Equity.com: has the Internet made the stock market a level playing field?

Alberto Marcos Fernández
Bachelor's Degree in Philosophy, Politics and Economics
Final Bachelor Thesis

Tutor: Eva Østergaard-Nielsen
Department of Political Science and Sociology

May 2024

Acknowledgements

I wanted to dedicate a few lines to express my gratitude to those people that have contributed to the completion of this bachelor thesis.

Firstly, I wanted to recognise Eva Østergaard-Nielsen (Ph.D. in Political Science), my tutor, for her help throughout the whole process of writing my thesis.

I also wanted to dedicate some words of appreciation to Déborah Sulem (Ph.D. in Statistics) for selflessly helping me in the selection process of an appropriate statistical model.

Lastly, I wanted to thank my family, and my partner for not only patiently enduring my long-winded speeches on household finance and stock market participation but also being there for me every step of this wonderful journey.

Statement of Authorship

I hereby declare that I am the sole author of this bachelor thesis and that I have not used any sources other than those listed in the bibliography and identified as references. I further declare that I have not submitted this thesis at any other institution in order to obtain a degree.

Alberto Marcos Fernández

Abstract

The Internet has proved to be a powerful tool affecting many aspects of our daily lives. The present thesis is concerned about the impact it has had on increasing stock market participation among economically underprivileged households. The Survey of Consumer Finances, is used to estimate *if* the Internet has increased low-income household stock market participation, *to whom* it has benefitted the most and to see whether disparities in *how efficiently* households invest with respect to economic factors could be related to these technological breakthroughs. Results suggest the Internet has increased the stock market participation of low-income households. Among these, the ones that have seen a steeper increase are economically underprivileged respondents reporting a low degree of risk aversion, while their highly-risk averse counterparts are lagging behind. Results also show risk-return efficiency is positively associated with wealth, a finding compatible with hypothesized and unobserved differences in financial advice that households receive across income.

keywords: **household finance, Internet, stock market participation**

Table of contents

1. Introduction

"How has the advent of widespread online access to financial information and financial platforms in the USA affected low-income households' participation in the stock market?"

The arrival of the Internet was claimed to be a disruptive force that would change everything. In this sense, one of its most notable promises was the arrival of an *everything-democratization*. The present thesis aims to examine the extent to which that potential has materialized in the field of household finance. In other words, has widespread access to online banks and to financial information through the Internet resulted in increased stock market participation of low-income households? Additionally, I will try to discern if these changes have created any new inequalities among disfavored households. However, since participation at all costs is not something positive either, I will also try to examine if this supposed level stock market playing field presents an efficiency (risk-return optimization) gap in the way that households invest with respect to economic variables (income and wealth).

The section immediately below is the theoretical framework. Due to the fact that the present thesis has a strong prescriptive component (specially the part about investing efficiency), I have decided to start the conceptual framework with theoretical models about the market (CAPM) and investors (utility maximizers) behavior. They act as normative benchmarks to evaluate households' attitudes toward investing. On the contrary, the section following the theoretical models is descriptive. It discusses the empirical findings about the economic, demographic and behavioral characteristics that have been shown to affect stock market participation. Even though these characteristics should (prescriptively) not influence stock holding attitudes (as it will be discussed in the theoretical models), they should be controlled for when estimating the impact of the Internet on households' willingness to invest in the stock market, as they have been shown to be empirically significant. Lastly, I will comment on the two main hurdles that low and middle-income households have historically faced when trying to enter the stock market and I will briefly discuss the potential of the Internet to eliminate them. Next, I will focus on the hypotheses to be tested, the methodology followed for this purpose and the datasets used. Afterward, the results of the statistical models will be displayed and I will discuss them. The last section will be the concluding remarks.

1.1 Relevance

Why is it important that lower and middle-income households be included in the stock market? Before delving into the matter that concerns us today, I deem it adequate to dedicate a few

lines reflecting on its relevance. Modern postindustrial societies present a new array of never-seen-before problems such as aging populations, debt increase, and overall lack of financial sustainability of their welfare models. Even though these new issues do not represent an existential threat, they are dire and should be addressed. In consequence, be it because we want to rethink the public welfare system or because we want to create a private safety net, the transition (which does not need to be complete) toward a resilient capitalization model seems desirable.

Among the sources of savings through which the citizens of a society could protect their savings from inflation and secure their retirement income, a portfolio that includes (at least a fraction of) holdings in equity has historically preserved purchasing power in the long run (Boudoukh and Richardson, 1993). In a nutshell, democratizing a sensible approach to equity investment could serve as a means to preserve the level of welfare that developed countries have historically enjoyed.

2. Literature review

I have decided to divide the literature review into two components: a market model and household finance theory. In the former, I will present the Capital Asset Pricing Model, from which the Efficient Market Hypothesis can be derived (assessing the theoretical justification for the existence of the risk premium). The latter part of the conceptual framework will delve into the theoretical and rational justifications that would warrant market asset holding for the majority of households (rational household finances). In this section I will also evaluate the degree of compliance that households present to the aforementioned normative rational model (behavioral household finances).

2.1 The CAPM and the Efficient Market Hypothesis:

In the simplest terms, a market is "any region in which buyers and sellers are in such free intercourse with one another that the prices of the same goods tend to equality easily and quickly" (Cournot, 1838: p.55). The financial market is the specific place (be it physical or digital) where the demand for raising finances and the supply of liquidity of private agents meet each other.

In order to draw the path toward the second part of the literature review (i.e. the role of the individual in the market), it is necessary to understand in the first place how prices of financial assets are set. Depending on our description one shall prescribe different investing strategies. For simplicity, I have decided to conceptualize financial markets price-setting dynamics through William Sharpe's (1964) work on the Capital Asset Pricing Model (CAPM). *For further information about the decision-making process on the election CAPM see Appendix A)*.

Borrowing Markowitz's (1952) model of utility maximization through risk-adjusted returns in portfolio design, Sharpe proposes a market in which investors seek the maximum expected returns (E_R) for a certain level of volatility (the term standard deviation will be interchangeably used to refer to volatility, and both will be represented with σ_R), proxied for risk. If individuals seek high expected returns and low standard deviations, one could graphically plot different portfolio compositions of risky assets that provide them with equal utility. Risk-return tradeoffs can be plotted in an upward-sloping *indifference curve* (*see Fig.1*). Indifference curves are dominated by the ones located on their right, which represent asset selections that provide a higher utility. Consequently, individuals will select the asset combination from the realm of possibilities (the shaded area in Fig.1) that is tangent to the right-most possible indifference curve (the option that represents the highest amount of utility). This corresponds to point *F* in Fig.1.

The market in which Sharpe's investors interact is initially composed of a risky asset (he will ulteriorly add more of them) and a riskless one. Consequently, the risk-return possibilities that any

investor could get could be plotted as seen in Fig. 2. Provided that the equity risk premium is positive, different amounts of wealth devoted to the risky asset form a Capital Market Line across the risk-return plane. Its intersection with the horizontal axis indicates the Pure Interest Rate of the riskless asset, the case in which an investor lends their whole wealth at the riskless asset rate. Additionally, values along the line would be associated with an increasingly higher proportion of wealth invested in the risky asset. The Capital Market Line does not have an end because, at least theoretically, investors could get limitless leverage, finance their operations at the riskless rate of interest in order to invest in the risky asset.

At this point, it is possible to add more risky assets into the market. The shaded area in Fig.1 is the set of possible risk-returns combinations resulting from allocating different proportions of wealth to different risky assets. It has a U-shaped right side (connecting points A to X) called the *Investment Opportunity Curve*. It comprises every efficient combination of assets possible. They are efficient, for one cannot go rightward and get higher returns without going upward and assuming a higher standard deviation*, (for a detailed explanation on the reason behind this U-shaped Investment Opportunity Curve see Appendix B*).

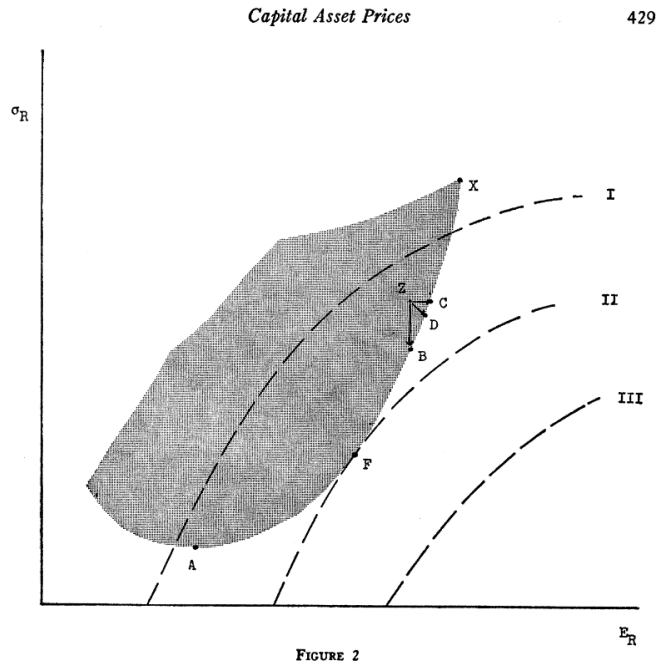

Fig. 1: Indifference curves representing equally desirable risk-return combinations (Sharpe, 1964: 429).

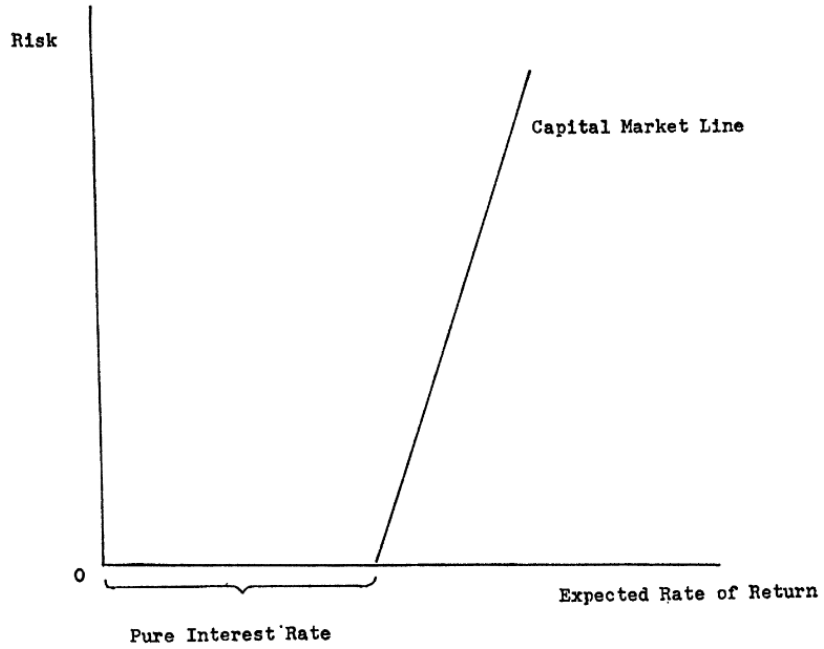

Fig. 2: Capital Investment Line: different risk-return combinations available in a market conformed by a risky and a riskless asset (Sharpe, 1964: 426).

As a consequence, investors will seek to invest in an efficient asset combination (i.e. one located on the Investment Opportunity Curve). Their possible returns will be located along a Capital Market Line (such as the one in Fig. 2). However, since they can now select from different risky assets, the slope of the line will also be variable (*see Appendix C for a detailed explanation on the Capital Investment Line*). In the case of a market of the characteristics of Fig. 3, the dominant risky asset combination will be Ø. The reason behind it is that there exists no Capital Market Line with a lower slope connecting the Pure Interest Rate of the riskless asset with any risky asset combination. Therefore, there exists no asset combination that provides lower increases in standard deviation for additional expected returns.

Once the stage for investors' interaction in the market is set, two conditions ought to be satisfied to achieve the price equilibrium. First, investors have to get financing at the same rate, the Pure Interest Rate. Second, homogeneity of prospects is assumed (i.e. investors have a common view about assets' risk-return characteristics and their correlation coefficients).

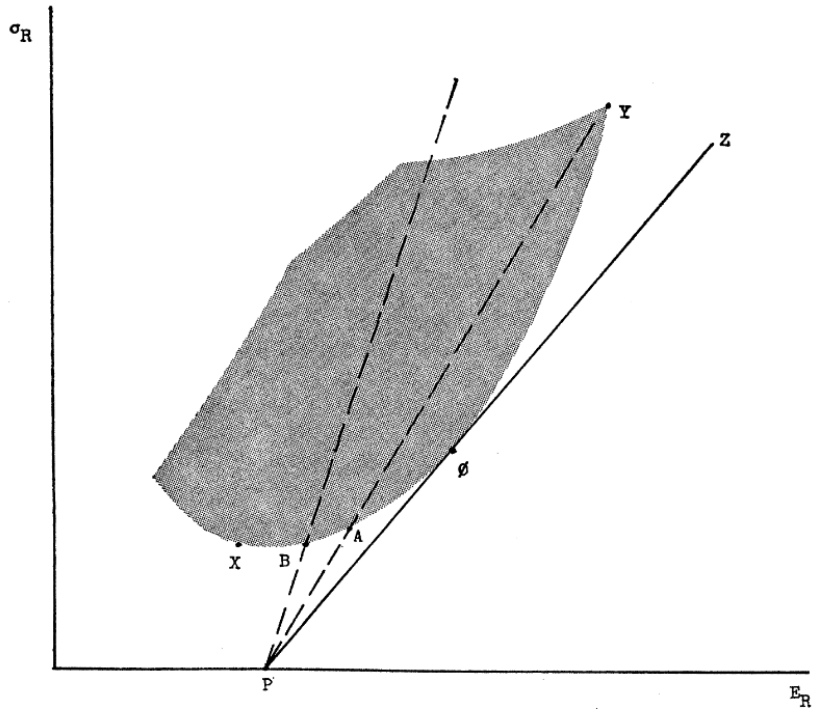

Fig. 3: The figure shows the difference in slope depending on the risky asset combination selected among the ones present in the Investment Opportunity Curve. Combination examples shown would be A, B and Ø (Sharpe, 1964: 433).

Under these conditions investors will invest according to the only interpersonal variable left, their indifference curves. An illustration of one of such markets would be Figure 4, three investors' (A, B and C) indifference curves can be seen superimposed along the most efficient Capital Market Line. In this instance, investor A would be the most risk averse for their indifference curve lies the closest to the pure interest rate which means lending all their wealth at the Pure Interest Rate. Conversely, investor C, the least risk averse would prefer to borrow at the riskless asset interest rate to invest in Ø. The middle point is represented by investor B, who would invest all their wealth in point Ø risky asset combination.

As point Ø is more demanded than other less efficient asset combinations, its price will increase. As a result, since prices and expected returns are negatively correlated, it will lose its attractiveness. Consequently it will move leftward, as other relatively more demanded asset combinations would do. Conversely, less demanded ones' prices would fall, becoming comparatively more attractive. Accordingly, their points (e.g. points G and F) would move rightward. The result from these two displacements would be a lining up of points so that every risky asset in the market has at least one combination on the Capital Market Line and the market equilibrium is achieved. Now, every investor can hold an efficient asset combination regardless of their indifference curve (i.e. their

risk aversion). Furthermore, these asset allocations are perfectly positively correlated, as they are just a combination of the fraction of wealth allocated into the riskless asset (an amount that could be negative if an investor decides to borrow at the riskless rate, like investor C of Fig. 3 would) and a portion of their wealth (which can be 0 for extremely risk averse individuals) in the risky efficient combination *g* (*see Fig. 6*).

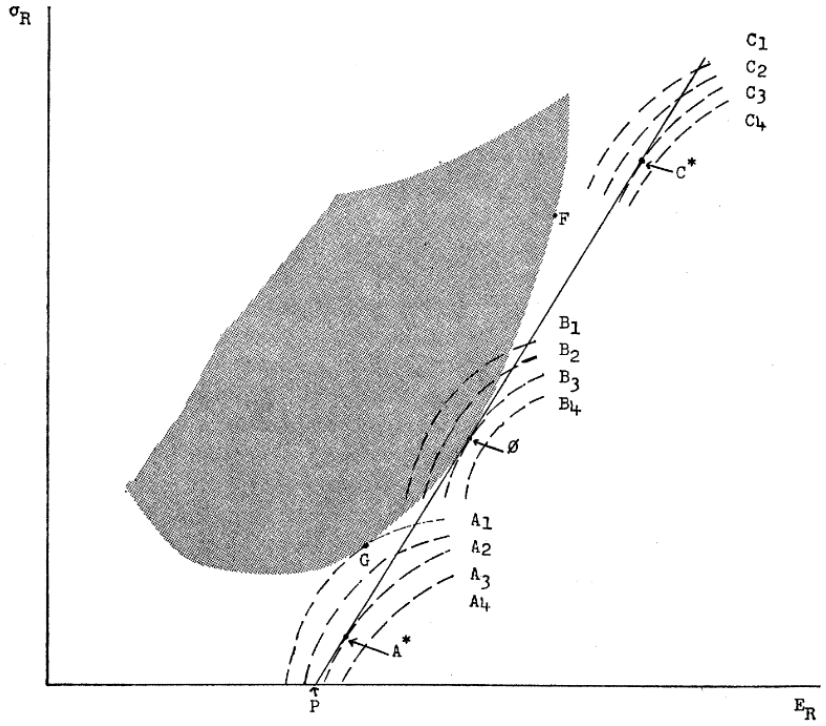

Fig. 4: In this market, indifference curves of three investors (i.e. A, B and C) are superimposed on the optimal Capital Market Line connecting the Pure Interest Rate to point Ø (*see Appendix C for an explanation about the optimal risk-return slope*) (Sharpe, 1964: 433).

In light of the above, Sharpe states that a set of combinations where every risky asset in the market is held exists along the Investment Opportunity Curve. Nevertheless, he still has not concluded anything about the reason why each efficient combination is linearly related to all the other ones along the Capital Market Line. In order to try to elucidate this reason, Sharpe differentiates between two sources of risk: a systematic and an unsystematic (or idiosyncratic) component in a way that resembles Markowitz's (1952) Modern Portfolio Theory. Accordingly, the standard deviation of an asset (i.e. volatility, the proxy for risk) depends on two variables. On the one hand, idiosyncratic risk, which is the one related to the specific asset, and can be diluted (i.e. averaged out) through diversification without decreasing expected returns (getting closer to the Capital Market Line, from point *i* to *g* in Fig. 6). On the other hand, systematic risk is common to all assets and, consequently cannot be averaged out through diversification. This is the risk to which every efficient asset

allocation on the Capital Market Line is linearly related (as idiosyncratic risks have been eliminated through diversification as all combinations on the Capital Market Line include the efficient *g* risky asset combination). Sharpe hypothesizes that this linear relationship is the result of the interaction of each asset allocation with an unobserved variable. He concludes with one plausible candidate to fulfill the role of this hidden variable to which all efficient asset allocations are linearly related. He hypothesizes this unobserved factor is dependence on the overall level of economic activity. In other words, the exposure to economic swings is the reason that compensates risky assets, the reason behind the existence of the positive *risk premium*.

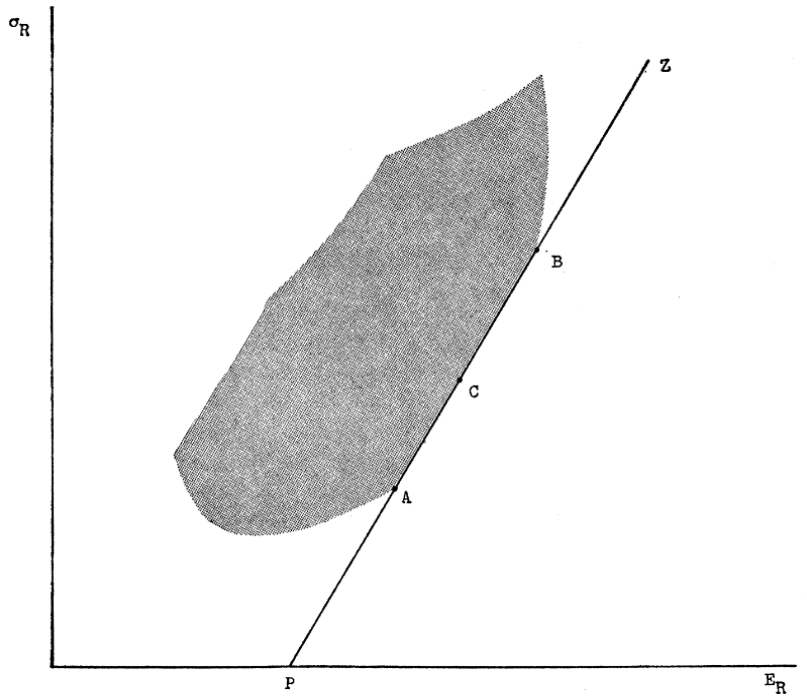

Fig. 5: Market equilibrium, the Investment Opportunity matches the Capital Market Line

Sharpe's analysis has remarkable implications. If his conclusions hold true for the stock market, then all assets could be held efficiently, since their market capitalization (price) would have adjusted to the risk they entail (which also reflects their return). In other words, the *g* risky asset allocation in Fig. 6 would be the Market Portfolio (which comprises all available assets in the market weighted by their market capitalization). This corollary is known as the Efficient Market Hypothesis (EMH). According to the EMH an investor can only increase their returns if they increase their exposure to the market (β_i). The factor behind this efficiency is competition among market participants, which leads to a situation where prices of individual securities already reflect the effects of information based both on events that have already occurred and on events which, as of now, the market expects to take place in the future (Fama, 1995, p. 76).

$$E_{Ri} = E_{Rriskless} + \beta_i(E_{Rrisky} - E_{Rriskless})$$

As a result, investors should only expect higher returns in the form of risk premia instead of trying to anticipate the behavior of assets' prices. The only prescription possible for an investor in compliance with the EMH would be a buy-and-hold strategy of the Market Portfolio (in combination with the riskless asset or leverage according to one's risk aversion) without trying to time the market (provided that the equity premium is positive and that the market is efficient, one would not be able to anticipate market highs or lows and would miss out on returns which are positive on average) or picking winner stocks (efficient markets accurately asset prices according to all information available).

I will operate from this conception of markets in my analysis of risk-taking attitudes of low-wealth households of the USA since the advent of the Internet (*see Appendix A for a summary of the reasons that have led me to choose this model and for a more detailed explanation of the Efficient Market Hypothesis*).

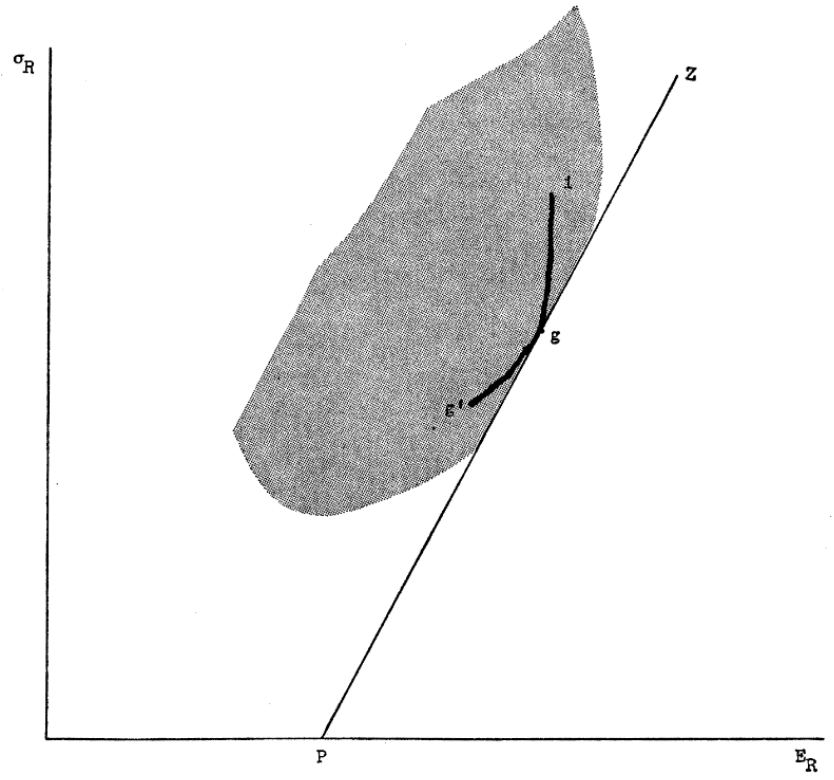

Fig. 6: Market equilibrium, the Investment Opportunity matches the Capital Market Line

2.2 Household finance

According to Campbell (2006, p. 1553), household finance studies the way in which households use financial instruments to attain their objectives. It can be positive or normative (ibid, p. 1554), the first describes what agents do and the second prescribes what they should do (a distinction that I will also make). However, household finance presents a specific problem: households make decisions which are hard to reconcile with standard models or advice. In the face of this problem Campbell (ibid, 1554) proposes two alternative paths. On the one hand, nonstandard preferences that are not captured in our traditional normative models could be taken into account (behavioral household finance). On the other hand, one can abandon the rational framework and consider the possibility that households may not express their preferences optimally (standard household finance). I will embrace this distinction and try to establish a normative definition of a rational household while considering at the same the empirically proven irrational factors that deter them from investing in equity.

Haliassos and Bertaut (1995) studied factors barring individuals from participating in stock markets. I deem it the best starting point on household finance since we share the same database (the Survey of Consumer Finances). For this reason, their work will be of great aid when considering which factors to include in my model. It will serve both as a rational normative model for individual decisionmaking and as a valuable heuristic tool as I will take into consideration the empirical control variables that Haliassos and Bertaut use. This way, I will try to get the most accurate depiction of household risk taking behavior possible.

2.2.1 Normative household finance: Utility maximization model

Haliassos and Bertaut present an *equity market non-participation puzzle*. They argue individuals are expected to invest in stocks no matter their risk aversion provided that they have excess liquidity. Assuming investors are utility maximizers, they propose the following First Order Conditions (FOC) for a expected-utility model:

$$Max\ U(c_0) + \beta E_0 U(c_1),$$

$$s.t.\ Y_0 = c_0 + s_1 + s_2,$$

$$c_1 = s_1 R_1 + s_2 R_2 + Y_1$$

Where c is real consumption, Y_t real income in period t, s_1 and s_2 real holdings of stocks and of the riskless asset which return R_1 and R_2 respectively. The second FOC is:

$$U'(Y_0 - s_1 - s_2) = \beta E_0 R_i U'(s_1 R_1 + s_2 R_2 + Y_1)$$

According to this equation, individuals invest up to the point where the amount of utility they give up in $t = 0$ is the same they get in $t = 1$ once returns on (s_1 and s_2) are added and the discount rate of future cash flows (R_i), the equity premium (E_0) and sensitivity of their asset allocation to risk factors (β) are taken into account. Combining these FOC, the following must be true:

$$E_0(R_1 - R_2)E_0 U'(s_2 R_2 + Y_1) = 0$$

This equation implies that investors maximize along two axes: $E_0(R_1 - R_2)$ and $E_0 U'(s_2 R_2 + Y_1)$. Investors seek the points on these axes in which either their marginal utility from investing $i.e.\ E_0 U'(s_2 R_2 + Y_1) = 0$ or their expected returns cannot be higher if more risk is assumed $i.e.\ E_0(R_1 - R_2) = 0$. According to this model, $s_1 = 0$ cannot be a corner solution if:

1. Second period bliss is ruled (i.e. a situation where higher returns does not increase utility is ruled out), in which $EU'(s_2 R_2 + Y_1) = 0$.
2. Risk premium is positive ($E_0 R_1 > R_2$).
3. There is no correlation between stocks returns and income.

Haliassos and Bertaut argue the *equity market non-participation puzzle* is not eliminated if the direct effect of risk aversion is considered. It is not risk that matters, but how stock holdings covariate with marginal utility. For the case of $s_1 = 0$, stocks have zero covariance with marginal utility. Therefore, FOC do not hold under this scenario. In other words, an individual that presented an acute risk aversion would be expected to invest an arbitrarily small quantity in stocks up until the point where a higher stockholding would yield a negative marginal utility (but still $s_1 \neq 0$).

2.2.2 Descriptive household finance: control variables

Once theoretical models are set in place, rational nuances and deviations from the utility maximization model under efficient markets can be considered. This descriptive approach is important to improve the explanatory power of the statistical models below, as there are instances in which households' behavior violate rational models. For instance, one of the most blatant ones is that they sometimes leave money on the table. In other words, we are not talking about counterintuitive trade-offs in which it is unclear whether it is households that behave sub-optimally or it is academics that fail to model their complex utility functions. On the contrary, they seem to miss the opportunity to

cash free money. For instance, they do not refinance their mortgages when interest rates fall (Agarwal, Rosen & Yao, 2016).

For this reason, I have decided to include control variables in the statistical models for (seemingly) irrational behavior patterns that households may present and that are also present in the Survey of Consumer Finances. If the objective is to evaluate the effect of the Internet on household investing, other factors than those which comply with rational expectations ought to be considered. If not, we could ascribe more explanatory power to the Internet than what is warranted at a fundamental level. It could be the case that part of our estimated Internet effects corresponded to an unconsidered factor correlated to our Internet variables that has proven to be empirically linked to stock market participation. Bearing that in mind, it should be no surprise that literature on factors affecting households' appetite for stocks is vast and diverse in nature. Campbell (2006) and Haliassos and Bertaut (1995) document the effect that several economic variables (income, occupational risk and net worth) have on stock holding controlled for demographic characteristics (age, education, ethnicity, gender, marital status and existence of offspring). Cocco (2005) and Flavin and Yamashita (2002) shed light on the negative effects of home ownership and mortgage debt on willingness to invest in equity, while Davis, Kubler and Willen (2006) also consider negative effects that other debt costs have on equity market participation. On the behavioral finance front, Van Rooij, Lusardi and Alessie (2011) and Guiso and Jappelli (2005), have shown the existence of a link between financial literacy and financial awareness respectively on stock market participation. Lastly, Campbell (2006) and Haliassos and Bertaut (1995) also consider financial risk aversion in their work.

2.3 The poor, the rich, the Internet… and the advisor?

2.3.1 Transaction, information and psychological costs

Once the outline on how households should and do behave has been set in place, the independent variables of interest (i.e. online access to financial institutions and information) can be introduced into the mix in order to address the *equity market non-participation puzzle*. Traditionally the stock market has been gatekept to low-income individuals. High transaction and information costs used to bar underprivileged households from equity investing. Vissing-Jorgensen (2002, p. 33) found a per period cost of 50 dollars in year 2000 prices could explain the nonparticipation of half of the nonparticipants [in the stock market]. However, the advent of Financial Technology has revolutionized the field. Decreases in transaction costs stem from lower costs for Internet trades, an increased competition resulting from the presence of Internet-based brokerage firms and the decline of other rates and fees associated with stock purchases such as margin rates and service fees (Bogan,

2008, p. 194). On top of that, the Internet has also made widespread access to financial information possible, which has also reduced information costs (although individuals still need to be able to process it, hence the relevance of financial literacy). However, these two effects (lower transaction and information costs) have not translated into a steady increase of the share of households that invest in stocks. According to the Survey of Consumer Finances data, the share of households with investments in equity is still recovering from the decline it suffered in the aftermath of the Great Financial Crisis of 2008 (*see Fig. 7*) . Consequently, we should look at how the proportion of high to low-income households has changed over recent years (as both have been affected by the stock market crash) to see if any new patterns have emerged.

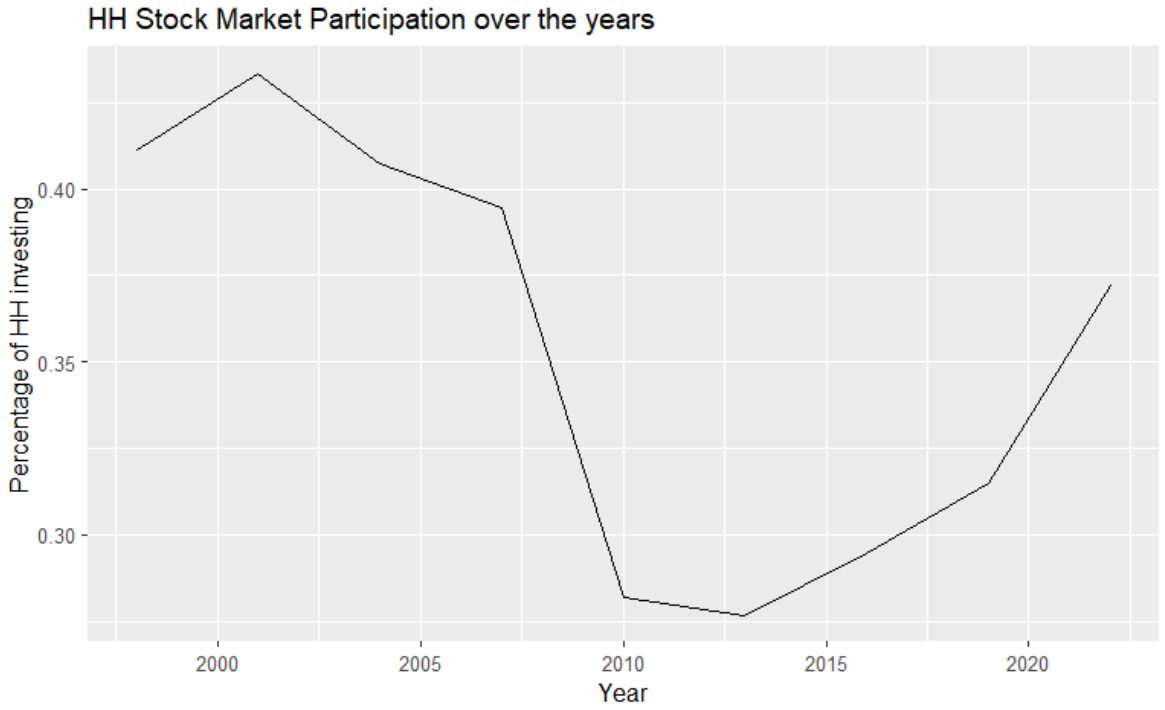

Fig. 7: Percentage of household participation in the stock market (elaborated with data from the Survey of Consumer Finances).

Going straight into the matter that concerns us today, Bogan (2008) examined the Health and Retirement Study (HRS) and concluded that results support the idea that US households that were using computers/Internet increased participation substantially more than households that did not use computers/Internet (p. 208). The evidence in other areas of the world also points in this direction. In China, according to Hong, Lu and Pan (2020, p.29), "(t)he entry of tech firms into the financial industry has substantially broken down the physical barrier and unshackled the mental constraints for individual investors participating in the financial markets". Similarly, Fernández-López, Rey-Ares and Vivel-Búa (2018, p. 882) analyzed a dataset containing information on 14 European countries and found that Internet usage is a driver of stock market participation decision.

Bogan's (2008) is, to the best of my knowledge, the most similar study to mine as far as the topic (Internet effect on stock market participation) and the sample analyzed (US citizens) are concerned. Nevertheless, I believe my work may add value by complementing Bogan's work since it compares the Internet effect on household stock market participation with respect to income. On top of that, the SCF is more diverse than the HRS, the latter only considers US citizens over the age of 50, while the former's respondents vary widely in age (although this comes at a cost: Bogan conducts a panel analysis of her data, while the SCF only allows for a repeated cross-sectional analysis, which is subject to cohort effects, *see the Methodology section below for more information*). In addition, as years pass, we have more data available. Nowadays we have post 2008 GFC data, and we can see if this period of uncertainty had any effect on the balance of stock market participation of households as a function of wealth.

Nevertheless, transaction costs do not account for all the difference in household stock market participation across wealth in the US, there is more to this story than meets the eye. Information and transaction barriers were not the only ones that low net worth households faced. One aspect which is usually overlooked is the capabilities that the wealthy have to outsource their decision to participate in the stock market, which is something critical. As Foerster, Linnainmaa, Melzer and Previtero (2017) showed, judging by the data on investing returns, financial advisors do not provide their clients with higher returns than a lifecycle fund. In other words, there is no strong financial incentive to choose a financial advisor over one of the simplest financial products available in the market. This fact leads them to hypothesize financial advisors add value by mitigating psychological costs rather than by providing financial benefit (ibid, p.28). This could mean that many clients of financial advisors would not invest on their own if they had to carry the psychological burden of being exposed to risky assets on their own.

This factor seems to have remained little changed with the advent of the Internet. A survey conducted by Vanguard (Costa and Henshaw, 2022) asked investors with human advisors about the type of advising relationship they would search for in the future. 76% answered they preferred a human advisor, 17% opted for a combination of human and digital services. Only 4% would choose a digital advisor or service and 3% would manage their investments on their own. One could think this is a consequence of self selection bias, that clients are loyal to their selected option (i.e. human advice) because they have chosen it themselves. However, they also found that this loyalty did not apply to investors with digital advisors. When asked about how willing they were to work with a human financial advisor in the future, 88% were willing, 6% were indifferent and another 6% was unwilling to change. From both surveys, it seems reasonable to conclude that almost everybody would benefit from having a financial advisor, but only a few can afford to pay them. In conclusion, it seems that the least wealthy still have to overcome a hidden third barrier, a psychological one, to participate in the market.

2.3.2 How? The question about incentive structures and efficiency

Trying to assess the degree of efficiency with which individuals invest is a matter of utmost importance. Limiting an analysis to only observing *if* low and middle-income households hold risky assets (equity market participation) without trying to look at *how* they do it would be narrow-minded. Moreover, it would miss aspects of investing that are important to the average investor (if we consider them to be risk-return optimizers), for they also care about preserving their assets from market volatility. In view of that, I will assess the extent to which individuals adhere to what would be considered an efficient risk-taking strategy. Efficiency in this context, as it is derived from the CAPM model, will be composed of two variables: diversification and absence of market-timing behavior. On the diversification front, according to the CAPM points on the Capital Market Line are sets of efficient portfolios. They include all the assets in a given market. Accordingly, individuals seeking to maximize returns and minimize risk should hold a diversified portfolio. On the market-timing side, if according to the Efficient Market Hypothesis asset prices already reflect all available information to market participants, trying to predict future prices should be futile and result in missing (on average) positive returns.

However, the Survey of Consumer Finances, its data only reaches so far for assessing diversification. As Campbell (2006, p. 1556) explains, "it is only disaggregated enough to address questions of asset allocation, and cannot shed light on diversification, because it does not report holdings of individual assets". For instance, it provides information on the number of companies in which participants held stock (*nstocks*) but not on each individual holdings. As a result, if the objective of diversification is holding uncorrelated assets (averaging out idiosyncratic risk in the process), the degree of diversification can only be approximated. It leaves out important information about diversification across countries, sectors or market capitalization, all of which are important for diversifying risk (Zaimovic, Omanovic and Arnaut-Berilo, 2021). The same occurs with overtrading. We only have access to the amount of trades the surveyed individual carried out (*ntrad*). We do not know, on the contrary, if they bought or sold, which securities they traded or according to which criteria they decided to do so. Accordingly, we have no way to determine if they are timing the market, liquidating assets to cover their liquidity needs or buying equity to increase their holdings.

The Anti-Robinhood? Financial advice, wealth and efficiency

In the light of the above, we observe that the paths through which individuals access the stock market vary depending on their level of wealth. Individuals with low resources are, in principle, more likely to enter the market on their own since they cannot afford financial advice. At first glance, if there existed differences in efficiency across different wealth stages, they ought to disappear if controls for financial information (I control for them in the models, *see the models discussion section below*), as well as for demographic and behavioral variables are put in place. Nevertheless, there are

growing concerns about the structure of incentives in which disadvantaged households enter when they take up investing. One of such preoccupations is the proliferation of gamified broker services, which may lead individuals to engage in high risk trades and excessive trading activity, which harms their expected returns. An example of this trend is Robinhood, an online financial services company that offers commission-free trades of a wide range of assets that has gained notorious popularity in recent years. Since, according to their website (Robinhood, 2024), one of their sources of revenue comes from third parties that pay the app for order of flow (i.e. for routing the clients' investment orders using their infrastructure), they try to foster engagement through a gamified user experience in order to increase trades in the app. Since it is commission free, its main user base is formed by vulnerable individuals. According to Ritholtz Wealth Management CEO Josh Brown (CNBC, 2021), the median Robinhood customer has a $240 balance in their account. Robinhood's approach to investing led the Commonwealth of Massachusetts to sue the app in 2020 for violations of the Act and Regulations by aggressively marketing itself to Massachusetts investors without regard for the best interests of its customers (Commonwealth of Massachusetts, 2020, p. 1). They adduced the platform was using strategies such as gamification to encourage and entice continuous and repetitive use of its trading application, failed to follow its own written supervisory procedures regarding the approval of options trading and breached the fiduciary conduct standard required by the Act and Regulations (ibid. p.2).

How does this relate to the question about efficiency? According to the Secretary of Massachusetts, Robinhood blurred the line between broker-dealers providing investment advice and investment advisers (Robinhood Financial LLC v. Secretary of Commonwealth, 2023, p. 3) as the company dispensed ill-suited investment advice to customers encouraging them to engage in risky trading practices using its online trading platform (ibid, p.1). Consequently, SCF respondents self-reporting having been advised financially may have been subject to opposite incentives structures. On one side of the spectrum we would have traditional fee-based financial advisors with no conflict of interests. On the opposite side we would have a financial firm that provides advice and at the same time rewards individuals who interact daily with the application (which is the opposite that the buy-and-hold approach prescribed earlier would suggest) and shows colorful confetti raining down screens after executing trades (Commonwealth of Massachusetts, 2020, p.4). If the individuals receiving these different forms of advice are classified equally in the SCF, we would expect to observe significant effects of economic variables (income and net worth) even after controlling for other factors. This could be the case judging by the SCF data on access to professional financial advice (*see Fig. 8*), which do not present very different figures for the top and the bottom 10% of households. I am well aware that the case of Robinhood is anecdotal as far as our historical series, which goes all the way back to 1998, is concerned. On the contrary, I wanted to use this evident and well-known scandal to put forward a problem that is much older: underprivileged households have traditionally

received advice coming from individuals subject to conflicts of interests. Their limited funds do not allow them to purchase advice from conflict-of-interest-free professionals.

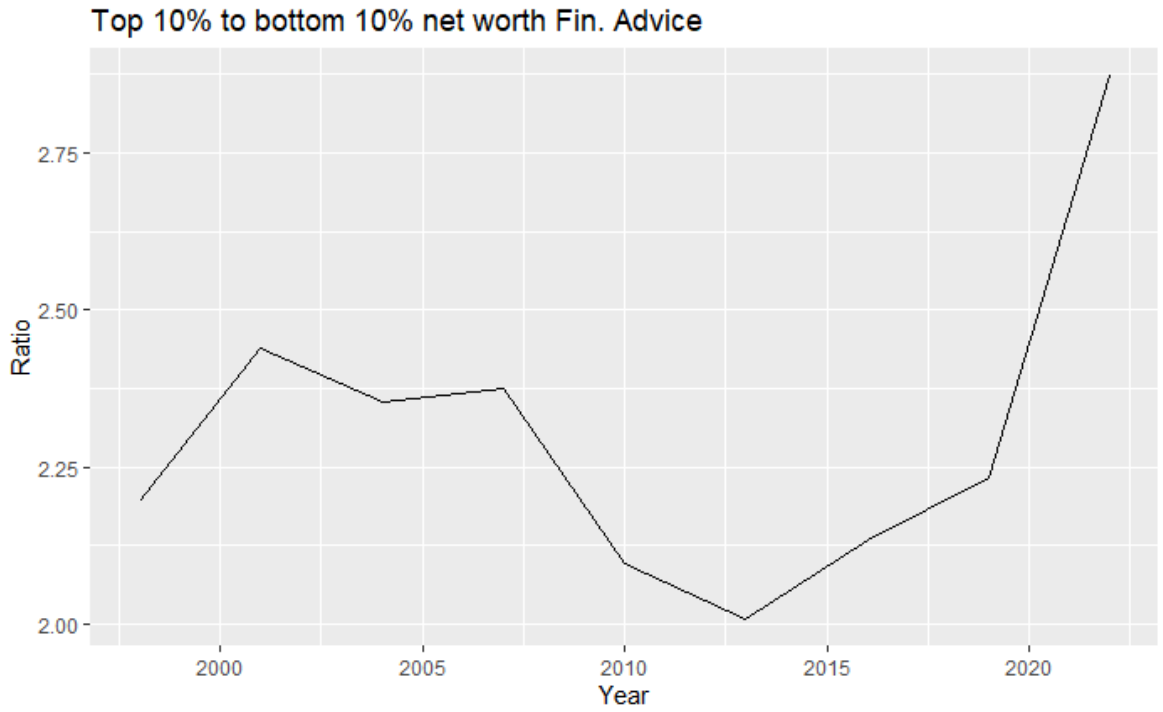

Fig. 8: How many times a household in the top 10% of net worth distribution is more likely to receive professional financial advice (i.e. having a value of one for a least *ifinplan* or *ifinpro* variables) when compared with one in the bottom 10% (elaborated with data from the Survey of Consumer Finances).

3. Methodology

The methodology followed in the present study is quantitative. More specifically, I will conduct a repeated cross-sectional analysis on each of the editions of the Survey of Consumer Finances since 1998 (the year in which Internet variables *iinternet* and *internet* were added to the survey). It is a survey conducted by the Board of Governors of the Federal Reserve System (BGFRS) on families of the United States in order to understand their financial condition and study the effects of changes in the economy. It collects data about a wide variety of topics including saving, investment, debt payments, pension coverage, business ownership, use of financial institutions, credit discrimination and financial markets (Board of Governors of the Federal Reserve System, 2022). I must acknowledge one limitation that this approach presents. Since we are talking about a repeated cross-sectional analysis, we are not observing the same individual over the years as we would if it was a panel analysis. Consequently, results may be subject to cohort effects, unique influences that a particular group of individuals, who experience a common event within the same time period, has on the outcomes being studied. Even though this consideration does not seem to affect the estimated results, it is one worth reminding.

According to the BGFRS (2019) the SCF is based on a dual-frame sample design. One set of the survey cases is selected from a standard multistage area-probability design to provide good coverage of characteristics that are broadly distributed in the population. The other set of the survey cases is selected from tax data by the Statistics of Income Division of the Internal Revenue Service (SOI) and disproportionately selects relatively wealthy families for technical purposes. Nevertheless, responses are weighted afterwards to compensate for unequal probabilities of selection in the original design and for nonresponses (failure to obtain an interview). Additionally, a multiple imputation procedure yields five values for each missing household characteristic in order to approximate the distribution of the missing data (estimating missing values according to the distribution of the missing data conditioned on the variables observed) (idem). In a nutshell, the SCF is a cross-sectional dataset that allows for consistent estimations of effects of changes in the economy due to its complex nature. I have been able to account for these data transformation techniques of the SCF thanks to the contributions of Anthony J. D'Amico (2014), who wrote the code to download the SCF microdata and take the SCF design into account when running regressions. Chang W (2015), Lumley (2019 and 2024), R Core Team (2023 and 2024), Signorell (2024), Tsegelskyi (2022) and Wickham's (2016, 2023 and 2024) have also been indispensable for the estimation of the models.

I have selected the SCF as my case of study as it gathers information from respondents across very diverse fields. This unmatched multidisciplinarity allows me to estimate the effect of my independent economic (income, wealth) and technological variables (online banking usage and access

to financial information through the Internet) while being able to control for other economic, demographic and behavioral aspects. On top of that, its complex design allows for unbiased estimates, which is also extremely desirable.

4. Hypotheses

The theoretical framework has provided us with:

1. A model describing the way in which the stock market works (CAPM).
2. A model describing the way in which investors should behave (utility maximization).
3. An understanding of the relevant demographic, economic and behavioral controls.
4. An idea of which barriers have been taken down with the advent of the Internet (transaction and information costs) and which ones have remained in place (psychological costs).
5. An idea of which incentive structures individuals may face depending on their wealth

We are now in a position to make some hypotheses about the effect that the Internet may have had on the stock market participation of disadvantaged households:

Hypothesis 1: "The drop of information and transaction barriers has enabled underprivileged households participation in the stock market".

In a nutshell, I want to examine if the *equity market participation puzzle* that Haliassos and Bertaut (1995) derived from their utility maximizer investors model can be partially explained by transaction and information costs. Although we cannot affirm that participation in the stock market necessarily leads to utility maximization behavior (*see the equation below*), it is a necessary requirement for compliance with the FOC (as the FOC did not hold the case of $s_1 = 0$; individuals were expected to invest at least an arbitrarily small quantity in stocks, *see section 1.2.1*):

$$E_0(R_1 - R_2)E_0U'(s_2R_2 + Y_1) = 0$$

Hypothesis 2: "Economically underprivileged households engage in less efficient risk-return investment practices compared to their wealthier ones even after controlling for financial professional financial advice".

As I have exposed above, it may be the case that differences in risk-return efficiency persist even after controlling for demographic and behavioral factors. The advice that low-income individuals receive may come from agents with conflicts of interest. Due to the fact that the SCF provides no way to discriminate between advisors' incentive structures, its financial advice variable may capture very dissimilar advice realities. Were that the case, economic variables would preserve their statistical significance even after controlling for demographic, behavioral and other sources of investment advice.

Hypothesis 3: “Low-wealth-risk-averse households are underrepresented in the stock market”.

The Internet has brought about lower transaction and information costs. However, psychological costs seem to have remained relatively unchanged. The Internet may have enabled the participation in the stock market of those low-wealth households that only faced economic and information constraints and were already willing to invest. In other words, it might have transformed latent demand into actual demand. Nonetheless, according to this hypothesis, the advent of the Internet might have not increased the demand coming from the subset of low-wealth-highly-risk-averse households (that face important psychological barriers to enter the stock market). Since they have no way to outsource the decision to participate in the equity market as their wealthy counterparts do (through financial planners), they may have been left behind with respect to other low-income households with higher risk tolerance.

5. Model discussion

Hypothesis 1: "The drop of information and transaction barriers has enabled unprivileged households participation in the stock market". The dependent variable is stock market participation. Since it is binary, I have decided to conduct a logistic regression on the independent variables of interest "*access to low transaction costs*" and "*access to low information costs*" and on the demographic, economic and behavioral controls. I have also tested if the assumptions for the validity of the logistic model (independence of observations, linear relationship of continuous variables with the log-odds of the dependent variable, absence of influential outliers and absence of multicollinearity) are met afterwards (*see Appendix D if you wish to check the code utilized*). This model will be run twice for every SCF edition, for the whole dataset and for the subset of individuals that fall in the lowest quartile of the income in order to compare the effects that independent variables of interest have across the income spectrum.

- *Dependent variable*: **stock market participation** (*dummy_deq*). This variable takes a value of 1 if the individual has at least $1 of value of equity in either directly held stocks, stock mutual funds or combination mutual funds (i.e. *deq*>0). It is also equal to 1 if the respondent of the SCF has no money invested but has traded in the last year (i.e. if the binary variable *htrad*=1), which is somebody who has participated in the stock market but has liquidated their positions.
- *Independent variables*:
    - *Independent variables of interest:* **Access to low-cost information on investments**, measured as using the Internet for investing decisions (*iinternet*=1) and **access to low-transaction-cost investment vehicles**, proxied for doing business with financial institutions via the Internet (*internet*=1). I have decided to use the latter variable as the proxy for access to low transaction costs since financial institutions in the USA usually offer their clients one-stop apps or websites that act as bank accounts, payment and transfer systems, financial management tools, loan services and *investment accounts*. Although this does not mean that clients that once signed up for online banking did so for investment purposes originally, it may be the case that they started investing once they learned about this service or that clients seeking to invest in the stock market are overrepresented among the application users.
    - Control variables: in compliance with the literature, I have controlled for age group (*agecl*), debt and leverage (*debt2inc* and *levratio*), education (*educ*), family structure (*famstruct*, which provides information on marital status and presence of offspring), gender of the respondent (*hhsex*), *home ownership* (*housecl*), income percentile

(*income_percentile*), amount of time dedicated to shopping around for investment decisions, which is proxied for financial awareness (*ishop grdl* for great amount of time; *ishop mode* for moderate with no search, *ishopnone*, as the reference category), ethnicity (*race*), aversion to financial risk (*nofinrisk*), occupational risk (*occat1*) and net worth (*networth_percentile*). I have also controlled for alternative sources to the Internet for investing decisions, with *idont* (never invest) as the reference category (*icall* for calling around, *finplan* for lawyer, accountant, financial planner, *finpro* for banker, broker, real estate broker, builder, dealer, insurance agent, *ifriends work* for a friend or work/business contacts, *imagznews* for magazines and newspapers and books, *imailadtv* for mail, tv, radio, advertisements, telemarketer and *iself* for personal research). On top of that, I have removed from the sample the individuals that reported not being able to save (*sabres 1*=1), as these individuals violate Haliassos and Bertaut assumption (existence of excess liquidity, *see section 2.2.1*) for the First Order Condition to be valid (*check the SCF Combined Extract Data for a further definition of the control variables*).

Hypothesis 2: "Economically underprivileged households engage in less efficient risk-return investment practices compared to their wealthier ones". The dependent variable is "*efficiency of stock market participation*". Since it is binary (*see the reason behind this variable design decision in the dependent variable section below*), I have decided to conduct a logistic regression on the independent variables of interest "*income*" and "*wealth*" and on the demographic, economic and behavioral controls. I have also tested the assumptions for the validity of this logistic model (*see Appendix D if you wish to check the code utilized*).

- *Dependent variable*: **efficiency of stock market participation** (*efficiency*). This variable is also binary. It depends on two factors: *diversification* and *absence of market timing behavior*. As far as the former is concerned, according to Campbell (2006, p. 1556), it is hard to assess diversification with the SCF data as it does not report holdings of individual assets. Regarding equity, we can only access the number of different companies in which respondents directly hold stock (*nstocks*) and the total amount invested in equity funds, which partially covers the question for diversification (*see section 2.3.2*). The part of market-timing behavior is even more elusive, realistically we can only punish excessive amounts of trades but have no way to evaluate trading behavior outside of that. Even though the data is limited, after assessing diversification and punishing outliers with a high number of trades, I have noticed that many values fall on the extremes. For instance, for the first implicate of the 2022 SCF data, when efficiency is calculated in a scale ranging from 1 to 10, 71.66% of the values fall inside [0, 2)

∪ (8, 10] ranges (*see Fig. 9 below*). Consequently, I believe it is safe to assume that not much information is lost when computing the variable as binary for the logistic model (*check Appendix E for more information of the calculation method of the efficiency binary variable*).

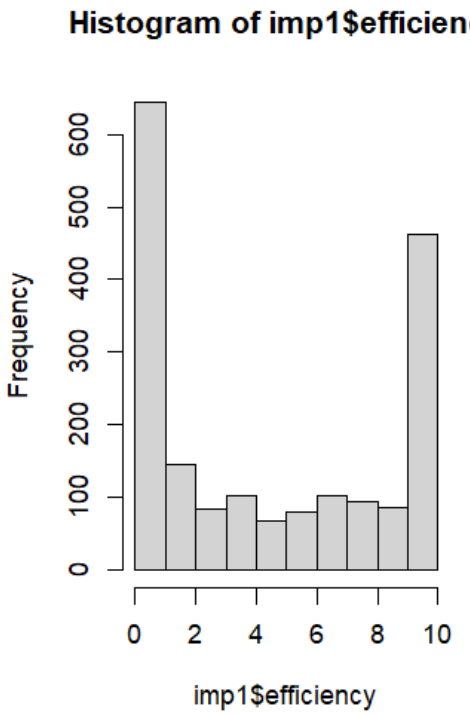

Fig. 9: Distribution of efficiency scores ranging from 1 to 10 for (elaborated with data from the first implicate of the 2022 edition of the Survey of Consumer Finances).

- *Independent variables*:
  - *Independent variables of interest:* **Income**, measured as income percentile (*income_percentile*) and **net worth**, measured as net worth percentile (*networth_percentile*). I have decided to use measures of position (i.e. percentiles) instead of direct measures of income or net worth because the former are the only ones that have a linear distribution with the log-odds of the dependent variable (a necessary condition for the validity of the logit model).
  - Control variables: I have decided to control for the same variables as in the first hypothesis. The exceptions are net worth and income, which now are the interest variables).
  -

Hypothesis 3: "Low-wealth-risk-averse households are underrepresented in the stock market". I will evaluate the joint evolution of two ratios. The Low Risk Aversion Ratio:

$$\frac{P(wealthy\ \&\ low\ risk\ aversion\ households\ to\ participate\ in\ the\ market)}{P(low{-}wealth\ \&\ low\ risk\ aversion\ households\ to\ participate\ in\ the\ market)}$$ (or LRAR),

and the High Risk Aversion Ratio:

$$\frac{P(wealthy\ \&\ high\ risk\ aversion\ households\ to\ participate\ in\ the\ market)}{P(low{-}wealth\ \&\ high\ risk\ aversion\ households\ to\ participate\ in\ the\ market)}$$ (or (HRAR)

to see if they have followed different trends over the years since the introduction of the Internet variables (*internet* and *internet*).

- *Dependent variable*: **the gap between LEAR and HEAR**.

Independent variable: **the year of the Survey of Consumer Finances edition**.

6. Results discussion

Hypothesis 1: "The reduction of information and transaction barriers has enabled unprivileged households participation in the stock market". If we take a look at the evolution of the odds ratio of the "*access to low transaction costs*" variable (*internet*), it has steadily decreased throughout the years. Fig. 10 summarizes the effect of the variable.

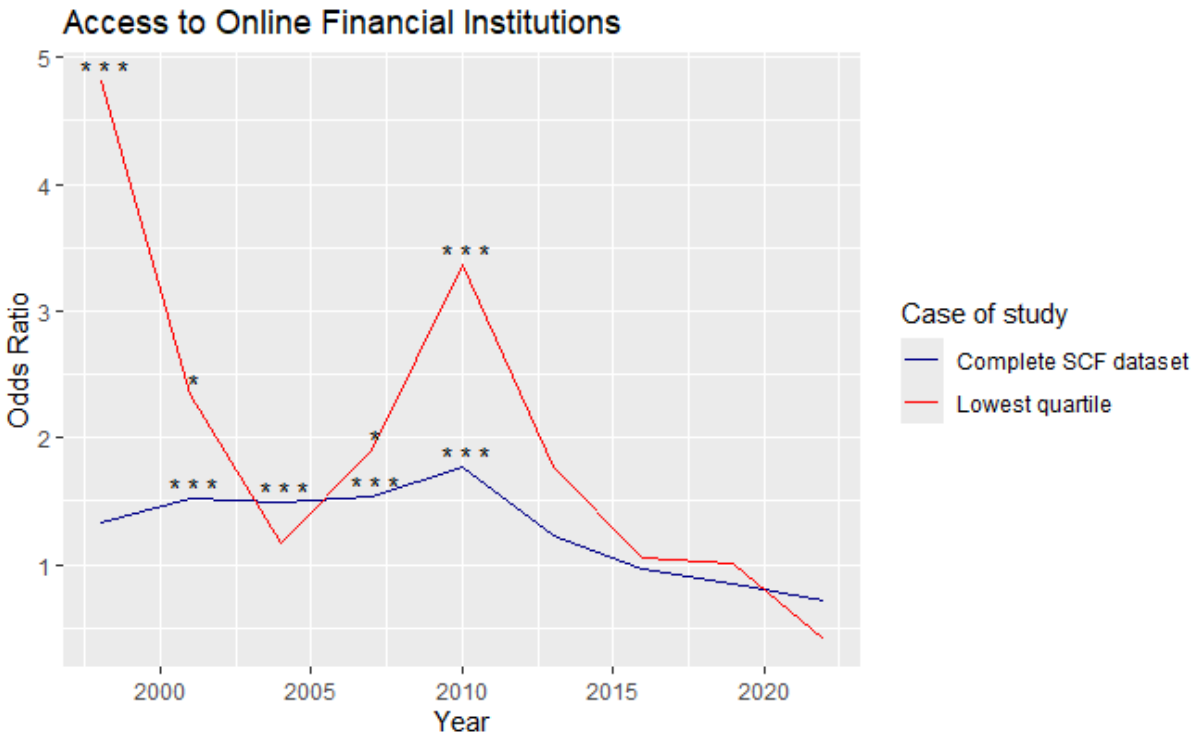

Fig. 10: Odds ratios and significance levels of "access to online financial institutions" variables across different editions (elaborated with data from the Survey of Consumer Finances).

The odds ratio of participating in the stock market for online financial services users used to be significant and well over one. For instance, the odds of participating in the stock market were about 50% higher for the Complete Dataset Series for years 2001, 2004 and 2007 and about 75% higher in the 2010 edition (and significant at the 0.1% level for all four editions). However, since 2010 two noticeable things have changed: odds ratios have steadily declined and their statistical significance has abruptly fallen. In spite of that, I believe the effect observed between 2001 and 2010 is still in place, I believe nothing has fundamentally changed. If we take a look at the percentage of households that have access to financial institutions digitally, it has reached really high quotas in recent years (which, for the case of the year 2022 almost reaches 90%, *see Fig. 11 below*). The business financial model transition from physical branches to online-based services has resulted in a widespread adoption of

their digital platforms. Consequently, I believe that what used to be a product used by the subset of households interested in accessing investment services, has become the default option. As fewer observations with no digital access to financial institutions via the Internet remain, it becomes more difficult to estimate significant results.

Additionally, the data suggests that individuals who fall in the lowest quartile have benefitted the most from this revolution. According to the data, for all the years in which coefficients are statistically significant, the odds ratios for lower-income households with access to financial services of participating in the stock market services users are significantly higher than the ones for the whole SCF dataset. This points in the direction of a leveling of the stock market playing field by virtue of lower transaction costs, which is consistent with Bogan's (2008) results.

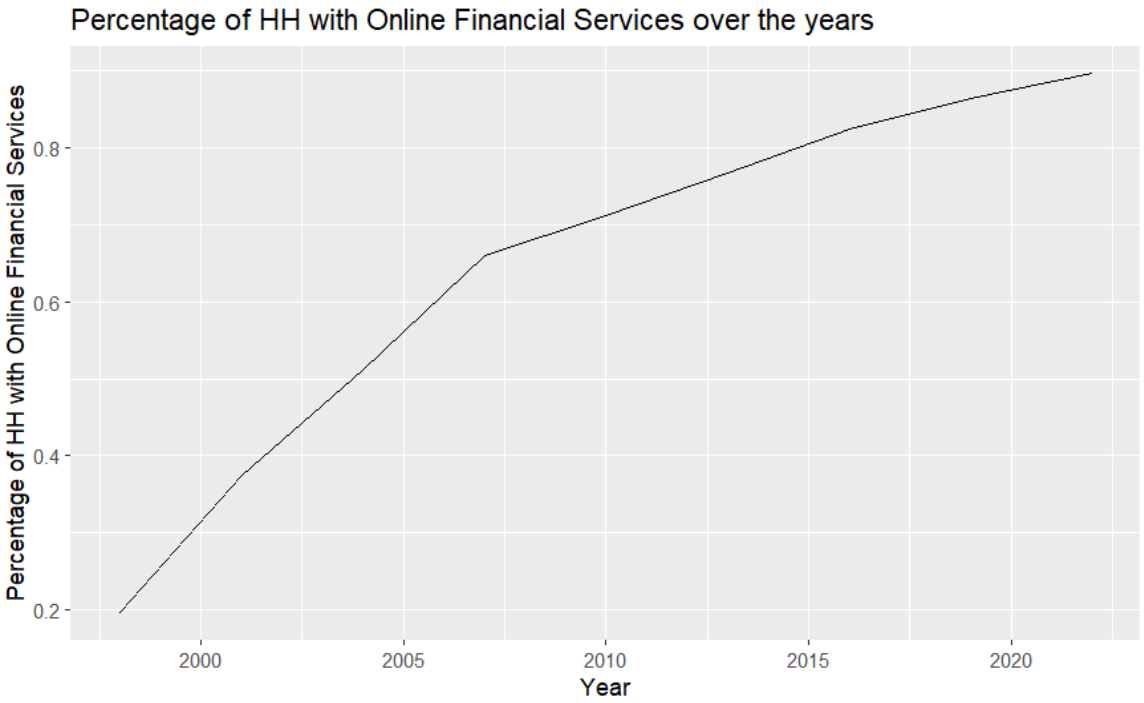

Fig. 11: Percentage of households with access to online financial services (elaborated with data from the Survey of Consumer Finances).

The second independent variable of interest, "access to online financial information" (*internet*), is also positively associated with participation in the stock market (*see Fig. 12*) for the case of the complete SCF dataset as its odds ratios are over one in all editions (except for the 2001 edition, where it is slightly below one, although not statistically significant at any of the levels considered).

Moreover, for the editions in which it is highly statistically significant, its odds ratios are well over one.

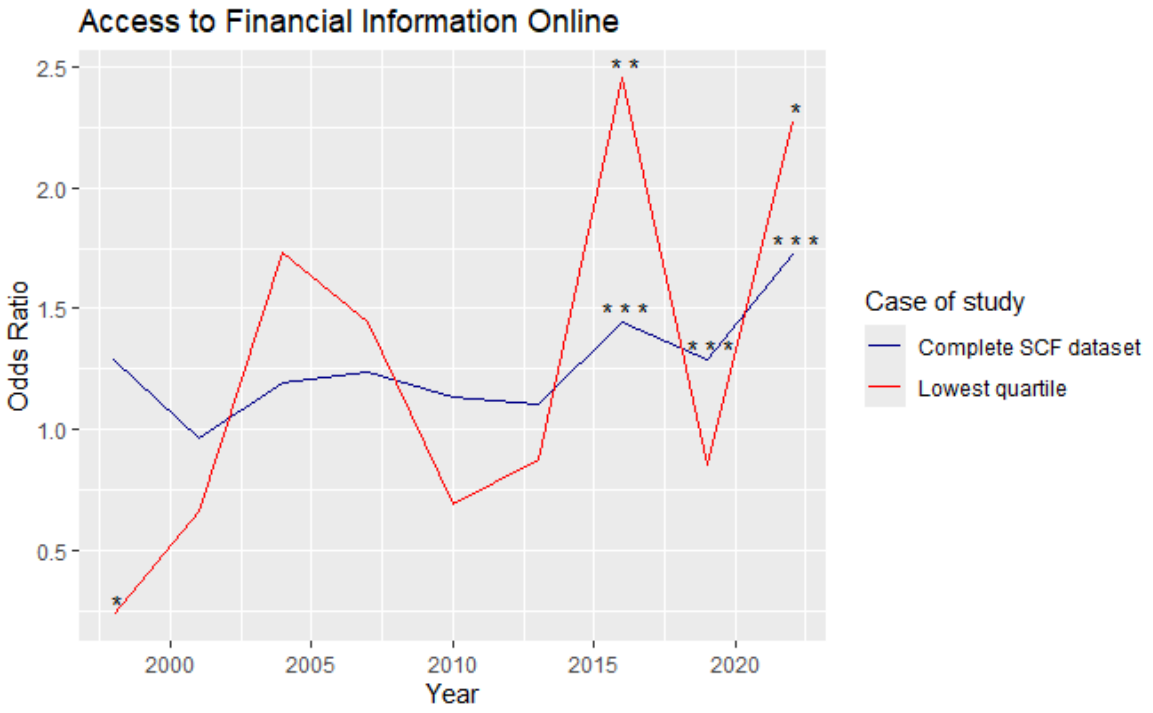

Fig. 12: Odds ratios and significance levels of "access to online financial information" variables across different editions (elaborated with data from the Survey of Consumer Finances).

In the case of access to online financial information, a direct comparison between the complete SCF sample and its lowest income quartile does not present a difference in odds ratios as conclusive as the one for "online access to financial institutions". For instance, while it is true that in recent years online financial information seems to have enabled disadvantaged households to participate in the stock market, the same is not true for editions prior to 2016. Additionally, even though higher results for households that fall in the lowest income quartile seem significant for the 2016 (at the 1% level) and 2022 (at the 5% level) editions, the gap is not as consistent across editions as it was for the previous variable of interest. On top of that, the data of the 1998 edition, which is deeply negative and statistically significant at the 5% level, further questions recent subtle differences across income. This 1998 outlier seems hard to reconcile both with subsequent years' odds ratios and with the conceptual framework. For this reason, further investigation ought to be conducted in order to assess whether it is just an irrelevant outlier or if it is a result which is worth considering as it reflects fundamental properties of this lower-income subset of observations. Nevertheless, I am personally slightly inclined to believe that the impact of the variable on stock market participation has been more positive for the lowest quartile than for the whole dataset. The opposite trend to the one affecting "access to financial institutions" variable may be taking place. Since the inception of the

"access to online information" variable (*iinternet*) in the SCF, we have seen a steady uptake in the percentage of households that use the Internet for their investment decisions (*see Fig. 13*). With fewer households using it in first editions, it is hard to estimate significant odd ratios, even more if we take into account that they use the Internet in conjunction with other sources. According to this possible explanation, it is only now that more people use the Internet to inform their investment decisions (as opposed to what has happened in recent years with the variable for "access to financial institutions") that we see higher and more significant odds ratios for underprivileged households. Consequently, I feel tempted to qualitatively assign a higher degree of importance to the results coming from recent editions, which show higher odds ratios for the years in which results coming from the bottom quartile are significant. Nevertheless, due to the evident lack of editions in which this happens, I am forced to reject this informed guess in virtue of scientific prudence. However, it is an intuition worth keeping in mind when new editions of the SCF are published in the future.

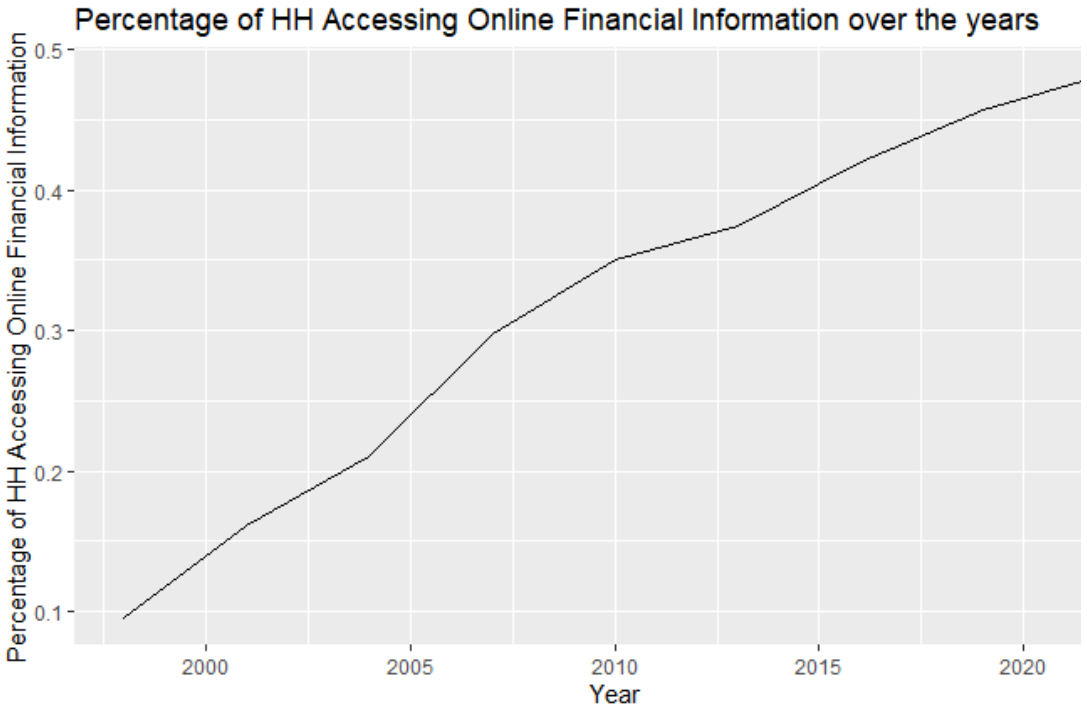

Fig. 13: Percentage of households with access to online financial information (elaborated with data from the Survey of Consumer Finances).

Regarding control variables, in line with Campbell (2006) and Haliassos and Bertaut (1995), education, income, net worth and lack of aversion to financial risk have shown to be positively and very significantly related to stock market participation. The same is true for the reference ethnicity category *white* when compared to other categories, which likely reflects unobserved social characteristics. Also in agreement with Haliassos and Bertaut (idem), marriage and children are

uncorrelated to stock market participation, except for the categories composed of a single person without children (both under and over 55 years old), which show a positive and significant relationship. The same is true for occupational risk and gender, where self-employment (which is arguably the riskiest category considered) is related to a statistically significant lower odds of participating in the stock market while gender shows no significant relationship. The only disagreement between both papers and the present model is age. While both papers find a statistically significant positive relationship of stock ownership with age, I have found a non-linear strong relationship between both variables. Coefficients are mostly negative, meaning the reference class (respondents under 35) are more likely than other age groups to participate in the stock market (except, perhaps, for respondents over 75, who have surpassed the reference group in several editions at the 5% significance level). Coefficients generally turn more negative until the fourth age category and then pick back up, so the relationship is negative for the first three groups and positive for the remaining two. In absence of a conceptual framework that could enlighten this disparity, I cannot find any reason to justify its persistence over the years.

The coefficient for housing class is, in line with Cocco (2005) and Flavin and Yamashita (2002), is positive and highly significant, signaling that house ownership reduces investors' appetite for risk. Financial awareness, proxied for the amount of time dedicated to shopping around for investment decisions, is also in line with literature (Guiso and Jappelli, 2005). Moderate and high amounts of time are increasingly and positively related to stock market participation. Another discrepancy with literature comes with debt variables and with financial literature (the latter is available only since 2016, so I have estimated three alternative models for it), for which I have not been able to estimate significant coefficients. On the contrary, according to Davis, Kubler and Willen (2006) debt costs have a negative effect on market participation, while Van Rooij, Lusardi and Alessie (2011) estimate its relationship with financial literature is positive. Among other sources of information for investing, the one coming from financial planners/accountants/lawyers is, in concordance to Foerster, Linnainmaa, Melzer and Previtero (2017) highly related to stock market participation. The same is true for magazines/newspapers/books, while calling around is significantly associated with a lower probability of participating in the stock market for several editions. The same control variable relationships hold for the lowest income quartile model, although they are usually less significant (since we run it on a fourth of the original observations). The only difference the second model presents is a very high multicollinearity (GVIF) value for age and family structure variables. However, as they are control variables, this fact seems irrelevant for estimating the effect of the independent variables of interest on stock market participation (*see Appendix D for the GVIF code*).

Hypothesis 2: "Economically underprivileged households engage in less efficient risk-return investment practices compared to their wealthier ones". The case for this hypothesis is harder to defend. The only conclusive result would be finding evidence that rejected our assumptions and found no relationship between efficiency (our dependent variable) and dependent economic variables of interest (net worth and income) after factoring in our demographic, economic and behavioral controls. On the contrary, finding a significant relationship between economic variables of interest and efficiency would not be an univocal synonym with affirming my hypothesis. There may be unaccounted-for households characteristics affecting investing efficiency. Furthermore, in line with Campbell's (2006, p. 1556) complaint on difficulty of estimating diversification model, its fitness is limited when compared to the one for the first hypothesis test (*see comparisons with the null model in the Appendix D*). This results in many variables' effects that could be fundamentally associated with investing efficiency becoming statistically insignificant. This happens due to the reduced pool of individuals investing in stocks and the lack of variables related to efficiency (*see section 2.3.2*).

As far as our independent variables of interest are concerned, net worth is significant and positively related to investing efficiency as its odds ratios are well over one for all SCF editions (*see Fig 14*). The case for income is unclear, for the most part its odds ratios are insignificant and slightly below one, the exception being 2007 and 2010, for which each increase in income percentiles is significantly (at the 5% level) associated with around 1% less probability of investing efficiently (*see Fig. 15*).

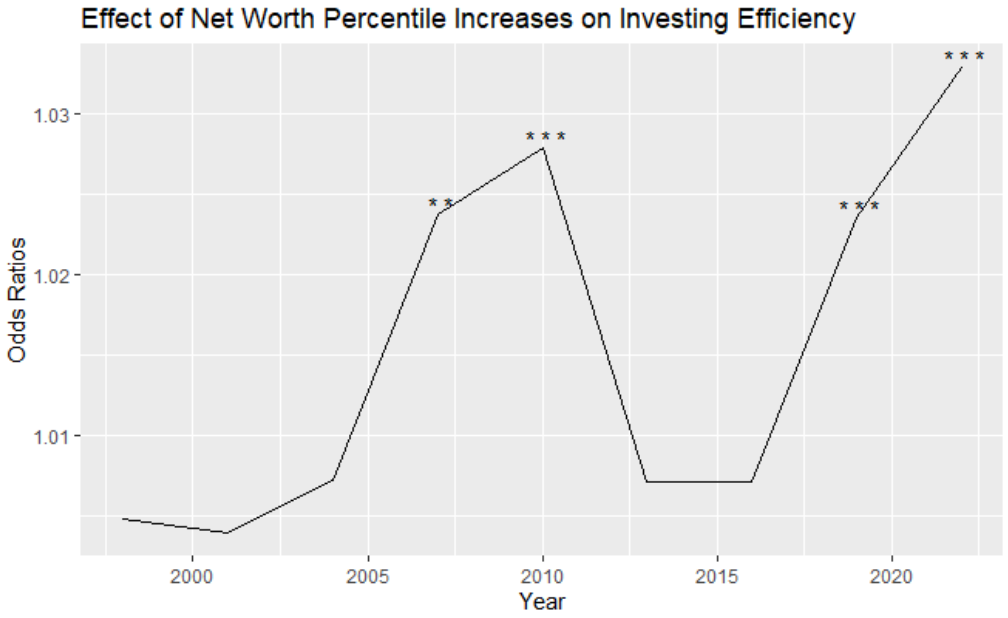

Fig. 14: Efficiency odds ratios and significance levels of single increments in net worth percentiles across different editions (elaborated with data from the Survey of Consumer Finances).

Although these findings may seem difficult to reconcile, I believe they could very well be compatible with the explanation hypothesized. If differences between investing efficiency were, in part, a consequence of unobservable differing incentive structures (since they all fall under the same investment advice categories), it would make sense that net worth be more important than income. If proper advice (which is the one that would be more likely to result in efficient outcomes) is fee-based, the only way for these fees to be justifiable would require that these payments be not so onerous that they eat away all profits it provided in excess of alternative sources of investing recommendations. In other words, fee-based advice can only be desirable if the fixed payments it requires suppose such a small part of a portfolio that they do not significantly diminish its returns. Accordingly, since the size of an equity portfolio is more related to net worth than it is to income (as it is one of the subdivisions of the wealth of a household), it makes sense for the latter to be more related to fee-based financial advice and, ultimately, to efficient investing practices. However, even though this is a plausible scenario, the limited evidence provided by the SCF and the existence of possible alternative explanations (like unobserved social effects of net worth) do not allow for conclusive findings.

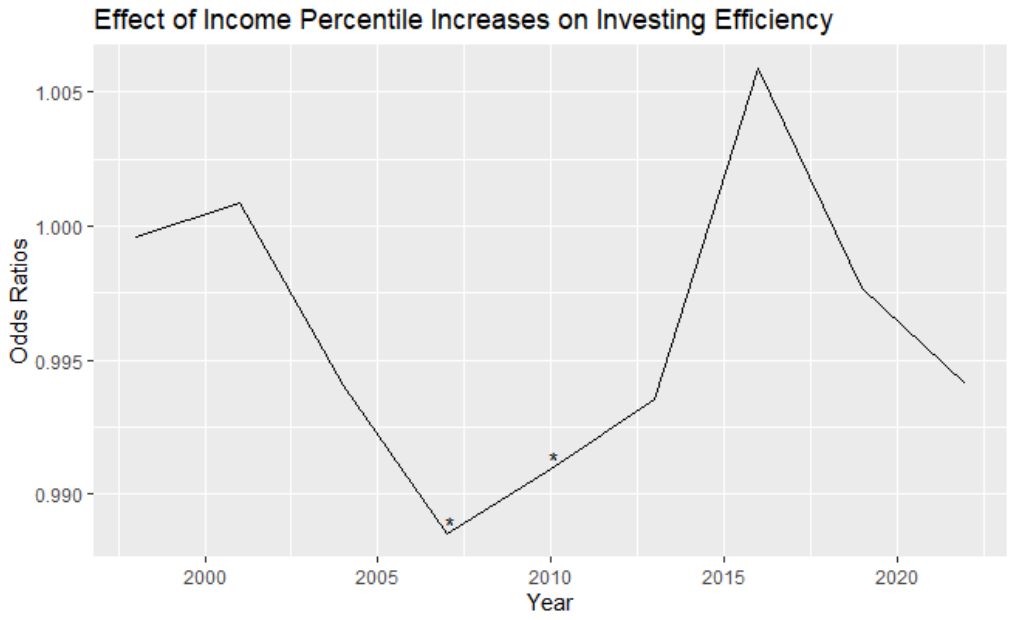

Fig. 15: Efficiency odds ratios and significance levels of single increments in income percentiles across different editions (elaborated with data from the Survey of Consumer Finances).

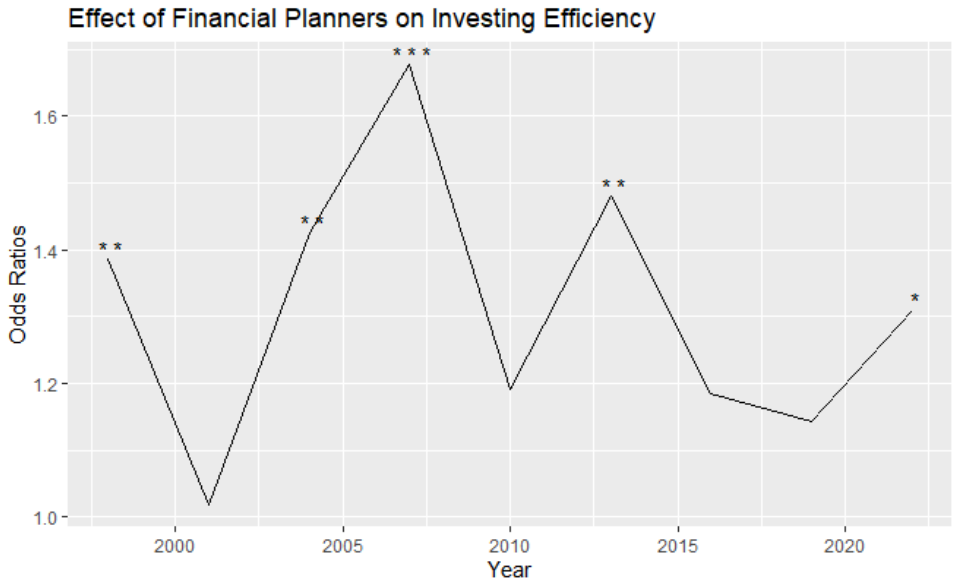

Fig. 16: Odds ratios and significance levels of "access to online financial information" variables across different editions (elaborated with data from the Survey of Consumer Finances).

Regarding control variables, for the reasons provided above, their significance is limited. However, access to investment information from a financial planner/accountant/lawyer is the only highly significant source, and it is positively related to efficiency (*see Fig. 16*). This points in the direction I previously hypothesized even though, again, it does not necessarily confirm it. Education is also directly and significantly related to the dependent variable. Aversion to financial risk is negatively correlated to efficiency and is highly significant for several editions.

Hypothesis 3: "Low-wealth-risk-averse households are underrepresented in the stock market". Since what I do in this hypothesis is indirectly evaluating the impact of the Internet, I have decided to extend the analysis to older editions of the SCF. The new first edition considered in the present hypothesis test will be the 1989 one, the oldest preserving the current format. For the third hypothesis I will evaluate how participation in the stock market has changed between highly and low-income households controlling for risk aversion.

Figure 17 depicts the evolution of stock market participation disaggregated by net worth and risk aversion. Respondents that fall in the first three net worth categories (*nwpctlecat*<=3) have been classified as low-net-worth while remaining observations have been categorized as high-net-worth (*nwpctlecat*>3). It seems that the third hypothesis holds true. In spite of fluctuations, low-income households that present low risk aversion are participating more in 2022 in the stock market than in any previous edition. Furthermore, they are the only subgroup that has surpassed pre-Great Financial

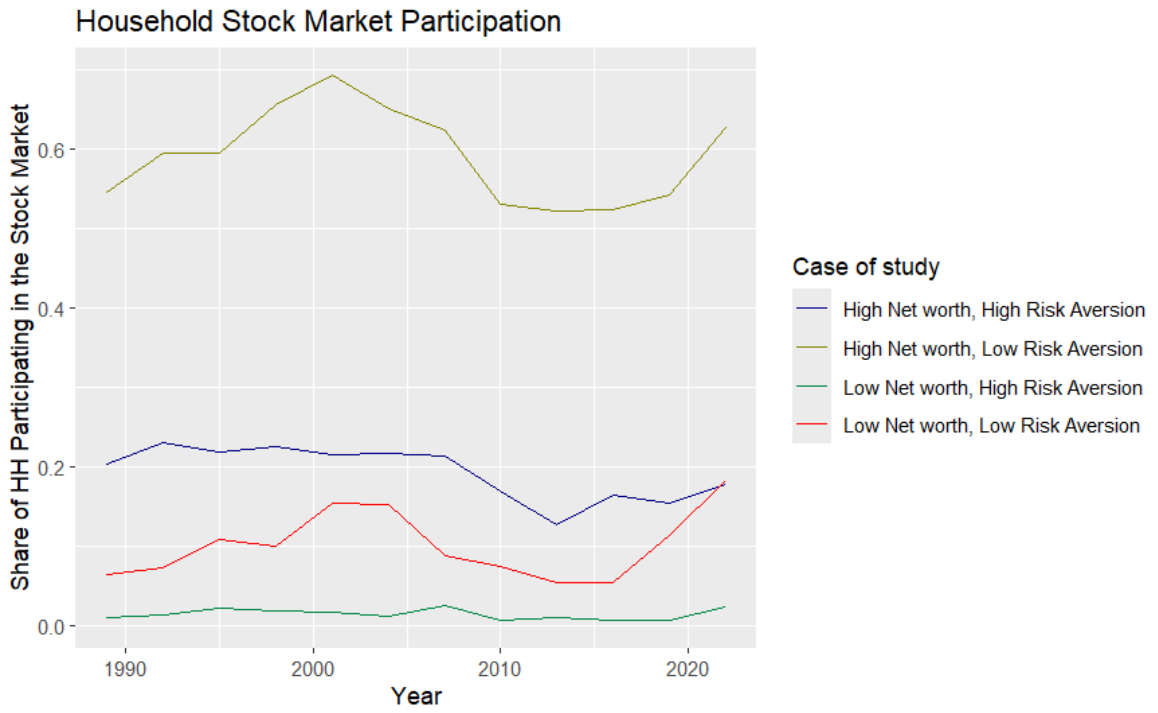

Fig. 17: Share of households from each subgroup considered which participate in the stock market across editions (elaborated with data from the Survey of Consumer Finances).

Crisis Participation highs in stock market participation share. Not only that, but the last edition is the first one in which market participation of low-net-worth-low-risk-aversion households has surpassed that of high-net-worth-high-risk-averse individuals. The contrast with their highly-risk-averse counterpart is striking. Their levels of participation have not only been much lower than those of low-income-low-risk-averse households, but have also remained well below their 2.62% all-time-high reached before the 2008 GFC. If we take a look at Figure 18, we can see that low-income are closing their gap with their high-income counterparts as both LRAR and HRAR have decreased since the beginning of the historical series (although they surged back up in the aftermath of the 2008 GFC). However, the third hypothesis is true according to the trends as HRAR presents higher values throughout the whole historical series (i.e. lower-net-worth households are underrepresented in the market when compared with their high-income counterparts). On a positive note for low-income-high-risk-averse households, HRARs have declined slightly more than LRARs: the latter have declined 59.08% while the former have fallen by 62.06%. Nevertheless, these differences in ratios declines ought to be taken with a grain of salt, as both figures remained mostly unchanged until 2016 for LRAR and 2019 for HRAR. Additionally, since HRAR figures are much lower both in the numerator and the denominator (as they represent households that are unlikely to invest due to their risk aversion), they are much more volatile than LRAR figures and, as a result, harder to interpret. In spite of that, the fact that low-net-worth-high-risk-averse households have catched-up their wealthier homologues has not precluded them from growing slower than their low-risk-averse counterparts (*see*

Fig. 19). Even though they have fallen behind at a slower pace than high-net-worth-highly-risk-averse households, they are still being left behind.

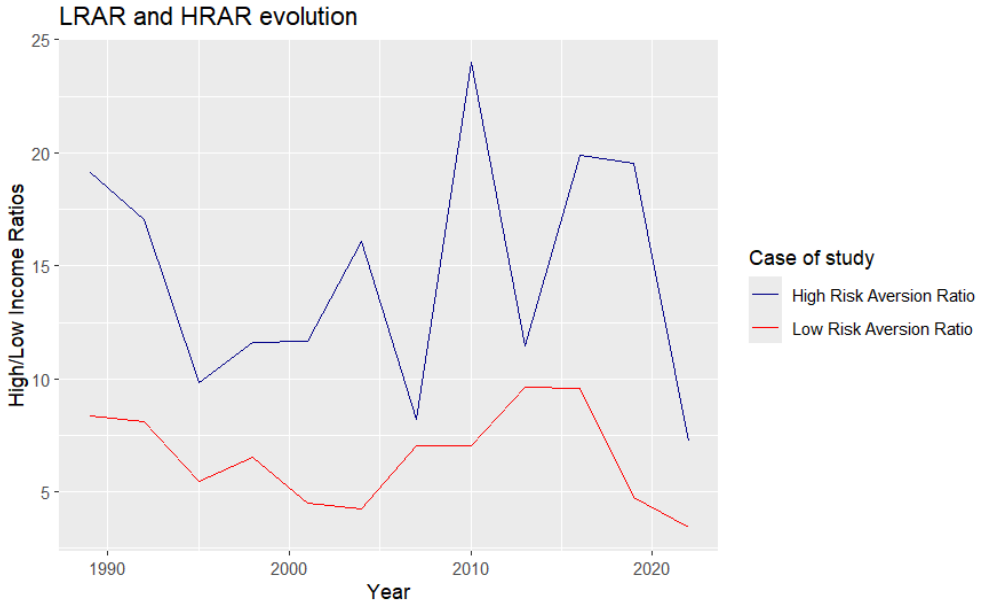

Fig. 18: LRAR and HRAR evolution across editions (elaborated with data from the Survey of Consumer Finances).

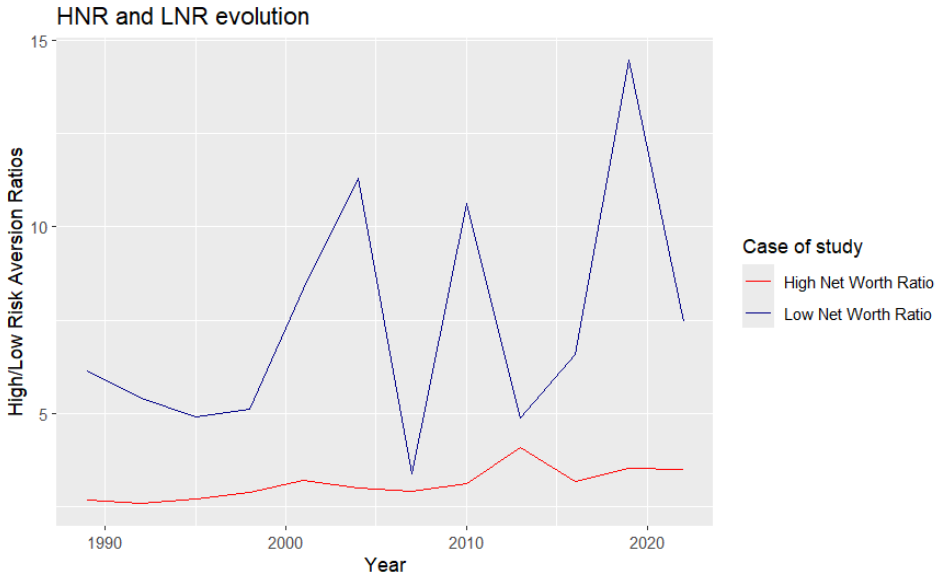

Fig. 19: High-to-low risk aversion ratios across editions (elaborated with data from the SCF).

7. Conclusion

The present bachelor thesis was born from a compromise with effective equality. Having learned the great potential for wellbeing that the stock market has historically had over the long term, I wanted to assess whether the technologies that promised to solve accessibility problems for the disfavored have fulfilled expectations.

In the theoretical framework, borrowing Haliassos and Bertaut's (1995) definition, I have defined what a normative investor ought to look like. I have also contextualized them in a market which complies with Sharpe's (1964) Capital Asset Pricing Model. The subsequent section has versed on the changes that the Internet brought about: lower transaction and information costs (Bogan, 2008). I have also discussed what the Internet has likely left untouched: psychological costs (Foerster, Linnainmaa, Melzer and Previtero, 2017).

This part of the framework has informed the first and third hypotheses. The first one tried to test if lower information and transaction costs were associated with a higher probability of participating in the stock market. I have run a logistic model on the Survey of Consumer Finances and the effects have resulted statistically significant in the last editions for information costs (access to investment information through the Internet) and in the first ones for transaction costs (online access to financial institutions). In recent surveys, widespread adoption of online financial services may be behind the drop both in significance and odds ratio of the latter variable. The third hypothesis was concerned with which households have benefitted the most from this revolution. As it turns out, the low-net-worth-low-risk-averse share of households participating in the stock market has grown relatively more than the ones of every single other group studied, even surpassing high-net-worth-highly-risk-averse households in share of stock market participation participation for the 2022 SCF edition. The opposite side of the coin is exemplified by low-income-highly-risk-averse households, which are lagging well behind every other group. This is consistent with the conceptual framework, where I have hinted that psychological costs are likely to have remained unchanged. This would deter risk-averse individuals with no income to outsource psychological costs to financial planners from entering the stock market, but not low-income-low-risk-averse households, which have to bear a lower psychological cost.

The remaining part of the conceptual framework delved into the relationship between economic variables and differing incentive structures. I hypothesized that since economically underprivileged individuals are more likely to access the market through institutions that present misaligned incentive structures, they may invest less efficiently. As the difference in incentive structures is not reflected in the SCF variables, I expected differences in efficiency to hold for the

economic variables of interest even after controlling for demographic, behavioral, occupational risk variables and other investment advice sources. I estimated a logit model testing how economic variables affected risk-return efficiency. Findings are compatible with the misaligned incentives scenario requirements. In spite of the lack of sufficiently disaggregated data on risky holdings of households portfolios, a complaint put forward by Campbell (2006), and the fact that the model had to be run on fewer observations (as I can only observe efficiency in individuals that already invest) drew the model fitness down, some variables of interest hold statistically significant relationships in the direction put forward in the second hypothesis. However, since I cannot control for alternative explanations (e.g. social effects highly correlated to net worth), I can only conclude that results are compatible with the second hypothesis but that they cannot prove it.

This study also has left some questions unanswered that could be turned into future lines of investigation. For instance, it has not been able to shed light on the non-linear relationship estimated in the first model between age and stock market participation. Additionally, it has not been able to account for the significant negative coefficients that interest variables presented in some editions of the SCF (1998 edition coefficient for the *iinternet* variable in the first hypothesis and the coefficients for the *income_percentile* variable in 2007 and 2010 editions in the second hypothesis).

All in all, I would conclude that although new inequalities have emerged, the Internet has made the stock market a more egalitarian place overall. However, we cannot turn a blind eye to the challenges that still remain. For this reason, I am convinced that the most valuable lines of investigation that this work could inspire are the ones for which there is no answer yet: how can we make sure that everyone, no matter their risk preference or their wealth, can effectively access the stock market in the best terms possible? I believe that the future has plenty of things in store, technological revolutions that have yet to be unveiled and which may help people access proper financial advice in the same way that the Internet has allowed for widespread access to the stock market. I hope for a future in which technology fights on the side of the disadvantaged and allows them to flourish economically.

8. References

Agarwal, S., Rosen, R. J., & Yao, V. (2016). Why do borrowers make mortgage refinancing mistakes?. *Management Science*, *62*(12), 3494-3509.

Board of Governors of the Federal Reserve System. (2019). Codebook for 2019 Survey of Consumer Finances. Retrieved from: https://www.federalreserve.gov/econres/files/codebk2019.txt

Board of Governors of the Federal Reserve System. (2022, February 25). Survey of Consumer Finances. Retrieved from https://www.federalreserve.gov/scf/scf.htm

Board of Governors of the Federal Reserve System. Survey of Consumer Finances (April 05, 2024) [Data files]. Retrieved from https://www.federalreserve.gov/econres/files/

Bogan, V. (2008). Stock market participation and the internet. *Journal of Financial and Quantitative Analysis*, *43*(1), 191-211.

Boudoukh, J., & Richardson, M. (1993). Stock returns and inflation: A long-horizon perspective. *The American economic review*, *83*(5), 1346-1355.

Campbell, J. Y. (2006). Household finance. *The journal of finance*, *61*(4), 1553-1604.

Chang W (2015). _downloader: Download Files over HTTP and HTTPS_. R package version 0.4, https://CRAN.R-project.org/package=downloader

CNBC. (2021, July 29). *Robinhood debuts on Nasdaq – Jim Cramer and others share their take* [Interview]. Retrieved from https://www.cnbc.com/2021/07/29/robinhood-stock-nasdaq-ipo-jim-cramer-and-others-share-their-take.html

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

Commonwealth of Massachusetts, Office of the Secretary of the Commonwealth, Securities Division. (2020). *Administrative complaint against Robinhood Financial LLC*. Retrieved from https://www.sec.state.ma.us/divisions/securities/download/MSD-Robinhood-Financial-LLC-Complaint-E-2020-0047.pdf

Costa, P., & Henshaw, J. E. (2022). Quantifying the investor's view on the value of human and robo advice. Vanguard Mexico. Retrieved from https://www.vanguardmexico.com/content/dam/intl/americas/documents/mexico/en/2022/02/mx-sa-2056467-quantifying-the-investors-view-on-the-value-of-human-and-robo-advice.pdf

Cournot, A. A. (1838). *Recherches sur les principes mathématiques de la théorie des richesses* (Vol. 48). L. Hachette.

Damico, A. (2014): Code for Downloading SCF microdata. Retrieved from https://github.com/DjalmaPessoa/usgsd/blob/master/Survey%20of%20Consumer%20Finances/download%20all%20microdata.R

Damico, A. (2014): Analysis Example of the SCF. Retrieved from: https://github.com/DjalmaPessoa/usgsd/blob/master/Survey%20of%20Consumer%20Finances/analysis%20examples.R

Daróczi G, Tsegelskyi R (2022). _pander: An R 'Pandoc' Writer_. R package version 0.6.5, https://CRAN.R-project.org/package=pander

Fama, E. F. (1970). Efficient capital markets. *Journal of finance*, *25*(2), 383-417.

Fama, E. F. (1995). Random walks in stock market prices. *Financial analysts journal*, *51*(1), 75-80.

Fernández-López, S., Rey-Ares, L., & Vivel-Búa, M. (2018). The role of internet in stock market participation: just a matter of habit?. *Information Technology & People*, *31*(3), 869-885.

Fidelity Investments. (n.d.). What are Mutual Funds? Retrieved from https://www.fidelity.com/learning-center/investment-products/mutual-funds/what-are-mutual-funds

Flavin, M., & Yamashita, T. (2002). Owner-occupied housing and the composition of the household portfolio. *American Economic Review*, *92*(1), 345-362.

Foerster, S., Linnainmaa, J. T., Melzer, B. T., & Previtero, A. (2017). Retail financial advice: does one size fit all?. The Journal of Finance, 72(4), 1441-1482.

Guiso, L., & Jappelli, T. (2005). Awareness and stock market participation. *Review of Finance*, *9*(4), 537-567.

Haliassos, M., & Bertaut, C. C. (1995). Why do so few hold stocks?. *The Economic Journal*, 105(432), 1110-1129.

Hong, C. Y., Lu, X., & Pan, J. (2020). FinTech Adoption and Household Risk-Taking: From Digital Payments to Platform Investments (No. w28063). National Bureau of Economic Research.

Jensen, M. C. (1968). The performance of mutual funds in the period 1945-1964. *The Journal of finance*, 23(2), 389-416.

Lumley T (2019). _mitools: Tools for Multiple Imputation of Missing Data_. R package version 2.4, https://CRAN.R-project.org/package=mitools

Lumley (2024) "survey: analysis of complex survey samples". R package version 4.4.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7(1), 77-91.

McFadden, D. (2021). Quantitative methods for analysing travel behaviour of individuals: some recent developments. In *Behavioural travel modelling* (pp. 279-318). Routledge.

R Core Team (2023). _foreign: Read Data Stored by 'Minitab', 'S', 'SAS', 'SPSS', 'Stata', 'Systat', 'Weka', 'dBase', ..._. R package version 0.8-86, https://CRAN.R-project.org/package=foreign

R Core Team. (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Robinhood. (n.d.). *How we make money*. Retrieved from https://robinhood.com/us/en/about-us/how-we-make-money/

Robinhood Financial LLC v. Secretary of Commonwealth. (2023). *Docket No. SJC-13381*. Supreme Judicial Court of Massachusetts.

Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *The Journal of Finance, 19*(3), 425-442.

Signorell A (2024). _DescTools: Tools for Descriptive Statistics_. R package version 0.99.54, https://CRAN.R-project.org/package=DescTools

Van Rooij, M. C., Lusardi, A., & Alessie, R. J. (2011). Financial literacy and retirement planning in the Netherlands. *Journal of economic psychology*, *32*(4), 593-608.

Vissing-Jorgensen, A. (2002). Towards an explanation of household portfolio choice heterogeneity: Nonfinancial income and participation cost structures.

Wickham H (2016). _ggplot2: Elegant Graphics for Data Analysis_. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org

Wickham H, François R, Henry L, Müller K, Vaughan D (2023). _dplyr: A Grammar of Data Manipulation_. R package version 1.1.4, https://CRAN.R-project.org/package=dplyr

Tidyr Wickham H, Vaughan D, Girlich M (2024). _tidyr: Tidy Messy Data_. R package version 1.3.1, https://CRAN.R-project.org/package=tidyr

Yao, R., & Zhang, H. H. (2005). Optimal consumption and portfolio choices with risky housing and borrowing constraints. *The Review of Financial Studies*, *18*(1), 197-239.

Zaimovic, A., Omanovic, A., & Arnaut-Berilo, A. (2021). How many stocks are sufficient for equity portfolio diversification? A review of the literature. *Journal of Risk and Financial Management*, *14*(11), 551.

9. Appendix A. Decision-making process on the election CAPM

The CAPM provides a succinct and elegant explanation of the way in which financial markets function. Even though further considerations and implementations could be added to the model (e.g. the intertemporal dimension of the ICAPM), I have deliberately chosen to leave those out of the conceptual framework for simplicity's sake.

Derived from the CAPM, academics have suggested that markets are efficient, constituting what has become the Efficient Market Hypothesis (EMH). However, testing the factual validity of the EMH is troublesome. An indirect approach that favors this conception of financial markets is the well-documented fact that actively managed funds have underperformed passive buy-and-hold strategies for decades. Jensen (1968) first documented that fund managers were not able to predict security prices, and no individual fund performed better than what could be attributed to luck. The same holds true today, where actively managed funds are expected to underperform index funds strategies.

Nevertheless, the direct statistical validity of the EMH could even be impossible to determine. As Fama (1970) suggests, any test of market efficiency must be tested jointly with a model for expected normal returns. Put differently, returns can only be considered abnormal in relation to an alternative expected returns model. Consequently, one cannot challenge the Efficient Market Hypothesis alone but has to put the model from where it is derived into question. All in all, if the EMH and its expected returns model were found to be jointly false, one cannot discern whether the Market is Inefficient, the expected returns model is mistaken or both are incorrect.

10. Appendix B. Investment Opportunity Curve Shape and Asset Correlation

In a market conformed of two risky assets, the expected returns of an investor will be determined by the following equation (where α is the proportion of wealth invested in asset A):

$$E_{Ri} = \alpha E_{Ra} + (1 - \alpha) E_{Rb}$$

The standard deviation their portfolio will experiment will be determined by another equation (where r_{ab} is the correlation of assets a and b):

$$\sigma_i = \sqrt{\alpha^2 \sigma_a^2 + (1 - \alpha)^2 \sigma_b^2 + 2 r_{ab} \alpha (1 - \alpha) \sigma_a \sigma_b}$$

Consequently, depending on the proportion of wealth that we devote to each asset we will be in a different point of the AZB curve of Fig. 20, or on the line AB if assets follow a perfect positive correlation.

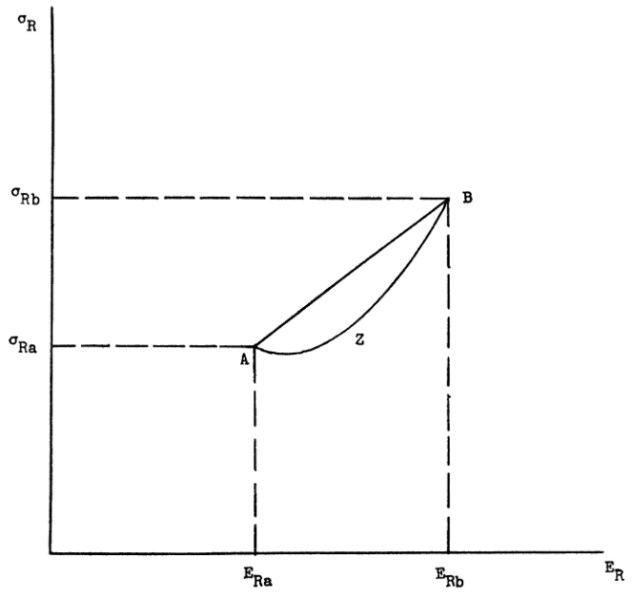

Fig. 20: Different risk and returns depending on r_{ab} (Sharpe, 1964: 430).

11. Appendix C. Description of the Capital Market Line

The returns will be determined by the same equation as in Appendix C. However, instead of being a combination of the returns of asset A and B, they are a combination of the returns of the riskless and the specific selection of risky assets chosen (which are in turn determined by the risky assets that comprise this combination).

$$ER_i = \alpha ER_{riskless} + (1 - \alpha)\, ER_{risky}$$

As for the standard deviation, the previous equation still applies. However, since the standard deviation of the riskless asset is nonexistent by definition, the expression can be simplified as follows:

$$\sigma_i = \sqrt{\alpha^2 \sigma_{riskless}^{\ 2} + (1 - \alpha)^2 \sigma_{risky}^{\ 2} + 2r_{riskless\, risky}\, \alpha(1 - \alpha)\sigma_{riskless}\sigma_{risky}}$$

Since $\sigma_{riskless} = 0$, it could be expressed as follows:

$$\sigma_i = (1 - \alpha)\ \sigma_{risky}$$

In the light of the above, we can notice that the Capital Market Line previously described will differ depending on the specific combination of assets selected (which will affect E_{Ri} and σ_i) and on the proportion of wealth devoted to risk to riskless assets (*see Fig. 3*). The former will influence the slope of the line as it affects its intersection point on the Investment Opportunity Curve (e.g. points A, B and Ø shown in Fig. 3) while the latter will impact the specific risk-returns values along the line. On top of that, as it was the case for Fig. 2, one could borrow money at the Pure Interest Rate to increase their perceived risk and returns along any given Capital Market Line (e.g. if one borrowed to invest in the combination A of risky assets, they could reach point Y in Fig. 3).

12. Appendix D. R Code Clarifications

The following link will allow you to access the online folder containing the RStudio documents necessary for computing the statistical tests commented in the thesis and the models estimated with these documents in order to save the reader from having to compute them again. However, they could compute it themselves if they wished to do so. Additionally, it also contains the code for other miscellaneous calculations that have been commented throughout the document, information on Pseudo R^2 estimators by McFadden (2021), the comprobations of the logit model assumptions and, for the second hypothesis, a comparison with the null model containing only the intercept:

Link to the R Code Online Folder: R Files Folder

Before opening the hypothesis testing documents, you should first download the SCF microdata in your device (run the Install_SCF_Code). Afterwards, you will be set to make the comprobations of the hypothesis test that you wish.

13. Appendix E. Calculation of the Efficiency Binary Variable

For the calculation of the efficiency binary variable I have computed three different calculations: a stock diversification grade, an overall diversification grade and an overtrading grade. As far as the stock diversification grade is concerned, it is clear that the more diversified that a portfolio is, the better. Theoretically, the optimum is set in the Market Portfolio (according to the CAPM). However, a Total Market Portfolio is impractical as most of the idiosyncratic risk is eliminated with far less holdings. However, where do we draw the line? According to the literature review conducted by Zaimovic, Omanovic and Arnaut-Berilo, (2021), there is no agreement in what is considered effective diversification. Consequently, I have decided to follow a brute force approach and average the number of stock holdings that the studies they consider suggest, which yields 52.34 different holdings. Accordingly, for the stock grade I have given the maximum punctuation to any *nstocks* value above this mean and assigned linearly decreasing punctuations to values below this number. For the calculation of the overall diversification grade I have computed a weighted (by the amount of money invested in each asset class) average between the stock diversification grade and the mutual fund grade coming from mutual funds. Since according to Vanguard (n.d.) the average fund has well over 100 different holdings, I have given the maximum grade to each dollar invested in mutual funds. Regarding the overtrading grade, I have started gradually punishing investors that trade more than once a week up until values over 750, which receive all the lowest scores. Finally, I have multiplied the overall diversification grade and the overtrading grade and divided the result by 10, which yields values between 0 and 10. I have assigned a score of ones to values over 5 and a 0 to the rest, which results in the efficiency binary variable.