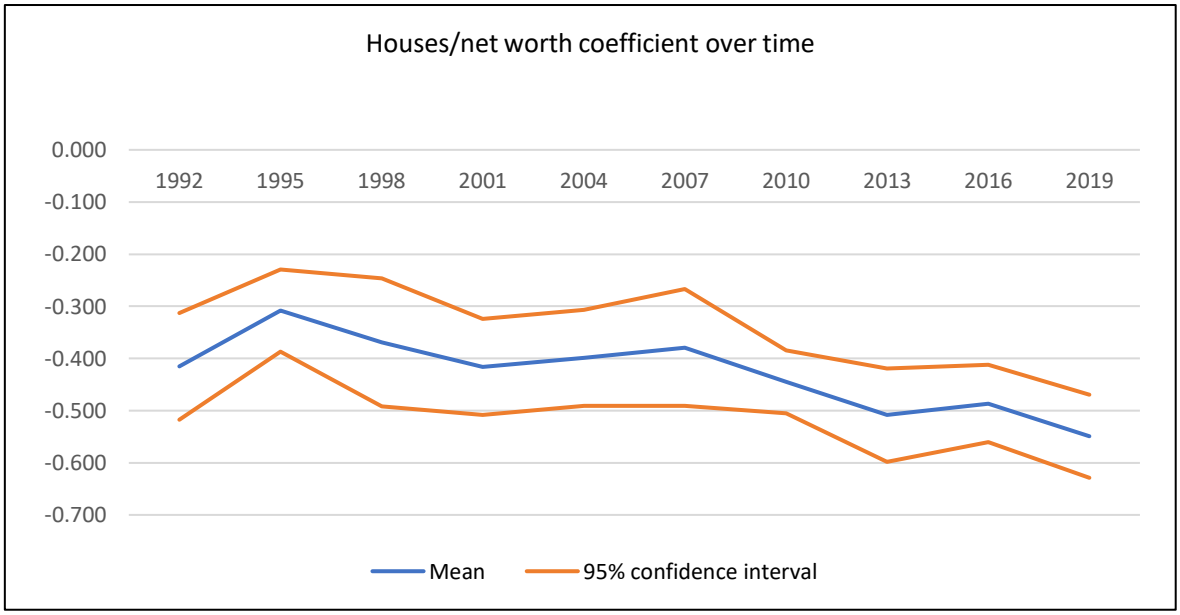


Graph 5.2: beta and confidence interval of Net worth. Probit results with bootstrap weights.

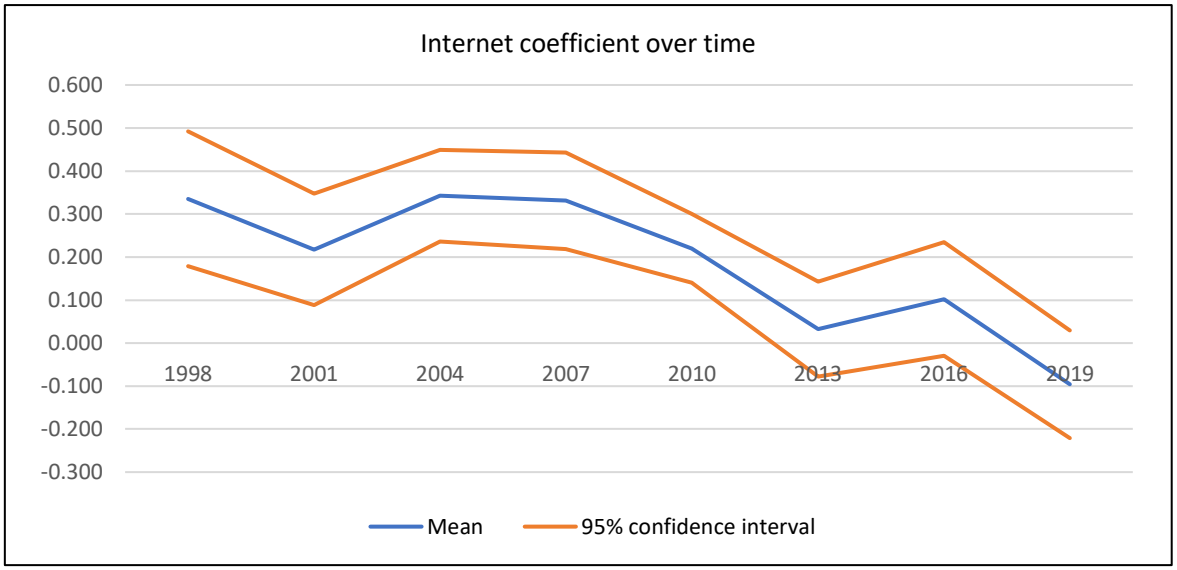
Analogous year comparisons (1992-2016 and 1992-2019) show that the coefficient of ratio Houses/Net worth decreases (Graph 5.3). In a pooled analysis this change is significant at the 10 percent level.

The internet variable, which was significant until 2010, has shown a decrease in the effect on participation over time (Graph 5.4), so much so that in 2019 the coefficient was negative. Therefore, despite the initial difference between users and non-users of internet, the effect has gradually diminished. The reasons for this phenomenon could be many:

- it is conceivable that initially the users were people with different characteristics compared to the general population and more inclined to shareholding. It cannot be excluded that the variables of the model are not able to control these characteristics; for example, the degree variable measures the level of education, but it does not distinguish possible specializations.
- The financial investments sector has undergone major changes over the past 30 years. While broker-assisted services were widespread in the past, today most transactions take place with online banking or online brokers.
- The old intermediation systems have had to evolve in order to remain competitive, because as the adoption and trust of low-cost systems has grown, the differential in transaction costs has decreased.
- New investment systems have gradually been introduced, such as ETFs.



Graph 5.3: coefficient and confidence interval of Houses/Net worth. Probit results with bootstrap weights.

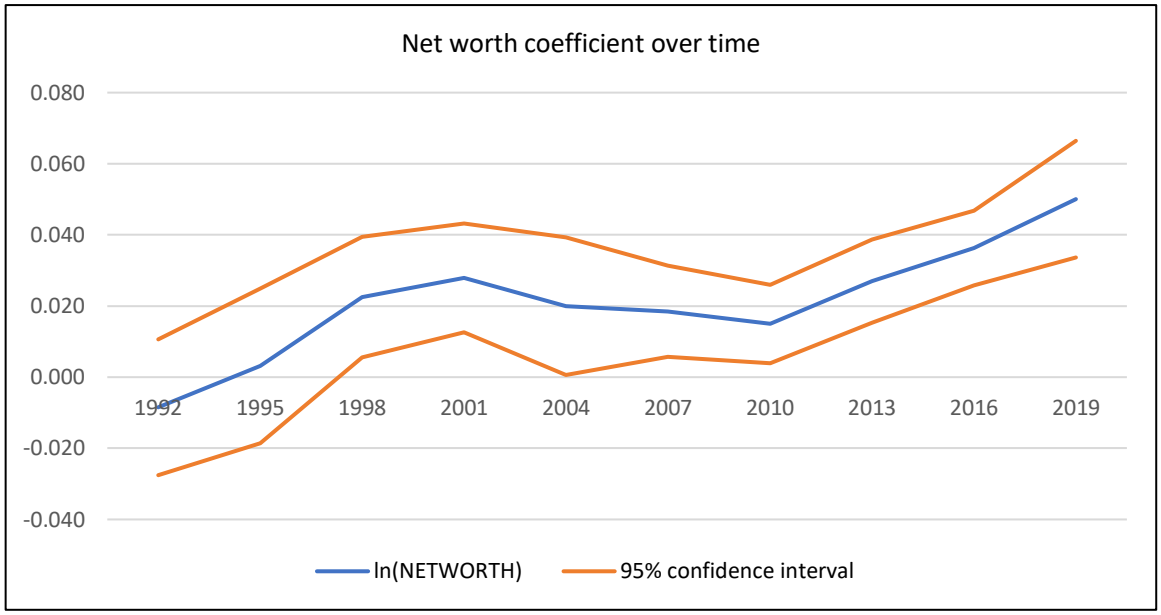


Graph 5.4: coefficient and confidence interval of Internet. Probit results with bootstrap weights.

5.5. Linear regressions: share of equity in the financial portfolio, bootstrap weights

As in the previous sub-chapter, the results that emerge from the linear regressions with bootstrap weights are available in Annex 5.

Altogether, in the case of these regressions, there was a reduction in the significance of the predictors. For no variable the null hypothesis that the coefficient is zero can always be rejected. However, in eight out of ten cases Business/Net worth, Net worth, Hight and Substantial financial risk were significant, while Married and Houses/net worth in at least six out of ten. Due to the great heterogeneity of households' choices and the small sample size further reduced compared to the probits (many families are not participants), it is believed that the results provided in paragraph 5.2 are overall more interesting. Nevertheless, for the variables that in the previous analyses proved to be significant, the effects on average are consistent with those already provided. In any case, it is interesting to note the temporal evolution of the coefficient relating to net worth (Graph 5.5). Over time it has shown an evident growing trend. Although it is not possible to prove it, the graph together with the evolution of the equity indices suggests that investors with participations in the financial markets have become increasingly wealthier due to their share holdings, thus increasing the risky component in their portfolios.



Graph 5.5: coefficient and confidence interval of Net Worth. Linear regression results with bootstrap weights.

5.6. Considerations on the use of bootstrap weights

Overall, as found by Hanna and Shin (2017), the z-values associated with the coefficients calculated with the population and bootstrap weights were generally higher than those obtained only with the population weights. Consequently, the P-values using bootstrap weights are lower. It follows that if a coefficient is significant in the case with population weight only, it will also be significant in the other case. Therefore, at least with regards to the study carried out, testing the significance of the coefficients by not using bootstrap weights is generally more conservative.

5.7. Interaction terms: saving for retirement

Both in a probit and in a linear regression model with interaction dummies, the hypothesis that the effects related to saving for retirement are the same between 1992 and the other years cannot often be rejected. Considering the aforementioned fact, possible terms of interaction between saving for retirement and other variables were analysed in pooled regressions. The outputs related to the probit and linear regression analyses are shown in Annexes 6 and 7 respectively.

5.7.1. Effects on participation

The results obtained can be summarized as follows:

- the effect on participation of the dummy saving for retirement remains significant and positive.
- Although the interaction with age is significant only at the level of 10 percent, it reduces the positive effect mentioned above. The result seems reasonable: for retired people the savings should decrease or even be eliminated to compensate for the lack of working salary.
- The term of interaction between income and saving for retirement is significant. Increasing income reduces the positive effect of the retirement predictor; but, without considering the other interaction factors, it would be necessary an income higher than a million dollars to overcome the positive effect on participation of "saving for retirement".
- Growth in net worth increases the predictor effect, but the term of interaction is not statistically significant.

5.7.2. Effects on the financial risky share component

The effects on the fraction of financial capital dedicated to equity of saving for retirement, considering interaction terms, are the following:

- When considered individually, the predictor "saving for retirement" is not significant. The same is also true for its interaction with income and net worth.
- As it is reasonable to expect, with aging, the positive effect on the risky component associated to retirement decreases. The term of interaction between age and saving for retirement is significant.

Checking for additional interaction variables such as college degree, gender, or marital status showed no more interesting results than those already stated.

6. Conclusion and implications

This research work studies the determinants of shareholding participation and fraction of equity over the total financial capital of American households. Using a set of cross-sectional data covering the time span 1992-2019 in a setting of non-strictly exogenous explanatory variables, it was possible to consider the effects of the predictors in the long term. All the previous studies analysed were based on shorter time intervals.

The results show that net worth positively influences both dependent variables studied. Interestingly, the positive effects generated by the financial endowment did not decrease, but rather increased over time. In the literature an attempt to rationalize the limited participation has been made with the presence of transactions and information cost. On the basis of these arguments, a reduction of equity holdings costs should lead to the entry into the financial market of less wealthy individuals. The data does not seem to suggest this trend: between 1992 and 2019 online-brokers entered the market (leading to a dramatic reduction of transaction cost) and tools such as ETFs (that reduce information costs because they are based on diversification) became widespread. Nevertheless, equity participation in the first, second, third and fourth wealth quartiles has varied by approximately 0%, 11%, 16% and 20% respectively.

The research confirms how variables previously studied and linked to risky investments maintain their effects even using recent data. Risk aversion and ethnicity are negatively linked to both participation and risk profile, while the opposite is true for having a degree. By checking for the level of indebtedness, it has been shown that investments in primary residence, real estate and private businesses reduce average participation and exposure to financial risk. In particular, the greatest effect was found for the possession of businesses which can be considered as a substitute for stocks. Finally, in line with other empirical studies, age has a hump-shaped effect on the risky component of the portfolio.

The introduction of internet has brought about great changes in the financial markets such as the reduction of transaction costs and the possibility of having real time data. Internet has also allowed the birth of new operating methods such as high frequency trading. In the empirical models analysed, by introducing a variable that measured households' use of Internet to make choices related to the economic-financial sphere, it was observed that the families that used it were more likely to participate in the stock markets. However, this

effect has waned over time and since 2013 it has not been statistically significant. Regarding the influence of internet use in determining the overall investment profile in stocks, it was not found to be statistically significant.

Finally, particular attention was paid to the study of the variable that measures if the households save for retirement and if the effects of this predictor depend also on other household characteristics. It has been shown that this variable induces an increase in the probability of participation, however this positive relationship is negatively influenced by age and income. With regard to the implications on the fraction of equity investments, the variable saving for retirement, alone, is no longer significant in a regression with terms of interaction.

Overall, it is possible to state that the study carried out allows to predict with a reasonable degree of accuracy whether an American family holds shares. On the contrary, allocation choices show greater heterogeneity; indeed, investment decisions can only partially be explained through the characteristics analysed. Many other factors not studied in this research play a significant role in portfolio choices such as family background, past experiences, sociological and historical reasons, financial advice. Being able to incorporate all these factors into a model could greatly enhance our understanding of financial allocation behaviours.

Implications

A first general consideration that can be extrapolated from the study is that the commitment to a long-term goal such as saving for retirement induces behaviour more in line with theoretical models. This consideration can take on great importance because, considering the historical returns of the financial market, failure to participate in it involves opportunity costs reasonably not understood by part of the population. Especially in countries with raising concerns about the stability of the pension system and with low equity and bond participation, fiscally facilitating private pension plans could improve households' financial returns, increase investments diversification and reduce home bias. Finally, some conclusions can be drawn for financial planners, as already highlighted by Hanna et al. (2008). Knowledge of the household's exposure to non-financial investments such as real estate and business can help in determining the optimal investment profile and eventually warn the customer about the risks associated with under-diversification.

Limitation and suggestion for future research

The study carried out could be affected by endogeneity, in particular simultaneous causality effects may not be negligible. To correct these problems, it would be useful to study a model that jointly estimates investment decisions in stocks, real estate, businesses and other financial assets. The paper focused on overall equity holdings, a separate study for each typology of possible participation methodology (direct stock holdings, investment funds and retirement accounts) could yield interesting results.

References

Agnew, J., Balduzzi P. & Sundén A. (2003). Portfolio Choice and Trading in a Large 401(k) Plan. American Economic Review, 93 (1): 193-215.

Aizcorbe, A.M., Kennickell, A.B. & Moore, K.B. (2003). Recent changes in US family finances: evidence from 1998 and 2001 survey of consumer finances. Federal Reserve Bulletin 89, 1–32.

Allison, P.D. (1999). Multiple Regression: A Primer. Thousand Oaks, CA: Pine Forge Press. p. 142.

Ameriks, J. & Zeldes, S. (2001). How Do Household Portfolio Shares Vary with Age? Working Paper, Columbia University, Graduate School of Business.

Barber, B.M. & Odean, T. (2001). Boys will be boys: gender, overconfidence, and common stock investment. Quarterly Economic Review 116, 261–292.

Benzoni, L., Collin-Dufresne, P. & Goldstein, R.S. (2007). Portfolio Choice over the Life-Cycle when the Stock and Labour Markets are Cointegrated. Journal of Finance, 52, 2123-2168.

Bertaut, C. & Starr-McCluer, M. (2000). Household portfolios in the United States. Finance and Economics Discussion Series 2000-26, Board of Governors of the Federal Reserve System (U.S.).

Bertaut, C. (1998). Stockholding Behaviour of U.S. Households: Evidence From the 1983-1989 Survey of Consumer Finances, Review of Economics and Statistics. 80, 263-275.

Bodie, Z. & Crane, D.B. (1997). Personal Investing: Advice, Theory, and Evidence from a Survey of Tiaa-Cref Participants (May 28, 1997).

Bogan, V. (2008). Stock Market Participation and the Internet. Journal of Financial and Quantitative Analysis, 43(1):191–211.

Calvet L.E., Campbell, J.Y. & Sodini, P. (2009). Fight or Flight? Portfolio Rebalancing by Individual Investors, The Quarterly Journal of Economics, Volume 124, Issue 1, February 2009, Pages 301–348,

Calvet L.E., Campbell, J.Y. & Sodini, P. (2009). Measuring the Financial Sophistication of Households. American Economic Review, 99 (2): 393-98.

Campbell, J.Y. (2006). Household finance. Journal of Finance, 61(4), 1553-1604.

Cho, I. (2014). Homeownership and investment in risky assets in Europe. Review of European Studies, 6(4), 254-267.

Cocco, J.F. (2005). Portfolio choice in the presence of housing. The Review of Financial Studies, 18(2), 535-567.

Coleman, S. (2003). Academy of Accounting and Financial Studies Journal; Arden Vol. 7, Fasc. 2: 99-114.

Cupák, A., Fessler, P., Hsu, J.W. & Paradowski, P.R., (2020). Confidence, Financial Literacy and Investment in Risky Assets: Evidence from the Survey of Consumer Finances (January 16, 2020).

Darrell, D. & Zame, W. (1989). The Consumption-Based Capital Asset Pricing Model. Econometrica, vol. 57, no. 6, pp. 1279–97.

Deaton, A. (2019). The Analysis of Household Surveys: A Microeconometric Approach to Development Policy. 2019. Washington, DC: World Bank. © World Bank.

Efron, B. & Tibshirani, R.J. (1994). An Introduction to the Bootstrap, Taylor & Francis Ltd.

Fratantoni, M.C. (1998). Homeownership and Investment in Risky Assets, Journal of Urban Economics, Elsevier, vol. 44(1), pages 27-42, July.

Georgarakos, D. & Inderst, R. (2014). Financial Advice and Stock Market Participation, May.

Guiso, L. & Sodini, P. (2013). Household finance: An emerging field. In Handbook of the Economics of Finance, pages 1397–1532. Elsevier.

Guiso, L., Jappelli, T. & Haliassos, M. (2003). Household Stockholding in Europe: Where Do We Stand, and Where Do We Go? (January 2003).

Halko, M.L., Kaustia, M. & Alanko, E. (2012). The gender effect in risky asset holdings. Journal of Economic Behavior & Organization, Elsevier, vol. 83(1), pages 66-81.

Hanna, S.D. & Shin, S.H. (2017). Accounting for Complex Sample Designs in Analyses of the Survey of Consumer Finances. Journal of Consumer Affairs 50(2).

Hanna, S.D., Wang, C. & Lindamood, S. (2008). Household Ownership of Stocks, Business Assets, and Investment Real Estate: An Endogenous Probit Model.

Hvide, H., Meling, T., Mogstad, M. & Vestad, O.L. (2022). Broadband Internet and the Stock Market Investments of Individual Investors (August 23, 2022). University of Chicago, Becker Friedman Institute for Economics Working Paper No. 117, 2022.

Jeffrey R. Brown, Zoran Ivković, Paul A. Smith, Scott Weisbenner (2008). Neighbours Matter: Causal Community Effects and Stock Market Participation. The Journal of Finance, vol 63, Pages 1509-1531.

Jianakoplos, N.A. & Bernasek, A. (1998). Are women more risk averse? Economic Inquiry, 36(4), 620-630.

Kennickell, A. B., Douglas A.M. & Woodburn R.L. (1996). Weighting design for the 1992 Survey of Consumer Finances. Board of Governors of the Federal Reserve System.

Kennickell, A.B. & Woodburn R.L. (1999). Consistent weight design for the 1989, 1992 and 1995 scfs, and the distribution of wealth. Review of Income and Wealth, International Association for Research in Income and Wealth, vol. 45(2), pages 193-215, June.

Kullmann, C. & Siegel, S. (2005). Real Estate and its Role in Household Portfolio Choice (November 1, 2005). EFA 2003 Annual Conference Paper No. 918, Sauder School of Business.

Lindamood, S., Hanna, S.D. & Bi, L. (2007), Using the Survey of Consumer Finances: Methodological Considerations and Issues. Journal of Consumer Affairs.

Merton, R.C. (1969). Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case. The Review of Economics and Statistics, vol. 51, no. 3, pp. 247–57.

Merton, R.C. (1971). Optimum Consumption and Portfolio Rules in a Continuous Time Model. Journal of Economic Theory, 3, 373-413.

Minderhoud, I., Molenaar, R. & Ponds, E.H.M. (2011). The Impact of Human Capital on Life-Cycle Portfolio Choice: Evidence for the Netherlands (February 17, 2011).

Montalto, C.P. & Sung, J. (1996). Multiple Imputation in the 1992 Survey of Consumer Finances. Journal of Financial Counseling and Planning, 7.

Pence, K. (2015). SCFCOMBO: Stata module to estimate errors using the Survey of Consumer Finances. Statistical Software Components S458017, Boston College Department of Economics, revised 04 Oct 2015.

Shin, S.H, & Kyoung Tae, K. (2018). Income Uncertainty and Household Stock Ownership During the Great Recession. Journal of Financial Counseling and Planning; Columbus Vol. 29, Fasc. 2: 383-395.

Shum, P. & Faig, M. (2006). What explains household stock holdings? Journal of Banking and Finance, 30, 2579-2597.

Winship C. & Radbill L. (1994). Sampling Weights and Regression Analysis. Sociological Methods & Research. 23(2):230-257.

Ye, Y., Pu Y. & Xiong A. (2022). The impact of digital finance on household participation in risky financial markets: Evidence-based study from China.

Web resources:

https://data.worldbank.org/indicator/SI.POV.GINI?locations=US

https://www.ilsole24ore.com/art/fondi-pensione-rendimenti-piu-elevati-tfr-si-arriva-fino-all-11percento-AEVDY1eB

https://www.federalreserve.gov/econres/files/codebk2019.txt

https://www.federalreserve.gov/econres/files/bulletin.macro.txt

Link to the Stata ado files needed to make statistical inference:

micrombine: http://fmwww.bc.edu/RePEc/bocode/m/micombine.html

scfcombo: http://fmwww.bc.edu/RePEc/bocode/s/scfcombo.ado

List of acronyms

CCAPM: Consumption-based Capital Asset Pricing model

EH: Equity Holder

ETF: Exchange Traded Funds

FED: Federal Reserve System

HNWI: High Net Worth Individual

IRA: Individual Retirement Account

NEH: Not Equity Holder

PEU: Primary Economic Unit

RII: Repeated-Imputation Inference

SCF: Survey of Consumer Finances

Annex 1: definition of variables

AGE: age of the reference person.

AGE^2: squared age of the reference person.

ASSET: sum of financial assets and nonfinancial assets held by household*[a].

BLACK/AFRICAN AMERICAN: dummy variable equal to one if the ethnicity of respondent person was black or African American, zero otherwise.

BUSINESS: total value of business(es) in which the household has either an active or nonactive interest. Businesses include both actively and not actively managed business(es). No financial items enter in the computation of this variable.

Value of active business(es) calculated as net equity if business(es) were sold today, plus loans from the household to the business(es), minus loans from the business(es) to the household not previously reported, plus value of personal assets used as collateral for business(es) loans that were reported earlier.

Value of nonactive business(es) is calculated as the market value of the business(es)*.

COLLEGE DEGREE: dummy variable equal to one if the reference person has a college degree, zero otherwise.

COMUTF represent combination mutual funds also called balanced mutual funds, such funds invest in a combination of bond and equity. For the purpose of the analysis they will be considered as composed by 50% of stocks and 50% of bonds. This assumption is necessary since in the codebooks of SCF there is no mention about their compositions which is not investigated on details in the SCF survey.

DEBT/ASSET: ratio between DEBT and ASSET, observations showing a null denominator were discarded. Refer to the definition of ASSET and DEBT for further clarifications.

DEBT: total value of debt held by household, includes principal residence debt (mortgages and HELOCs), other line of credit, debt for other residential property, credit card debt, installment loans, and other debt*.

EQUITY: total value of financial assets held by household that are invested in stock. Includes:

1. directly held stock
2. stock mutual funds: full value if described as stock mutual fund, 1/2 value of combination mutual funds
3. IRAs/Keoghs invested in stock: full value if mostly invested in stock, 1/2 value if split between stocks/bonds or stocks/money market, 1/3 value if split between stocks/bonds/money market
4. other managed assets w/equity interest (annuities, trusts, MIAs): full value if mostly invested in stock, 1/2 value if split between stocks/MFs & bonds/CDs, or "mixed/diversified," 1/3 value if "other"
5. thrift-type retirement accounts invested in stock full value if mostly invested in stock 1/2 value if split between stocks and interest earning assets.
6. NOTE: The allocation rules for mixed investments in 3), 4), and 5) do not apply to 2004 since new questions in 2004 directly ask the share of stock in those assets*.

EQUITY PARTICIPANT: a household is defined to be a participant in stock market if he/she holds an amount greater than zero in equity, for further clarification refer to EQUITY. This is the dependent variable of probit models. If the household is a participant the dummy variable is one, zero otherwise.

EQUITY/FINANCIAL: ratio of EQUITY and FINANCIAL. Refer to the definition of EQUITY and FINANCIAL for further clarifications. This is the dependent variable of linear regression models.

FINANCIAL (FIN): total value of financial assets held by household. Consists of liquid assets, certificates of deposit, directly held pooled investment funds, stocks, bonds, quasi-liquid assets, savings bonds, whole life insurance, other managed assets, and other financial assets. See the definition of each asset for further details*.

HISPANIC: dummy variable equal to one if ethnicity of respondent was Hispanic, zero otherwise.

HOUSES/NET WORTH: ratio between HOUSES and NET WORTH, observations showing a null denominator were discarded. Refer to the definition of NET WORTH and HOUSES for further clarifications.

HOUSES: total value of primary residence. Excludes the part of a farm or ranch used in a farming or ranching business*.

INCOME: household income for previous calendar year. Includes des wages, self-employment and business income, taxable and tax-exempt interest, dividends, realized capital gains, food stamps and other support programs provided by the government, pension income and withdrawals from retirement accounts, Social Security income, alimony and other support payments, and miscellaneous sources of income*. If the income was less than $100 (income dominium is $[0; \infty)$) by convention it was set to $ 100, altogether out of a sample of 50410, 392 household show an income less than $100 in implicate 1.

INTERNET: dummy variable equal to one if the household have used internet or online services:

1. to make decisions about borrowing or credit
2. or to make decisions about saving and investments
3. or to do business with institutions (institutions were mentioned by the respondent, it could be for example a commercial bank); this question was removed since 2016
4. to use online banking services, this question has been introduced since 2016.

IRAKH: total value of IRA/Keogh accounts*.

Ln(INCOME): natural logarithm of INCOME. Refer to the definition of INCOME for further clarifications.

ln(NET WORTH): natural logarithm of NET WORTH. Refer to the definition of NET WORTH for further clarifications.

MARRIED: dummy variable that account for the marital status of the reference person, equal to one if he/she is married or living with partner, zero otherwise.

NET WORTH (NW): The difference between assets and debt. Refer to the definition of ASSET and DEBT for further clarifications[a].

NNRESRE: total value of net equity in non-residential real estate held by household*.

NO FINANCIAL RISK: dummy variable equal to one if the respondent report that is not willing to take financial risk, zero otherwise.

OMUTF: value of investment in mutual funds that can not lead back to other categories. A codebook note said that they consist almost entirely of hedge funds and only a small amount is of exchange traded funds. Not available for 1989, 1992, 1995, 1998, and 2001.

ONE CHILD: dummy variable equal to one if there is one child in the household, zero otherwise. Includes natural children, step-children, and foster children of household reference person or spouse/partner.

ORESRE: total value of other residential real estate held by household, includes land contracts/notes owed to the household and properties other than the principal residence, including 1-4 family residences, time shares, and vacations homes*.

OTHER ETHNICITY: dummy variable equal to one if ethnicity of respondent was not white, Hispanic, black or African American, zero otherwise.

REAL ESTATE /NET WORTH: ratio between BUSINESS and NET WORTH, observations showing a null denominator were discarded. Refer to the definition of NET WORTH and BUSINESS for further clarifications.

REAL ESTATE/NET WORTH: ratio between REAL ESTATE and NET WORTH, observations showing a null denominator were discarded. Refer to the definition of NET WORTH and REAL ESTATE for further clarifications.

REAL ESTATE: this variable equal the sum of ORESRE and NNRESRE. Refer to the definition of ORESRE and NNRESRE for further clarifications.

RETEQ: total value of equity in quasi-liquid retirement assets. Includes the total value of the following: IRAs, Keoghs, and thrift-type plans invested in stocks or stock mutual funds.*

SAVING FOR EDUCATION: dummy variable equal to one if the respondent report that children's education, education of grandchildren or own education, spouse/partner's education or in general education (not known for whom) was one of his/her most important reasons for saving, zero otherwise.

SAVING FOR HOME: dummy variable equal to one if the respondent report that buying own house (primary residence not second home) was one of his/her most important reasons for saving, zero otherwise.

SAVING FOR RETIREMENT: dummy variable equal to one if the respondent report that retirement, old age or funeral expenses was one of his/her most important reasons for saving, zero otherwise.

STMUTF: value of stock mutual funds held by the household.

STOCKS: total value of the stocks directly own by the household.

SUBSTANTIAL FINANCIAL RISK: dummy variable equal to one if the respondent report that is willing to take substantial financial risks expecting to earn substantial returns, zero otherwise.

THREE OR MORE CHILDREN: dummy variable equal to one if there are more then two children in the household, zero otherwise. Includes natural children, step-children, and foster children of household reference person or spouse/partner.

THRIFT: total value of account-type pension plan from respondent and spouses's current job*.

TWO CHILDREN: dummy variable equal to one if there are two children in the household, zero otherwise. Includes natural children, step-children, and foster children of household reference person or spouse/partner.

WOMAN: dummy variable equal to one if the gender of the reference person was female, zero otherwise.

* The variables indicated with an asterisk are those used in the Federal Reserve Bulletin article, the definitions are provided by the University of California, Berkeley

[a] For a useful representation of household asset, debt, financial and not financial asset refers to the following link: https://www.federalreserve.gov/econres/files/Networth%20Flowchart.pdf

Annex 2: VIF tables

VIF is the acronym of variance inflation factor, the VIF of a parameter is the ratio between the variance of a model that includes multiple independent parameters and the variance of model constructed using only the parameter object of interest. It quantifies the severity of multicollinearity in an ordinary least square regression analysis, a VIF higher that 10 implies a severe multicollinearity, the VIFs values for the first six regression models estimated in section "share of equity in financial portfolio" are available in this annex, based on this indicator the model is not affected by severe multicollinearity problems; clearly the hight VIF value for Age and Age^2 could be ignored because is due to the quadratic relationship between them.

1992 Unweighted	VIF
AGE	44.54
AGE2	43.36
ln(NET WORTH)	5.14
DEBT/ASSET	3.61
ln(INCOME)	2.65
HOUSES/NW	2.56
a2016	2.53
a2019	2.46
a2010	2.46
a2013	2.39
a2007	2.23
a2001	2.21
a2004	2.2
a1998	2.11
a1995	2.01
BUSINESS/NW	1.5
TWO CHILDREN	1.37
THREE + CHILDREN	1.29
COLLEGE DEGREE	1.25
MARRIED	1.22
SAVING EDUCATION	1.2
ONE CHILD	1.19
SAVING RETIREMENT	1.17
NO FIN RISK	1.16
REAL ESTATE/NW	1.15
SAVING HOME	1.11
HIGH FIN RISK	1.04

1992 Weighted	VIF
AGE	42.6
AGE2	42.11
DEBT7ASSET	3.83
ln(NET WORTH)	3.21
HOUSES/NW	2.58
a2019	2.5
a2016	2.47
a2007	2.42
a2004	2.32
a2013	2.32
a2001	2.32
a2010	2.29
a1998	2.18
a1995	1.96
Ln(INCOME)	1.9
TWO CHILDREN	1.36
MARRIED	1.28
THREE + CHILDREN	1.25
COLLEGE DEGREE	1.22
SAVING RETIREMENT	1.21
ONE CHILD	1.21
SAVING EDUCATION	1.21
NO FIN RISK	1.18
BUSINESS/NW	1.15
REAL ESTATE/NW	1.14
SAVING HOME	1.12
HIGH FIN RISK	1.03

1992 Unweighted	VIF
AGE	44.77
AGE2	43.52
ln(NET WORTH)	5.21
DEBT/ASSET	3.61
ln(INCOME)	2.66
HOUSES/NW	2.57
a2016	2.54
a2019	2.47
a2010	2.46
a2013	2.39
a2007	2.24
a2001	2.21
a2004	2.2
a1998	2.11
MARRIED	2.09
a1995	2.01
WOMAN	1.94
BUSINESS/NW	1.5
TWO CHILDREN	1.38
THREE + CHILDREN	1.3
COLLEGE DEGREE	1.26
SAVING EDUCATION	1.2
ONE CHILD	1.2
SAVING RETIREMENT	1.18
NO FIN RISK	1.17
REAL ESTATE/NW	1.15
SAVING HOME	1.11
BLACK AFRICAN	1.08
HISPANIC	1.05
HIGH FIN RISK	1.04
OTHER ETHNICITY	1.02

1992 Weighted	VIF
AGE	42.98
AGE2	42.43
DEBT/ASSET	3.83
ln(NET WORTH)	3.26
HOUSES/NW	2.58
a2019	2.52
a2016	2.48
a2007	2.42
a2004	2.33
a2013	2.33
a2001	2.32
a2010	2.3
MARRIED	2.27
a1998	2.19
WOMAN	2.05
a1995	1.96
Ln(INCOME)	1.9
TWO CHILDREN	1.38
THREE + CHILDREN	1.26
COLLEGE DEGREE	1.23
ONE CHILD	1.22
SAVING RETIREMENT	1.21
SAVING EDUCATION	1.21
NO FIN RISK	1.18
BUSINESS/NW	1.15
REAL ESTATE/NW	1.14
SAVING HOME	1.12
BLACK AFRICAN	1.08
HISPANIC	1.05
HIGH FIN RISK	1.03
OTHER ETHNICITY	1.03

1998 Unweighted	VIF
AGE	44.19
AGE2	43.35
ln(NET WORTH)	5.2
DEBT/ASSET	3.57
ln(INCOME)	2.66
a2016	2.52
HOUSES/NW	2.51
a2019	2.48
a2010	2.38
a2013	2.35
a2007	2.1
a2004	1.96
a2001	1.9
INTERNET	1.51
BUSINESS/NW	1.5
TWO CHILDREN	1.37
THREE + CHILDREN	1.29
COLLEGE DEGREE	1.28
MARRIED	1.22
SAVING EDUCATION	1.2
ONE CHILD	1.19
NO FIN RISK	1.18
SAVING RETIREMENT	1.17
REAL ESTATE/NW	1.14
SAVING HOME	1.1
HIGH FIN RISK	1.04

1998 Weighted	VIF
AGE	42.18
AGE2	42.03
DEBT/ASSET	3.8
ln(NET WORTH)	3.24
HOUSES/NW	2.55
a2019	2.44
a2016	2.37
a2013	2.2
a2007	2.16
a2010	2.14
a2004	1.98
a2001	1.91
Ln(INCOME)	1.9
INTERNET	1.57
TWO CHILDREN	1.36
MARRIED	1.27
THREE + CHILDREN	1.25
COLLEGE DEGREE	1.25
ONE CHILD	1.2
SAVING EDUCATION	1.2
SAVING RETIREMENT	1.2
NO FIN RISK	1.19
BUSINESS/NW	1.15
REAL ESTATE/NW	1.14
SAVING HOME	1.11
HIGH FIN RISK	1.03

Annex 3: correlation matrices

Correlation matrix for the period 1998-2019. The matrix correlates the choice to own equity with the variables used in the probit model; it is calculated using implicate 1 and not using sampling weights. Only families that met the requirements imposed as a constraint in the probit analysis were considered, that is: FINANCIAL>0, NET WORTH>0, DEBT/ASSET<1, HOUSES/NW<4, REAL ESTATE/NW>-4, REAL ESTATE/NW<4, BUSINESS/NW<4, BUSINESS/NW>-4. For space reasons each variable is represented by a letter, the legend is present after the two correlation matrices.

	A	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
A	1.00																							
C	0.07	1.00																						
D	0.03	0.99	1.00																					
E	0.40	0.07	0.05	1.00																				
F	-0.02	0.10	0.11	-0.04	1.00																			
G	0.05	-0.23	-0.25	0.04	-0.19	1.00																		
H	0.00	-0.21	-0.22	0.01	-0.15	-0.14	1.00																	
I	0.26	0.01	0.04	0.18	0.04	0.16	0.14	1.00																
J	0.54	0.35	0.31	0.48	-0.06	0.01	0.00	0.36	1.00															
K	0.46	0.13	0.10	0.45	-0.03	0.07	0.05	0.38	0.74	1.00														
L	-0.06	0.09	0.10	-0.09	0.06	0.09	0.05	0.05	-0.13	-0.15	1.00													
M	0.08	0.04	0.04	0.08	0.00	0.00	-0.01	0.08	0.15	0.08	-0.06	1.00												
N	0.14	0.03	0.01	0.18	-0.02	0.04	0.06	0.18	0.45	0.37	-0.20	-0.03	1.00											
O	-0.07	-0.36	-0.36	-0.11	0.08	0.12	0.09	0.03	-0.35	-0.17	0.63	0.08	-0.15	1.00										
P	-0.45	0.10	0.13	-0.35	0.02	-0.06	-0.02	-0.20	-0.40	-0.37	0.08	-0.09	-0.19	0.02	1.00									
Q	0.05	0.04	0.04	0.06	-0.01	0.01	0.03	0.02	0.11	0.10	-0.05	0.03	0.12	-0.01	-0.16	1.00								
R	0.21	0.05	0.01	0.12	0.01	-0.01	-0.06	0.11	0.16	0.12	0.03	0.03	0.03	0.00	-0.16	-0.01	1.00							
S	-0.01	0.18	0.18	0.01	0.05	0.16	0.18	0.07	-0.05	0.00	0.03	0.00	0.00	0.06	-0.02	0.00	-0.23	1.00						
T	-0.09	0.18	0.16	-0.07	0.01	0.01	0.00	-0.04	-0.17	-0.09	-0.09	-0.03	-0.05	0.03	0.03	0.00	-0.13	-0.05	1.00					
U	0.25	0.19	-0.21	0.25	0.03	0.07	0.06	0.15	0.17	0.20	0.03	0.04	0.08	0.13	-0.26	0.05	0.09	0.01	0.00	1.00				
V	-0.22	0.07	0.09	-0.16	0.02	-0.07	-0.07	-0.68	-0.31	-0.32	0.00	-0.07	-0.18	0.00	0.23	-0.05	-0.10	-0.03	0.03	-0.13	1.00			
W	-0.19	0.09	0.08	-0.15	0.04	-0.01	0.01	-0.18	-0.27	-0.18	-0.01	-0.02	-0.10	0.06	0.13	-0.01	-0.09	0.01	0.07	-0.06	0.18	1.00		
X	-0.19	0.15	0.14	-0.16	0.02	0.06	0.09	0.00	-0.20	-0.14	0.02	-0.02	-0.07	0.06	0.15	0.00	-0.08	0.07	0.07	-0.04	0.01	-0.09	1.00	
Y	0.00	0.07	0.07	0.07	0.02	0.03	0.01	0.03	-0.01	-0.01	0.00	0.01	0.00	0.01	0.00	0.01	-0.02	0.04	0.01	0.05	-0.03	-0.07	-0.06	1.00

Legend: dark blue<=-0.5; -0.5<light blue<-0.2; 0.2<light red<0.5; dark red>0.5

Correlation matrix for the period 1998-2019. The matrix correlates the fraction of equity held respect the total financial endowment with the variables used in the regression model; it is calculated using implicate 1 and not using sampling weights. Only families that met the requirements imposed as a constraint in the linear regression analysis were considered, that is: FINANCIAL>0, NET WORTH>0, DEBT/ASSET<1, HOUSES/NW<4, REAL ESTATE/NW>-4, REAL ESTATE/NW<4, BUSINESS/NW<4, BUSINESS/NW>-4, YES_EQUITY==1.

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
B	1.00																							
C	0.03	1.00																						
D	0.02	0.99	1.00																					
E	0.14	0.07	0.05	1.00																				
F	-0.02	-0.11	-0.12	-0.05	1.00																			
G	0.01	-0.26	-0.27	0.03	-0.20	1.00																		
H	0.02	-0.21	-0.22	0.03	-0.14	-0.14	1.00																	
I	0.04	0.00	-0.02	0.10	0.05	0.16	0.14	1.00																
J	0.17	0.43	0.39	0.40	-0.08	-0.03	0.02	0.28	1.00															
K	0.14	0.17	0.14	0.36	-0.03	0.05	0.06	0.30	0.75	1.00														
L	-0.08	-0.24	-0.24	-0.13	0.07	0.12	0.07	0.02	-0.39	-0.26	1.00													
M	-0.01	0.04	0.03	0.04	-0.01	-0.01	-0.01	0.05	0.08	0.03	-0.07	1.00												
N	0.01	0.04	0.02	0.14	-0.02	0.04	0.07	0.15	0.49	0.39	-0.25	-0.06	1.00											
O	-0.07	-0.45	-0.44	-0.15	0.09	0.14	0.09	0.00	-0.56	-0.29	0.73	0.10	-0.21	1.00										
P	-0.15	0.08	0.10	-0.24	0.02	-0.04	-0.02	-0.11	-0.25	-0.23	0.09	-0.04	-0.13	0.06	1.00									
Q	0.08	-0.04	-0.04	0.04	0.00	0.01	0.03	0.01	0.11	0.10	-0.05	0.02	0.12	-0.02	-0.11	1.00								
R	0.04	0.01	-0.03	0.03	0.01	-0.03	-0.05	0.05	0.00	0.00	0.02	0.01	-0.02	0.01	-0.08	-0.02	1.00							
S	0.00	-0.17	-0.17	0.03	0.04	0.17	0.18	0.08	-0.04	0.00	0.07	0.00	0.01	0.08	-0.02	-0.01	-0.26	1.00						
T	-0.05	-0.17	-0.14	-0.05	0.00	0.00	-0.01	-0.05	-0.17	-0.09	-0.07	-0.02	-0.05	0.05	0.02	-0.01	-0.12	-0.04	1.00					
U	0.01	-0.18	-0.20	0.18	0.03	0.06	0.05	0.09	0.03	0.09	0.05	0.01	0.02	0.12	-0.15	0.02	0.06	0.00	0.00	1.00				
V	-0.07	0.06	0.07	-0.10	0.00	-0.09	-0.09	-0.68	-0.24	-0.26	0.02	-0.05	-0.15	0.03	0.15	-0.05	-0.04	-0.04	0.03	-0.08	1.00			
W	-0.05	-0.10	-0.10	-0.11	0.04	0.00	0.01	-0.13	-0.21	-0.13	0.04	0.00	-0.08	0.11	0.09	-0.01	-0.05	0.01	0.07	-0.01	0.14	1.00		
X	-0.04	-0.11	-0.11	-0.10	0.02	0.05	0.05	-0.01	-0.13	-0.08	0.06	0.01	-0.04	0.10	0.08	0.00	-0.04	0.03	0.03	0.01	0.01	-0.05	1.00	
Y	-0.02	-0.08	-0.07	0.07	0.03	0.04	0.01	0.02	-0.03	-0.02	0.01	0.01	0.00	0.02	0.01	0.01	-0.02	0.04	0.01	0.05	-0.02	-0.05	-0.04	1.00

Legend: dark blue<=-0.5; -0.5<light blue<-0.2; 0.2<light red<0.5; dark red>0.5

Yes equity	A	One child	F	Ln(income)	K	No financial risk	P	Internet	U
Equity/net worth	B	Two children	G	Houses/ net worth	L	High financial risk	Q	Woman	V
Age	C	Three or more children	H	Real-estate/ net worth	M	Save for retirement	R	Black/African American	W
Age^2	D	Married	I	Business/ net worth	N	Save for education	S	Hispanic	X
College degree	E	ln(net worth)	J	Debt/asset	O	Save for home	T	Other ethnicities	Y

Annex 4: probit analyses with bootstrap weights, years from 1992 to 2019

Weighted probit regressions for participation, including ethnicity, gender and internet (if data were available). Period: 1992-2019. Models estimated considering data of all the five implicates and bootstrap weights through the Stata ado "scfcombo". One, two or three asterisks means that the predictor is significant at the ten, five or one percent level respectively.

	1992		1995		1998		2001		2004	
	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
AGE	$2.69x10^{-2}$**	0.010	$-1.53x10^{-2}$	0.149	$6.45x10^{-3}$	0.449	$-1.26x10^{-2}$	0.174	$2.72x10^{-2}$***	0.007
AGE^2	$-3.46x10^{-4}$***	<0.001	$5.65x10^{-5}$	0.585	$-2.02x10^{-4}$**	0.012	$-3.83x10^{-5}$	0.658	$-3.64x10^{-4}$***	<0.001
Education (ref: not have a degree)										
COLLEGE DEGREE	0.2345***	<0.001	0.3183***	<0.001	0.0082	0.902	0.2168***	<0.001	0.0845	0.205
Children (reference: without children)										
ONE CHILD	-0.1845***	0.001	-0.0478	0.413	-0.0214	0.723	0.0370	0.581	-0.1632**	0.030
TWO CHILDREN	-0.0639	0.366	0.0442	0.582	0.0136	0.871	-0.0597	0.459	-0.2172***	0.003
THREE OR MORE CHILDREN	-0.2711***	0.005	-0.0641	0.539	-0.1760*	0.058	0.1147	0.244	-0.2106	0.122
Married (ref: not married/living with partner)										
MARRIED	0.1291*	0.061	0.0704	0.353	0.0191	0.772	-0.0679	0.320	0.0918	0.268
ln(NET WORTH)	0.4009***	<0.001	0.3733***	<0.001	0.4544***	<0.001	0.5077***	<0.001	0.4604***	<0.001
Ln(INCOME)	0.1660***	<0.001	0.2136***	<0.001	0.1956***	<0.001	0.1689***	<0.001	0.1813***	<0.001
HOUSES/NET WORTH	-0.4150***	<0.001	-0.3082***	<0.001	-0.3690***	<0.001	-0.4165***	<0.001	-0.3991***	<0.001
REAL ESTATE/NET WORTH	-0.6255***	<0.001	-0.3721***	<0.001	-0.5993***	<0.001	-0.9137***	<0.001	-0.4961***	0.007
BUSINESS/NET WORTH	-0.9831***	<0.001	-1.4072***	<0.001	-1.2011***	<0.001	-1.5486***	<0.001	-1.4227***	<0.001
DEBT/ASSET	1.7870***	<0.001	1.2750***	<0.001	1.5033***	<0.001	1.6069***	<0.001	1.3816***	<0.001
Risk attitude (ref: medium financial risk)										
NO FINANCIAL RISK	-0.4567***	<0.001	-0.4651***	<0.001	-0.5260***	<0.001	-0.6551***	<0.001	-0.6720***	<0.001
SUBSTANTIAL FINANCIAL RISK	-0.0314	0.812	0.0689	0.515	0.0988	0.369	0.3048***	0.006	0.1749	0.143
Saving reasons										
SAVING FOR RETIREMENT	0.1455***	0.008	0.4195***	<0.001	0.2021***	<0.001	0.2043***	<0.001	0.1658***	0.007
SAVING FOR EDUCATION	-0.1198	0.160	0.0462	0.532	0.0203	0.826	-0.0100	0.902	0.1033	0.369
SAVING FOR HOME	-0.1212	0.346	0.1977	0.156	0.2041	0.109	0.1527	0.256	0.2706**	0.039
Sex (reference: male)										
WOMAN	0.2814***	<0.001	0.2172**	0.013	0.1716**	0.014	0.0054	0.946	0.1785**	0.047
Ethnicity (ref: white not Hispanic)										
BLACK/AFRICAN AMERICAN	-0.2339**	0.015	-0.1188	0.184	-0.0773	0.386	-0.0483	0.583	-0.2752***	<0.001
HISPANIC	-0.3828**	0.036	-0.0445	0.677	-0.4779***	<0.001	-0.1074	0.323	-0.5691***	<0.001
OTHER ETHNICITY	-0.4703***	<0.001	0.0129	0.925	-0.0804	0.519	0.0163	0.946	-0.2673**	0.034
Internet usage (ref: not used)										
INTERNET	-	-	-	-	0.3356***	<0.001	0.2175***	0.001	0.3426***	<0.001
constant	-6.9314***	<0.001	-6.2580***	<0.001	-7.0125***	<0.001	-6.7400***	<0.001	-7.4781***	<0.001
Sample size	N=3399		N=3753		N=3753		N=3926		N=3923	

	2007		2010		2013		2016		2019	
	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
AGE	-2.02×10^{-3}	0.834	1.97×10^{-3}	0.760	1.31×10^{-2}	0.109	-6.97×10^{-4}	0.923	-2.89×10^{-2}***	<0.001
AGE^2	-4.01×10^{-5}	0.652	-8.56×10^{-5}	0.137	-2.46×10^{-4}***	0.001	-7.90×10^{-5}	0.219	1.45×10^{-4}*	0.053
Education (ref: not have a degree)										
COLLEGE DEGREE	0.1720***	0.005	0.2328***	<0.001	0.2945***	<0.001	0.2182***	<0.001	0.1993***	<0.001
Children (reference: without children)										
ONE CHILD	-0.1067	0.154	-0.0015	0.976	-0.1469**	0.022	-0.0642	0.275	-0.0386	0.535
TWO CHILDREN	-0.0736	0.375	0.0451	0.378	-0.0685	0.384	0.0346	0.601	0.0894	0.226
THREE OR MORE CHILDREN	-0.2330**	0.032	-0.1880**	0.029	-0.2378***	<0.001	-0.2776***	<0.001	-0.1496	0.133
Married (ref: not married/living with partner)										
MARRIED	0.1185	0.201	-0.0209	0.727	0.0081	0.900	0.0160	0.785	0.0719	0.228
ln(NET WORTH)	0.3821***	<0.001	0.4365***	<0.001	0.4376***	<0.001	0.4583***	<0.001	0.4972***	<0.001
Ln(INCOME)	0.1600***	<0.001	0.2010***	<0.001	0.3035***	<0.001	0.2947***	<0.001	0.2171***	<0.001
HOUSES/NET WORTH	-0.3789***	<0.001	-0.4445***	<0.001	-0.5087***	<0.001	-0.4866***	<0.001	-0.5492***	<0.001
REAL ESTATE/NET WORTH	-0.2685**	0.033	-0.5623***	<0.001	-0.5744***	<0.001	-0.6944***	<0.001	-0.8999***	<0.001
BUSINESS/NET WORTH	-1.2460***	<0.001	-1.2450***	<0.001	-1.5433***	<0.001	-1.3011***	<0.001	-1.5930***	<0.001
DEBT/ASSET	1.2583***	<0.001	1.6969***	<0.001	1.6774***	<0.001	1.4702***	<0.001	1.7662***	<0.001
Risk attitude (ref: medium financial risk)										
NO FINANCIAL RISK	-0.5155***	<0.001	-0.5651***	<0.001	-0.4990***	<0.001	-0.5088***	<0.001	-0.4920***	<0.001
SUBSTANTIAL FINANCIAL RISK	-0.0615	0.708	-0.2352**	0.025	0.0453	0.690	-0.4079***	<0.001	-0.2193**	0.042
Saving reasons										
SAVING FOR RETIREMENT	0.3217***	<0.001	0.1675***	<0.001	0.2049***	<0.001	0.3274***	<0.001	0.2652***	<0.001
SAVING FOR EDUCATION	-0.0467	0.662	0.0510	0.528	0.0565	0.351	0.0545	0.515	-0.1545	0.112
SAVING FOR HOME	-0.0042	0.978	0.1032	0.349	-0.1126	0.408	0.0654	0.577	-0.1328	0.246
Sex (reference: male)										
WOMAN	0.0742	0.429	0.1461*	0.065	0.2331***	0.002	0.1019	0.130	0.1896***	0.003
Ethnicity (ref: white not Hispanic)										
BLACK/AFRICAN AMERICAN	-0.1842*	0.064	-0.1678**	0.011	-0.1291	0.136	-0.2461***	<0.001	-0.1874***	0.001
HISPANIC	-0.2887***	0.002	-0.3623***	<0.001	-0.4480***	<0.001	-0.3253***	<0.001	-0.5380***	<0.001
OTHER ETHNICITY	0.0447	0.762	-0.1893**	0.019	-0.5141***	<0.001	-0.3528***	<0.001	0.0170	0.885
Internet usage (ref: not used)										
INTERNET	0.3310***	<0.001	0.2201***	<0.001	0.0327	0.562	0.1021	0.130	-0.0958	0.134
constant	-5.9038***	<0.001	-6.9116***	<0.001	-8.0245***	<0.001	-7.9699***	<0.001	-6.7065***	<0.001
Sample size	N=3872		N=5289		N=4966		N=5428		N=5087	

Annex 5: regression analyses with bootstrap weights, years from 1992 to 2019

Weighted linear regressions for equity share, including ethnicity and gender and internet (if data were available). Period: 1992-2019. Models estimated considering data of all the five implicates and bootstrap weights through the Stata ado "scfcombo". One, two or three asterisks means that the predictor is significant at the ten, five or one percent level respectively.

	1992		1995		1998		2001		2004	
	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
AGE	$1.15x10^{-2}$***	<0.001	$1.14x10^{-2}$***	0.002	$1.84x10^{-3}$	0.537	$1.51x10^{-3}$	0.555	$1.52x10^{-3}$	0.652
AGE^2	$-1.12x10^{-4}$***	<0.001	$-1.08 x10^{-4}$***	0.002	$-2.16x10^{-5}$	0.441	$-1.71x10^{-5}$	0.493	$-1.59x10^{-5}$	0.6
Education (ref: not have a degree)										
COLLEGE DEGREE	0.0231	0.233	0.0338**	0.032	0.0204	0.28	0.0297*	0.062	0.0427***	0.001
Children (reference: without children)										
ONE CHILD	-0.0250	0.196	-0.0159	0.365	0.0324*	0.079	-0.0330*	0.052	-0.0099	0.656
TWO CHILDREN	-0.0234	0.26	0.0241	0.219	0.0101	0.591	0.0061	0.719	-0.0108	0.556
THREE OR MORE CHILDREN	-0.0370	0.148	0.0069	0.821	0.0217	0.391	0.0510**	0.034	0.0204	0.411
Married (ref: not married/living with partner)										
MARRIED	-0.0862***	0.001	-0.0738**	0.017	-0.0199	0.384	-0.0310	0.116	-0.0452**	0.025
ln(NET WORTH)	-0.0085	0.385	0.0032	0.776	0.0225***	0.009	0.0279***	<0.001	0.0200**	0.043
Ln(INCOME)	0.0180	0.132	0.0088	0.494	0.0065	0.475	-0.0023	0.825	-0.0047	0.616
HOUSES/NET WORTH	-0.0256	0.113	-0.0089	0.577	-0.0306**	0.03	-0.0361**	0.02	-0.0309**	0.017
REAL ESTATE/NET WORTH	-0.0451	0.165	-0.0581*	0.077	-0.0716***	0.009	-0.0833***	0.004	-0.0606**	0.014
BUSINESS/NET WORTH	-0.0929**	0.032	-0.0291	0.592	-0.0776***	0.009	-0.1408***	0.002	-0.0573	0.164
DEBT/ASSET	0.0486	0.483	0.1639***	0.005	0.1719***	0.001	0.1285**	0.032	0.1257**	0.015
Risk attitude (ref: medium financial risk)										
NO FINANCIAL RISK	-0.0616***	0.004	-0.0166	0.427	-0.0664***	<0.001	-0.0506***	0.006	-0.0558***	<0.001
SUBSTANTIAL FINANCIAL RISK	0.0962**	0.031	0.0355	0.307	-0.0087	0.746	0.0789***	0.004	0.0597**	0.024
Saving reasons										
SAVING FOR RETIREMENT	0.0396**	0.025	0.0153	0.39	0.0249	0.115	0.0214	0.101	0.0183	0.186
SAVING FOR EDUCATION	-0.0023	0.931	0.0171	0.494	-0.0161	0.489	0.0007	0.973	0.0055	0.791
SAVING FOR HOME	-0.0825*	0.075	0.0385	0.301	-0.0363	0.392	0.0151	0.683	-0.0199	0.617
Sex (reference: male)										
WOMAN	-0.1128***	<0.001	-0.0467	0.101	0.0077	0.765	-0.0365*	0.093	-0.0038	0.88
Ethnicity (ref: white not Hispanic)										
BLACK/AFRICAN AMERICAN	0.1011***	0.005	-0.0825***	0.008	-0.0254	0.362	-0.0483**	0.033	-0.0239	0.447
HISPANIC	0.0262	0.645	-0.0949**	0.033	0.0270	0.437	-0.0064	0.846	-0.0161	0.609
OTHER ETHNICITY	0.0055	0.833	-0.0241	0.542	0.0340	0.336	-0.0197	0.622	-0.0326	0.378
Internet usage (ref: not used)										
INTERNET	-	-	-	-	0.0288*	0.062	0.0221*	0.071	0.0092	0.491
constant	0.0992	0.357	0.0528	0.716	0.0914	0.391	0.1923*	0.094	0.2549***	0.002
Sample size	N=1909		N=2243		N=2450		N=2710		N=2650	

	2007		2010		2013		2016		2019	
	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
AGE	$7.68x10^{-3}$***	0.001	$4.55x10^{-3}$**	0.011	$3.66x10^{-3}$	0.157	$1.04x10^{-3}$	0.726	$-2.46x10^{-4}$	0.907
AGE^2	$-8.14x10^{-5}$***	<0.001	$-4.40x10^{-5}$***	0.008	$-3.75x10^{-5}$	0.108	$-1.87x10^{-5}$	0.469	$-1.87x10^{-6}$	0.916
Education (ref: not have a degree)										
COLLEGE DEGREE	0.0211	0.177	0.0310***	0.003	0.0056	0.638	0.0244**	0.036	0.0067	0.62
Children (reference: without children)										
ONE CHILD	-0.0061	0.744	0.0119	0.516	0.0038	0.825	0.0158	0.298	0.0110	0.527
TWO CHILDREN	-0.0246	0.134	0.0061	0.723	0.0387*	0.055	-0.0053	0.783	0.0296	0.116
THREE OR MORE CHILDREN	-0.0181	0.417	0.0510**	0.021	0.0386*	0.083	0.0093	0.703	0.0341*	0.083
Married (ref: not married/living with partner)										
MARRIED	-0.0419*	0.07	-0.0699***	<0.001	-0.0391**	0.015	-0.0378**	0.046	-0.0421***	0.009
ln(NET WORTH)	0.0185***	0.005	0.0150***	0.008	0.0270***	<0.001	0.0363***	<0.001	0.0500***	<0.001
Ln(INCOME)	-0.0072	0.33	0.0055	0.335	-0.0077	0.319	-0.0098	0.167	-0.0157*	0.07
HOUSES/NET WORTH	-0.0344*	0.064	-0.0157	0.124	-0.0345***	0.004	-0.0538***	<0.001	-0.0477***	0.004
REAL ESTATE/NET WORTH	-0.0147	0.547	-0.0262	0.247	-0.0609**	0.013	-0.0758***	0.003	-0.0564*	0.079
BUSINESS/NET WORTH	-0.1134**	0.011	-0.0942***	<0.001	-0.1196***	<0.001	-0.1376***	<0.001	-0.1398***	<0.001
DEBT/ASSET	0.0726	0.294	0.0703*	0.069	0.0847	0.104	0.1770***	<0.001	0.1976***	0.006
Risk attitude (ref: medium financial risk)										
NO FINANCIAL RISK	-0.0528***	0.005	-0.0455***	0.001	-0.0673***	<0.001	-0.0294**	0.032	-0.0195	0.336
SUBSTANTIAL FINANCIAL RISK	0.0631***	0.008	0.1393***	<0.001	0.1039***	0.001	0.0715***	0.009	0.0648**	0.018
Saving reasons										
SAVING FOR RETIREMENT	-0.0047	0.796	0.0070	0.392	-0.0096	0.397	0.0193	0.111	0.0316***	0.002
SAVING FOR EDUCATION	0.0090	0.726	0.0253	0.25	-0.0328*	0.095	0.0188	0.389	0.0044	0.866
SAVING FOR HOME	-0.1228***	0.004	-0.0615*	0.079	-0.0356	0.42	0.0001	0.997	-0.0117	0.833
Sex (reference: male)										
WOMAN	-0.0255	0.341	-0.0834***	<0.001	-0.0303	0.146	-0.0669***	0.001	-0.0318	0.105
Ethnicity (ref: white not Hispanic)										
BLACK/AFRICAN AMERICAN	-0.0519**	0.049	-0.0088	0.727	-0.0626***	<0.001	-0.0089	0.686	0.0072	0.728
HISPANIC	-0.0140	0.696	-0.0048	0.854	-0.0172	0.577	-0.0258	0.253	0.0149	0.556
OTHER ETHNICITY	-0.0559*	0.081	0.0087	0.665	-0.0090	0.77	0.0022	0.943	-0.0193	0.395
Internet usage (ref: not used)										
INTERNET	0.0075	0.614	0.0098	0.564	-0.0194	0.26	-0.0460*	0.079	-0.0450*	0.08
constant	0.2156**	0.020	0.1136	0.114	0.1944**	0.034	0.1657*	0.055	0.0550	0.58
Sample size	N=2748		N=3346		N=3167		N=3536		N=3330	

Annex 6: interaction terms: saving for retirement, probit

Weighted probit for participation in equity markets considering interaction term for saving for retirement, including ethnicity and gender. Period: 1992-2019. Pooled data. One, two or three asterisks means that the predictor is significant at the ten, five or one percent level respectively. Models estimated considering data of all the five implicates and sample weights. Sample size N=43396.

WEIGHTED Probit for equity participation	Coefficient	Standard error	P-value	95% confidence interval	
AGE	$-6.64x10^{-4}$	$3.63x10^{-3}$	0.855	$-7.77x10^{-3}$	$6.44x10^{-3}$
AGE^2	$-1.01x10^{-4}$***	$3.41x10^{-5}$	0.003	$-1.68x10^{-4}$	$-3.41x10^{-5}$
Education (reference: not have a degree)					
COLLEGE DEGREE	0.2205***	0.0221	<0.001	0.177	0.264
Children (reference: without children)					
ONE CHILD	-0.0698**	0.0276	0.011	-0.124	-0.016
TWO CHILDREN	-0.0255	0.0304	0.401	-0.085	0.034
THREE OR MORE CHILDREN	-0.1695***	0.0376	<0.001	-0.243	-0.096
Married (reference: not married/living with partner)					
MARRIED	0.0520*	0.0298	0.081	-0.006	0.110
ln(NET WORTH)	0.4372***	0.0112	<0.001	0.415	0.459
Ln(INCOME)	0.2304***	0.0188	<0.001	0.194	0.267
HOUSES/NET WORTH	-0.4328***	0.0199	<0.001	-0.472	-0.394
REAL ESTATE/NET WORTH	-0.5783***	0.0551	<0.001	-0.686	-0.470
BUSINESS/NET WORTH	-1.3218***	0.0803	<0.001	-1.479	-1.164
DEBT/ASSET	1.5782***	0.0724	<0.001	1.436	1.720
Risk attitude (reference: medium financial risk)					
NO FINANCIAL RISK	-0.5425***	0.0211	<0.001	-0.584	-0.501
SUBSTANTIAL FINANCIAL RISK	-0.0377	0.0493	0.444	-0.134	0.059
Saving reasons					
SAVING FOR EDUCATION	0.0107	0.0349	0.76	-0.058	0.079
SAVING FOR HOME	0.0522	0.0498	0.295	-0.045	0.150
Sex (reference: male)					
WOMAN	0.1663***	0.0306	<0.001	0.106	0.226
Ethnicity (reference: white not Hispanic)					
BLACK/AFRICAN AMERICAN	-0.1716***	0.0306	<0.001	-0.232	-0.112
HISPANIC	-0.3859***	0.0359	<0.001	-0.456	-0.316
OTHER ETHNICITY	-0.1789***	0.0471	<0.001	-0.271	-0.087
SAVING FOR RETIREMENT					
no	0.0000				
yes	0.8541***	0.2965	0.004	0.273	1.435
RETIREMENT*AGE					
no	0.0000				
yes	$-3.02x10^{-3}$*	$-1.6x10^{-3}$	0.059	$-6.16x10^{-3}$	$1.16x10^{-4}$
RETIREMENT*ln(INCOME)					
no	0.0000				
yes	-0.0617**	0.0284	0.03	-0.117	-0.006
RETIREMENT*ln(NET WORTH)					
no	0.0000				
yes	0.0193	0.0155	0.211	-0.011	0.050
Survey year (reference: 1992)					
1995	-0.0022	0.0487	0.964	-0.098	0.093
1998	0.1752***	0.0456	<0.001	0.086	0.265
2001	0.2247***	0.0458	<0.001	0.135	0.314
2004	0.1725***	0.0474	<0.001	0.080	0.265
2007	0.2593***	0.0458	<0.001	0.170	0.349
2010	0.3240***	0.0412	<0.001	0.243	0.405
2013	0.2828***	0.0430	<0.001	0.198	0.367
2016	0.2887***	0.0442	<0.001	0.202	0.375
2019	0.2574***	0.0425	<0.001	0.174	0.341
constant	-7.2624***	0.1901	<0.001	-7.635	-6.890

Annex 7: interaction terms: saving for retirement, linear regression

Weighted linear regression for equity share considering interaction term for saving for retirement, including ethnicity and gender. Period: 1992-2019. Pooled data. One, two or three asterisks means that the predictor is significant at the ten, five or one percent level respectively. Models estimated considering data of all the five implicates and sample weights. Sample size N=28089.

	Coefficient	Robust SE	P-value	95% confidence interval	
AGE	4.03×10^{-3}***	1.14×10^{-3}	<0.001	1.80×10^{-3}	6.25×10^{-3}
AGE^2	-3.78×10^{-5}***	1.01×10^{-5}	<0.001	-5.77×10^{-5}	-1.79×10^{-5}
Education (reference: not have a degree)					
COLLEGE DEGREE	0.0244***	0.0052	<0.001	0.014	0.035
Children (reference: without children)					
ONE CHILD	-0.0009	0.0086	0.917	-0.018	0.016
TWO CHILDREN	0.0036	0.0080	0.650	-0.012	0.019
THREE OR MORE CHILDREN	0.0198*	0.0105	0.059	-0.001	0.040
Married (reference: not married/living with partner)					
MARRIED	-0.0467***	0.0080	<0.001	-0.062	-0.031
ln(NET WORTH)	0.0206***	0.0030	<0.001	0.015	0.027
Ln(INCOME)	-0.0018	0.0039	0.648	-0.010	0.006
HOUSES/NET WORTH	-0.0305***	0.0052	<0.001	-0.041	-0.020
REAL ESTATE/NET WORTH	-0.0520***	0.0098	<0.001	-0.071	-0.033
BUSINESS/NET WORTH	-0.1020***	0.0129	<0.001	-0.127	-0.077
DEBT/ASSET	0.1201***	0.0189	<0.001	0.083	0.157
Risk attitude (reference: medium financial risk)					
NO FINANCIAL RISK	-0.0468***	0.0066	<0.001	-0.060	-0.034
SUBSTANTIAL FINANCIAL RISK	0.0667***	0.0121	<0.001	0.043	0.091
Saving reasons					
SAVING FOR EDUCATION	0.0077	0.0095	0.417	-0.011	0.026
SAVING FOR HOME	-0.0196	0.0204	0.337	-0.059	0.020
Sex (reference: male)					
WOMAN	-0.0416***	0.0086	<0.001	-0.058	-0.025
Ethnicity (reference: white not Hispanic)					
BLACK/AFRICAN AMERICAN	-0.0219**	0.0100	0.029	-0.042	-0.002
HISPANIC	-0.0109	0.0113	0.335	-0.033	0.011
OTHER ETHNICITY	-0.0125	0.0130	0.334	-0.038	0.013
SAVING FOR RETIREMENT					
no	0.0000				
yes	-0.0157	0.0605	0.795	-0.134	0.103
RETIREMENT*AGE					
no	0.0000				
yes	-9.864×10^{-4}***	3.995×10^{-4}	0.014	-1.77×10^{-3}	-2.03×10^{-4}
RETIREMENT*ln(INCOME)					
no	0.000				
yes	4.757×10^{-4}	0.0065	0.941	-0.012	0.013
RETIREMENT*ln(NETWORTH)					
no	0.000				
yes	6.161×10^{-3}	0.0041	0.132	-1.86×10^{-3}	1.42×10^{-2}
Survey year (reference: 1992)					
1995	0.0661***	0.0125	<0.001	0.042	0.091
1998	0.1132***	0.0121	<0.001	0.089	0.137
2001	0.1463***	0.0118	<0.001	0.123	0.169
2004	0.0854***	0.0122	<0.001	0.061	0.109
2007	0.0805***	0.0124	<0.001	0.056	0.105
2010	0.0490***	0.0115	<0.001	0.026	0.072
2013	0.0665***	0.0112	<0.001	0.045	0.088
2016	0.0574***	0.0126	<0.001	0.033	0.082
2019	0.0567***	0.0116	<0.001	0.034	0.079
constant	0.0800*	0.0473	0.091	-0.013	0.173

Appendix 1: imputation and implicates

Financial surveys are often subject to non-response rates or inconsistency in revealed data, such a thing could depend on the degree of invasiveness and complexity of the requested questions. In the first case the respondent may not want to declare the exact amount or even a range of values relating to the value of the investments in the requested assets, in the second it may not be exactly known, such as the value of the investment in an unlisted business or in a real estate property.

To limit the impact of missing values or to obscure information that could lead to the identification of respondents, the SCF staff impute realistic values of these data. The fact that a variable in the full public dataset has been imputed or not can be analysed by observing the shadow variables (prefixed by a J) related to the relative variable (X).

The process that allows the estimation of the values, which take into account the uncertainty of this estimate, involves the creation of five distinct datasets called implicates, each of them containing all information of the respondents and his household. For this reason, the global dataset has a fivefold row cardinality compared to the number of observations.

Failure to consider this data structure by using the entire dataset and considering each observation as attributable to a different household leads to evident distortion of the real variability of the variables, considerably underestimating it. However, even considering only one implicate is not recommended, it generally leads to a greater likelihood of finding statistically significant effects as pointed out by Lindamood et al. (2007). The SCF codebook is a useful source of information regarding the proper way to deal with this issue, in particular the proper technique to use is called in the literature as repeated-imputation inference (RII); an easily and understandable introduction to this method, which is applicable to univariate descriptive statistic and models estimated by both least square and maximum likelihood, is describe by Montalto and Sung (1996). **Table a** summarise the formulas, provided by the paper mentioned above, needed to account for imputed values.

Table a: Montalto and Sung, formulas to consider for multiple implicates.

Descriptive statistic	Explanation	Formula
Point estimate	Average of the implicates point estimate	$\overline{Q_m} = \frac{\sum_1^m Q_i}{m}$
Within imputation variance	Average of the variance in each implicate	$\overline{V(W)_m} = \frac{\sum_1^m VAR_i}{m}$
Between imputation variance	Variance due to imputation of missing values	$\overline{V(B)_m} = \frac{\sum_1^m (Q_i - \overline{Q_m})^2}{m-1}$
Total variance	Weighted sum of within and between variance	$\overline{V_m} = \overline{V(W)_m} + (1 + m^{-1})\overline{V(B)_m}$

Appendix 2: sampling and bootstrap weights

The staff of Survey of Consumer Finances provide two different types of weight: sampling weights and bootstrap weights.

Sampling weights

The main data set, for each survey wave, contains the final nonresponse-adjusted sampling weights. Regarding their construction the 2019 codebook state: "*the weight (X42001) is a partially design-based weight constructed at the Federal Reserve using original selection probabilities and frame information along with aggregate control totals estimated from the Current Population Survey."*

Sampling weights are computed separately for each implicate and survey wave, they represent the number of households the interviewed PEU would represent in the total population, therefore the household specific inflation factor.

The population defined by the sum of all the weights of each implicate is approximately equal to the total number of US households, so the sum of weights has changed over time: in 1992 was approximately equal to 95.9 million families, whereas in 2019 was 128.6 million.

The reasons for the need to introduce weights are:

- unequal household probability of selection, the survey design has been studied in order to be representative of the American population, but to do so some clusters have been over-represented to keep reasonable the sample size.
- unequal response rates. Familiar characteristics affect response rates, especially high wealth individuals tend to refuse to participate in surveys more or not to provide accurate information to specific questions that could harm their privacy.
- Post-stratification, adjusting the sample distribution for key variables of interest such as age, ethnicity, sex, to make it conform to a known population distribution.

Bootstrap weights

The complex design of the Survey of Consumer Finances introduces a further problem when performing statistical analysis on the data: the correct calculation of the standard errors; in addition to considering the variability introduced through the implicates, it is also necessary to consider the sample variance.

The 2019 codebook states: "*because we are unable to release any of the basic sample information about the cases in the data set, users are unable on their own to compute reasonable estimates of the sampling variances of their estimates using standard packages. To facilitate such estimation, we provide a file of replicate weights and multiplicity factors corresponding to X42001. Using detailed information about the original sample design, we selected 999 sample replicates from the final set of completed cases in a way intended to capture the important dimensions of sample variation.*"

The file (specific for each survey wave) is provided to allow sampling variability estimates; considering the lack of knowledge about the sample information is not possible to use standard technique (such as balanced repeated replication or linearization) to obtain good estimate of variances, the only straightforward way is to use bootstrap technique (Kennickell et al., 1996).

Bootstrap (invented by Efron in 1979) is a resampling statistical technique to approximate the sample distribution of a statistic, it allows to approximate the mean and variance of an estimator, calculate p-values and confidence intervals. Basically, using computational power is possible to resample with re-entry K time the sample (such as SCF) coming from the population object of study and for each round compute and store the statistic of interest; the K values obtained can be used to compute measures of dispersion.[a] For further technical inside about bootstrap see Efron and Tibshirani: An Introduction to the Bootstrap (1993).

Specifically, the SCF bootstrap file contain 999 replicates, calculated only for the first implicate, that were generated to preserves important properties of both the area-probability and list samples. At each bootstrap sample is associated a vector of weights, needed for the reason explained before, that where computed using exactly the same procedures for the main weight development (Kennickell et al., 1996). The estimation of

the standard error for a statistic V can be estimated using the following formula (I) which combine both sample and imputation variance:

$$SV_{TOT} = [\frac{6}{5} * SV_{IMP}^2 + SV_{SAMP}^2]^{1/2} \quad (I)$$

Where the sampling variance SV_{SAMP}^2 is given by:

$$SV_{SAMP} = \frac{1}{998} * \sum_{r=1}^{999} \left(V_r - mean(V)\right)^2 \quad (II)$$

[a] consider a sample $\boldsymbol{x} = (x_1, \dots, x_n)$, from this sample is possible to extract with re-entry K samples $(\boldsymbol{x_1}, \dots, \boldsymbol{x_K})$ and to compute for each of them the estimator E of the random variable V object of study $E(\boldsymbol{x}) = \hat{V}$. Using $E(\boldsymbol{x_1}), \dots, E(\boldsymbol{x_K})$ is possible to estimate the bootstrap variance of V.