UNIVERSITA' DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI "M. FANNO"

CORSO DI LAUREA MAGISTRALE IN ECONOMICS AND FINANCE

TESI DI LAUREA

"Financial Literacy and Portfolio Diversification: An Analysis Using Spanish Portfolio Micro Data"

RELATORE:

CH.MO PROF. GUGLIELMO WEBER

LAUREANDO: MARCO GASTALDELLO

MATRICOLA N. 2004260

ANNO ACCADEMICO 2021 - 2022

DICHIARAZIONE DI AUTENTICITA'

Dichiaro di aver preso visione del "Regolamento antiplagio" approvato dal Consiglio del Dipartimento di Scienze Economiche e Aziendali e, consapevole delle conseguenze derivanti da dichiarazioni mendaci, dichiaro che il presente lavoro non è già stato sottoposto, in tutto o in parte, per il conseguimento di un titolo accademico in altre Università italiane o straniere. Dichiaro inoltre che tutte le fonti utilizzate per la realizzazione del presente lavoro, inclusi i materiali digitali, sono state correttamente citate nel corpo del testo e nella sezione 'Riferimenti bibliografici'.

I hereby declare that I have read and understood the "Anti-plagiarism rules and regulations" approved by the Council of the Department of Economics and Management and I am aware of the consequences of making false statements. I declare that this piece of work has not been previously submitted – either fully or partially – for fulfilling the requirements of an academic degree, whether in Italy or abroad. Furthermore, I declare that the references used for this work – including the digital materials – have been appropriately cited and acknowledged in the text and in the section 'References'.

Firma (signature) ...Marco Gastaldello...

Acknowledgements

I would like to express my deepest gratitude to Professor Weber and Francesco Maura for their valuable and constructive guidance and useful critiques during the planning and the development of this thesis work. Their willingness to give their time so generously has been very much appreciated.

I could not have undertaken this journey without my parents. Thank you for your endless support and encouragement throughout my study. I hope to repay all of that in my future career. I would like to extend my sincere thanks to my little brother Luca who has been patient with me in his childhood and with whom now I've built a wonderful complicity.

Words cannot express my gratitude to my girlfriend Adelaide. You have always stood by me, from the bachelor's degree where we shared the rollercoaster of emotions together, up to now with your kind words and unconditional support. I love you. Now comes the time I've been waiting for!

Thank you to my second family, Cristina, Doriana, Raffaele and Fabrizio for your overwhelming generosity and for taking me in as one of your own. Another special thanks to Pierluigi, I know that I can always count on you whenever I need it. Moreover, our dinners together have been wonderful moments of sharing and laughter which I hope will continue.

Last but not least, I would also like to thank all of my friends and other family members for encouraging and supporting me whenever I needed them.

Table of Contents

List of Figures

List of Tables

Introduction

In recent decades we have witnessed a substantial increase in research of economists around the topic of financial literacy and its implications related to the financial behaviors and decisions of households and individuals. Over the course of their lives, households will have to face huge financial decisions such as selecting houses and related mortgages, saving for education and also carving out enough for retirement. These decisions could have both an instantaneous impact on households' finances and also a long-horizon impact on their expenditure capacity. The sign of the long run impact, i.e. whether the household made good or bad financial choices, depends on a multiplicity of factor among which financial literacy plays a relevant role.

The range of possibilities people have to choose from is increasing. Simultaneously, the recent pandemic has turned upside down the job world, making it more and more dynamic and competitive with significant repercussions on the income uncertainty of households. Monetary and macroeconomic dynamics push individuals towards greater budgeting and careful considerations on issues such as pension savings and housing.

One of the most interesting financial choices of households/individuals is the one of portfolio allocation and one of the best strategies to properly manage and reduce the idiosyncratic risk of its own portfolio is diversification.

In an ever-changing industry where the spectrum of choice is increasingly wide and sophisticated, it is crucial to identify factors that can guide economic agents towards informed and economically healthier decisions. From an investment perspective, one of the most important problems studied by economists is the so-called stock holding puzzle. Haliassos (2002) studies in depth this phenomenon and some of the causes behind it, namely risk aversion, risky labor income, participation costs and borrowing constraints. The stockholding puzzle reflects the tendency of a large proportion of households to hold no stocks regardless of a historical average return premium on equity with respect to riskless assets. This behavior is in contrast with the classical economic theory (Samuelson 1969) which tells us that each agent should invest a fraction of its wealth in risky assets. Other economic studies (Calvet et al. (2007), Kumar and Goetzmann (2001) show that investors hold undiversified portfolios composed of a limited number of assets, highlighting the diversification aspect of households' investment portfolio.

Among potential drivers of such an issue, scarce financial literacy has emerged as a plausible factor related to poor diversification given its influence on financial behavior. Lusardi and Mitchell (2007) show how individuals who are financially illiterate are less prone to plan for

retirement and consequently build up a lower amount of wealth. Other findings like those of Alessie, Lusardi and van Rooji (2011) and Christelis, Jappelli and Padula (2008) highlight the link between low financial literacy and investors' stock market participation.

This thesis investigates the relationship between financial literacy (FL) and the portfolio diversification of Spanish individuals. More in details, it aims at examining the impact of financial literacy on portfolio diversification and investigates whether financial literacy affects the heterogeneity of households and individuals' portfolios in terms of categories of assets held. The impact of financial literacy on portfolio returns and riskiness will not be an object of the discussion given the large and well-aligned literature (in terms of findings) around the topic, but rather the key of this analysis will be on the relationship between FL and the numerosity of assets included in the portfolio.
I use the Spanish Survey of Financial Competences (ECF) of 2016 to analyze empirically this relationship. I use eleven (11) questions from the ECF questionnaire to construct an individual financial literacy index. Then, I study the relationship between the financial knowledge of Spanish people and their portfolio diversification (in terms of the variety of assets held).
The thesis is organized as follows:

Chapter 1 frames the topic of financial literacy giving a rundown of the main arguments provided by the literature as well as some numbers in the international context. In this section, I focus on the delicate process of measuring financial literacy by outlining the main issues and by illustrating the major initiatives at international level to harmonize such a procedure which still has wide margins of discretion.

Chapter 2 deals with the theme of diversification. I explain the measure of diversification used for the analysis and why I focus on this type of index rather than a more sophisticated measure. Finally, I discuss the issue of the relationship between diversification and financial literacy.

Chapter 3 describes the data used in the main analysis, i.e. the ECF survey. I present the main characteristics of the dataset and describe the sample used for the analysis. Then, I describe the problem of missing data in this dataset and the strategy used to overcome this problem.

Chapter 4 presents the empirical analysis and the main results of the thesis. First, I describe the variables used in the analysis, their construction and their main characteristics. Then, I illustrate the main results of the empirical analysis and the robustness checks that justify the identification strategy.

CHAPTER I: Financial Literacy

Literature review

The research interest of economists around financial literacy rose in the last two decades, in conjunction with the higher relevance and connection between households' financial choices and households' well-being. Moreover, financial markets, financial products and financial investments have become more accessible to the population thanks to the large diffusion of the internet, personal computers and web-based investment services. However, the complexity of the financial markets has increased too, and therefore the probability of bad choices increases when agents are not well financially educated.

The importance of financial literacy has increased considerably with the evolution of financial markets which have become more and more accessible for individuals. Overall, the context has positive implications as it provides people with the chance of being financially proactive and of being responsible for their wealth management and financial planning. On the other hand, it could lead to negative outcomes and potentially disruptive consequences since individuals need specific financial knowledge and ability to plan in the long horizon.
All these problems together raise the interest of economists and researchers in financial literacy and its consequences on the well-being of agents, individuals and households. I consider "financial literacy" an issue because many studies found evidence that financial illiteracy is a widespread phenomenon and a large fraction of the population ignores even the most elementary economic concepts. (Anthes (2004), Lusardi and Mitchell (2007)).

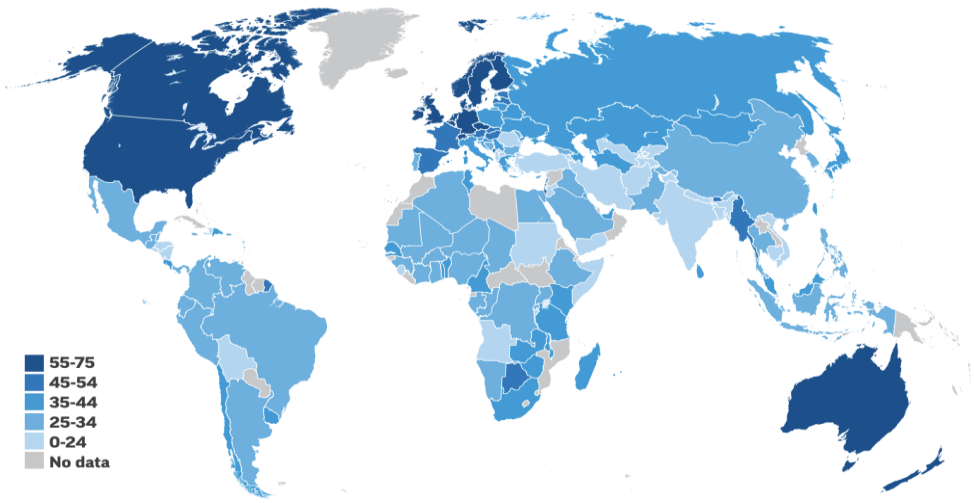

Figure 1: *% of Adults who are financially literate (Source: S&P Global FinLit Survey 2014)*

Figure 1 provides an overview of the degree of financial literacy of adults across the world. The dark blue identifies the regions where the share of the adult population with an acceptable level of financial literacy is high, while the light blue identifies those countries where financial literacy is low. The first group is concentrated in North America, Northern Europe and Oceania, while, at the bottom of the scale, there are many Middle East countries, as well as some South American and African ones. The group with the intermediate graduation of blue include countries spread in Middle and South America, Southern Europe and in the majority of Asia and Africa.

The definitions of financial literacy are heterogeneous, and it still does not exist a uniquely used and recognized one. The OECD INFE[1] defines it as: *"a combination of awareness, knowledge, skill, attitude and behavior necessary to make sound financial decisions and ultimately achieve individual financial wellbeing"*. The next table, retrieved from Hung et al. (2009) working paper, collects some of the most popular definitions that have been given to such a concept. The definition provided by Lusardi and Mitchell (2007c), focused on the term *familiarity* with *"the most basic economic concepts needed to make sensible saving and investment decisions"* is the one which is able to capture the essence of the concept in the simplest way. As one can see there are many other interpretations which are given from slightly different perspectives. Lusardi and Tufano (2008) define just *debt literacy*, a component of the overall financial knowledge. Lusardi's one (2008a and 2008b) revolves around the knowledge of the concepts behind the Big Three questions on financial literacy. Mandell's (2007) definition of FL is a more complete one as it considers both "*the ability to evaluate the new and complex financial instruments"* and "*make informed judgments"*.

[1] The Organization for Economic Co-operation and Development / International Network on Financial Education

Source	Conceptual Definition[a]
Hilgert, Hogarth, & Beverley (2003)	Financial *knowledge*
FINRA (2003)	"The *understanding* ordinary investors have of market principles, instruments, organizations and regulations" (p. 2).
Moore (2003)	"Individuals are considered financially literate if they are competent and can demonstrate they *have used knowledge* they have learned. Financial literacy cannot be measured directly so proxies must be used. Literacy is obtained through practical *experience* and active *integration of knowledge*. As people become more literate they become increasingly more financially sophisticated and it is conjectured that this may also mean that an individual may be more competent" (p. 29).
National Council on Economic Education (NCEE) (2005)[b]	"*Familiarity* with basic economic principles, knowledge about the U.S. economy, and *understanding* of some key economic terms" (p. 3).
Mandell (2007)	"The *ability* to evaluate the new and complex financial instruments and *make informed judgments* in both choice of instruments and extent of use that would be in their own best long-run interests" (pp. 163-164).
Lusardi and Mitchell (2007c)	[*Familiarity*] with "the most basic economic concepts needed to make sensible saving and investment decisions" (p. 36).
Lusardi and Tufano (2008)	Focus on debt literacy, a component of financial literacy, defining it as "the *ability to make simple decisions* regarding debt contracts, in particular how one *applies basic knowledge* about interest compounding, measured in the context of everyday financial choices" (p. 1).
ANZ Bank (2008), drawn from Schagen (2007)	"The *ability* to *make informed judgements* and to take effective decisions regarding the use and management of money" (p. 1).
Lusardi (2008a, 2008b)	"*Knowledge* of basic financial concepts, such as the working of interest compounding, the difference between nominal and real values, and the basics of risk diversification" (p. 2).

Table 1: *Conceptual definitions of financial literacy - Hung et al. (2009) - RAND Corporation Publications Department, Working Papers, Vol.708*

Besides its definition, what is clear is that financial literacy has implications both on a systemic level and on a micro level. The diffusion of a reasonable level of financial literacy is now being globally accepted as a critical element for a sound financial regulatory and supervisory framework, as well as a fundamental factor of financial stability. In 2018, the Vice-President of Deutsche Bundesbank, Claudia Buch stated that *"financial illiteracy is an important channel through which financial instabilities can arise"*. Financially uninformed investors may expose

themselves to excessive financial risk, and this is the main source of financial instability cited by Claudia Buch. However, other forms of financial instability derive from poor financial literacy: retirement planning and excessive credit card debt are only two of these problems. Batsaikhan and Demertzis (2018) also emphasize the potential for financial literacy to boost the EU economic inclusion agenda as they found evidence of a link between inequality, poverty, social exclusion, immobility and scarce financial knowledge.

On a micro level, the role of financial literacy in investors' economic behavior and financial attitudes has been studied by an increasing number of researchers. Angrisani et al. (2016) emphasize the potential of FL to help shape personal future economic health as people who demonstrate higher levels of financial knowledge are expected to act in a more financially wise manner and hence to be characterized by a more solid financial status, less exposed to problems of financial distress. Lusardi and Mitchell (2014) contribute to show that more financially literate households save more and are more likely to accumulate wealth for themselves and to pass on their resources to the next generation. They came to this conclusion after seeing that FL was a key explanatory factor in the US wealth gap since it is positively correlated with wealth building.

Financial Literacy measurement

Measuring agents' financial literacy has been one of the main issues of economists in the past years. First, it is necessary to delimit financial knowledge and its components and then define the proper questionnaire. Indicators for such a quality are very different from each other. The first attempts to measure it belong to the CFA (Consumer Federation of America) and its Consumer Knowledge surveys in the early 1990s. These questionnaires were submitted to different layers of the population and incorporated many queries about personal finance.

Later in the years, Lusardi & Mitchell (2006) introduced the following three questions that today are the standard for future research and that addressed some of the most crucial aspects of finance and financial literacy: **(a) compound interest**, **(b) inflation** and **(c) risk diversification**.

a. *Suppose you had $100 in a savings account and the interest rate was 2% per year. After 5 years, how much do you think you would have in the account if you left the money to grow?*

- More than $102
- Exactly $102
- Less than $102

b. *Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, how much would you be able to buy with the money in this account?*

- More than today
- Exactly the same
- Less than today

c. *Buying a single company's stock usually provides a safer return than a stock mutual fund.*

- True
- False

Summing the other two questions added by the NFCS[2], about **(d) mortgage** and **(e) bond pricing** we obtain the so-called "Big Five" of financial knowledge which are summarized in the next graph.

d. *A 15-year mortgage typically requires higher monthly payments than a 30-year mortgage, but the total interest paid over the life of the loan will be less.*

- True
- False

e. *If interest rates rise, what will typically happen to bond prices?*

- They will rise
- They will fall
- They will stay the same
- There is no relationship between bond prices and the interest rate

[2] National Financial Capability Study of the Financial Industry Regulatory Authority (FINRA) Foundation

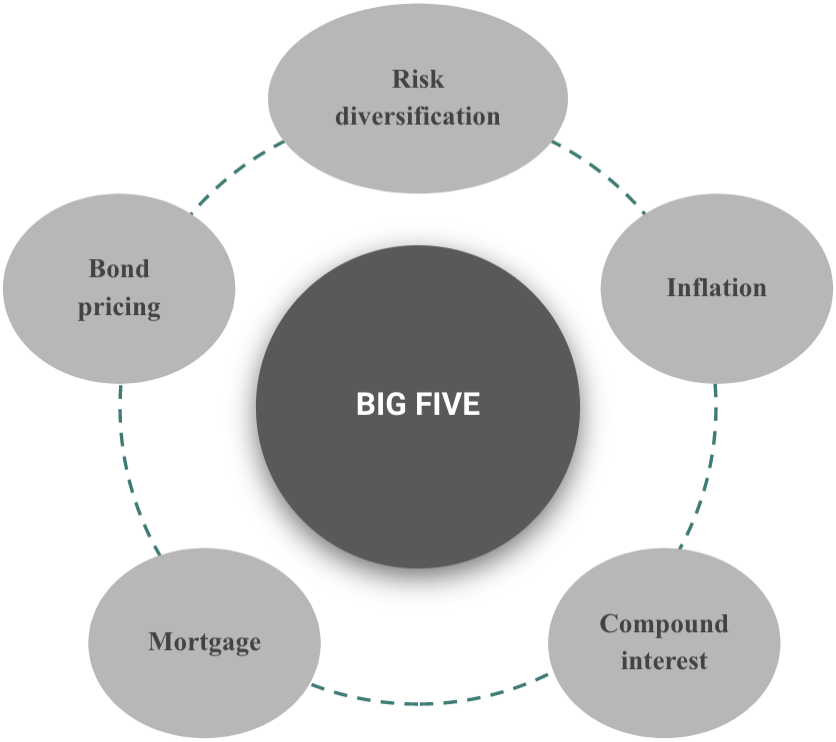

Figure 2: *The Big Five questions for Financial Literacy measurement (Source: Personal representation)*

Despite this set of questions seems to grant a secure way to trace financial literacy, there is, unfortunately, minor evidence on whether such an approach is the best among other possible ones, hence leaving the door open for discretion in the measurement process. Hung et al. (2009) analyzed 18 different studies and concluded that the leading method to manage the notion of FL is the sum or the ratio of the correct answers on some kind of performance test.
Another aspect to consider is the availability of information at disposal of the researchers since not all the surveys are structured in the same way and therefore don't provide the same kind of observations on individuals' financial knowledge. The latter augments the magnitude of FL measures' variability since inevitably some adaptations will take place in the definition of the score for financial knowledge.

OECD/INFE has developed a guideline for setting up a measure of FL that consists in summing up three different scores: a financial knowledge score, a financial behavior score and a financial attitude score. The maximum grade of a respondent is 21, however, the average score of the OECD countries participating in the study is just 12.7 (from 2020 data). These results show how financial illiteracy is still sizable.

Measuring financial literacy is a process in constant evolution as demonstrated by the continuous adjustments of the OECD questionnaires that are updated from year to year. For example, in 2022 questions related to digital financial literacy were added. This is a consequence of the transformation of financial markets and the growth of new types of investments such as cryptocurrency or peer-to-peer lending.

The fact is that here financial literacy is intended to capture a broader spectrum of aspects with respect to what is strictly concerned with the understanding of simple economic principles and in a sense, it sneaks away from what originally intended. The financial knowledge component of OECD is probably the one that reconciles with it.

Drivers and impact of Financial Literacy

Many research papers agree that financial literacy skills, regardless of how they are derived, are affected by age, educational attainment, income and labor status.

More in detail:

- age seems to have a U-shaped relationship with FL as the young and the elderly display less financial literacy while the most knowledgeable group is the one of middle–aged;
- education is positively correlated with financial literacy;
- income is positively correlated with financial literacy;
- self-employed are in general more financially literate than employed workers, while both are more financially literate than unemployed people.

The previously stated drivers do not encounter any opposition even when studied in minor countries like how Morgan and Trinh (2016) did for Cambodia and Vietnam.

Some other elements like gender, place of residence and ability collect still contrasting evidence.

Another factor that strongly affects individual financial literacy is gender: the vast majority of the financial literacy literature documented a gender gap in financial. Males generally show higher levels of financial knowledge than females, controlling for age and education. Nevertheless, there is some growing sign that these differences may be due to cultural factors, especially in some regions of the world. Hospido et al. (2021) assess that gender gaps are significantly smaller in Spanish autonomous regions with more egalitarian economic regimes in terms of custody and marriage, suggesting that social norms may be important to explain these disparities.

Country of birth represents another factor that may influence the level of individual financial literacy, however, there is not a common conclusion across studies. Lusardi and Mitchell (2011) recall that there exist some regional and ethical differences, e.g.: city inhabitants are more informed than rural areas residents, while in the US there is a clear poorer financial literacy among Afro-Americans and Hispanics. Davoli and Rodriguez (2020) track down that in the US there's a 4% additional likelihood of responding correctly to the "Big Three" questions when the individual's country of ancestry level of financial literacy is a one-standard deviation above the mean showing that also origins play a role in determining the financial literacy score.

Last, individual skills and ability may play a role in the determination of financial literacy. Indeed, learning about financial markets and financial products is complicated and requires time and effort. Barboza et al. (2016) attest that those holding a top-level GPA display superior financial literacy scores. It has been shown that also PISA[3] math score positively correlates with financial literacy. Lusardi (2012), indeed, explores the role of numeracy as determinant of financial literacy and finds a considerable correlation between the two. Lastly, Fort et al. (2016) indicate how bank information policies have a substantial impact on FL but on a limited fraction of the population represented by over-sixties low educated individuals.

Policy interventions on financial education

Managing personal finances is an everyday task, however individuals are not generally trained to do so. Furthermore, there are some aspects and decision related to portfolio allocation and financial choices that are "one in a lifetime" moments, therefore individuals and households do not have any experience or specific training for that. Policy makers and experts have been progressively suggesting that financial literacy education might be the remedy for many of the financial difficulties experienced by individuals. The ultimate goal of financial education is to pave the way for financial empowerment of individuals through new acquired knowledge of financial terms and products. Several research studies support the notion that financial training would impact effectively on financial conduct.

Bernheim (1998) asserts that the lack of some form of guidance and preparation could explain why Americans tend to take poor financial actions, therefore education policy initiatives are needed to compensate for it. However, even after numerous field studies and experiments, economists do not agree on what is the ideal structure of any training courses on personal finance and its timing. This is because the investigations returned mixed evidence on "financial

[3] Programme for International Student Assessment

training" efficiency, some of them seemed to be promising but with margins of improvement while others seemed to provide scarce results.

For instance, Becchetti et al. (2011) conducted an experiment on a group of Italian high school students where some were interviewed twice and taught a course in finance in between, while the control group was just interviewed once. The impact of such training did not get empirical support once comparing the two groups FL scores, but provided some good news as it increased the interest on economic articles and the ability to comprehend them. Brown et al. (2016) spotted some improvement in the debt behavior of a large representative of American youth as they recorded that mathematics and financial learning reduce non-student debt dependence and enhance repayment scheduling. Surprisingly they also found that economic studies encourage a higher level of indebtedness and a larger probability of debt reimbursement issues, probably because of overconfidence.

The financial learning process is influenced by direct experience too. West and Friedline (2016) looked at the differences in economic behavior of lower-income millennials and discovered that those who were more financially capable had almost twice the chance to successfully face unexpected expenses and to have an emergency fund and 34% less likely to have an unsustainable level of debt. Moreover, being exposed to investing at school-age is considered to have a positive effect on financial knowledge as reported by a study of Hysmith (2020) in which he analyzes the impact of SMIF (Student Managed Investment Fund) involvement on financial literacy, satisfaction and general healthy financial habits.
Davoli and Hou (2021) show that educational systems' organization matters as they register that people studying in the soviet side of Germany manifest up to 20% lower financial literacy scores with respect to those not exposed to such educational system.

To sum up, financial education-related policy intervention has to address the basics of finance which are essential for daily financial management. Moreover, it should cover cognitive and non-cognitive factors since it has recently emerged that the latters might outweigh rational reasoning. What is less clear is the structure of the underlying programs, given such mixed evidence from the literature. Hence there's still the need for more information and evidence on educational programs.

CHAPTER II: Individuals' portfolios

The concept of diversification

"Don't put all your eggs in one basket" is a famous idiom whose meaning resembles the one behind portfolio diversification. In the world of investments, diversification is pursued to reduce the idiosyncratic risk of households' portfolios, allocating the wealth in such a way not to risk all in a single security or asset class.

The father of the studies on portfolio selection is Harry Markowitz who developed the so-called "Modern Portfolio Theory". This theory allows investors to set up a portfolio whose assets' composition maximizes expected returns for a specific given level of risk. One of the pillars of MPT is the assumption that all investors are risk-averse (which is not far from the reality) implying that for a certain level of return, they will always choose the less risky option.
The fundamental elements in Markowitz theory are **expected return** and **variance (risk).** The key takeaway is that each individual aims at maximizing his/her own portfolio returns while reducing the risk of its investments. One of the main results of Markowitz's theory underlines that a properly balanced portfolio includes assets whose returns are the least perfectly correlated. This condition allows to achieve the highest possible return bearing the smallest overall portfolio riskiness.
The correlation among the different securities and assets included in the portfolio plays a central role. Essentially, correlation is the measure of how two or more objects are related to one another. The coefficient of correlation between two random variables is defined as the ratio between:

$$\frac{\sigma_{\{y_1,y_2\}}}{\sigma_{\{y_1\}}\cdot\sigma_{\{y_2\}}}.$$

Where y_1 and y_2 represent the two random variables/financial assets included in the analysis.

In the context of investments, correlation between assets is positive when they move in the same direction (the value of both assets increases/decreases at the same time) and it is negative when they move contrary to each other (the value of one increases/decreases while the other one decreases/increases). Markowitz (1952) enforces the central role of this statistical parameter by proving that the definition of an optimal portfolio is a process that involves statistical measures

(expected return, variance and correlation) and in which the weight of a single asset's performance is less crucial than how it affects the whole investment portfolio.
For instance, the construction of an optimal portfolio is not simply the combination of n assets all with the same weight, even if in some cases it has been shown that such a naive 1/N rule seems to perform better than differently built portfolios.

The consequences of a bad portfolio choice and/or composition may be disruptive for a household, leaving its components in a poor financial situation. Friedman (1953) famous analogy[4] on the process that leads a billiard player to hit the ball could indicate in a sense that none of us need to be a financial expert to get to a well-diversified portfolio after a path of experiential learning. Literature provides some minor evidence on the experience effect on some economic behaviors. However, the fact is that some decisions like retirement savings, mortgages, real estate contracts, are generally one-in-a-lifetime choices, with little or no space for learning effects. Therefore, the space and the consequences of potential financial education programs is huge, given that it increases the efficiency and the results of households' financial choices. Then, it is crucial to handle the investment process following all the recipes to reach an optimal combination of assets.

As previously described, portfolio diversification is one of the main suggestions given by Markowitz portfolio theory as well as financial advisors. Therefore, agents may decide for a fully diversified portfolio, which means to hold a fraction of each product available on the financial market. However, it has been shown by Wagner and Lau (1971) that few stocks are sufficient to reach a reasonable degree of diversification when these are highly rated by S&P. Moreover, it is economically wise to reduce the number of securities in the portfolio to optimize transactional, informational and management costs considering that benefits from differentiation augments at a decreasing rate. Evans and Archer (1968) demonstrate that a limited number of securities is sufficient to capture most of the diversification advantages. Driessen and Laeven (2006) elicit the difference in benefits between investors in developing and developed countries, showing how the former have a large advantage from international diversification when considering short-sales constraints. They also find proof that the size of such advantages evolves as country risk varies overtime.

[4] The key takeaway was that billiard players do not take into account mathematical and physics law and all the complicated calculus behind before hitting the ball. They just do it as the experience has taught them to do.

There have been many upgrades of the original Markowitz perspective and the naive equally weighted portfolio like the ones of Jorion (1986), MacKinlay and Pastor (2000) and Kan and Zhou (2007) trying to improve performances, however the always optimal portfolio strategy hasn't been implemented yet.

Financial literacy and diversification

The investment behavior of individuals is affected by a variety of shortcomings. Hastings et al. (2013) point out some of them:

- low stock market participation,
- poor debt management: inability to refinance in changing interest rates environment,
- home bias: the predisposition in the choice and predominance of domestic assets within the portfolio, discarding international diversification,
- disposition effect, tendency to sell the winners and buy the losers (Odean 1998).

Under these circumstances, many studies have questioned whether financial literacy may have an impact on the investment process of individuals, both in a positive and negative direction. MacArdle et al. (2009) investigate the existence of a connection between cognitive skills and wealth accumulation in pre and post retirement and concluded that more capable individuals were more likely to build up wealth. The advantage of being cognitively skilled could increase as workers are increasingly required to govern their pension planning and household wealth in general. Indeed, an emerging threat that is affecting some areas of the world, especially the European continent, is the viability of public pension schemes as the age of the population is moving up. Hence private pension plans will probably represent a valid alternative tool to realize retirement savings. Christelis et al. (2008) reinforce the importance of cognitive abilities such as math, memory and verbal fluency as they find they are positively associated with stock market participation, both directly and indirectly via investment fund and pension accounts.
The last consideration that needs to be done is that households' financial decisions and trading habits depend not only on personal abilities but also on the socio-cultural environment into which they are lowered and on congenital traits (Gomes et al. (2020)).

Other researchers focused on the relationship between financial literacy and portfolio allocation. Guiso and Viviano (2015) demonstrated that high-literacy people have greater timing abilities when investing in the market and are more prone to follow the prescriptions of standard models. Several analyses exhibit that higher financial literacy goes hand in hand with superior cash management and more efficient planning for retirement.

Moreover, more financially literate individuals hold statistically a wider range of assets on their balance sheets. Van Rooji et al. (2011) use data about Dutch households to show that individuals with the previously mentioned characteristics tend to be invested in the equity market and to own a more diversified and cost-efficient portfolio. The cost-efficient hypothesis was confirmed also by Muller and Weber (2010) in a previous study where they underlined how the odds of investing in low-cost fund alternatives when choosing across different mutual funds.

Lack of financial literacy prevents effective use of various financial assets. If someone does not understand the functioning and the nature of the financial markets, he/she will lack the motivation and skills to diversify their investment strategies. Guiso and Jappelli (2008) were among the first to question the occurrence of an influence of financial literacy on portfolio compositions. They based their research on a UniCredit Customers' Survey and concluded that the degree of diversification strongly correlates with their measure of financial literacy. Abreu and Mendes (2010) expanded such findings relying on a Portugal sample and considering also the impact of source of information on the numerosity of assets to buy.

Again, De Winne and Petkeviciute (2021) based their study on more than 75.000 Belgian's trading records between 2003 and 2012. They found that financial literacy has a strong relationship with portfolio diversification, regardless of the measure used both for financial literacy and diversification. The most interesting result of their study is that the non-stocks holdings seem to be most positively impacted by financial knowledge. Furthermore, Hsiao and Tsai (2018) study, based on a Taiwanese National Survey, encountered some evidence of a higher participation rate in the derivatives markets of more financially literate individuals, most probably due to the reduction of the entry costs that FL helps them to reach.

Bonaparte (2021) focused on the crypto market participation of millennials showing that they are more likely to invest in these kinds of speculative assets as they tend to have shorter investment horizons than other generations. This evidence, together with the typical lower degree of financial knowledge of young people, should be a warning as the implications for their wealth accumulation "journey" might be serious. Koh et al. (2018) worked on a very detailed analysis distinguishing between complex and non-complex holdings of Singaporeans and the impact of financial literacy on such investments. They found that more financially literate investors are more likely to be simultaneously invested in equities, fixed-income and cash. Moreover, those obtaining a higher FL score are more likely to respect the fraction of equity on the overall asset portfolio that is generally suggested by financial advisors, that is 100 minus the age of the investor.

The portfolio returns of the more financially knowledgeable appear to benefit too: Von Gaudecker (2015) recorded an additional 50 basis point return for groups that score high financial literacy or rely on external financial advice. To finish, Bianchi (2018) investigated the relationship between returns and asset choice finding that more financially savvy investors have a risky approach when returns are higher, they actively intervene in the reallocation of their portfolio and are more likely to select assets with higher return than those that they sell.

This introduction about the main issues and advantages of financial literacy in relationship with portfolio diversification gives an overview of the relevance of this type of knowledge for households and individuals across life. However, the heterogeneity of the measure of financial literacy decreases the comparability of the studies and increases the difficulty of results interpretation and policy-intervention design.

Next chapters will present the main idea and the empirical results of this thesis, that deals with financial literacy and portfolio diversification. Based on a Spanish sample, I will try to answer the following questions: does financial literacy influence the variety of asset classes included in investors' portfolios? Is this impact of sizable magnitude or not? Are there any further possible reasons for such results? My definitions of both diversification and financial literacy differ slightly from the ones of the existing literature. The first is to be meant as the heterogeneity of asset families held by an individual, not as the number of stocks or just the ownership of mutual funds. The second will be measured as the sum of correct answers to a mix of standard questions and some other ones related with numeracy.

CHAPTER III: Data and methodology

ECF Questionnaire (aim, history, structure)

I use the Spanish Survey of Financial Competences (hereafter ECF) to study the relationship between financial literacy and individual portfolio diversification.
The Survey of Financial Competences is a joint initiative of the Banco de España (BdE) and the CNMV[5] under the Financial Education Plan. The purpose of this cooperation is to boost the financial education of Spanish citizens and help them develop the necessary skills needed to improve their financial choices.

The Financial Education Plan was signed back on 19 May 2008. It was started "*with the aim of helping improve the financial education of citizens, providing them with the tools, knowledge and skills to adopt appropriate, informed financial decisions throughout their life"* as stated in the CNMV official website. It is implemented according to the OECD guidance and High-Level Principles on public strategies for financial training dictated by OECD/INFE (International Network for financial Education) which were strongly supported in the G20 Summit of 2012. The Financial Education Plan includes a variety of educational projects other than the ECF, e.g. the website www.finanzasparatodos.es where whoever is interested can grasp information and learn financial notions independently from his/her education background.

Figure 3 shows that the number of participants to the mission of FEP is growing over time, highlighting the growing interest and necessity of financial knowledge in the population as well as the relevance of the effort of policymakers in spreading financial competences.

[5] Comisión Nacional del Mercado de Valores

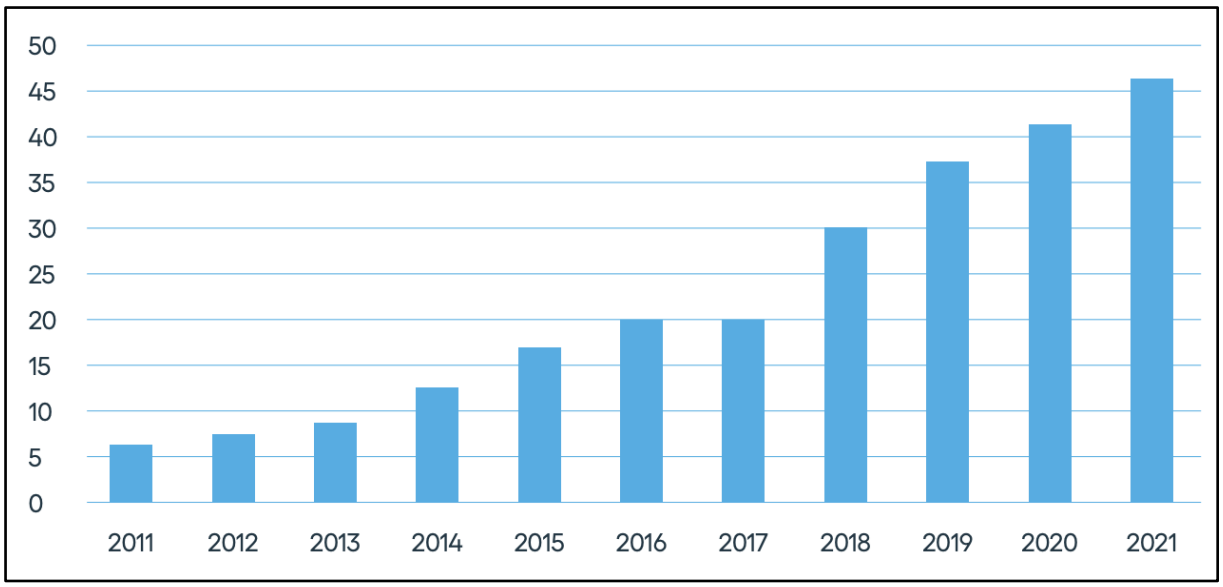

Figure 3: *Increase in number of FEP collaborators (Source: FEP 2022-2025)*

As previously said, the Survey of Financial Competences (ECF) is one of the initiatives of this national Plan. Included also in the National Statistics Plan, this survey aims to offer an exhaustive portrait of the financial literacy and financial knowledge of Spaniards aged 18 to 79. It is part of a cross-continental campaign coordinated by the INFE allowing to compare the financial skills of Spanish respondents with the financial skills of other countries.
Even if the ECF is part of the INFE study, it presents some unique peculiarities.
First, it extracts not only individual respondents' financial competences but also the one at household level when the interviewee is not aware of household finance. Second, it includes measures of cognitive skills using adapted specific questions from other international studies. These questions measure reading comprehension, graph understanding and numeracy. Third, it contains a module on housing limited to what concerns the main residence. Finally, it provides data about the expectations of respondents on matters like probability of finding a job, evolution of main residence value and so on.

The ECF survey has two waves, the first run in 2016 and the second in 2021. However, the data of the second one are not yet available, therefore this thesis uses the 2016 survey data release. The goal of such a questionnaire is to track the financial knowledge of the adult population in Spain allowing also for regional specification.

The National Statistics Institute is responsible for the sample selection, which is representative of the whole Spanish territory and of each of its comunidades autónomas. It is relevant that the sample is balanced "by region" because in Spain regional authorities are in charge of education policies that generally have strong effects on financial literacy.

Moreover, the heterogeneity of the socio-economic traits of the population is high across regions in Spain, and also this may influence individual financial knowledge.
The sample is obtained through a two-steps process, stratified in the first passage by region and municipality size. The first-step units correspond to the census sections while the second-step ones are the people between 17 and 79 years old who reside in those sections. The interviews of the first edition were conducted between the last quarter of 2016 and the first semester of 2017 via Computer Assisted Personal Interviews (CAPI).
The sample is composed of 8952 individuals, however further cleaning and control procedures drop about 400 interviews and the sample reduces to 8554 validated interviews.

The ECF questionnaire is organized in the following eight sections:

- Demographic characteristics and labor market status of interviewees
- Portfolio of respondents and sources of information exploited
- Source of income of unemployed (both retired and inactive)
- Saving habits
- Financial literacy
- Economic literacy
- Main residence
- Household budgeting and financial weakness.

The first six sections collect data from individuals, i.e. survey respondents, while the information related to household budgeting and financial weakness, in the last two sections, are at the household level.

Descriptive statistics

This section presents the basic demographics of the ECF sample: age, education, employment status and income.

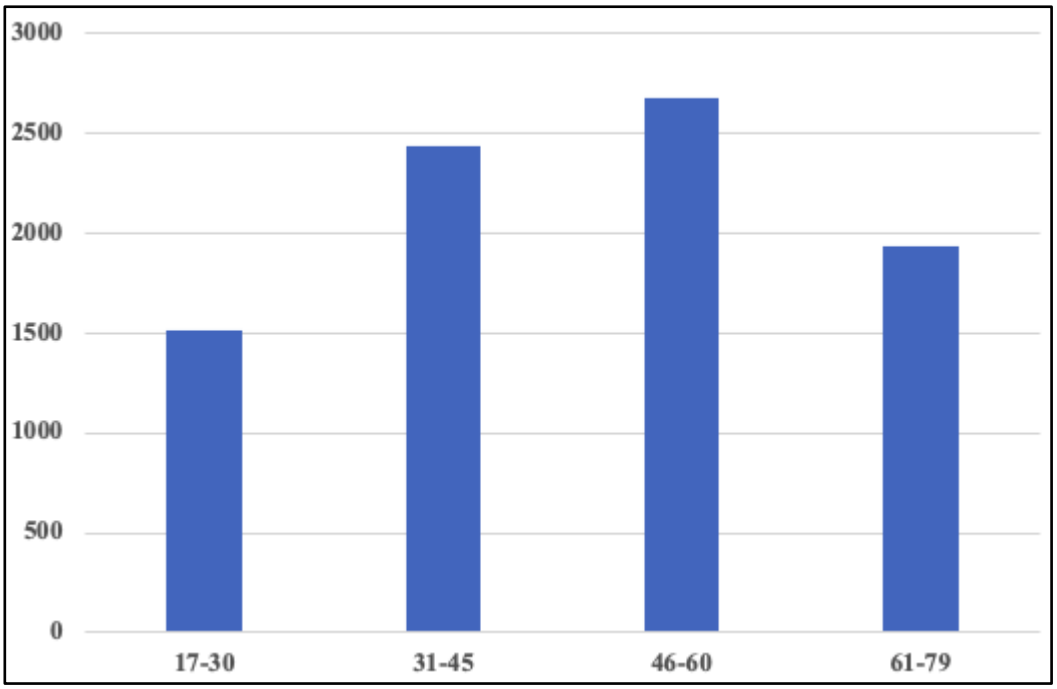

Figure 4: *Age distribution*

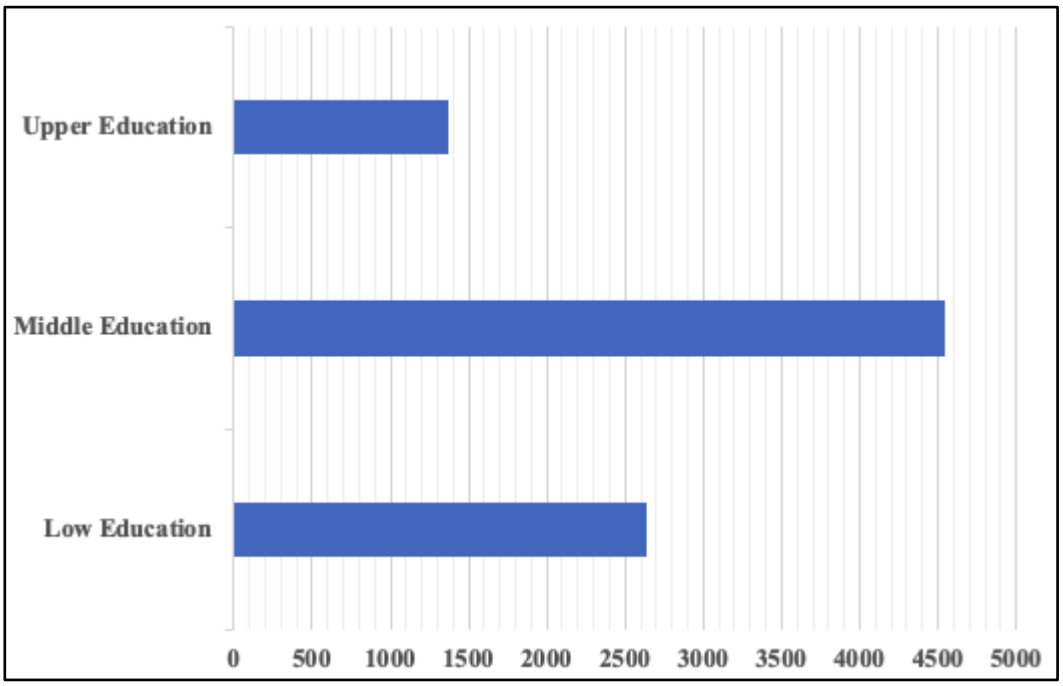

Figure 5: *Education level of the sample*

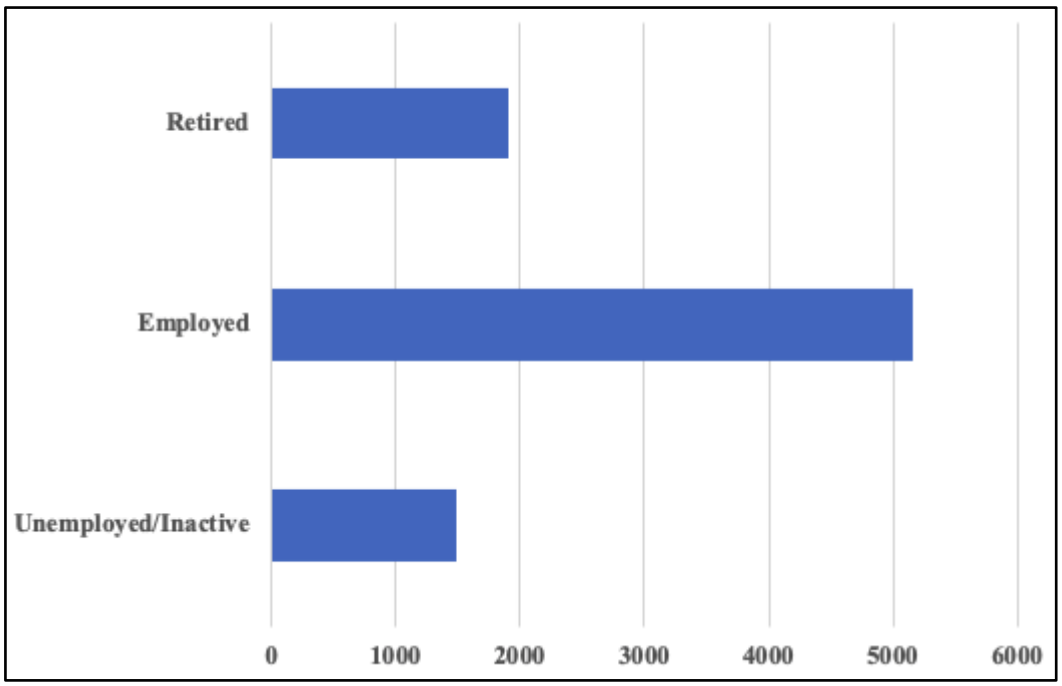

Figure 6: *Occupational Status*

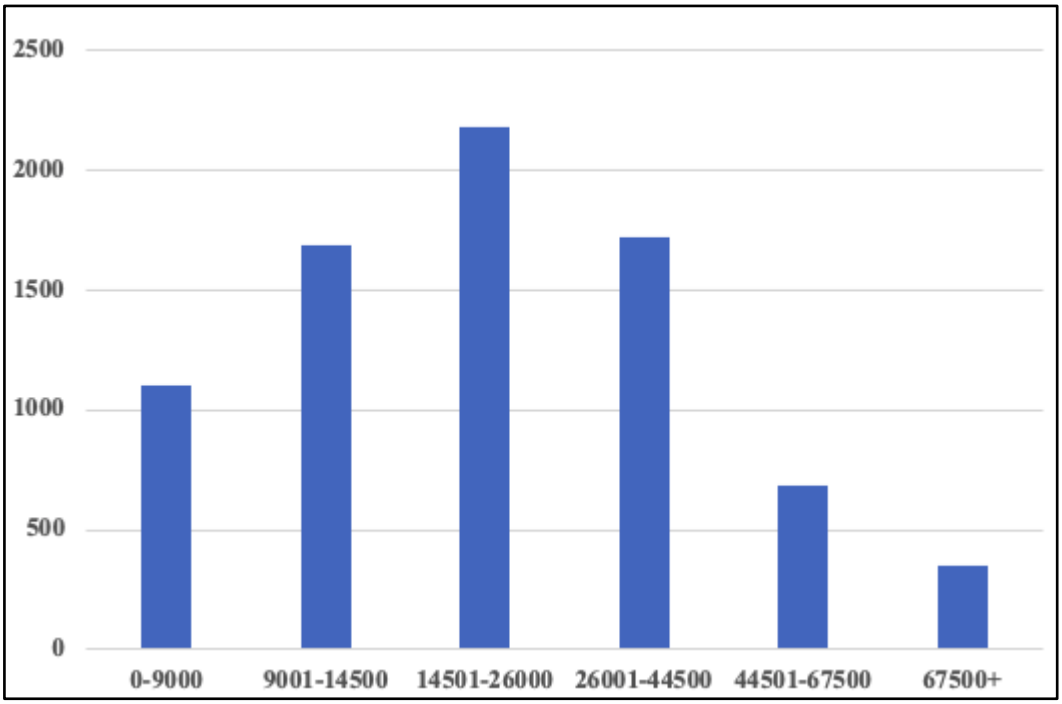

Figure 7: *Income distribution*

The sample is well balanced in terms of gender as it is composed of 4264 males and 4290 females. The average age of the respondents is 47 years: **Figure 4** shows that almost 60% of the participants is between 31 to 60 years old (31-45 are 28%, while 46-60 are 31% of the entire sample). I divided education into three categories, low, middle and upper education, corresponding to the following level of studies: low means primary school or lower, middle identifies those with a secondary school diploma while upper education represents individuals with university degree or doctorate.

Figure 5 shows that there is a significant percentage share of individuals with just primary education (30.8%), while, in line with the major European countries, more than half belongs to the middle-educated category and 16% of interviewees has at least a university degree.

Figure 6 displays the employment status of respondents, grouped in three macro-categories: employed and self-employed, unemployed and retired. However, unemployment includes a variety of other situations like disability (and so on), therefore the high share of unemployed respondents is not a problem.

Last, **Figure 7** presents the data describing income distribution among the sample. The survey does not ask the exact monthly/yearly income, instead the respondents choose between six income intervals (0-9000, 9001-14500, 14501-26000, 26001-44500, 44501-67500, 67500+). This strategy allows the reduction of non-responses, which is a serious problem for this type of question given that individuals are generally reluctant to give information about their income. Almost 75% of respondents' income lies between the 9000 and 44500 euros, with the most populated income tier being the 14501-26000 (28.2%). The low-income respondents represent 14.2% of the sample, while the two highest income intervals account for the remaining 13.4%, with top paid individuals being less than 5% of the entire sample.

Table 2 provides the same kind of information grouping for autonomous regions and expressing all in percentage. Andalusia, Madrid and Valencia are the autonomous regions with the larger weight in the sample composition, while Catalonia is the one with the lowest number of interviewees. The balance in the gender of respondents is stable across all the autonomous regions. The age distribution is very similar in all the Spanish regions.

The aspect that seems to be more heterogeneous is the education level of respondents. For instance, the portion of those having just a primary education varies from 24% in Navarre to 40% in Andalusia. The middle-educated subgroup portion ranges between 43% and 60%. Lastly, the top educated subgroup is most populated in regions like Asturias (23%) and least populated in the Canary Islands (11%). There is variability also in the occupational status of respondents across regions. As previously stated, the majority belongs to the employed class with percentages ranging from 56% to 66%. Then there is the retired category that accounts for 14% to 30% of the cases and finally the unemployed whose weight is the lowest in Madrid (9%) and highest in Andalusia (24%). Lastly, the income distribution of respondents varies a lot across regions with more than 80% of the respondents belonging to the first four income levels (from 0 to 44500). These demographics are given a general overview for the imputed version of the dataset, too (see Annex 1).

	Andalusia	Aragón	Asturias	Balearic Islands	Canary Islands	Cantabria	Castile-León	Castile-La Mancha	Catalonia	Valencia	Extremadura	Galicia	Madrid	Murcia	Navarre	Basque Country	La Rioja
Relative weight	0,11	0,05	0,05	0,03	0,04	0,03	0,06	0,06	0,1	0,08	0,06	0,07	0,09	0,04	0,03	0,05	0,04
GENDER																	
Male	0,49	0,51	0,52	0,53	0,5	0,49	0,51	0,51	0,5	0,49	0,48	0,49	0,5	0,5	0,46	0,53	0,49
Female	0,51	0,49	0,48	0,47	0,5	0,51	0,49	0,49	0,5	0,51	0,52	0,51	0,5	0,5	0,54	0,47	0,51
AGE																	
17-30	0,21	0,19	0,14	0,15	0,22	0,14	0,13	0,19	0,17	0,19	0,19	0,15	0,19	0,2	0,17	0,18	0,18
31-45	0,27	0,28	0,28	0,31	0,23	0,32	0,26	0,3	0,3	0,3	0,26	0,26	0,3	0,31	0,3	0,3	0,29
46-60	0,31	0,28	0,29	0,31	0,37	0,33	0,34	0,31	0,29	0,31	0,34	0,31	0,32	0,29	0,3	0,31	0,34
61-79	0,22	0,25	0,29	0,23	0,17	0,21	0,28	0,2	0,24	0,2	0,22	0,28	0,19	0,2	0,23	0,21	0,2
EDUC																	
Low Education	0,4	0,28	0,33	0,26	0,38	0,28	0,28	0,33	0,26	0,31	0,39	0,29	0,29	0,3	0,24	0,25	0,27
Middle Education	0,43	0,55	0,44	0,57	0,51	0,57	0,54	0,54	0,56	0,54	0,48	0,5	0,57	0,57	0,6	0,6	0,6
Upper Education	0,16	0,17	0,23	0,17	0,11	0,15	0,18	0,13	0,18	0,15	0,13	0,21	0,14	0,13	0,16	0,15	0,13
STATUS																	
Unemployed/Inactive	0,24	0,17	0,17	0,2	0,21	0,12	0,17	0,21	0,17	0,17	0,21	0,21	0,09	0,17	0,14	0,12	0,14
Employed	0,56	0,59	0,6	0,65	0,56	0,61	0,63	0,6	0,64	0,58	0,61	0,6	0,58	0,64	0,6	0,58	0,66
Retired	0,2	0,24	0,23	0,14	0,22	0,26	0,2	0,19	0,2	0,24	0,18	0,19	0,32	0,19	0,27	0,3	0,2
INCOME																	
0-9000	0,18	0,08	0,13	0,14	0,3	0,08	0,13	0,13	0,1	0,15	0,18	0,11	0,1	0,15	0,08	0,07	0,09
9001-14500	0,24	0,16	0,17	0,23	0,22	0,18	0,2	0,24	0,16	0,26	0,25	0,2	0,14	0,24	0,17	0,12	0,14
14501-26000	0,22	0,29	0,24	0,22	0,2	0,26	0,27	0,28	0,3	0,22	0,24	0,29	0,23	0,3	0,25	0,25	0,27
26001-44500	0,15	0,26	0,2	0,17	0,13	0,3	0,2	0,21	0,18	0,19	0,16	0,2	0,23	0,16	0,23	0,26	0,24
44501-67500	0,04	0,1	0,09	0,1	0,06	0,05	0,08	0,06	0,09	0,06	0,05	0,07	0,12	0,06	0,13	0,14	0,1
67500+	0,02	0,05	0,05	0,02	0,03	0,05	0,04	0,02	0,04	0,04	0,02	0,03	0,09	0,03	0,03	0,07	0,04

Table 2: *Demographics of sample grouped by autonomous regions*

Imputation

The ECF has a high number of missing answers among questions that are particularly relevant for this study. There are three main possible ways of handling this issue. The first and predominant approach is the so-called *listwise deletion method*, i.e. excluding observations with missing values and using only complete data. This way of proceeding causes an efficiency cost due to the loss of observations. If I had to consider just the complete records, I will end up with just 60% of the original sample, with a possible problem of self-selection or incidental truncation of the observation. Another aspect to consider is that typically the observations which lack completeness are particularly informative for the outcome of the dependent variable and their exclusion might cause systematic bias to the analyses.

The *omitted variable method* is the second common approach. It consists of the exclusion of those variables that register some missing values across the sample. In the current empirical analysis, this cannot be considered an option since the variables with missing data are needed for the definition of dependent variable, thus eliminating them from the dataset would block me from any kind of model implementation.

The third alternative is to apply some kind of *imputation* methodology. One of the possibilities is to assume the missing values to take some ad-hoc value. In particular, it could consist of imputing zeroes for all missing values or to apply a binomial random imputation algorithm to randomly allocate zeroes and ones to missing data to discrete variables. Instead, in many cases researchers have made use of mean imputation when considering continuous variables, namely their missing values are replaced by the mean of the observed values.

Regarding the last option, a more complete approach is the multiple imputation method. It basically looks for matching observations from the similar but complete observations. I decide to approach this problem using the technique of multiple imputation, that allows the estimation of the missing value by comparing the responses given by their peers and their characteristics with the ones of the non-respondents. In many other economic studies, the analyses are implemented on top of already existing imputed data that are made available by the organization in charge of the particular survey or questionnaire. Differently from those, I have conducted this operation personally to monitor the correct execution of the replacement of missing values, to avoid misspecification in the single variable imputation method in order to end up with a feasible complete dataset.

More in detail, the need for this kind of procedure is due to a significant portion of missing data related to the kind of assets held in individuals' investment portfolios. This is the case since before asking whether individuals hold a specific asset class, they are asked whether they have heard of it or not. When the question: "*Have you ever heard of ..?"* is answered negatively or not answered at all, the next question about whether the interviewee held such a kind of asset in his/her portfolio is not asked.

If I had to consider just the complete responses to such questions, I would come up with a sample of 5011 individuals, which would mean a reduction of a bit more than 40% of the entire sample. This magnitude of drop would most probably lead to a self-selection of the sample and consequent biases in the models' estimates.

It might be argued that a reasonable imputation would have been to simply consider as "No" the answer to those questions to substitute missing values. However, with the support of the literature and given the level of complexity in the financial context, the imputation process has been run without any constraint. Indeed, it happens that in the name of diversification, inexperienced or poorly advised people can end up investing in assets that are unknown to them, so it is not unreasonable to have people who say that they do not know a particular security class but that at the same time are invested in it!

Moreover, Table 3 shows that a large share of the answers related to the ownership of some specific financial products is positive (i.e., most respondents answer "yes"). In particular, 90% or more of the interviewees are aware of products like: mortgages, personal/occupational schemes, shares, personal loans, credit cards, life insurance and medical insurance. Only 3 out of 4 individuals have heard of saving/term deposits. Investment funds are recognized by 85% of the sample and public/private fixed income by only 78%. Table 4 gives us a picture of the rate[6] of affirmative and negative responses to the questions "*At present do you personally or jointly have any ... ?"* for each asset class. Credit cards, mortgages, life insurance, saving/term deposits are the most popular financial products by holdings and affirmative response rate. Riskier financial asset classes like shares, investment funds and fixed income record lower participation rates. In line with existing literature, education and income have a positive impact on the rate of people who hold such investments. Indeed, when the degree of education increases, the column "Yes" shows higher percentages and the same is valid for higher income tiers. On the other hand, more educated and high-income respondents rely to a lesser extent on asset classes like mortgages and personal loans.

[6] Percentage is calculated on those who were asked that question, not on the full sample.

To reinforce the choice of proceeding with an ad-hoc imputation procedure, I could argue that the percentage of ignorance of fixed income products is not plausible as mutual funds are even more complex instrument and most of the respondents know the existence of this financial product. Considering that in the ECF survey fixed income product holdings are meant to be personal holdings and not via investment funds, we should expect that those knowing investment fund (that could be entirely stocks, only fixed income or hybrid) should be aware also of such category of asset class.

Have you heard of ...? (Yes/No)		Saving/Term Deposits		Mortgage		Personal/Occupational scheme		Investment fund		Shares		Public/private FI		Personal Loan		Credit Card		Life Insurance		Medical Insurance	
		Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Total	Totals	6339	2204	8424	130	7695	855	7303	1244	7716	835	6686	1864	8122	429	8339	215	8326	227	8088	466
Total %		0,74	0,26	0,98	0,02	0,90	0,10	0,85	0,15	0,90	0,10	0,78	0,22	0,95	0,05	0,97	0,03	0,97	0,03	0,95	0,05
GENDER																					
Male	4264	0,73	0,26	0,98	0,01	0,91	0,09	0,87	0,13	0,92	0,08	0,80	0,20	0,95	0,05	0,98	0,02	0,98	0,02	0,94	0,06
Female	4290	0,75	0,25	0,99	0,01	0,89	0,11	0,84	0,16	0,89	0,11	0,76	0,24	0,95	0,05	0,97	0,03	0,97	0,03	0,95	0,05
AGE																					
17-30	1516	0,71	0,29	0,98	0,02	0,82	0,18	0,81	0,19	0,88	0,12	0,62	0,38	0,92	0,08	0,98	0,02	0,97	0,03	0,94	0,06
31-45	2433	0,79	0,21	0,99	0,01	0,91	0,08	0,87	0,13	0,92	0,08	0,81	0,19	0,97	0,03	0,98	0,02	0,98	0,02	0,97	0,03
46-60	2674	0,75	0,25	0,99	0,01	0,94	0,06	0,89	0,11	0,92	0,08	0,85	0,15	0,97	0,03	0,98	0,02	0,98	0,02	0,96	0,04
61-79	1931	0,69	0,30	0,98	0,02	0,89	0,11	0,82	0,18	0,88	0,12	0,77	0,23	0,93	0,07	0,95	0,05	0,95	0,05	0,91	0,09
EDUC																					
Low Education	2633	0,67	0,33	0,98	0,02	0,84	0,16	0,79	0,20	0,86	0,14	0,69	0,31	0,92	0,08	0,97	0,03	0,96	0,04	0,93	0,07
Middle Education	4545	0,79	0,21	0,99	0,01	0,93	0,07	0,89	0,11	0,93	0,07	0,83	0,17	0,97	0,03	0,98	0,02	0,98	0,01	0,97	0,03
Upper Education	1376	0,71	0,29	0,98	0,02	0,90	0,10	0,84	0,16	0,90	0,10	0,80	0,20	0,94	0,06	0,95	0,05	0,96	0,04	0,91	0,09
STATUS																					
Unemployed/Inactive	1486	0,55	0,45	0,95	0,05	0,75	0,25	0,67	0,33	0,75	0,25	0,59	0,41	0,87	0,12	0,92	0,08	0,92	0,08	0,85	0,15
Employed	5157	0,75	0,25	0,99	0,01	0,91	0,09	0,87	0,13	0,92	0,08	0,78	0,22	0,96	0,04	0,98	0,02	0,98	0,02	0,96	0,04
Retired	1911	0,88	0,12	1,00	0,00	0,98	0,02	0,96	0,03	0,98	0,02	0,93	0,07	0,98	0,02	0,99	0,01	1,00	0,00	0,99	0,01
INCOME																					
0-9000	1100	0,55	0,44	0,95	0,05	0,74	0,26	0,66	0,34	0,73	0,27	0,59	0,41	0,88	0,12	0,92	0,08	0,93	0,07	0,87	0,13
9001-14500	1685	0,65	0,35	0,98	0,02	0,85	0,15	0,79	0,21	0,86	0,14	0,69	0,31	0,93	0,07	0,97	0,03	0,96	0,04	0,91	0,09
14501-26000	2177	0,76	0,24	0,99	0,01	0,92	0,08	0,89	0,11	0,93	0,07	0,80	0,20	0,97	0,03	0,99	0,01	0,98	0,02	0,96	0,04
26001-44500	1723	0,85	0,15	1,00	0,00	0,98	0,02	0,95	0,05	0,98	0,02	0,90	0,10	0,98	0,02	0,99	0,01	1,00	0,00	0,99	0,01
44501-67500	685	0,89	0,11	1,00	0,00	0,99	0,01	0,98	0,02	0,99	0,01	0,95	0,04	0,99	0,01	1,00	0,00	1,00	0,00	0,99	0,01
67500+	350	0,90	0,10	1,00	0,00	0,99	0,01	0,99	0,01	0,99	0,01	0,95	0,05	1,00	0,00	1,00	0,00	0,99	0,01	0,99	0,01

Table 3: *Knowledge rate per asset class (original dataset)*

Do you currently hold...? (Yes/No)		Saving/Term Deposits		Mortgage		Personal/Occupational scheme		Investment fund		Shares		Public/private FI		Personal Loan		Credit Card		Life Insurance		Medical Insurance	
		Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Total	Totals	2249	4063	2709	5708	1837	5841	783	6477	1183	6517	114	6548	1500	6610	4782	3540	2559	5700	1648	6423
Total %		0,36	0,64	0,32	0,68	0,24	0,76	0,11	0,89	0,15	0,85	0,02	0,98	0,18	0,82	0,57	0,43	0,31	0,69	0,20	0,80
GENDER																					
Male	4264	0,27	0,48	0,32	0,67	0,19	0,70	0,09	0,75	0,12	0,77	0,01	0,75	0,16	0,79	0,54	0,43	0,27	0,70	0,19	0,76
Female	4290	0,26	0,47	0,31	0,66	0,23	0,66	0,09	0,76	0,16	0,75	0,01	0,78	0,19	0,75	0,57	0,39	0,33	0,63	0,19	0,74
AGE																					
17-30	1516	0,18	0,53	0,07	0,91	0,03	0,79	0,02	0,79	0,04	0,84	0,00	0,62	0,10	0,82	0,43	0,55	0,17	0,78	0,17	0,75
31-45	2433	0,28	0,50	0,50	0,49	0,21	0,70	0,07	0,79	0,12	0,79	0,01	0,80	0,23	0,74	0,62	0,36	0,44	0,53	0,23	0,74
46-60	2674	0,28	0,47	0,41	0,58	0,35	0,58	0,11	0,77	0,17	0,74	0,02	0,84	0,22	0,75	0,64	0,34	0,36	0,62	0,20	0,76
61-79	1931	0,28	0,41	0,15	0,83	0,17	0,71	0,14	0,67	0,18	0,69	0,03	0,74	0,11	0,82	0,47	0,48	0,14	0,81	0,15	0,76
EDUC																					
Low Education	2633	0,19	0,48	0,20	0,77	0,10	0,74	0,06	0,73	0,08	0,77	0,01	0,68	0,11	0,81	0,40	0,56	0,17	0,78	0,12	0,80
Middle Education	4545	0,30	0,49	0,43	0,56	0,30	0,63	0,10	0,79	0,15	0,77	0,01	0,82	0,23	0,74	0,67	0,32	0,42	0,56	0,24	0,72
Upper Education	1376	0,30	0,41	0,15	0,83	0,16	0,75	0,15	0,69	0,19	0,71	0,03	0,76	0,12	0,82	0,50	0,45	0,14	0,81	0,16	0,76
STATUS																					
Unemployed/Inactive	1486	0,14	0,41	0,17	0,78	0,09	0,66	0,04	0,63	0,05	0,69	0,01	0,58	0,12	0,75	0,29	0,63	0,14	0,77	0,06	0,79
Employed	5157	0,25	0,50	0,33	0,66	0,20	0,71	0,08	0,78	0,12	0,80	0,01	0,77	0,19	0,76	0,56	0,42	0,30	0,67	0,17	0,78
Retired	1911	0,41	0,47	0,40	0,60	0,36	0,62	0,17	0,79	0,27	0,71	0,03	0,90	0,17	0,81	0,76	0,23	0,42	0,57	0,35	0,64
INCOME																					
0-9000	1100	0,10	0,45	0,16	0,79	0,07	0,67	0,02	0,64	0,03	0,70	0,00	0,58	0,11	0,77	0,29	0,63	0,15	0,77	0,06	0,81
9001-14500	1685	0,16	0,49	0,24	0,74	0,11	0,75	0,05	0,74	0,06	0,81	0,01	0,68	0,17	0,75	0,41	0,56	0,19	0,76	0,10	0,81
14501-26000	2177	0,25	0,51	0,32	0,67	0,19	0,73	0,07	0,81	0,10	0,82	0,01	0,79	0,20	0,77	0,58	0,41	0,30	0,67	0,16	0,80
26001-44500	1723	0,38	0,47	0,41	0,59	0,29	0,68	0,14	0,81	0,20	0,78	0,02	0,88	0,19	0,79	0,71	0,28	0,39	0,59	0,27	0,71
44501-67500	685	0,43	0,46	0,49	0,51	0,45	0,54	0,18	0,79	0,31	0,68	0,02	0,93	0,22	0,77	0,82	0,17	0,48	0,51	0,38	0,61
67500+	350	0,47	0,42	0,47	0,53	0,56	0,43	0,25	0,73	0,43	0,56	0,05	0,90	0,24	0,75	0,85	0,15	0,55	0,43	0,49	0,50

Table 4: *Holdings rate per asset class (original dataset)*

Imputation procedure

After generating five imputed datasets, I ran numerous post-imputation tests and checked the resemblance in the proportion of affirmative answers in the "asset holding" questions to avoid unrealistic changes in the rate of affirmative responses.
The imputation procedure observes the steps and the suggestion by Stef van Buuren, Professor of Statistical Analysis of Incomplete Data at the University of Utrecht.

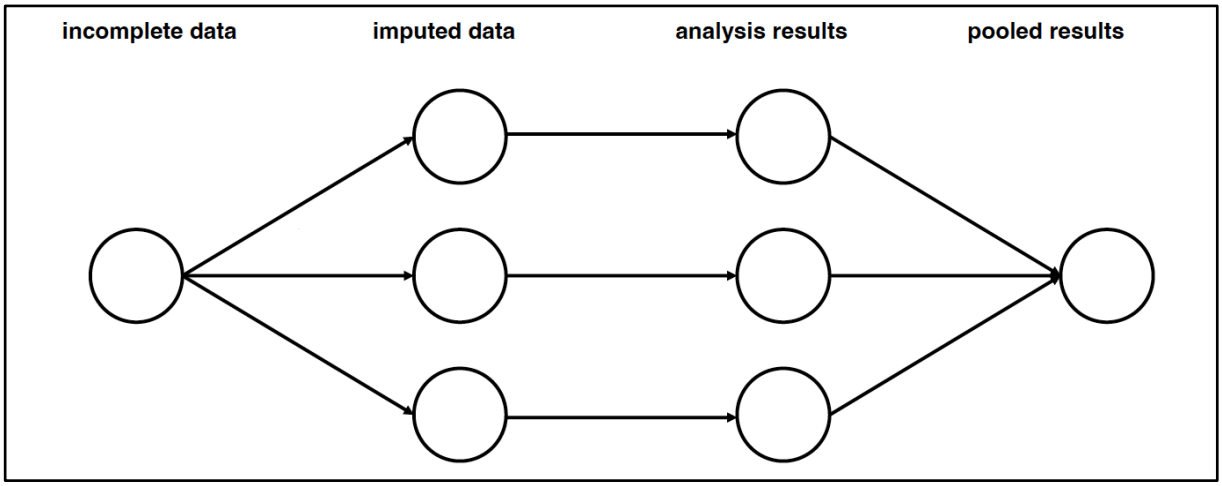

Figure 8: *Main steps of multiple imputation with mice (Source: mice documentation)*

For completeness, here follows a brief description of the multiple imputation process which is graphically summarized in Figure 8. First, the chosen statistical algorithm analyzes the original dataset and generates a range of plausible values for each missing data. The output of this phase is the creation of multiple copies of the dataset, where the missing values are replaced by imputed values. Such values are recorded with their variability as we can never know the true values behind missing data. The different imputed estimates will have non identical values across the various datasets, and this is crucial since these dissimilarities are due to the uncertainty about the value to be imputed. The next phase is to fit the model of interest into each of the imputed datasets. Ultimately, the outcomes need to be pooled together in order to account for the variability due to the insertion of imputed values. Further information on the selected *mice* package utilized for the imputation are available in Appendix 2.

CHAPTER IV: Empirical evidences

Models' setup

This Chapter presents the empirical strategy adopted and its results. First, I describe the variables included in the analysis, the dependent variable, the independent variable of interest and the controls. Then, I focus on the empirical estimates, the identification approach, the results and the robustness checks.

Dependent variable

This thesis aims at verifying the relationship between individuals' portfolio diversification and financial literacy. Therefore, the dependent variable of the empirical analysis is an index that measure portfolio's diversification. In the literature, various measures have been used as a proxy of the diversification degree, such as the number of stocks owned by the investors, the holding of investment funds or the number of different asset classes in the portfolio. I have designed the diversification index such that it captures the variety of asset classes held by individuals, following the last approach of the three above. This is probably the less precise measure of diversification; however, data limitation does not allow other type of score/index.

The diversification index described above includes insurance products as type of financial investment, because they generally include also investment funds or similar and can be assimilated to proper form of investments. Moreover, the recent trends in the demographic evolution, especially in Europe, is raising doubts about the resilience of the public pension system and increasing the importance of alternative solutions to cover up particular events. As the population ages, public pension systems need to cover an increasing amount of people for a longer time, due to longer life expectancy. In the majority of European countries, social security systems are on a pay-as-you-go (PAYG) basis, namely taxpayers provide funds to the retirees. As stated above, however, the number of retired people is increasing, while the number of workers is stable or even decreasing. Therefore, the income of the state pension systems is decreasing, while the payments are skyrocketing. The result is that retirees are either going to see lower payoffs or the State is going to increase taxes. In both cases, the resources available to hedge against poor health circumstances (severe illness, loss of physical autonomy and

similar situations) will be affected. These are the main reasons that got me to include insurance products into the diversification index.

Sometimes, the literature related to financial literacy uses dummy variable indicators as a proxy of the diversification index, identifying whether the individual holds a certain amount of stocks or not. However, these type of indexes presents some limitations. The first one relates to the arbitrary decision of the number of stocks necessary to be considered diversified. Secondly, when an individual holds such an amount of stocks, he/she is not automatically sure to be diversified because of the correlation between different securities (Markowitz, 1952): the higher the correlation, the lower is the diversification of the portfolio. Lastly, considering just stocks is a limitation given the wide range of alternative opportunities offered by banks, financial advisors and financial markets that can be useful in the construction of a balanced portfolio. Hastings et al. (2013) provide a deep analysis of the impact of non-stock holdings as derivatives, bonds, treasury bills and many others. Hastings et al. (2013) mention also the potential impact of buying "on margin". I do not consider this aspect of individual portfolio management because I expect that a very small share of households is familiar with this type of operation, given that they are very complicated and imply high volatility and very high risk.

The dependent variable, i.e. the **Diversification Index,** is defined as the sum of the answers to the questions *"Do you currently hold … (asset class)?"*. These questions cover the following seven financial assets:

- term/savings accounts
- personal/occupational scheme
- investment fund
- shares
- public/private fixed income
- life insurance
- medical insurance

The possible answers are *"yes"* and *"no"*, therefore, the minimum score is zero (when the interviewee holds none of the above stated assets), while 7 is the highest value of the diversification index (when the interviewee holds all of them). As previously discussed, I include life insurance and medical insurance in the previous list since the insurance industry and insurance products play a critical role nowadays in households' portfolios, especially among the elderly. Moreover, the Spanish healthcare organization is very inclusive. In particular, Spain has both public and private healthcare systems. The public healthcare system

provides free basic healthcare to those who contribute to the Spanish social security system and their families.

It also offers free healthcare to retirees, including those from other EU countries. Public medical assistance is also provided to foreign employees, self-employed foreigners working in Spain and their families since they usually contribute to Spanish social security, too. The economic terms of such care are the same of Spanish residents, namely free or subsidized medical care. For this reason, every stipulation of a medical insurance contract can be considered a plus from an investment viewpoint.

The following graphs provide a general overview of the frequency distribution of the dependent variable. In particular, I show the distribution of the answers across the components of the Diversification index scores, both for the original data frame (considering not asked questions as response equal *"no"*) and the average of the 5 imputed datasets. The grey columns in Figure 9 are the result of the combination of the columns shown in Figure 10, where Figure 10 gives a detailed representation of the distribution of respondents' diversification scores across each of the 5 imputed datasets. Figure 9 highlights the consistency of the imputation process. Indeed, the distribution of the diversification scores among the imputed datasets is similar to the one of the original dataset. In percentage terms, the largest difference is registered in the "zero asset" subgroup where around 5% of people move away from that circumstance, but overall there seems to be an acceptable shift in the overall distribution of the sample. These tiny changes in the distribution of respondents across different diversification index scores can be reconducted to the ignorance of some small share of investors about where their money is invested. Secondly, a large portion of the sample holds no assets at all: 40% in the original datasets and 35% (average) in the imputed ones. As the number of assets within the portfolio increases, the number of people belonging to that subgroup decreases. Only 16,8% (original dataset) and 18% (imputed) of interviewees holds 3 or more assets!

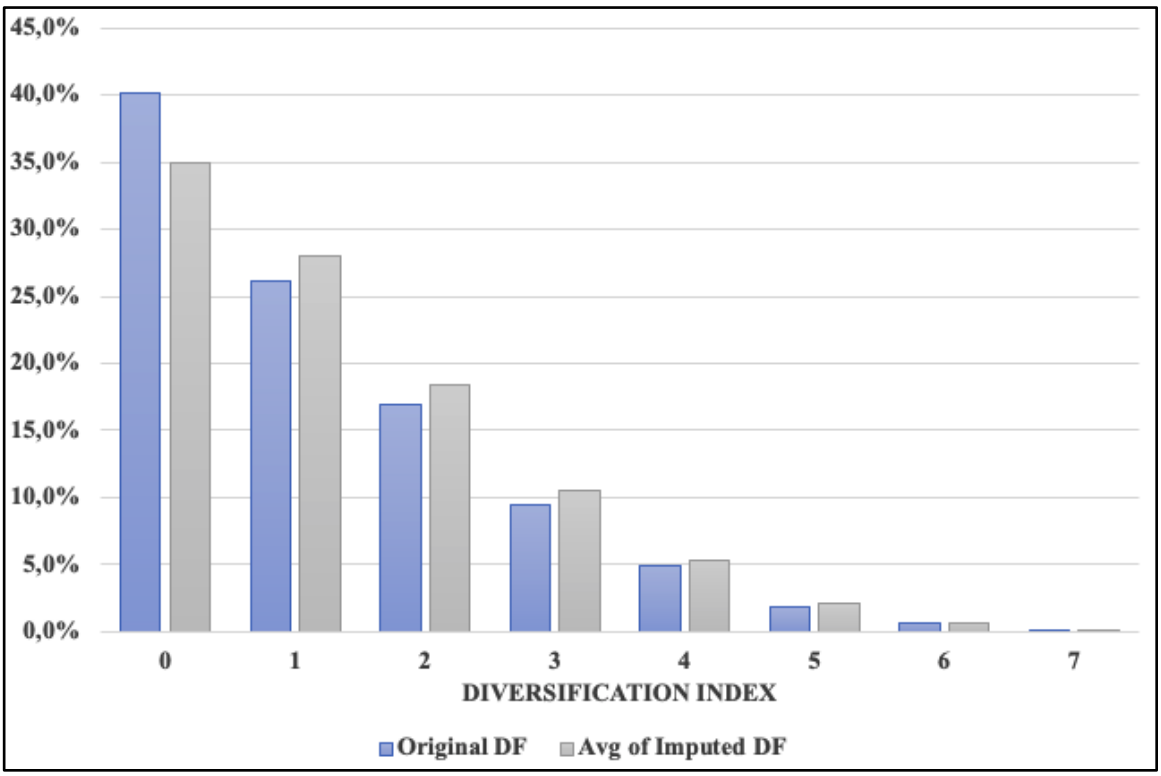

Figure 9: *Respondents' Div. Index scores in original data frame vs average of imputed ones*

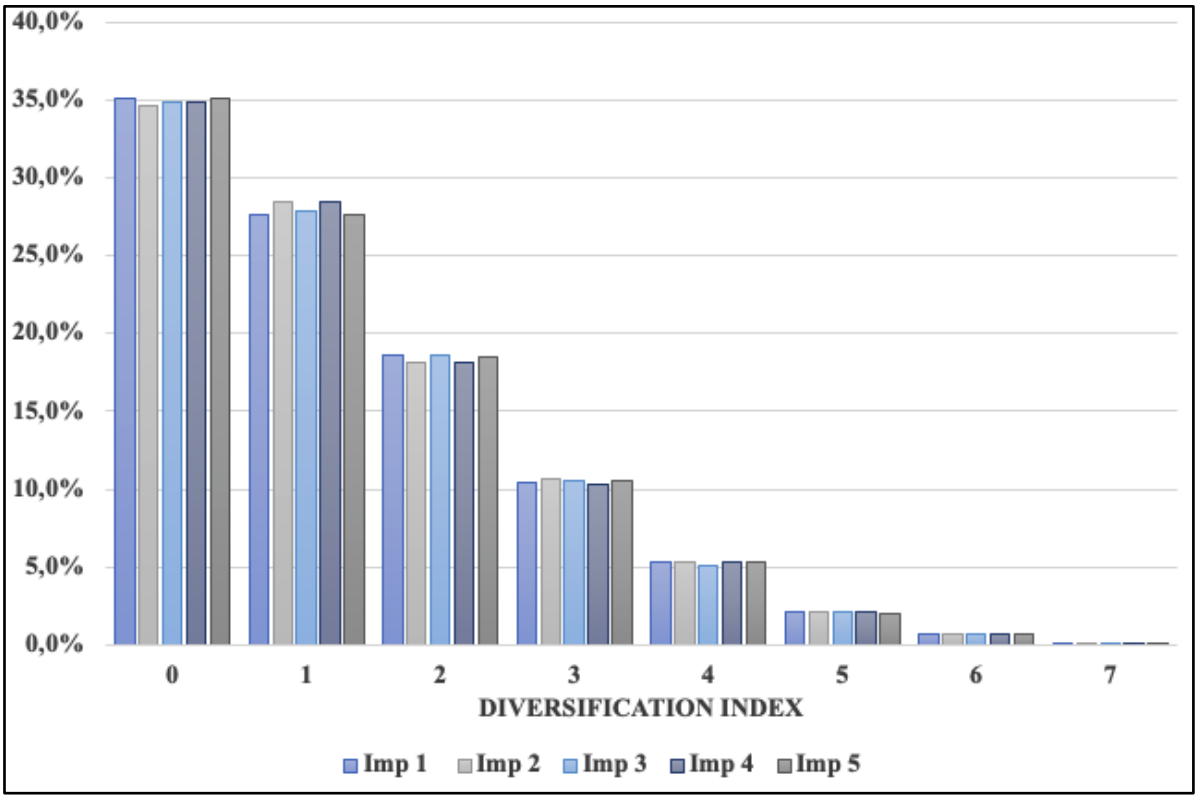

Figure 10: *Frequency distribution of respondents across Diversification Index scores in the 5 imputed datasets*

The investments holding of the sample are differently crowded as the following figures illustrate. Figure 11 shows the number of holders per asset class across the different imputed datasets while Figure 12 contains the average (across the 5 imputed datasets) participation ratio for each investment asset class. Notice that in the ECF questionnaire the category *"public/private fixed income"* refers to the direct investments in such instruments, thus excluding indirect holdings via investment funds or mutual funds. Hence the definition of this category of investments may explain why only a very small share of interviewees holds this type of asset. On the other hand, the category with the highest participation rate is the saving/term deposits (33% of individuals), followed by life insurance (31%), pension schemes (23%) and medical insurance (20%). Then come shares, investment funds and fixed income direct holdings. Not surprisingly, the products that involve higher volatility and intrinsic risk are the one with the lowest participation rate, while the safer once (term deposit) and those that cover from some risk (insurances) register higher participation rates.

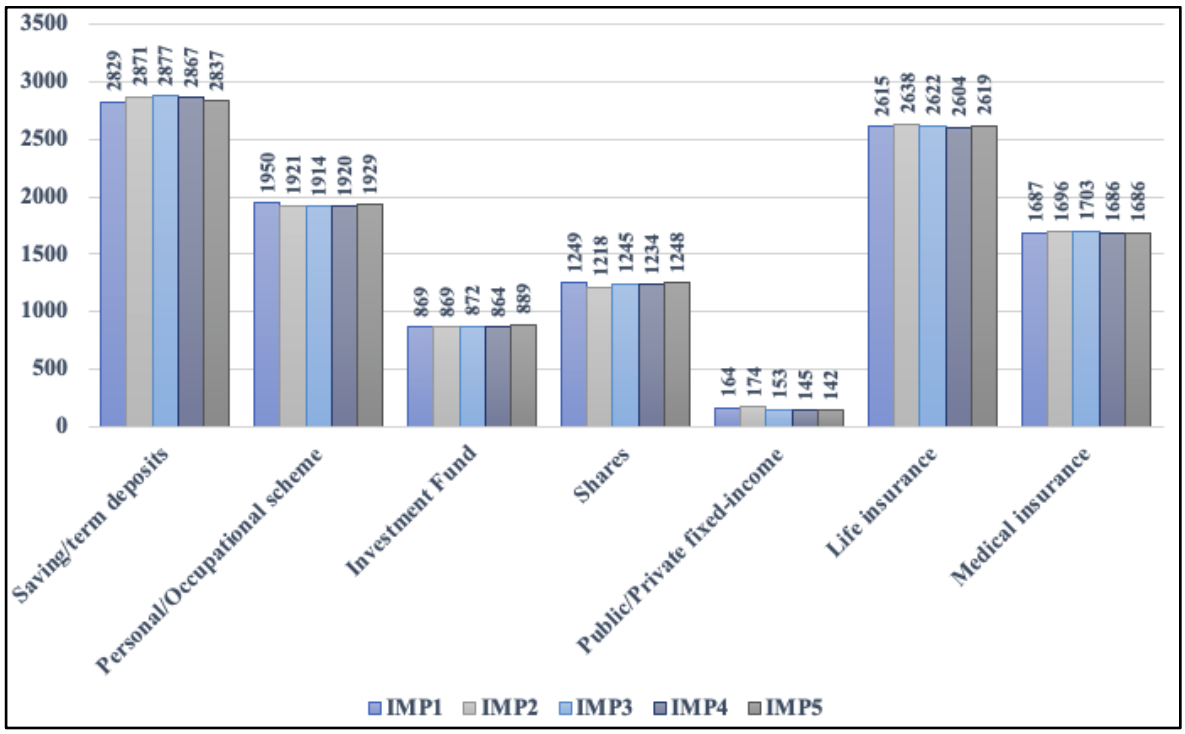

Figure 11: *Holders per investment asset class*

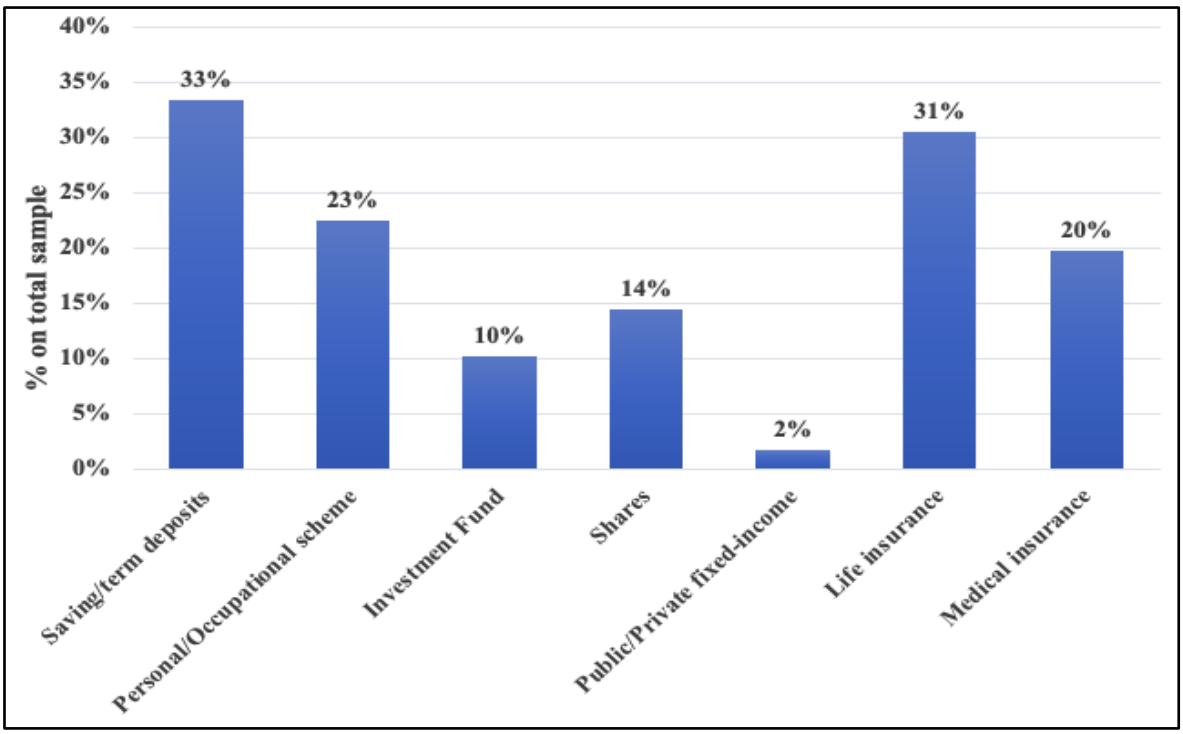

Figure 12: *Holding rate per asset class*

Explanatory variables

This paragraph presents and describes the control variables of the empirical analysis. As largely documented in the literature, individual demographics play a significant role in household financial decision, therefore, variable such as age, income and so on should be included in the empirical analysis. In particular, the ECF survey provides information about the following demographics:

- Gender
- Age
- Place of birth
- Labor status
- Income tier
- Educational attainment
- Area of specialization (in case of upper education)

The ECF survey asked (yearly) income boundaries and not the specific net income of the individuals. It organizes the possible answers in six categories, as the following table shows:

Income Tier 1	0-9000 €
Income Tier 2	9001-14500 €
Income Tier 3	14501-26000 €
Income Tier 4	26001-44500 €
Income Tier 5	44501-67500 €
Income Tier 6	67500+ €

The independent variable of interest of the empirical analysis is financial literacy, defined as *"we mean peoples' ability to process economic information and make informed decisions about financial planning, wealth accumulation, pensions, and debt"* by Lusardi and Mitchell (2014). In this thesis, I construct a measure of financial literacy summing up a set of eleven questions.

Moving to the definition of explanatory variables, the one at the core of the investigation is the financial literacy score. The majority of the eleven selected questions is a multiple-choice question, with only one correct answer: if the respondent's answer is correct, then he/she scores 1, otherwise 0.

The financial literacy score (FL Score hereafter) is therefore a number that varies from 0 to 11, where 0 means that the individual is financially illiterate and 11 that he/she has a high financial knowledge.

Notice that such score includes answers to the **"**Big Three**"** questions about inflation, compound interest and diversification, considered the cornerstone of the measurement of financial literacy in the economic literature. These queries were included in a slightly different framing in the ECF questionnaire. The chosen questions used to evaluate the level of financial literacy are a mix of standard questions largely used in this literature and some new ones. I enlarge the factors affecting our FL score since instruments like the "Big Three" may be narrow and unable to seize the whole span of human abilities related to personal finance. In fact, Huston (2010) suggests crafting an instrument based on four financial areas of knowledge rather than only three, including money basics, borrowing, investing and hedging/risk management. Others like Carpena et al. (2011) emphasize the necessity for questions that do not demand high numeracy skills, as compound interest or inflation.
Then, I construct the financial FL score using the following set of questions. It includes the "big three" slightly reformulated, as well as other topic such as graph interpretation and the risk/return relationship.

1. Imagine that 5 siblings receive a gift of €1000. If this is equally divided into 5 shares, how much would each sibling obtain?

2. Now imagine that the brothers have to wait for one year to get their share of the €1,000 and inflation stays at X percent. In one year's time will they be able to buy:
    a. *More with their share of the money than they could today*
    b. *The same amount*
    c. *Less than they could buy today*
    d. *It depends on the types of things that they want to buy*

3. You lend €25 to a friend/acquaintance one evening and he gives you €25 back the next day. How much interest has he paid on this loan?

4. Imagine that someone puts €100 into a no fee, tax free savings account with a guaranteed interest rate of 2% per year. They don't make any further payments into this account and they don't withdraw any money. How much would be in the account at the end of the first year, once the interest payment is made?

5. And how much would be in the account at the end of five years [add if necessary: remembering there are no fees or tax deductions]? Would it be:
    a. *More than €110*
    b. *Exactly €110*
    c. *Less than €110*
    d. *Impossible to tell from the information given*

6. An investment with a high return is likely to be high risk. *(True/False)*

7. High inflation means that the cost of living is increasing rapidly. *(True/False)*

8. It is usually possible to reduce the risk of investing in the stock market by buying a wide range of stocks and shares. *(True/False)*

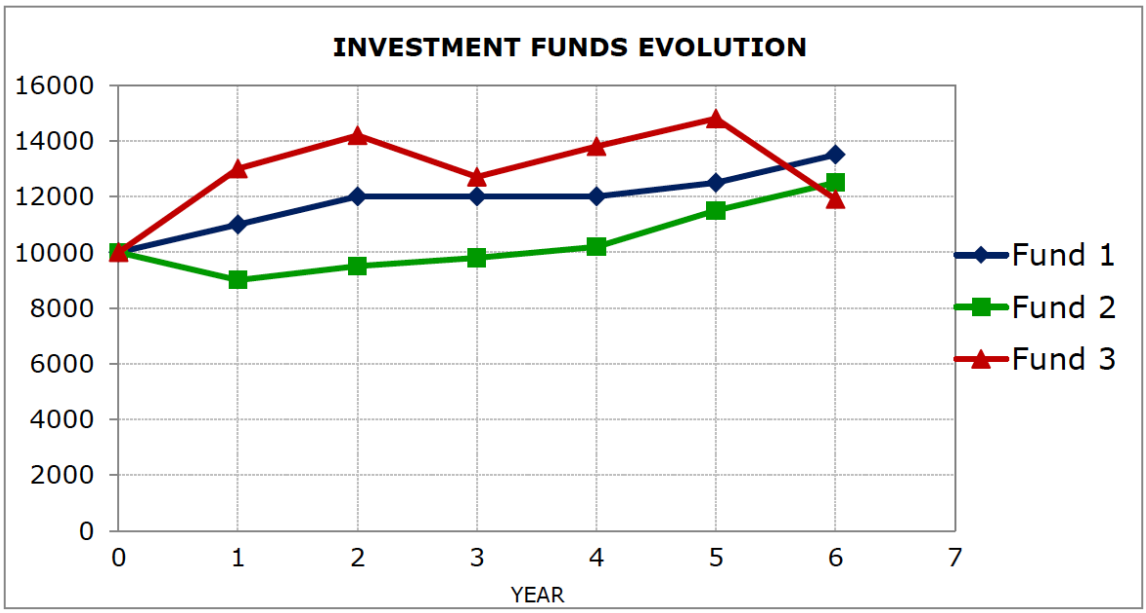

Figure 13: *Graph for question 9 and 10*

9. Assuming that the commission and expenses are the same for all the funds, which fund obtained the best return after 6 years?
    a. *Fund 1*
    b. *Fund 2*
    c. *Fund 3*

10. Which would have been the fund with the best return if the investment had to be withdrawn at the end of the 3 years?
    a. *Fund 1*
    b. *Fund 2*
    c. *Fund 3*

11. A 15-year mortgage normally requires higher monthly payments than a 30 years mortgage, but the total interest paid over the duration of the mortgage will be lower. *(True/False)*

Alongside the FL Score variable, I also introduce a dummy variable **High FL** to check whether being highly knowledgeable in financial matters gives any advantage in terms of degree of diversification of the investment portfolio. This dummy variable takes value 1 if FL Score is 10 or 11 and zero otherwise. The coefficient of this variable will testify if having a high FL knowledge yields some larger portfolio diversification with respect to those who respond correctly to less than 10 FL questions.

The next three figures give a picture of the degree of financial knowledge of the sample. Figure 14 shows that the large majority of respondents is able to answer correctly to six or more questions. 65% of the sample scores between 6 and 9 on financial literacy questions, while only a quarter of respondents is able to answer correctly to 10 or all questions. Last, 8 interviewees out of 100 show serious gaps in financial knowledge as they're not able to answer correctly to more than 5 questions on financial topics.

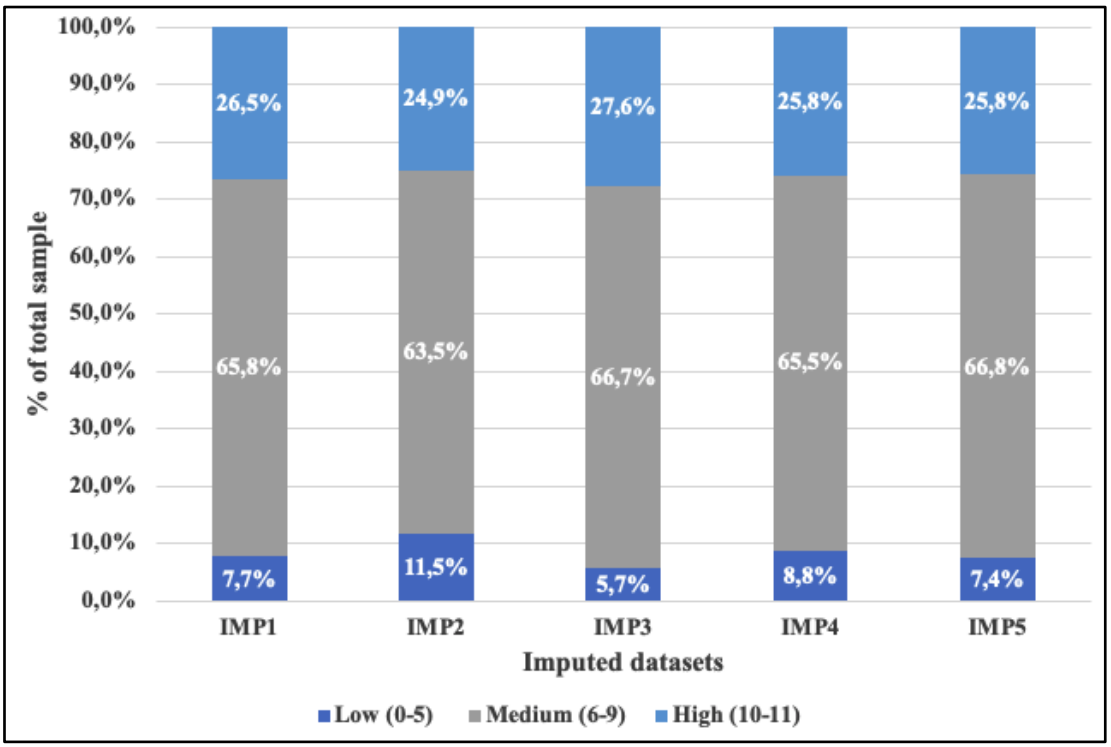

Figure 14: *Financial Literacy Score*

Figures 15 and 16 provide an overview of the distribution of financial literacy score by educational attainment and level of income. In line with previous studies, Figure 15 shows a positive relationship between education and financial literacy score. In particular, as the education level increases, the percentage of respondents scoring less than 6 on FL Score decreases significantly. On the other hand, the share of interviewees reaching the top on FL knowledge score is 5 times the one in the low educated subgroup, and twice the one in the middle-educated subgroup. This confirms that education is a primary driver of financial knowledge. Figure 16 is constructed in the same way as Figure 15 and shows the distribution of FL Score by income levels. As the income increases, the proportion of those with high financial literacy increases, while the financial illiterates generally concentrate among the low-income individuals.

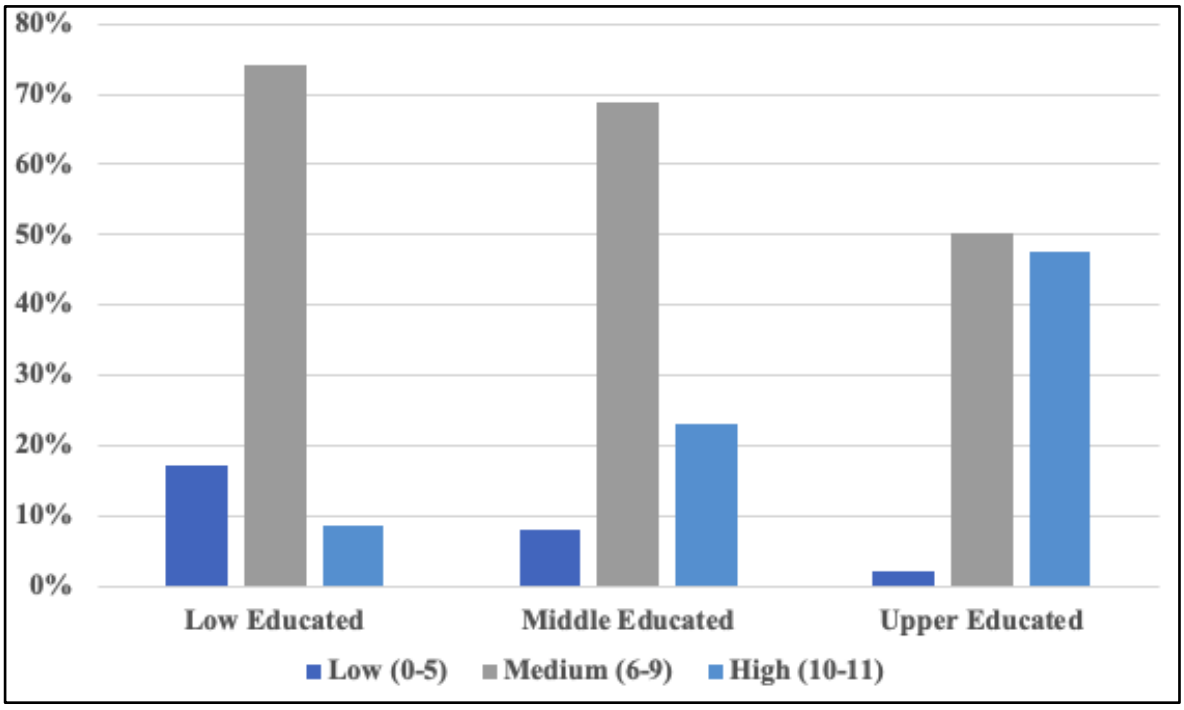

Figure 15: *Financial Literacy Score by Education level*

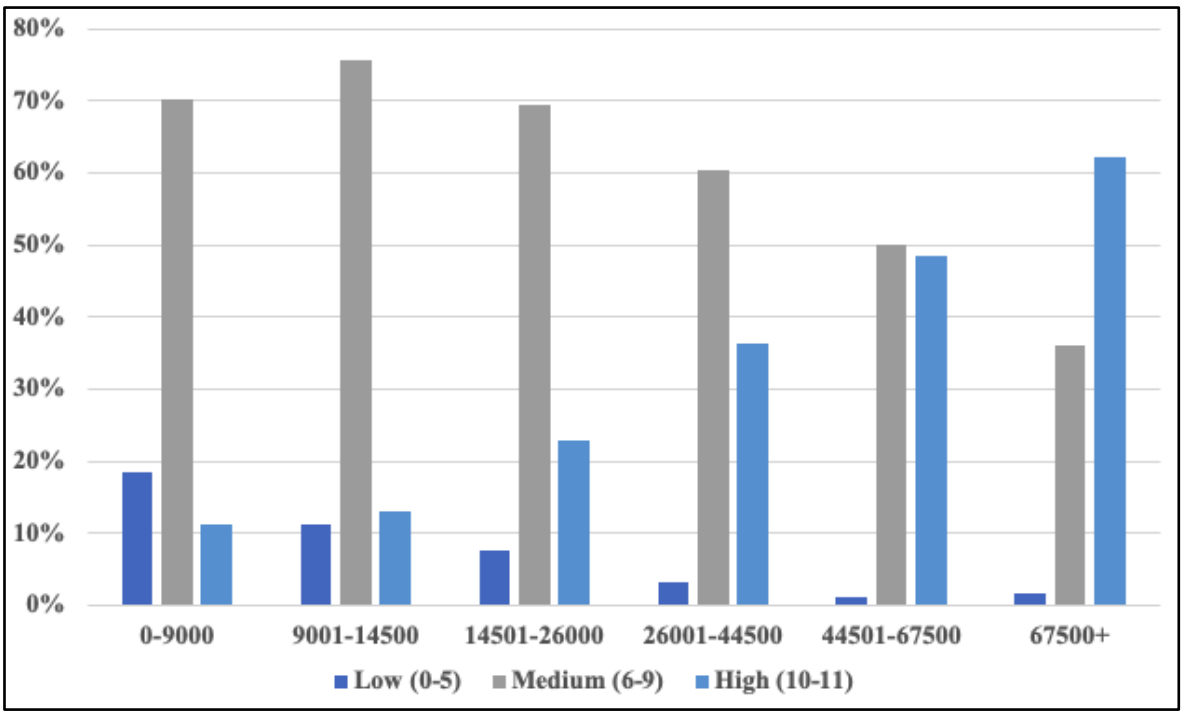

Figure 16: *Financial Literacy Score by Income level*

Besides the two crucial variables that have just been described, I characterized the succeeding controls: education, STEM[7], Economics and Status on top of already existing control variable like gender, age, place of birth and income tier.

Education is a categorical variable with three possible values:

• Low Educated: if the respondent has reached at best a primary level education;

• Middle Educated: if the respondent completed secondary or post-secondary education;

• Upper Educated: if the respondent has obtained a University degree or attended Doctorate studies.

Financial literacy and portfolio choices may be correlated to the job sector or education specialization of the individual, therefore, controlling only for the level of job or education is not sufficient. Among the high educated group, i.e. those who completed university or with even higher educational degree, I create a dummy (Economics) for those whose attended specialization belongs to the Economics, Management and Business field.

[7] STEM = Science, Technology, Engineering and Mathematics.

Finally, the categorical variable *Status* divides the respondents in three occupational categories: the first one (base) includes those who are unemployed or inactive for several reasons (incapacity, illness), the second one collects the employed (both self-employed and employees) and the last one groups the retired people.

Estimation results

This section presents the results of the empirical models that study the relationship between financial literacy and individual portfolio diversification.
First, I study the determinants of the FL score to understand its comparability with the literature, checking for external validity. Second, I move to the empirical study of individual portfolio diversification using linear regression. However, this empirical approach does not address the endogeneity of financial literacy. Therefore, I moved to an IV approach, that provide more consistent and reliable results. The last paragraph of this Section will discuss further robustness checks.

FL Score determinants

As mentioned above, I first analyse the variable of interest, FL Score, and its possible determinants. Table 5 presents the results of a linear model, where the dependent variable is FL Score and the covariates are the individual demographics (gender, a polynomial in age, income, employment status, education).

The following equation represents the estimated model:

$$FL\,Score_i = \beta_0 + \beta_1 Male_i + \beta_2 Age_i + \beta_3 Age_i^2 + \beta_4\, Spanish_i + \gamma' Status_i + \delta' Income_i + \phi' Education_i + \beta_5 Economics_i + u_i$$

	Linear model on FL Score
(Intercept)	5.535***
	0.610 (0.000)
Male	0.599***
	0.076 (0.000)
Age	0.034**
	0.012 (0.021)
Age^2	0.000*
	0.000 (0.064)
Native	0.236***
	0.070 (0.002)
Employed	0.054
	0.048 (0.266)
Retired	−0.055
	0.097 (0.573)
Income 9001-14500	0.225**
	0.087 (0.021)
Income 14501-26000	0.633***
	0.108 (0.000)
Income 26001-44500	0.988***
	0.134 (0.000)
Income 44501-67500	1.306***
	0.151 (0.000)
Income 67500+	1.468***
	0.153 (0.000)
Middle Educated	0.613**
	0.187 (0.024)
Upper Educated	1.256***
	0.223 (0.003)
Economics	0.359***
	0.103 (0.001)
Num.Obs.	8554
Num.Imp.	5
R2	0.218
R2 Adj.	0.217

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5: *Factors affecting financial literacy score*

First, remember that the dependent variable, FL Score, is a number between 0 and 11, and the higher it is the higher is the financial knowledge of the individual. The results are in line with the main findings of the literature presented in Chapter 1, such as Guiso and Jappelli (2008), Lusardi and Mitchell (2011).

First, Table 5 highlights a gender gap in financial literacy: males show higher financial literacy (0.6 more than females). This is a common results in financial literacy questionnaire, that show how the interest in financial topics' and financial decision is still dominant among men rather than women. Age shows a positive and statistically significant effect on FL Score as each additional year yields a rise of 0.034 points to the FL Score. This suggests that older individuals have a higher knowledge of financial topics. This may be a consequence of two factors: experience and higher accumulated wealth than younger individuals. Another difference is the one between immigrants and natives, namely born in Spain. The latter record a higher FL score of about 0.24 points. Using these estimates, it is not possible to argue whether these differences are a consequence of particular socio-economic status or related to background culture and habits. Contrary to expectations, there is no evidence of a statistically significant role of the job status and current employment of the individual. Therefore, being employed, unemployed or retired makes no difference for the individual financial competences.

As usual, financial literacy is then associated with higher income tiers and educational attainment. Lusardi and Mitchell (2011b) find that US low-educated people are more likely to answer incorrectly or to not answer at all. The results in Table 5 are in line with these findings since the magnitude of the coefficients is increasing in the income tier, and all the coefficients are positive.

In particular high-income individuals have a greater degree of financial knowledge. Just consider that the difference in the FL score between those residing in the highest level of income and those in the lowest is almost 1.5 points. Moreover, middle and upper educated subgroups are associated with higher FL Score than low educated class. The effect of a university degree (or higher education) is nearly twice the effect of middle education, 1.256 against 0.613 of the intermediate education level.

The last coefficient reflects the importance of the Economics degree which grants a sizeable and significant impact on FL score (0.359 higher), even after controlling, as I did, for the general level of education.

OLS estimates

This section presents the results of two linear model that studies the relationship between financial literacy and individual portfolio diversification. These models use the same control variable of the estimates presented in Table 5, however, the dependent variable now is the diversification index, while the FL Score is now a covariate. The two models differ in the following way: the first one uses the standard 0-11 FL Score variable, while the second one uses High financial literacy, i.e. the dummy variable that identifies whether the individual FL Score is above 10 or not. The following two equations show the estimated linear models.

OLS 1

$$\text{Div. Index}_i = \beta_0 + \beta_1 \text{Male}_i + \beta_2 \text{Age}_i + \beta_3 \text{Age}_i^2 + \beta_4 \text{Spanish}_i + \gamma' \text{Status}_i + \delta' \text{Income}_i + \phi' \text{Education}_i + \beta_5 \text{Economics}_i + \beta_6 \text{FL Score} + u_i$$

OLS 2

$$\text{Div. Index}_i = \beta_0 + \beta_1 \text{Male}_i + \beta_2 \text{Age}_i + \beta_3 \text{Age}_i^2 + \beta_4 \text{Spanish}_i + \gamma' \text{Status}_i + \delta' \text{Income}_i + \phi' \text{Education}_i + \beta_5 \text{Economics}_i + \beta_6 \text{High FL} + u_i$$

Considering that the two models differ only in one variable, I do not expect big variation in the coefficients of the control variables.

Table 6 shows that there is no effect of individual gender on portfolio diversification, i.e. the variety of assets hold by males and females is more or less the same. This might be connected with the higher overconfidence that males tend to display when dealing with investment decisions. Barber and Odean (2001) show that in areas like finance, males have higher confidence than females. Therefore, they may concentrate their wealth into a few asset classes, but characterized by higher intrinsic risk, like stocks. On the other hand, females are typically more prudent and more risk averse, therefore they typically decide to invest only in those financial products with low risk and volatility, with a lower diversification.

Age has a positive impact on the Div. Index, as for each additional year, the coefficient indicates an increase of 0.064 in the number of assets held on individual portfolios. Age^2 has a very small negative effect, this is a common result for age polynomials, that capture the non-linear effects of age on the dependent variable.

Being an immigrant or a native Spanish individual makes no difference in terms of portfolio diversification.

The current job status, i.e. whether the individual is unemployed, employed or retired, play a role in portfolio diversification. In particular, being employed seems to increase diversification (0.221 higher Div. Index). These results may be explained as follow: employed individuals hold a higher income than unemployed or retired person, therefore they may take higher financial risk and invest also in those products with higher volatility.

Instead, coefficients of variables connected with income level and educational attainment show strong statistical significance. Income coefficients demonstrate their predominance in the determination of the level of diversification. Wider inflows of capital during the lifetime of an investor allow him/her to explore a larger spectrum of investments' possibilities, thus building up a more heterogeneous portfolio. Numerically, this possibility is increasing with income, as the magnitude of the positive coefficient rises when also the income tier is higher. For instance, the difference between low-middle income (< 26.000€/year) and top-income ranges between 1.25 assets when compared with the third income tier (14.501-26.000€) and 1.65 assets when compared with the lowest one (< 9.000€). Education is confirmed as another clear driver of diversification. An individual who completes university education or obtain even higher degree, like a doctorate, holds nearly 0.6 additional assets in his/her portfolio.

In these models, high education is represented also by another dummy, namely Economics which refers to the specialization field of the University degree. The results from the Economics variable indicate the existence of a positive and statistically significant relationship at the 1% level of confidence or better with my measure of Diversification Index. Those reporting to have attended Economics studies have been probably exposed to portfolio optimization theory and learnt about the benefits of diversification. With this set of knowledge, they have probably managed better their portfolio, diversifying the investments between risky and non-risky assets, on the basis of their own risk perception and risk aversion.

The coefficients of interest of this work are the ones of Financial Literacy, I provide two set estimated each with a different version of financial literacy: the scalar version FL Score which takes values between 0 and 11 and the dummy version indicating those answering correctly to 10 questions or more. In the first column (OLS 1) the coefficient of FL Score measures the association between an additional point in the FL Score and the degree of diversification. In the second column (OLS 2) the coefficient of High FL indicates the impact of being highly financially literate on the portfolio composition.

We expect those coefficients to be positive and statistically significant indicating that those having a higher financial knowledge are also more likely to hold more diversified assets, hence being able to navigate more financial scenarios with their investment strategies. The first model shows no evidence of a role of FL Score in the number of assets that an individual possesses given the tiny absolute value and the statistical insignificance of the coefficient. Oppositely, the coefficient of High FL seems to be significant at the 10% confidence level and have a modest positive effect in the number of assets held in portfolio (0.146). These two coefficients show that there is low or little evidence of any effects of financial literacy on portfolio composition and diversification of the Spanish individuals. However, as the next Section will explain, these results may be a consequence of an endogeneity problem of financial literacy.

	OLS 1	OLS 2
(Intercept)	−1.683***	−1.484***
	0.315 (0.002)	0.268 (0.001)
Male	0.024	0.023
	0.030 (0.432)	0.028 (0.411)
Age	0.064***	0.064***
	0.007 (0.000)	0.007 (0.000)
Age^2	−0.001***	−0.001***
	0.000 (0.000)	0.000 (0.000)
Native	0.063	0.061
	0.058 (0.300)	0.059 (0.313)
Employed	0.221***	0.222***
	0.035 (0.000)	0.034 (0.000)
Retired	0.024	0.028
	0.060 (0.688)	0.059 (0.641)
Income 9001-14500	0.119*	0.124*
	0.061 (0.066)	0.061 (0.060)
Income 14501-26000	0.394***	0.400***
	0.063 (0.000)	0.065 (0.000)
Income 26001-44500	0.823***	0.828***
	0.068 (0.000)	0.072 (0.000)
Income 44501-67000	1.195***	1.199***
	0.082 (0.000)	0.084 (0.000)
Income 67000+	1.653***	1.649***
	0.101 (0.000)	0.101 (0.000)
Middle Educated	0.317***	0.321***
	0.063 (0.000)	0.060 (0.000)
Upper Educated	0.578***	0.581***
	0.075 (0.000)	0.072 (0.000)
Economics	0.381***	0.375***
	0.073 (0.000)	0.073 (0.000)
FL Score	0.039	
	0.022 (0.123)	
High FL		0.146***
		0.046 (0.009)
Num.Obs.	8554	8554
Num.Imp.	5	5
R2	0.261	0.262
R2 Adj.	0.260	0.260

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 6: *Ordinary Least Squared models on Diversification Index*

Controlling for endogeneity

This section addresses the endogeneity problem of financial literacy and explain the empirical and identification strategy used to overcome this problem. As pointed out by Jappelli and Padula (2013), financial literacy suffers from a potential problem of endogeneity since the causal effect on financial outcomes is not clear. Actually in this study OLS estimates may be biased for a variety of reasons including: (i) individual unobserved characteristics (IQ, risk aversion, ability, patience) which have been shown to be connected with financial literacy and portfolio allocation, (ii) reverse causality, namely the variable Y (Div. Index) might influences variable X (FL Score) instead of the opposite, and lastly, (iii) measurement error probably enhanced by the imputation process. Moreover, the results presented in Table 6 are in line with Fort et al (2016), who explain that there is evidence of a downward estimation bias of the FL impact on whatsoever variable of interest rather than upward biased one.

These reasons motivate me in looking for alternative identification strategy. Following Jappelli and Padula (2011 and 2013) I use IV estimates to overcome the possible endogeneity problem of financial literacy. In line with their works, I decide to use the following two instruments for financial literacy: STEM and number of books in the house at the age 10 of the respondents.

The STEM instrument identifies whether the high education field attended belongs to one of the following areas:

• Engineering and Technology (Architecture, Electronics, Mechanics)
• Health Sciences (Medicine, Nursing, Pharmacy)
• Experimental sciences (Physics, Mathematics, Chemistry, Biology).

The second instrument, instead, measures the number of books in the house of the respondent when he/she was a child. The exact question used to construct the instrument is the following:

"Approximately how many books were there where you lived when you were 10 years old? Do not include magazines, newspapers or textbooks."

1. *None or very few (0-10 books)*
2. *Enough to fill a shelf (11-25 books)*
3. *Enough to fill a bookcase (26-100 books)*
4. *Enough to fill two bookcases (101-200 books)*
5. *Enough to fill more than two bookcases (over 200 books)*.

First, I assume that individuals who attended STEM courses possess the knowledge in numerical and mathematical terms sufficient to deal with the questions that make up the index of financial literacy.
This assumption is in line with the results of Fairfax (2018). The findings of this paper show that the US federal securities law regime founds its principle on the belief that investors hold the minimum cognitive skills required to understand the complexity of financial products and financial information. Therefore, this entire law regime assumes that individual who decide to invest in the financial markets are capable of basic financial reasoning and can understand, for example, a financial report or a KID[8] (Key Information Document).

Second, there is evidence from psychology and medicine research that people have gaps in some numerical skills like performing calculations that involve percentages (see Peters et al., 2007; Chen and Rao, 2007). These findings are relevant because calculations about compound interest serve as the basis of most financial decisions. Moreover, using credit cards, engaging in a mortgage contract, or embracing the advantages of "early saving" demands some understanding of interest compounding and how it works. STEM can easily be reconciled to a proxy of numerical abilities, generally associated to higher financial literacy.

Third, the two instruments selected are in line with the instruments used by other economists for financial literacy measures. For instance, Jappelli and Padula (2011) address the problem of endogeneity of financial literacy using the PISA test scores[9] as instruments for it. After doing this additional step, they find that the effect of the instrumented financial literacy on savings is higher under the IV framework than under the OLS setup. They also perform some robustness checks adding some additional instruments like books' number in the house in childhood or parents' occupation and they end up with similar results and magnitude of the OLS bias. Such findings suggest that OLS tend to underestimate the coefficient of financial knowledge, thus producing a downward bias in the estimates. Fernandes et al. (2014) and Morgan and Trinh (2016) address the endogeneity issue by using the average financial literacy score at provincial level as an instrumental variable for financial literacy. In the study of Disney and Gathergood (2013) the instrument used is time-dated financial education, or better debt literacy as their research focuses on the association between consumer credit and individual financial literacy. Fessler et al. (2020) introduce a dummy instrument for financial knowledge based on newspaper

[8] It is a standardized informative document containing all the necessary information about a product outlined in a clear form. It aims to provide the customer with all the notions to be able to evaluate a given investment product.
[9] PISA is the OECD's Programme for International Student Assessment. PISA measures 15-year-olds' ability to use their reading, mathematics and science knowledge and skills to meet real-life challenges. Source: OECD official website

reading habits. It takes value one when the newspapers read by respondents belong to high quality newspapers and zero otherwise. The main idea behind this identification strategy is that those reading high quality journalistic articles are more likely to have larger knowledge of financial matters with respect to those reading shoddy newspapers. Klapper et al. (2012) exploit the numerosity of newspapers in circulation and the concentration of universities by region in a study on a Russian sample.

Last, the number of books at home is already a standard query in PISA which captures family background and influence into a student's education that is one of the main drivers of financial literacy. I have included it in the IV estimates as it gives information about the sociocultural background from which a person has grown and it has been demonstrated that socio-cultural factors are relevant among financial literacy determinants, and the results of Table 5 are in line with these findings. Moreover, Arellano et al. (2014) confirm the validity of the instrument number of books as they find that those students declaring to have more than 25 books in their residence achieve nearly a 40% higher score in the financial literacy test.

The selection of these two instruments responds to the conditions about the validity of instruments. Indeed, these are variables that are presumably not related to the holding of financial products, but that impact on the capacity of answering correctly to financial literacy questionnaires. This reflects the respect of the fundamental requirements in the estimation with the method of the instrumental variables.

Specifically, on one side, the number of books at home in the childhood can be interpreted as an indicator of socio-cultural family background and can be assumed to influence the critical thinking and problem-solving skills of people, which are factors affecting financial literacy. Evans et al. (2010) document that books are a useful indicator for showing the effect of many family inputs, like family spending in education, the households' cultural capital and their educational preferences. They also demonstrate that the presence of a larger quantity of books at home is positively associated with educational attainment and with mathematical test scores, two major factors impacting on financial literacy. Moreover, Terman and Oden (1959) show that more books at home can benefit a child's IQ, thus being able to face more easily problems in areas not specifically taught at school.

On the other side, STEM works as a proxy of respondents' statistical and numerical abilities. Having these skills eases the effort needed to understand financial product functioning and their risk and potential return components. This contributes to the enlargement of their financial knowledge. Moreover, the financial information allegedly contains guidance on investment

suggestions and diversification/efficient portfolio composition. Hence, we can assume that being surrounded by many books in the childhood and having a STEM study background creates the right circumstances for a greater ability to assimilate financial notions and improve personal financial literacy. This will have a cascade effect on the quality of choices regarding portfolio composition, most likely ending with a greater portfolio diversification in terms of asset categories.

IV estimates

This section presents the results of two IV models. These models use the same control variable of the estimates presented in Table 6, with the difference that here Financial Literacy is instrumentalized with STEM and books' number, i.e. the two instruments described before. The following equations show the estimated IV models, highlighting the first and second stages setups.

IV 1 (1st and 2nd Stage equations)

$$FL\ Score_i = \beta_0 + \beta_1 Male_i + \beta_2 Age_i + \beta_3 Age_i^2 + \beta_4\ Spanish_i + \gamma' Status_i + \delta' Income_i + \phi' Education_i + \beta_5 Economics_i + \beta_6 STEM + \beta_7 Books\ Number + u_i$$

$$Div.Index_i = \alpha_0 + \alpha_1 Male_i + \alpha_2 Age_i + \alpha_3 Age_i^2 + \omega' Spanish_i + \eta' Status_i + \theta' Income_i + \psi' Education_i + \alpha_4 Economics_i + \alpha_5 FL\ Score + \epsilon_i$$

IV 2 (1st and 2nd Stage equations)

$$High\ FL_i = \gamma_0 + \gamma_1 Male_i + \gamma_2 Age_i + \gamma_3 Age_i^2 + \gamma_4\ Spanish_i + \varphi' Status_i + \tau' Income_i + \rho' Education_i + \gamma_5 Economics_i + \gamma_6 STEM + \gamma_7 Books\ Number + \varepsilon_i$$

$$Div.Index_i = \vartheta_0 + \vartheta_1 Male_i + \vartheta_2 Age_i + \vartheta_3 Age_i^2 + \vartheta_4\ Spanish_i + \gamma' Status_i + \delta' Income_i + \phi' Education_i + \vartheta_5 Economics_i + \vartheta_6 High\ FL + \omega_i$$

	1st Stage IV1	IV1	1st Stage IV2	IV2
(Intercept)	3.921***	−4.306***	−0.333***	−0.936***
	0.629 (0.002)	1.353 (0.019)	0.088 (0.006)	0.389 (0.045)
Male	0.476***	−0.249***	0.134***	−0.263***
	0.113 (0.010)	0.078 (0.003)	0.019 (0.000)	0.075 (0.000)
Age	0.046***	0.040***	0.010***	0.044***
	0.015 (0.025)	0.011 (0.001)	0.003 (0.004)	0.011 (0.001)
Age^2	0.000**	0.000***	0.000***	0.000***
	0.000 (0.055)	0.000 (0.008)	0.000 (0.018)	0.000 (0.002)
Native	0.063	0.006	0.026	−0.012
	0.069 (0.373)	0.072 (0.937)	0.018 (0.170)	0.070 (0.861)
Employed	0.071**	0.181***	0.016	0.189***
	0.040 (0.077)	0.042 (0.000)	0.013 (0.206)	0.044 (0.000)
Retired	−0.016	0.042	−0.024	0.087
	0.081 (0.840)	0.081 (0.607)	0.021 (0.246)	0.079 (0.276)
Income 9001-14500	0.132	0.034	−0.002	0.123
	0.077 (0.112)	0.079 (0.668)	0.018 (0.894)	0.081 (0.152)
Income 14501-26000	0.414***	0.129	0.070***	0.228***
	0.086 (0.001)	0.091 (0.161)	0.019 (0.001)	0.089 (0.017)
Income 26001-44500	0.677***	0.385***	0.145***	0.477***
	0.091 (0.000)	0.125 (0.003)	0.021 (0.000)	0.122 (0.000)
Income 44501-67000	0.904***	0.605***	0.210***	0.691***
	0.101 (0.000)	0.156 (0.000)	0.024 (0.000)	0.140 (0.000)
Income 67000+	1.035***	0.967***	0.299***	0.928***
	0.111 (0.000)	0.189 (0.000)	0.030 (0.000)	0.196 (0.000)
Middle Educated	0.440***	0.021	0.092***	0.084
	0.117 (0.011)	0.095 (0.823)	0.019 (0.000)	0.089 (0.348)
Upper Educated	0.749***	0.006	0.184***	0.043
	0.155 (0.003)	0.148 (0.967)	0.032 (0.000)	0.152 (0.780)
Economics	0.453***	0.165	0.164***	0.078
	0.105 (0.000)	0.108 (0.129)	0.035 (0.000)	0.119 (0.516)
Books Number	0.110***		0.029***	
	0.022 (0.001)		0.005 (0.000)	
STEM	0.141**		0.044**	
	0.074 (0.056)		0.023 (0.055)	
FL Score		0.646***		
		0.205 (0.010)		
High FL				2.348***
				0.597 (0.000)
Num.Obs.	8554	8554	8554	8554
Num.Imp.	5	5	5	5
R2	0.199		0.151	
R2 Adj.	0.198		0.149	

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01
F-test(IV1)=19.918; p.value= 1.2e-04
F-test(IV2)=17.492; p.value= 2e-05

Table 7: *IV on Div. Index. Books' Number and STEM are the instruments used.*

Table 7 presents the first and second stage IV results' for both models, the IV 1 with FL Score and the IV 2 with High FL. This set of estimates continues to confirm the results from the previous two sets.

In particular, Column 1 and Column 3 show the same signs and significance of coefficients as in Table 5 where FL is regressed for the main demographics. Obviously, Table 7 shows also the impact of the two chosen instruments. STEM seems to play a major role in absolute terms with respect to Books' number. In the 1st Stage of IV1, having a STEM degree is associated with a 0.141 higher FL Score. On the other hand, those individuals who were surrounded by more than 200 books in their childhood (corresponding to the highest possible option = 5) gain around half point in the FL Score assessment. In the 1st Stage of IV2 it is possible to observe similar behavior of both instruments' coefficients.

Regarding the estimates of the second stages of the two IV models, not surprisingly, the signs and the significance of the majority of the coefficients holds. Age still shows a positive and significant relationship with the diversification index, with a magnitude similar to the estimates of Table 6 (0.04). Under this setup, being a man impacts negatively the degree of portfolio diversification; thus, men tend to be less diversified than women, all other factors being equal. As previously described, being and immigrant or a Spanish individual does not make any difference. The same holds for retirement ("unemployed" is the base category), while being employed increase the portfolio diversification (0.181 higher Div. Index).

Income still plays a major role as determinant of portfolio diversification. Precisely, higher income tiers show statistical significance at 1%, with slightly lower coefficients' values than in the results presented in the linear models (Table 6). Again, the impact of income is increasing, on diversification, with income tier. With regards to educational attainment, the first stage shows that they are relevant factors in the determination of financial literacy, where higher education corresponds to higher financial literacy, as well as specializing in economics fields.

Last, I want to focus on the two coefficients of interest, i.e. FL Score and High FL. In particular, the standard variable FL Score is now statistically significant and carries a positive effect in absolute terms. For each additional correct answer, the respondents' Div. Index gains 0.646 in the overall score. In other words, individual with the higher FL score (11) are fully diversified among the financial products that compose the dependent variable. On the second IV model in column 2, High FL shows a sizable impact on portfolio diversification (2.348). This testifies how being very knowledgeable in financial matters plays an important role in the determination of the level of diversification of people's investment portfolios. Indeed, those subjects who have been able to face and answer correctly the large majority or all of the FL questions are presumably more informed and updated about the investment possibilities that the financial environment offers, the possible risks involved, the potential returns and also the taxation. This helps them choosing a more diversified and suitable composition of their investments.

I have performed some diagnostics' tests on instruments to check whether I use valid instruments or not and the overidentification hypothesis. These tests are standard tests in the IV framework, needed to prove that what has emerged from the IV1 and IV2 second stage outcomes can be accepted and considered valid. In order to have this assurance I have run Weak instruments test, Hausman test and Sargan test. The following table summarizes the results of these tests performed for each imputed dataset.

Test	Imputed dataset	Statistics	P value	Significance
Weak instruments F test	1	34.319	1.43e-15	***
	2	49. 639	< 2e-16	***
	3	29.688	1.42e-13	***
	4	39.304	< 2e-16	***
	5	36.662	< 2e-16	***
Wu Hausman	1	49.130	2.58e-12	***
	2	36.818	1.35e-09	***
	3	38.354	6.17e-10	***
	4	42.888	6.13e-11	***
	5	46.047	1.23e-11	***
Sargan	1	0.677	0.41	
	2	0.434	0.51	
	3	0.984	0.321	
	4	0.759	0.384	
	5	0. 487	0.487	

Table 8: *Diagnostics tests summary*

The weak instruments test is useful to exclude that the instruments have a low correlation with the endogenous explanatory variable. When an instrument is weak, it might cause a larger variance in the coefficient and some substantial bias in the estimates. Such a test consists of an F-test on the first stage regression. The null hypothesis is that the instruments are weak. For the two estimated models, the null hypothesis is always rejected, hence I can move on with the assumption that the instruments STEM and books' number are sufficiently strong. The Hausman test is necessary to support the choice of using IV procedure.

The null hypothesis here is that OLS and IV are equally consistent. If the statistics is significantly different from zero then the regressor, FL Score is not exogenous or predetermined with respect to e_i and the IV estimator (TSLS) should be used. If it is not significant, then the more efficient estimator is the OLS.

In this case, the null is rejected, hence, the variable of concern is uncorrelated with the error term, indicating that FL Score is endogenous. Rejecting the null of the Hausman test implies

that IV estimates are preferable to the OLS in this study since the OLS is not consistent under H_1, thus reinforcing the validity of results in Table 7.
The Sargan test checks the possible overidentification of the instruments. It is applicable only when there are more instruments than the potential endogenous explanatory variables. So, it tests that all instruments are exogenous and uncorrelated with the residuals. If the null is rejected, it means that there might be some invalid instrument since it is a global test. In my study the null hypothesis is accepted, hence the two instruments are not correlated with the error terms and thus they are consistent.

To summarize, diagnostics tests provide a positive robustness check for the validity and reliability of the estimates. First, the Sargan test has erased any doubts about the validity of the instruments. Second, the Weak instruments F-test has clearly expressed how these are not weak. Last, the Hausman test highlighted the need to use procedure IV. Moreover, the evidence of a downward bias (largely documented in the literature) in the OLS models' estimates when comparing them with the ones in the IV setup, provide an additional proof of the reasonable setup of the instrumental variable.
Thus, I conclude that it would be an error not to perform an IV regression since the bias of the OLS estimates would be much higher and the model itself less informative.

Conclusion

In this thesis, I conducted an analysis on the role of financial literacy as a potential factor affecting the degree of portfolio diversification, namely the number of asset classes held into individuals' investment portfolio.
The aim of this paper is to check whether individuals' financial knowledge, measured by an alternative financial literacy score, is able to influence the investment choices about the portfolio composition. The analysis on the data was performed using the Spanish Survey of Financial Competences (2017).

Starting from the original dataset, I selected observations of variables connected with general demographics, portfolio holdings, financial knowledge and additional variables related with STEM studies and personal background. Then I performed a multiple imputation process with the application of the mice[10] package in order to get rid of missing data. Some new variables were defined based on some other existing ones and imputed passively simultaneously with the multiple imputation process. 5 different imputed datasets were generated and all the empirical analysis was performed on each one of those and then pooled together.
Financial literacy score was the crucial variable of interest and was designed to range between 0 and 11 depending on the number of correct answers given to a selection of eleven questions. These queries consist of some already well-established questions in the financial literacy field, like the Big Three and some additional ones connected with numerical skills.
This measure of financial literacy behaves exactly as other FL measures in economic literacy, thus demonstrating to be a valid alternative. More in detail, it correlates positively with age suggesting that older individuals have a larger knowledge of financial notions, probably due to a wider experience. Spanish natives also seem to register higher FL Score; however, it is inadvisable to draw any conclusion without further examinations on the socio-cultural conditions of those being born away. Lastly, my FL Score is strongly positively associated with income and education, in line with other academical studies.

After this check, I set up two OLS models on my outcome variable of interest, i.e. Div. Index. The latter is defined as the sum of asset categories held in portfolio among the following ones: term/savings accounts, personal/occupational scheme, investment fund, shares, public/private fixed income, life insurance, medical insurance. Here the difference from the large majority of

[10] R Studio package which performs multiple imputation using Fully Conditional Specification (FCS) implemented by the MICE algorithm as described in Van Buuren and Groothuis-Oudshoorn (2011).

previous studies is that I decided to also include insurance products given their increasing importance in financial planning.

OLS 1 and OLS 2 diverge only for the different explanatory variables regarding financial knowledge. The first model has the continuous version of FL Score, while the second one has the dummy High FL which takes value 1 only when the individual is able to answer correctly to 10 or all of the eleven FL questions.

The OLS estimates indicate that the factors having a positive and significant effect on diversification are age, employment, income, education and Economics specialization. The two variables related with FL show tiny effects: in the first specification (FL Score) it is not even statistically significant, while in the second specification (High FL) it is significant with a positive effect of just +0.146 on the diversification index.

This little evidence could be due to endogeneity of financial literacy. To overcome this identification problem, I performed IV estimation using two instruments: STEM and the number of books at home in childhood. The second stage results indicate that most of covariates are in line with OLS findings, showing consistency across different estimation models. However, what changes is that IV models display a much stronger significance and magnitude of financial literacy on the degree of portfolio diversification. In particular, each additional point in the FL score corresponds to an increase of 0.646 in the number of assets in the portfolio. Similarly, those being highly financially literate hold almost an additional asset class (+ 2.348) in their portfolio with respect to those having lower financial knowledge.

All of the above suggests that financial literacy is a core determinant of portfolio diversification. Once again, the evidence indicates that IV estimates return higher coefficients for financial literacy denoting a general tendency of underestimated effects in the OLS procedure. This study reinforces the importance of financial knowledge for investors' diversification capability, suggesting that policy makers should explore solutions to improve it to help individuals avoid losses due to improper portfolio composition and under diversification.
To improve the diffusion of financial literacy among the population is really complicated and requires large investments from the central governments. These investments include the selection of the most appropriate educational programs, possibly targeted by age, employment status and/or households' income.

Organizing longitudinal experiments and randomized control trials might be a strategy. Policy makers should avoid those generic "one-size fits" programs and should prefer tailored education to the specific needs of population subgroups.
Moreover, another key point of these financial literacy educational programs is the measurement of financial knowledge: a good starting point is a shared, comprehensive financial literacy survey that covers the vast majority of these topics. The OECD INFE questionnaire is on the right track to do so, however there is still room for improvement in the Financial Literacy score buildup.

Ideally, such a kind of survey should be repeated every year or every two years to control the efficiency of the educational programs, ideally on a treated and a control group. Another advantage of the use of standardized surveys is that they allow for cross-country comparisons which might yield interesting suggestions to those countries which are behind in terms of financial literacy.

To conclude, there is an urgent need for cooperation for policy implementation and evaluation when dealing with the financial literacy issue. The suggestions above are only a first, possible, step of the diffusion process of financial knowledge among the population, especially among the low-income and low-educated households, protecting them from financial mistakes, including under diversification!

Appendix 1

	Absolute	%
GENDER		
Male	4264	49,8%
Female	4290	50,2%
AGE		
17-30	1507	17,6%
31-45	2436	28,5%
46-60	2675	31,3%
61-79	1937	22,6%
EDUC		
Low Education	1485	17,4%
Middle Education	5157	60,3%
Upper Education	1912	22,4%
STATUS		
Unemployed/Inactive	2632	30,8%
Employed	4546	53,1%
Retired	1376	16,1%
INCOME		
0-9000	1232	14,4%
9001-14500	1871	21,9%
14501-26000	2410	28,2%
26001-44500	1901	22,2%
44501-67500	752	8,8%
67500+	389	4,5%

Table 9: *Main demographics on Imputed dataset*

Appendix 2

Mice package

The MICE acronym stands for Multivariate Imputation via Chained Equations. The Chained Equation approach has the capability to handle several data types efficiently. Actually, the mice package allows the user to deal with missing data by creating multiple imputations (empty values are replaced by some new imputed ones) for multivariate absent observations. It is anchored to Fully Conditional Specification, namely it is able to detect missing values, recognize their type and "*complete"* them on a variable by variable basis by specifying for each one the related imputation model. The default methods used by this application are linear regression to predict continuous missing data and logistic regression for categorical variables.

Method	Description	Scale type	Default
`pmm`	Predictive mean matching	numeric	Y
`norm`	Bayesian linear regression	numeric	
`norm.nob`	Linear regression, non-Bayesian	numeric	
`mean`	Unconditional mean imputation	numeric	
`2L.norm`	Two-level linear model	numeric	
`logreg`	Logistic regression	factor, 2 levels	Y
`polyreg`	Multinomial logit model	factor, >2 levels	Y
`polr`	Ordered logit model	ordered, >2 levels	Y
`lda`	Linear discriminant analysis	factor	
`sample`	Random sample from the observed data	any	

Table 10: *Built-in univariate imputation techniques. Source: "mice: Multivariate Imputation by Chained Equations in R", p.16, Stef van Buuren and Karin Groothuis-Oudshoorn (2000).*

From the package documentation we are able to see the different options available for each kind of data to be imputed. As shown in the table above, the MICE algorithm can deal with combinations of continuous, binary, unordered and ordered categorical data. Another useful tool of MICE is that one can impute continuous two-level data by means of passive imputation without losing consistency between imputations. In my work, I made use of such a tool to define new variables from other already existing ones. The code script of this section of the process is available in the Annex 2.

The process of imputation is also modifiable by the user in the sense that it is possible to specify alternative methods for the imputation of each variable of interest. For instance, if there are many categorical variables, each with many categories, one can speed up the algorithm considerably by imputing them as numerical variables with pmm instead of polyreg. This way of proceeding wouldn't cause any mistakes since pmm imputes only values that are observed; hence the original categories of the variable are preserved.

Hereafter, I reported the R codes with which I went through the ECF questionnaire data, ran the multiple imputation and defined the different empirical models for my analysis.

```
### IMPUTATION ###

rm(list = ls())

library(pacman)
pacman::p_load(readr,dplyr, psych, haven,ggplot2,tidyverse,dslabs,gridExtra,
gt,Hmisc,knitr,rmarkdown,stargazer,ivreg,sandwich,mitml,
broom.mixed,mice,modelsummary)

# Set workind directory
setwd("~/Desktop/Tesi Magistrale/R project")

path_dta <- "~/Desktop/Tesi Magistrale/R project/Survey of Financial
        Competences (ECF)/Data/Accesso ai dati"

# Insert ECF dataset
Tab1 <- read_dta(paste(path_dta,"/ECF/ecf_2016_e.dta", sep = ""))

# Legend of non-positive values
# - 97 →Don't know
# - 98 →Not relevant (filter not applied)
# - 99 →No reply

# Demographics
# a0000 gender
# a0100 Born in Spain
# a04 age
# a1100 education
# a1200 specialization
# j1300 income
# a1500 labor status
```

```
Tab1[Tab1 < 0] <- NA

# Creating a dataset with the selection of variables of interest for the study
DF_Partenza <- select(Tab1,a0000,a0100,a04,a1100,a1200,a1500,j1300,
                      e0401,e0402,e0403,e0200a,e0200b,e0300,a1400,
                      e0500,e0600,e0700,e0800,e0900,e1001,e1002,e1003,
                      e1101,e1102,e1200,
                      b0100,b0208,b0308,b0201,b0301,b0202,b0302,b0203,b0303,
                      b0204,b0304,b0205,b0305,b0206,b0306,b0207,b0307,b0209,
                      b0309,b0210,b0310)

# Coding variables as categorical
DF_Partenza$j1300 = as.factor(DF_Partenza$j1300)

DF_Partenza$a1500 = as.factor(DF_Partenza$a1500)
DF_Partenza$a1500 <-
  fct_collapse(DF_Partenza$a1500,
               Unemployed = c("4","5","7","8","9","10"),
               Employed = c("1","2","11","12","21","22"),
               Retired = c("6"))
DF_Partenza$a1500 <- fct_relevel(DF_Partenza$a1500,"Unemployed")

# ASSET CLASSES
# *** indicates those included in Div Index
# b0100  Current accounts/saving accounts
# b0308  Saving/term deposits  ***
# b0301  Mortgage
# b0302  Personal/occupational scheme ***
# b0303  Investment fund ***
# b0304  Shares ***
# b0305  Public/private fixed income ***
# b0306  Personal loan
# b0307  Credit card
# b0309  Life insurance ***
# b0310  Medical insurance ***

# MULTIPLE IMPUTATION --> MICE Package #
# Adding New Variable for passive imputation
DF_Partenza <- DF_Partenza %>%
                  mutate(age_square=NA,Male=NA,Born_in_Spain=NA,
                         STEM=NA,ECON=NA,Education=NA,Div_Index=NA,
                         High_Diversification=NA,Q1=NA,Q2=NA,Q3=NA,Q4=NA,
                         Q5=NA,Q6=NA,Q7=NA,Q8=NA,Q9=NA,Q10=NA,Q11=NA,
                         FL_Score=NA,High_FL=NA,Text_Understanding=NA)
```

```
## 1st phase of imputation
DF_Imputed <- mice(DF_Partenza, maxit = 5, seed=500 )

# Definition of passive imputation method for each NEW variable
meth<-DF_Imputed$meth

meth["age_square"]<- "~I(a04^2)"
meth["Male"]<- "~I(ifelse(a0000==1,1,0))"
meth["Born_in_Spain"]<- "~I(ifelse(a0100==1,1,0))"

# Education attainment and two specialization paths for the upper educated
meth["Education"]<- "~I(ifelse(a1100<=2,0,ifelse(a1100>2&a1100<=6,1,2)))"
meth["STEM"]<- "~I(ifelse((a1100>=7) & (a1200==1| a1200==2| a1200==6),1,0))"
meth["ECON"]<- "~I(ifelse((a1100>=7) & (a1200==4),1,0))"

## Dependent variable
meth["Div_Index"] <- "~I(b0302+b0303+b0304+b0305+b0308+b0309+b0310)"

## Eleven FL questions and overall FL Score
meth["Q1"]<- "~I(ifelse(e0500==200,1,0))"
meth["Q2"]<- "~I(ifelse(e0600==3|e0600==3,1,0))"
meth["Q3"]<- "~I(ifelse(e0700==0,1,0))"
meth["Q4"]<- "~I(ifelse(e0800==102,1,0))"
meth["Q5"]<- "~I(ifelse(e0900==1,1,0))"
meth["Q6"]<- "~I(ifelse(e1001==1,1,0))"
meth["Q7"]<- "~I(ifelse(e1002==1,1,0))"
meth["Q8"]<- "~I(ifelse(e1003==1,1,0))"
meth["Q9"]<- "~I(ifelse(e1101==1,1,0))"
meth["Q10"]<- "~I(ifelse(e1102==3,1,0))"
meth["Q11"]<- "~I(ifelse(e1200==1,1,0))"
meth["FL_Score"] <- "~I(Q1+Q2+Q3+Q4+Q5+Q6+Q7+Q8+Q9+Q10+Q11)"
meth["High_FL"]<- "~I(ifelse(FL_Score>9,1,0))"

## 2nd phase of imputation - Includes the passive imputation
DF_Imputed <- mice(DF_Partenza,maxit = 5,  seed=500, meth=meth)

save(DF_Imputed, file="DF_Imputato.rda")
```

EMPIRICAL ANALYSIS

```
load("DF_Imputato.rda")
x <- DF_Imputato
DF_Imputato <- mice::complete(DF_Imputato, "long")

# Renaming some variables
DF_Imputato <- DF_Imputato %>% rename(Age=a04, Native=a0100, Gender=a0000,
Educ=a1100,
                        Study_Area=a1200, Status=a1500, Income=j1300,
                        Text1=e0401, Text2=e0402, Text3=e0403,
                        Graph_US1=e0200a, Graph_US2=e0200b,
                        Math_Logical=Numeracy, Books_Number=a1400)

(DF_Imputato <- DF_Imputato %>% group_by(.imp) %>% nest())

## LINEAR MODEL ON FL SCORE ##

Reg = DF_Imputato %>%
  mutate(FL_linear = map(data,
                  ~lm(FL_Score~Male+Age+age_square+Native+as.factor(Status)+
                       as.factor(Income)+as.factor(Education)+ECON,
                      data = .)))
FL_linear <- pool(Reg$FL_linear)

## OLS ##

Reg = DF_Imputato %>%
  mutate(lin_reg1 = map(data,
                  ~lm(Div_Index~Male+Age+age_square+Native+as.factor(Status)+
                       as.factor(Income)+as.factor(Education)+ECON+FL_Score,
                      data = .)))
OLS1 <- pool(Reg$lin_reg1)

Reg = Reg %>%
  mutate(lin_reg2 = map(data,
                  ~lm(Div_Index~Male+Age+age_square+Native+as.factor(Status)+
                       as.factor(Income)+as.factor(Education)+ECON+High_FL,
                      data = .)))
OLS2 <- pool(Reg$lin_reg2)
```

```
## IV ##

# IV1
Reg = Reg %>%
  mutate(iv_reg1 = map(data,
         ~ivreg(Div_Index~Male+Age+age_square+Native+as.factor(Status)+
                as.factor(Income)+as.factor(Education)+ECON+FL_Score|

                Male+Age+age_square+Native+as.factor(Status)+
                as.factor(Income)+as.factor(Education)+ECON+Books_Number+STEM,
                data = .)))
IV1 <- pool(Reg$iv_reg1)

# IV2
Reg = Reg %>%
  mutate(iv_reg2 = map(data,
         ~ivreg(Div_Index~Male+Age+age_square+Native+as.factor(Status)+
                as.factor(Income)+as.factor(Education)+ECON+High_FL|

                Male+Age+age_square+Native+as.factor(Status)+
                as.factor(Income)+as.factor(Education)+ECON+Books_Number+STEM,
                data = .)))
IV2 <- pool(Reg$iv_reg2)

## DIAGNOSTICS ##

IMP1 <- mice::complete(x,1)
IMP2 <- mice::complete(x,2)
IMP3 <- mice::complete(x,3)
IMP4 <- mice::complete(x,4)
IMP5 <- mice::complete(x,5)

IMP_Single <- list(IMP1,IMP2,IMP3,IMP4,IMP5)

IMP_Single <- lapply(IMP_Single, rename, Age=a04, Native=a0100, Gender=a0000,
                Educ=a1100, Study_Area=a1200,Status=a1500, Income=j1300,
                Text1=e0401, Text2=e0402, Text3=e0403, Graph_US1=e0200a,
                Graph_US2=e0200b, Math_Logical=Numeracy, Books_Number=a1400)

IMP1 <- IMP_Single[[1]]
IMP2 <- IMP_Single[[2]]
IMP3 <- IMP_Single[[3]]
IMP4 <- IMP_Single[[4]]
IMP5 <- IMP_Single[[5]]
```

```
IV_IMP1 <- ivreg(Div_Index~Male+Age+age_square+Native+as.factor(Status)+
        as.factor(Income)+as.factor(Education)+ECON+FL_Score|

        Male+Age+age_square+Native+as.factor(Status)+
        as.factor(Income)+as.factor(Education)+ECON+Books_Number+STEM,data=IMP1)

IV_IMP2 <- ivreg(Div_Index~Male+Age+age_square+Native+as.factor(Status)+
        as.factor(Income)+as.factor(Education)+ECON+FL_Score|

        Male+Age+age_square+Native+as.factor(Status)+
        as.factor(Income)+as.factor(Education)+ECON+Books_Number+STEM,data=IMP2)

IV_IMP3 <- ivreg(Div_Index~Male+Age+age_square+Native+as.factor(Status)+
        as.factor(Income)+as.factor(Education)+ECON+FL_Score|

        Male+Age+age_square+Native+as.factor(Status)+
        as.factor(Income)+as.factor(Education)+ECON+Books_Number+STEM,data=IMP3)

IV_IMP4 <- ivreg(Div_Index~Male+Age+age_square+Native+as.factor(Status)+
        as.factor(Income)+as.factor(Education)+ECON+FL_Score|

        Male+Age+age_square+Native+as.factor(Status)+
        as.factor(Income)+as.factor(Education)+ECON+Books_Number+STEM,data=IMP4)

IV_IMP5 <- ivreg(Div_Index~Male+Age+age_square+Native+as.factor(Status)+
        as.factor(Income)+as.factor(Education)+ECON+FL_Score|

        Male+Age+age_square+Native+as.factor(Status)+
        as.factor(Income)+as.factor(Education)+ECON+Books_Number+STEM,data=IMP5)

IV_SINGLE_DF <- list(IV_IMP1,IV_IMP2,IV_IMP3,IV_IMP4,IV_IMP5)

lapply(IV_SINGLE_DF, summary,diagnostics=TRUE)
```

`

References

Abreu, Margarida, e Victor Mendes. «Financial Literacy and Portfolio Diversification». *Quantitative Finance, 10*. (1 May 2010): 515-28. https://doi.org/10.1080/14697680902878105.

Agarwal, Sumit, Jeremy Tobacman, e Paige Skiba. «Payday Loans and Credit Cards: New Liquidity and Credit Scoring Puzzles?» *American Economic Review* 99 (1 aprile 2009): 412–17. https://doi.org/10.2139/ssrn.1327125.

Alsemgeest, L. «Arguments for and against financial literacy education: where to go from here?» *International Journal of Consumer Studies* 39 (2015): 155–61.

Angrisani, L., Santonicola, A., Iovino, P. *et al.* «IFSO Worldwide Survey 2016: Primary, Endoluminal, and Revisional Procedures». *OBES SURG* 28, 3783–3794 (2018). https://doi.org/10.1007/s11695-018-3450-2

Anthes, William L. «Financial Illiteracy in America: A Perfect Storm, a Perfect Opportunity». *Journal of Financial Service Professionals* 58, n. 6 (novembre 2004): 49–56.

Arellano, A., Cámara, N., & Tuesta, D. (2014). «The effect of self-confidence on financial literacy». *BBVA Research WP*, *14*, 28.

Banca d'Italia. «L'educazione finanziaria: le iniziative a livello internazionale», 2009.

Banco de España. «Banco de España - Statistics - Statistics by subject - Surveys of households and individuals - Survey of Financial Competences». https://www.bde.es/bde/en/areas/estadis/estadisticas-por/encuestas-hogar/relacionados/encuesta-de-comp/ECF_2016.html.

Barber, Brad & Odean, Terrance. (2001). «Boys Will Be Boys: Gender, Overconfidence, And Common Stock Investment». *The Quarterly Journal of Economics.* 116. 261-292. 10.2139/ssrn.139415.

Barboza, Gustavo, Chad Smith, e James G. Pesek. «Assessing Financial Literacy, Gender Gap and Cognitive Differences». *Journal of Financial Education* 42, n. 3–4 (2016): 205–42.

Batsaikhan, Uuriintuya, e Maria Demertzis. «Financial Literacy and Inclusive Growth in the European Union». *Bruegel-Policy Contributions*, maggio 2018, NA.

Becchetti, Leonardo, Stefano Caiazza, e Decio Coviello. «Financial Education and Investment Attitudes in High Schools: Evidence from a Randomized Experiment». *Applied Financial Economics* 23 (7 settembre 2011). https://doi.org/10.2139/ssrn.1923626.

Bernheim, Douglas. «Financial Illiteracy, Education, and Retirement Saving». *Wharton School Pension Research Council, University of Pennsylvania, Pension Research Council Working Papers*, 1 gennaio 1998.

Bianchi, Milo. «Financial Literacy and Portfolio Dynamics». *The Journal of Finance* 73, n. 2 (2018): 831–59.

Bonaparte, Yosef. «On the Portfolio Choice of Crypto Asset Class: Why the Millennials Own Crypto? (2021) ». Available at SSRN: https://ssrn.com/abstract=3829275.

Brown, Meta, John Grigsby, Wilbert van der Klaauw, Jaya Wen, e Basit Zafar. «Financial Education and the Debt Behavior of the Young». *The Review of Financial Studies* 29, n. 9 (2016): 2490–2522.

Buch, C. M. «Financial Literacy and Financial Stability», 2018.

Calvet, Laurent E., John Y. Campbell, and Paolo Sodini. "Down or Out: Assessing the Welfare Costs of Household Investment Mistakes." *Journal of Political Economy* 115, no. 5 (2007): 707–47. https://doi.org/10.1086/524204.

Carpena, Fenella & Cole, Shawn & Shapiro, Jeremy & Zia, Bilal. (2011). «Unpacking the Causal Chain of Financial Literacy».

Christelis, Dimitris & Jappelli, Tullio & Padula, Mario. (2008). «Cognitive Abilities and Portfolio Choice. *SSRN Electronic Journal*». 10.2139/ssrn.1275284.

Davoli, Maddalena, e Jia Hou. «Financial literacy, institutions and education: Lessons from the German reunification». *German Economic Review*, 8 maggio 2021. https://doi.org/10.1515/ger-2020-0027.

Davoli, Maddalena, e Nuria Rodriguez-Planas. «Culture and adult financial literacy: Evidence from the United States». *Economics of Education Review* 78 (1 giugno 2020): 102013. https://doi.org/10.1016/j.econedurev.2020.102013.

De Winne, Rudy; Petkeviciute, Aiste. «Financial Literacy and Multi-Asset Portfolio Diversification ». *International Conference of the French Finance Association (AFFI)* (Nantes, du 26/05/2021 au 28/05/2021)

Disney, Richard & Gathergood, John. (2013). «Financial literacy and consumer credit portfolios». *Journal of Banking & Finance*. 37. 2246–2254. 10.1016/j.jbankfin.2013.01.013.

Driessen, Joost, e Luc Laeven. «International portfolio diversification benefits: Cross-country evidence from a local perspective». *Journal of Banking & Finance* 31, n. 6 (1 June 2007): 1693–1712. https://doi.org/10.1016/j.jbankfin.2006.11.006.

Evans, J.L. and Archer, S.H. (1968). «Diversification and the reduction of dispersion: an empirical analysis». *The Journal of Finance*, 23: 761-767. https://doi.org/10.1111/j.1540-6261.1968.tb00315.x

Evans, M.D., Kelley, J., Sikora, J. & Treiman, D. (2010) «Family scholarly culture and educational success: books and schooling in 27 nations». *Research in Social Stratification and Mobility* [Online], 28, 171–197. doi:10.1016/j.rssm.2010.01.002

European Economic and Social Committee. «Financial education strategies and best practices within the European Union (Second Edition) », 2017.

Fairfax, Lisa M. «The Securities Law implications of Financial Illiteracy». *Virginia Law Review* 104, n. 6 (2018): 1065–1122.

Fernandes, Daniel & Lynch, John & Netemeyer, Richard. (2014). «Financial Literacy, Financial Education, and Downstream Financial Behaviors». *Management Science.* 10.1287/mnsc.2013.1849.

Friedman, Milton. «Essays in Positive Economics». (1953)

Fort, Margherita, Francesco Manaresi, Serena Trucchi, Elena Carletti, e Fabian Waldinger. «Adult financial literacy and households' financial assets». *Economic Policy* 31, n. 88 (2016): 743–82.

Gelman, Andrew, Gary King, e Chuanhai Liu. «Not Asked and Not Answered: Multiple Imputation for Multiple Surveys». *Journal of the American Statistical Association* 93, n. 443 (1998): 846–57. https://doi.org/10.2307/2669819.

Gomes, Francisco, Michael Haliassos, e Tarun Ramadorai. «Household Finance». *Journal of Economic Literature* forthcoming (17 marzo 2020). https://doi.org/10.1257/jel.20201461.

Guiso, L, e E Viviano. «How Much Can Financially Literacy Help? ». *Review of Finance* 19 (2015): 1347–82.

Guiso, Luig,i e Tullio Jappelli. «Financial literacy and portfolio diversification». *European University Institute, Economics Working Papers*, 1 January 2008.

Guiso, Luigi, e Paolo Sodini. «Household Finance: An Emerging Field». *Handbook of the Economics of Finance* 2 (1 April 2012). https://doi.org/10.1016/B978-0-44-459406-8.00021-4.

Haliassos, Michael. (2002). «Stockholding: Recent Lessons from Theory and Computations». Stockholding in Europe. *10.2139/ssrn.312899.*

Hastings, J.S., B.C. Madrian, e W.L. Skimmyhorn. «Financial Literacy, Financial Education and Economic Outcomes». *Annual Review of Economics* 5 (2013): 347–73.

Hospido, L., S Izquierdo, e M Machelett. «Diferencias de género en competencias financieras». *Boletín Económico* Quarter 1 (2021).

Hsiao, Yu-Jen , Tsai, Wei-Che. «Financial literacy and participation in the derivatives markets». *Journal of Banking & Finance,* Volume 88, 2018, Pages 15-29, ISSN 0378-4266, https://doi.org/10.1016/j.jbankfin.2017.11.006.

Hung, Angela & Parker, Andrew & Yoong, Joanne. (2009). «Defining and Measuring Financial Literacy». *RAND Corporation Publications Department, Working Papers. 708.* 10.2139/ssrn.1498674.

Huston, Sandra. (2009). «Measuring Financial Literacy». *Journal of Consumer Affairs*. 44. 10.2139/ssrn.1945216.

Hysmith, Ryan. «Student-Managed Investment Fund Participation and Financial Knowledge, Satisfaction, and Behavior». *Journal of Financial Education* 46, n. 2 (2020): 251–70.

Jappelli, Tullio & Padula, Mario. (2011). «Investment in Financial Literacy and Saving Decisions». *Journal of Banking & Finance*. 37. 10.1016/j.jbankfin.2013.03.019.

Jappelli, Tullio & Padula, Mario. (2013). «Investment in Financial Literacy, Social Security, and Portfolio Choice». *Journal of Pension Economics and Finance* 14, no. 4: 369–411. doi:10.1017/S1474747214000377.

Kiviet, Jan & Kripfganz, Sebastian. (2021). «Instrument approval by the Sargan test and its consequences for coefficient estimation». *Economics Letters.* 205. 109935. 10.1016/j.econlet.2021.109935.

Klapper, Leora & Lusardi, Annamaria & Panos, Georgios. (2012). «Financial Literacy and the Financial Crisis». *SSRN Electronic Journal*. 10.2139/ssrn.2038765.

Koh, Benedict & Mitchell, Olivia & Rohwedder, Susann. (2018). «Financial Knowledge and Portfolio Complexity in Singapore». *The Journal of the Economics of Ageing.* 17. 10.1016/j.jeoa.2018.11.004.

Kumar, Alok & Goetzmann, William. (2001). «Equity Portfolio Diversification». *Review of Finance.* 12. 10.1093/rof/rfn005.

Lusardi, Annamaria. «Numeracy, Financial Literacy, and Financial Decision-Making». *Numeracy* 5 (1 febbraio 2012). https://doi.org/10.5038/1936-4660.5.1.2.

Lusardi, Annamaria, e Olivia Mitchell. «Baby Boomer Retirement Security: The Roles of Planning, Financial Literacy, and Housing Wealth». *Journal of Monetary Economics* 54 (1 aprile 2006): 205–24. https://doi.org/10.1016/j.jmoneco.2006.12.001.

Lusardi, Annamaria & Mitchell, Olivia. (2014). «The Economic Importance of Financial Literacy: Theory and Evidence». *Journal of Economic Literature*. 52. 10.2139/ssrn.2243635.

Lusardi, Annamaria & Mitchell, Olivia. (2007). «Financial Literacy and Retirement Preparedness: Evidence and Implications for Financial Education». *Business Economics*. 42. 35-44. 10.2145/20070104.

Lusardi, A., & Mitchell, O. S. (2011). «Financial Literacy Around the World: an Overview». *Journal of Pension Economics and Finance, 10* (4), 497-508. http://dx.doi.org/10.1017/S1474747211000448

MacArdle, John, James Smith, e Robert Willis. «Cognition and Economic Outcomes in the Health and Retirement Survey». *NBER national bureau of economic research, NBER working paper No. 15266*, 1 agosto 2009. https://doi.org/10.2139/ssrn.1434603.

Mancebòn Torrubia, M.J., e D.P. Ximénez-de-Embùn. «Habilidades financieras de la población española adulta: diagnóstico y determinantes», 2020.

Mancebòn Torrubia, M.J., D.P. Ximénez-de-Embùn, e A Villar-Aldonza. «Habilidades financieras y hábitos financieros saludables: un análisis a partir de la Encuesta de Competencias Financieras». *Papeles de Economia Española* 166 (2020): 55–69.

Markowitz, Harry. «Portfolio Selection». *The Journal of Finance* 7, no. 1 (1952): 77–91. https://doi.org/10.2307/2975974.

Markowitz, Harry M. «Markowitz Revisited». *Financial Analysts Journal* 32, n. 5 (1 settembre 1976): 47–52. https://doi.org/10.2469/faj.v32.n5.47.

Maté, Anna, Irma Martinez-Garcia, e Gloria Suárez. «Financial education and savings and investment decisions: An analysis of the Survey of Financial Competences (ECF)». *SSRN Electronic Journal*, 20 aprile 2021.

Mitchell, Olivia, e Annamaria Lusardi. «Financial Literacy Around the World: An Overview». *Journal of Pension Economics and Finance* 10 (1 febbraio 2011): 497–508. https://doi.org/10.2139/ssrn.1810551.

Morgan, Peter J., e Long Q. Trinh. «Determinants and Impacts of Financial Literacy in Cambodia and Viet Nam». *Journal of Risk and Financial Management* 12, n. 1 (2016). https://doi.org/10.3390/jrfm12010019.

Muller, S, e M Weber. «Financial literacy and mutual fund investments: Who buys actively managed funds?» *Schmalenbach Business Review* 62 (2010): 126–53.

Odean, T. (1998) «Are Investors Reluctant to Realize Their Losses? ». *The Journal of Finance,* 53, 1775-1798. https://doi.org/10.1111/0022-1082.00072

Paradowski, Piotr, Andrej Cupak, Pirmin Fessler, and Joanne W. Hsu. «Confidence, Financial Literacy and Investment in Risky Assets: Evidence from the Survey of Consumer Finances». *Finance and Economics Discussion Series 2020-004. Washington: Board of Governors of the Federal Reserve System*, 2020. https://doi.org/10.17016/feds.2020.004.

Peng, Congmin, Po-Wen She, e Ming-Kun Lin. «Financial Literacy and Portfolio Diversity in China». *Journal of Family and Economic Issues*, 30 gennaio 2022. https://doi.org/10.1007/s10834-021-09810-3.

Samuelson, Paul A. «General Proof that Diversification Pays». *Journal of Financial and Quantitative Analysis* 2, n. 1 (1967): 1–13. https://doi.org/10.2307/2329779.

Samuelson, Paul A. «Lifetime Portfolio Selection By Dynamic Stochastic Programming». *The Review of Economics and Statistics* 51, no. 3 (1969): 239–46. https://doi.org/10.2307/1926559.

Sánchez Campillo, José, Dolores Moreno Herrero, e José Antonio Rodríguez Martín. «Efectos de la educación en el nivel de las contribuciones a los planes privados de pensiones de las familias en España». *Innovar* 26 (2016): 95–112.

Soskic, D. «Financial literacy and financial stability». 2011.

Terman L.M. & Oden M.H. (1959) «Genetic studies of genius», Vol IV. *The Gifted Child Grows up: Twenty-Five Years' Follow-up of a Superor Group.* Stanford University Press: Stanford, California.

Tu, Jun, e Guofu Zhou. «Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies». *Journal of Financial Economics* 99, n. 1 (1 gennaio 2011): 204–15. https://doi.org/10.1016/j.jfineco.2010.08.013.

Van Buuren, Stef. «mice: Reference Manual», s.d. https://cran.r-project.org/web/packages/mice/mice.pdf.

Van Buuren, Stef, e Catharina Groothuis-Oudshoorn. «MICE: Multivariate Imputation by Chained Equations in R». *Journal of Statistical Software* 45 (1 dicembre 2011). https://doi.org/10.18637/jss.v045.i03.

Van Rooji, M, A Lusardi, e R Alessie. «Financial literacy and stock market participation». *Journal of Financial Economics* 101 (2011): 449–72.

Von Gaudecker, Hans Martin. «How Does Household Portfolio Diversification Vary with Financial Literacy and Financial Advice? » *The Journal of Finance* 70, n. 2 (2015): 489–507.

Wagner, W. H., and S. C. Lau. "The Effect of Diversification on Risk». *Financial Analysts Journal* 27, no. 6 (1971): 48–53. http://www.jstor.org/stable/4470866.