FINANCIAL KNOWLEDGE ACQUISITION AND TRADING BEHAVIOR: EMPIRICAL EVIDENCE FROM AN ONLINE INFORMATION TOOL

Anthony Bellofatto, Marie-Hélène Broihanne, Catherine D'Hondt

REPRINT • 2024 / 13

UCLouvain

LiDAM

LFIN | Louvain Institute of Data Analysis and Modeling in economics and statistics

LFIN

Voie du Roman Pays 34, L1.03.01 B-1348 Louvain-la-Neuve

Tel (32 10) 47 43 04

Email: lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/lfin/publications.html

Financial Markets and Portfolio Management
https://doi.org/10.1007/s11408-024-00459-0

Financial knowledge acquisition and trading behavior: empirical evidence from an online information tool

Anthony Bellofatto[1] · Marie-Hélène Broihanne[2] · Catherine D'Hondt[3]

Accepted: 3 August 2024

Abstract
Using the advent of the MiFID regulation in Europe, we conduct a robust quasi-natural experiment in a large sample of retail investors to analyze how some of them changed their trading behavior after being granted access to an online information tool. Focusing on investors who asked for the information tool, we find that subjective financial literacy and education are both positively related to financial knowledge acquisition, as suggested in Lusardi et al. (2017)′s model. Using propensity score matched difference-in-differences regressions, we show that financial knowledge acquisition has mixed effects on trading behavior: better portfolio diversification but higher trading intensity and lower net returns. Our empirical evidence indicates financial knowledge acquisition being complementary to financial literacy (instead of a substitute). It also reveals that the investment profile does matter to explain heterogeneity in behavior across investors.

Keywords Retail investors · Financial knowledge acquisition · MiFID

JEL Classification D14 · D83 · G11 · G28 · G40

1 Introduction

Over the last two decades, technological innovations in the financial industry have led to substantial changes in financial services. One particular aspect is the widespread development of sophisticated online trading platforms allowing retail investors to have an easy access to ongoing financial information and professional

✉ Catherine D'Hondt
catherine.dhondt@uclouvain.be

[1] Arvestar Asset Management, Bruxelles, Belgium

[2] LaRGE Research Center, EM Strasbourg Business School, University of Strasbourg, Strasbourg, France

[3] Louvain Finance (LIDAM), Louvain School of Management - UCLouvain, Ottignies-Louvain-la-Neuve, Belgium

Published online: 26 November 2024

recommendations. These online trading environments affect how retail investors receive and search for information, as well as how they act on that information (Barber and Odean 2001a). In this paper, we address the specific question of whether retail investors seize the opportunity to acquire (more) financial knowledge using the information tool at their disposal on their broker's trading platform. For that purpose, we build on the theoretical models of endogenous financial knowledge acquisition (Jappelli and Padula 2013; Lusardi et al. 2017; Barthel and Lei 2021) to formulate hypotheses, and we test them through a robust quasi-natural experiment relying on the implementation of the European Markets in Financial Instruments Directive (MiFID).[1]

In terms of *Know Your Customer* requirements, MiFID distinguishes among the execution of orders, provision of financial advice, and supply of portfolio management services (see Fig. 1 in the appendix). Investors who ask for execution only have to complete the *appropriateness test* (*A*-test hereafter) when they want to trade "complex" instruments. The aim of the *A*-test is to ensure that the investor has the necessary experience and knowledge to understand the risks involved in complex financial instruments before investing. Investors who require financial advice and/or portfolio management services (in addition to order execution) have to complete the *suitability test* (*S*-test hereafter). In contrast to the *A*-test, the *S*-test is not limited to the investor's knowledge and experience but also includes assessment of investment objectives and financial situation.

In this paper, we use a very rich database that covers the January 2003–March 2012 period and allows us to analyze a large set of 21,233 retail investors who hold stock portfolios during at least 12 months both before and after the first of January 2008, that is, before and after the enforcement of MiFID. The data at hand allow us to examine whether retail investors changed their trading behavior after having completed the *S*-test to be granted access to an information tool on stocks available on the broker's trading platform.[2]

The online information tool under scrutiny is not a robo-advisor,[3] but it is close to it. On the one hand, as with robo-advice, to have access to this information tool,

[1] MiFID is central to the regulation of financial markets in the EU since it regulates the provision of investment services in financial instruments by banks and investment firms as well as the operation of stock exchanges and alternative trading venues. One of its core objectives is to ensure a high degree of harmonized protection for investors in financial instruments. MiFID I (2004/39/EC) was the first version of this directive that came into force in late 2007, while a revision of it (known as MiFID II (2014/65/UE)) was implemented in January 2018. For further details on the MiFID regulation, please visit the European Commission website (https://ec.europa.eu/info/law/markets-financial-instruments-mifid-ii-directive-2014-65-eu_en).

[2] The brokerage house that provided us with the data did not offer portfolio management services during the sample period under scrutiny in this paper.

[3] Robo-advice has emerged as a cost-effective alternative to traditional financial advisors. Formally, a robo-advisor is a digital platform that provides automated, algorithm-driven financial planning services with little to no human supervision (Investopedia). Such a platform typically collects information from clients about their financial situation and future goals through a quick risk profile survey. The data collected are then used to offer advice and automatically invest the client's assets, most often based on optimized mean-variance passive indexing strategies. The goal is to automatically create and manage an investment portfolio suitable to the client's risk-return profile.

investors must complete the *S*-test, which is a bespoke online risk-return profile survey. This MiFID-compliant assessment of suitability is intended to ensure that the instruments and services offered by the broker are appropriate to the investor's knowledge and experience, investment objectives and financial situation. Once the *S*-test is completed, investors have access to the information tool on the trading platform, but they are also provided with their investment profile. This profile results from the investor's score on the 11 questions in the *S*-test. High scores are associated with either "Aggressive" or "Dynamic" profiles, while low scores lead to either "Neutral" or "Conservative" profiles.

On the other hand, unlike robo-advice, the information tool provides investors with general information on stocks, analysts' reports, and professional recommendations, but no automated portfolios are created or offered to investors. This means that the tool does not deliver personalized advice (as a human financial advisor would do) and that retail investors are still fully in charge of their trades since there is no portfolio delegation. Consequently, taking the *S*-test is arguably a proxy for the investor's appetite for financial knowledge acquisition (and not for the demand of personalized advice). Since access to the information tool on the trading platform is free, investors who take their time to answer the 11 questions of the *S*-test show an inclination toward financial knowledge acquisition. By contrast, investors who do not take the *S*-test may either disregard financial knowledge acquisition or are not interested in knowing their investment profile.

To investigate whether investors with access to the information tool behave differently from their counterparts who neglected it, we rely on detailed information about investor trading activity, sociodemographic indicators and MiFID survey material. Notably, the latter allows us to distinguish between 'S-investors' who completed both the *S*-test and the *A*-test (i.e., investors who requested the information tool) from 'non S-investors' who took only the *A*-test (i.e., investors who disregarded the information tool).

Using the aforementioned field data, we build on the theoretical models of endogenous financial knowledge acquisition (Jappelli and Padula 2013; Lusardi et al. 2017; Barthel and Lei 2021) to formulate most of our hypotheses, and we test them through a robust quasi-natural experiment approach. Since the choice to take the *S*-test is endogenous, our empirical identification strategy relies on propensity score matched difference-in-differences (PSM-DiD) to check whether investors with access to the information tool (i.e., S-investors) behave differently from their matched counterparts who neglected it on purpose (i.e., matched non S-investors). Importantly, we control for heterogeneity of the treated and control groups since propensity score matching (PSM) is first used to associate S-investors with comparable non S-investors on a large set of covariates including variables characterizing investor trading behavior before taking the *S*-test.

Following our two-step approach, we first report PSM results showing that subjective financial literacy and education are both positively associated with the decision to take the *S*-test and thereby with financial knowledge acquisition. We also find that male investors, older investors, French-speaking investors, investors who declare executive responsibilities, and wealthier investors are more likely to complete the *S*-test. By contrast, investors with longer trading experience are less likely to do so. Based on the propensity scores, we ultimately have a homogeneous sample of 10,270 S-investors (i.e., treatment group) and 10,270 matched non S-investors (i.e., control group)[4] who no longer differ on the aforementioned characteristics.

Next, our PSM-DiD regressions provide evidence of significant changes in the behavior of S-investors, that is investors with access to the information tool: better portfolio diversification but higher trading intensity and lower net returns. Financial knowledge acquisition has therefore mixed effects on trading behavior. Furthermore, since S-investors are also provided with their investment profile, discriminating between these profiles to proxy for investor sophistication reveals that the investment profile does matter to explain heterogeneity in behavior across S-investors. However, the findings do not allow us to conclude that more sophisticated investors benefited the most from financial knowledge acquisition. Our main findings are confirmed in a couple of robustness checks.

This paper makes a twofold contribution to the literature. First, our empirical evidence provides additional insights to the stream of literature on endogenous financial knowledge acquisition and trading behavior. In the models with endogenous financial knowledge acquisition, individuals can accumulate, as with other forms of human capital, financial knowledge or literacy throughout the life cycle.[5] Both Jappelli and Padula (2013) and Barthel and Lei (2021) rely on individuals' investment in financial literacy, while Lusardi et al. (2017) address investment in financial knowledge. The latter better fits our paper since, by taking the *S*-test, S-investors have access to a new financial knowledge set that corresponds to the definition of Lusardi et al. (2017) on page 435: *"On a continuum from basic to sophisticated, financial knowledge incorporates the acquisition of particular skills as well as having information and awareness of financial products"*. Therefore, the *S*-test giving access to the information tool is seen as an opportunity for retail investors to acquire financial knowledge at a specific moment in their life (through the information tool offered by the broker). To the best of our knowledge, this feature makes our paper unique as the first to use direct observation of a decision made by retail investors to acquire financial knowledge. Indeed, the empirical tests of the predictions of the three aforementioned models have not directly incorporated the individuals' investment in financial knowledge.[6] More broadly, prior evidence on the relationship between financial knowledge acquisition and trading behavior has been found using survey data. Abreu and Mendes (2012) and Tauni et al. (2015) report a positive

[4] The 10,270 matched non S-investors includes 4663 duplicates since we allow replacement as explained in Sect. 4.1.

[5] In Jappelli and Padula (2013)'s model, financial literacy and wealth are determined jointly, which gives rise to a positive correlation between financial literacy and wealth over individuals' life-cycle. Lusardi et al. (2017) show that financial knowledge acquisition is a major determinant of retirement wealth inequality. Barthel and Lei (2021) build a model in which consumers' investment in financial literacy drives financial advice-seeking and financial literacy acquisition to be substitutes.

[6] Jappelli and Padula (2013) use the Survey of Health, Aging, Retirement in Europe (SHARE) but no measure of investment in financial knowledge. Lusardi et al. (2017) use simulations to obtain empirical insights from their theoretical model. Barthel and Lei (2021) work on the 2016 SCF Survey and consider investment in financial literacy for respondents who assess that they were "attending [an] investment seminar or investment club".

relationship between financial knowledge acquisition and trading activity, but their evidence is mostly descriptive and based on survey data, without considering actual trading records.[7] Compared to past research, our paper is therefore unique in that we rely on both the trading records of investors and their actual decision to deliberately acquire financial knowledge.

Second, financial knowledge acquisition is not limited to news, analysts' recommendations, and financial information in this paper. Indeed, as mentioned above, when investors choose to take the *S*-test, they obtain access to a twofold set of information that may enhance their financial knowledge. On the one hand, they are granted free access to the information tool available on the trading platform and, on the other hand, they are also informed about their investment profile. This combination allows us to provide innovative insights in the trading behavior depending on the profile. Our paper also contributes to the literature addressing whether financial knowledge acquisition and advice-seeking are real substitutes or complements. To the best of our knowledge, Barthel and Lei (2021) is the only paper investigating this issue under the assumption of endogenous financial knowledge acquisition. Focusing on a context wherein asking advice is costly and the advice quality varies, these authors find that investment in financial knowledge acquisition and demand for advice are substitutes.[8] In this paper, financial knowledge acquisition differs in two key aspects from Barthel and Lei (2021). First, the access to the information tool does not cost S-investors money; they are granted free access once the *S*-test is completed. In other words, it only costs time to complete the questionnaire. Second, the quality of advice is not relevant because of the nature of the tool, i.e., it delivers general information on stocks and professional recommendations but no personalized advice. Our empirical evidence is consistent with financial knowledge acquisition and the demand for advice being complementary, as suggested in Lusardi et al. (2017)'s model.

The *S*-test is used in this paper as a proxy for the investor's appetite for financial knowledge acquisition. However, we cannot check whether S-investors effectively used the information tool at their disposal or how they used it. Despite this shortcoming, our results show significant shifts in the behavior of S-investors, suggesting that these investors did actually acquire financial knowledge. Nevertheless, we

7 Abreu and Mendes (2012) use a survey conducted in 2000 by the Portuguese Securities Market Commission in which 1559 investors were interviewed. Tauni et al. (2015) analyze the survey results of 333 individual investors in Chinese futures markets. In both papers, survey data are based on questions such as "How often do you buy and sell financial assets?". Furthermore, these authors focus exclusively on trading frequency and do not analyze trading behavior in a broader sense or performance. Guiso and Jappelli (2006) is an unpublished paper that provides evidence based on actual behavior. Nevertheless, these authors rely on indirect measures of financial knowledge acquisition, since they use a questionnaire to measure the time investors spent acquiring financial knowledge from different sources (such as newspapers, internet, etc.).

8 Their model relies on two main assumptions that are (1) investment in financial knowledge acquisition increases the return on savings, and (2) it also increases the probability to receive "good" advice, which additionally increases the return on savings. Barthel and Lei (2021) solve their model by taking different functional forms for the probability that advice is "good" and show that financial knowledge acquisition is higher when no professional advice is bought.

cannot rule out that some investors potentially chose to take the *S*-test for purposes other than financial knowledge acquisition. Others could have regarded taking the *S*-test as a default option since access to the information tool is free and the additional time required to complete the online survey is not substantial. These potential limitations are discussed in Sect. 7.

The remainder of this paper is structured as follows. In Sect. 2, we formulate several hypotheses regarding the impact of financial knowledge acquisition on retail investor trading behavior. Section 3 describes our data and sample. Section 4 presents our methodology. We report our main empirical results in Sect. 5 and robustness checks in Sect. 6. We discuss the potential limitations of our work in Sect. 7. Section 8 concludes the paper.

2 Hypotheses

Our hypotheses regarding the investor's decision to take the *S*-test and its impact on the trading behavior mainly build on the literature on financial knowledge acquisition. As explained in the introduction, investors who take the *S*-test automatically obtain free access to a new financial knowledge set. The latter consists of two components: (1) access to ongoing financial information through the information tool on stocks, and (2) the investment profile attributed by the broker depending on the answers provided by the investor. We essentially consider the access to the information tool to formulate our hypotheses, as ongoing information on stocks, analysts' reports, and professional recommendations might arguably help retail investors acquire financial knowledge. We also outline some conjectures about the impact of knowing one's investment profile on one's trading behavior.

The theoretical literature on financial knowledge acquisition has presented rational expectations models. In such models, the proportion of informed individuals increases with price noise (Grossman 1976; Grossman and Stiglitz 1980). This means that information becomes increasingly valuable when the price system becomes less informative due to a high level of noise. One common feature in these models is that financial knowledge is exogenously determined. This does not fit our context wherein retail investors take the *S*-test on purpose to have access to the new financial knowledge set mentioned above.

Jappelli and Padula (2013), Lusardi et al. (2017) and, more recently Barthel and Lei (2021), offer a better suited theoretical framework wherein individuals do not start their economic lives with full financial knowledge; instead, financial knowledge is acquired endogenously over the life cycle. These models have in common the assumptions that (1) investing in financial knowledge acquisition has costs and benefits and that (2) financial knowledge (or financial literacy) is an asset that depreciates over time. In short, these three intertemporal models study the impact of financial knowledge acquisition on various outcomes (e.g., wealth accumulation, wealth inequality or the value of financial advice) over the individuals' life cycle.[9]

Lusardi et al. (2017)'s model is the most relevant to the present paper. This model examines how financial knowledge acquisition affects wealth inequality while incorporating many features such as the uncertainty of asset returns, income and medical expenditures, stochastic mortality risks, and groups of individuals with different education levels. Although its framework does not perfectly fit our purpose in this paper, this model drives many implications that are helpful to formulate most of our hypotheses.

First, an important finding of Lusardi et al. (2017)'s model is that some individuals may remain optimally financially ignorant, especially because not everyone benefits from greater financial knowledge acquisition. Hence, according to Lusardi et al. (2017), the demand for financial knowledge depends on the difference between the marginal costs and benefits of financial knowledge acquisition, which may differ across investors. In our context wherein completing the *S*-test automatically provides access to a new financial knowledge set, we argue that the *S*-test may attract more sophisticated investors because the tradeoff between the costs and benefits of financial knowledge acquisition is likely to be the lowest for them. In other words, we assume that financial knowledge acquisition attracts investors who are able to effectively use it, that is, investors with high financial literacy. This leads us to our first hypothesis:

H1: The decision to acquire financial knowledge by taking the *S*-test is positively related to financial literacy.

Another prediction of Lusardi et al. (2017)'s model is that people who invest in financial knowledge may earn higher expected returns than those who do not. Through the calibration of their model, these authors estimate that 30–40% of retirement wealth inequality is accounted for by financial knowledge acquisition alone. Consistent with this result, we assume that investors who completed the *S*-test earn higher returns on their stock portfolios than their counterparts who did not. This is our second hypothesis:

H2: Acquiring financial knowledge by taking the *S*-test leads to higher portfolio returns.

In the same vein, when investors acquire financial knowledge, they may better understand the benefits associated with diversification. Although Lusardi et al. (2017) do not directly address this relationship, there is empirical evidence showing that financial knowledge can improve retail investors' diversification (e.g., Calvet et al. 2009; Gaudecker 2015). Consequently, investors who completed the *S*-test are

[9] Jappelli and Padula (2013) propose a consumption model in which individuals who are endowed with an initial stock of financial literacy can increase their net returns from intertemporal trades by investing in financial literacy. Their model implies that financial literacy and wealth are positively correlated over the consumers' life cycle and shows that there exists an optimal investment in literacy. Barthel and Lei (2021) specifically address whether investment in financial literacy and financial advice-seeking are substitutes or complements when both advice costs money and the quality of advice fluctuates. Their model is similar to that of Jappelli and Padula (2013), except that individuals explicitly decide whether to purchase advice from a financial adviser.

expected to hold more diversified stock portfolios than their counterparts who did not, as postulated in our third hypothesis:

H3: Acquiring financial knowledge by taking the *S*-test leads to hold more diversified portfolios.

Financial knowledge acquisition may also generate side effects such as overconfidence that leads to higher trading volume and higher portfolio turnover (Guiso and Jappelli 2006). As previously mentioned, S-investors are also informed about their investment profile once the *S*-test is completed. Based on their answers to the questions included in the *S*-test, S-investors are assigned a profile, which is either "Dynamic" or "Aggressive" in the event of a high or very high score or "Conservative" or "Neutral" in the event of a very low or low score. The fact that S-investors are informed about their investment profile is likely to either reinforce or contradict their beliefs on their own skills and perceptions. Reinforcement in their beliefs is most likely to bolster overconfidence due to an illusion of both knowledge and control (Barber and Odean 2001a; Volpe et al. 2002). In contrast, according to Argentesi et al. (2010), contradictions in one's beliefs can lead to cognitive dissonance (Festinger 1957), thereby making the person uncomfortable.[10] However, since access to ongoing financial information is also known to encourage investors to trade more actively (e.g., Barber and Odean 2001b, 2002), overconfidence is expected to dominate cognitive dissonance. We therefore propose the following hypothesis:

H4: Acquiring financial knowledge by taking the *S*-test leads to an increase in trading intensity.

Our next hypothesis explores whether having access to a new financial knowledge set affects S-investors in their portfolio allocation choices. Ideally, we would like to check whether S-investors tend to follow the professional recommendations provided by the information tool and then invest in the promoted stocks. Unfortunately, we do not have any data about the informational content provided by the tool. We therefore need to find an alternative to identify the types of stocks the most likely to be promoted by this tool. According to Kumar (2009), stocks that could be perceived as lotteries show lower analyst coverage than other stocks. In addition, small firms are usually less recommended (Bhushan 1989) and lottery-like stocks are more likely to be small-caps than blue chips. Building on these findings, we posit that lottery-like stocks are less likely to be promoted by the information tool. Furthermore, lottery-like stocks are likely to attract similar socioeconomic clienteles to lotteries because of their characteristics, namely, low price, high idiosyncratic volatility and high idiosyncratic skewness. In other words, lottery-like stocks are more attractive to relatively less sophisticated investors (Kumar 2009). Taken together, all these points lead us to formulate the following hypothesis:

[10] In this situation, the investor may either avoid any information likely to create dissonance or process the available information so as to reduce it. Cognitive dissonance might explain why an investor may decide to ignore free information to preserve his/her beliefs. This could explain why, although investors obtain access to a new financial knowledge set, they can decide not to trade. A related theoretical direction is to assume, as the model of Andries and Haddad (2020) does, that retail investors exhibit information aversion that makes them inattentive to flows of news.

H5: Acquiring financial knowledge by taking the *S*-test leads investors to hold a lower proportion of lottery-like stocks in their portfolio.

Our last hypothesis is about the relationship between the profile of S-investors and their trading behavior. According to Barthel and Lei (2021), individuals with high education and financial literacy, who are wealthy and risk tolerant, homeowners, and those who have a relatively longer investment horizon are more likely to use financial advisers. These findings support past research (e.g., Bhattacharya et al. 2012; Collins 2012). Among S-investors, the investment profile is arguably a reliable proxy for the sophistication level since the higher the score is, the more aggressive the profile. As explained in the introduction, the aim of the *S*-test is to assess the suitability of the instruments and services offered by the broker depending on the investor's knowledge and experience, investment objectives and financial situation. Practically speaking, the *S*-test covers these three items and includes questions about financial literacy, risk tolerance, investment horizon, annual net income, return objective, etc. The four available profiles, ranked by increasing score, are Conservative, Neutral, Dynamic, and Aggressive. Hence, S-investors assigned a "Dynamic" or "Aggressive" profile are overall more sophisticated than those with the "Conservative" or "Neutral" profile. Since financial information would better benefit to those who are able to effectively use it (Lusardi et al. 2017), we conjecture that S-investors with a "Dynamic" or "Aggressive" profile are more likely to be positively affected by the information tool. In other words, these more sophisticated S-investors should make better trading decisions over time, leading to higher net returns, more diversified portfolios, lower trading intensity, and lower propensity to hold lottery-like stocks. This is well summarized in the following last hypothesis:

H6: S-investors who are more sophisticated (those with a "Dynamic" or "Aggressive" profile) benefit the most from financial knowledge acquisition than the other S-investors (those with the "Neutral" or "Conservative" profile).

3 Data and sample

Our primary dataset comes from a large online Belgian brokerage house and encompasses the trading records of more than 40,000 retail investors over the January 2003–March 2012 period (i.e., 111 months).[11] For the purpose of this paper, we focus on a selected set of 21,233 retail investors who hold stock portfolios during at least 12 months both before and after the first of January 2008, that is, before and after the implementation of MiFID. Our selection procedure is detailed and discussed in Sect. 3.1.

For each investor, we have (1) detailed trading records (ISIN code, timestamp, trade direction, executed quantity, trade price, explicit transaction costs, etc.); (2) sociodemographic indicators such as age, gender, level of education, and spoken language; and (3) MiFID survey data (see Table 8 in appendix for examples). The latter

[11] Several papers (e.g., D'Hondt and Roger 2017; Bellofatto et al. 2018; D'Hondt et al. 2020; Desagre and D'Hondt 2021; D'Hondt et al. 2021a, b) rely on the same database but focus on different samples, according to their research questions.

are notably used to differentiate 'S-investors' who completed both the *S*-test and the *A*-test from 'non S-investors' who took only the *A*-test. Accordingly, S-investors are those who received access to the information tool and their investment profile, while non S-investors are those who disregarded this new financial knowledge set. Note that since both S-investors and non S-investors completed the A-test (requirement for order execution on complex instruments[12]), this categorization exclusively relies on the *S*-test (requirement for financial advice).

In addition, we have historical stock prices from either Eurofidai[13] for European stocks or Bloomberg for non-European stocks. These data are necessary to value end-of-month individual stock portfolios and compute portfolio returns. We calculate monthly portfolio gross and net returns using an approximation of the modified Dietz method, with the aim of delivering a return close to the money-weighted rate of return (Shestopaloff and Shestopaloff 2007). Building on D'Hondt et al. (2020) and D'Hondt et al. (2021a), we compute portfolio returns per investor and month assuming that all the purchases (sales) executed in a given month take place on the first (last) day of the month. Mathematically, we have $R_t = \frac{\mathrm{EMV}_t - \mathrm{EMV}_{t-1} - P_t + S_t}{\mathrm{EMV}_{t-1} + P_t}$, where R_t is the portfolio gross return for month t, EMV_t is the end-of-month portfolio value at month t, and P_t and S_t are the aggregate monetary value of all purchases and sales executed during month t. When calculating net returns, we subtract from the numerator the aggregate transaction costs paid during month t.

3.1 Sample selection

As explained in the introduction, MiFID requirements came into force at the end of 2007. Since the brokerage data do not report the date at which each investor completed the MiFID questionnaires, we need to identify investors who were active enough both before and after the implementation of the MiFID requirements. Hence, we decided to select investors who hold a nonempty stock portfolio both for at least 12 months before January 1, 2008, and for at least 12 months after January 1, 2008. The MiFID questionnaires are then assumed to have been completed in early 2008, as in D'Hondt et al. (2021a).

Our selection leads to a sample composed of 21,233 retail investors, of whom 10,963 only asked for order execution and completed the *A*-test. By contrast, the other 10,270 investors also completed the *S*-test to be granted access to the information tool on stocks. As a result, we have 10,963 non S-investors (who only took the *A*-test) and 10,270 S-investors (who completed both the *A*-test and the *S*-test). Both

[12] Only guidelines and general rules for implementing the MiFID questionnaires are available. They contain neither a clear definition nor a specific list of complex instruments. The general consensus is that derivatives (such as options, futures, and warrants) and structured products are typically complex instruments. However, due to the precautionary principle, the broker that provided us with the data considered all securities that differ from common stocks as complex instruments. The consequence is that all the retail investors were invited to complete the *A*-test.

[13] www.eurofidai.org.

groups of investors execute their trades on their own, and there is no portfolio delegation. Since the access to the information tool on the trading platform is free, only a few minutes are needed to complete the *S*-test that consists of 11 questions covering the investor's investment objectives, financial situation, and financial knowledge and experience, as required by MiFID.[14] Consequently, investors who take their time to answer this additional online survey reveal their inclination toward financial knowledge acquisition. Based on their score at the *S*-test, they are also informed about their investment profile: "Dynamic" or "Aggressive" in the case of a high or very high score or "Conservative" or "Neutral" in the case of a very low or low score. As mentioned earlier, this profile will be used as a proxy for the investor's sophistication since the higher the score is, the more aggressive the profile.

The sample selection has potential consequences for the validity of our analysis. On the one hand, the assumption that investors completed the MiFID questionnaires before trading in the post-MiFID period is obviously reasonable since they were all already active before January 2008. On the other hand, by requiring a holding period of at least 12 months for stock portfolios in both pre- and post-MiFID periods, we are quite restrictive. Our motivation is to disregard dormant accounts and short-lived investors who could hold stocks for a few months only. As robustness checks, we also consider a less restrictive filter, i.e., a holding period of at least 6 months for stock portfolios in both the pre- and post-MiFID periods, as well as a more restrictive filter, i.e., a holding period of at least 24 months for stock portfolios in both the pre- and post-MiFID periods. In Sect. 6, we replicate our main analysis on these two alternative samples to check whether our findings depend on sample selection.

3.2 Trading activity of S-investors and non S-investors

Our sample of investors executed a total of 2,946,113 trades on stocks, of which 56.26% took place over the January 2003–December 2007 period. This aggregate amount of trades corresponds to a monetary volume of €26,485 million, with 57% of it being traded in the pre-MiFID period.

Descriptive statistics about investor trading activity over the entire period (i.e., January 2003–March 2012) are reported in Table 9 in the appendix. To further disentangle investor activity and behavior before and after the implementation of MiFID, Table 1 reports cross-sectional statistics before and after the first of January 2008. First, our selection leads to retain investors with enough average experience in both periods. At least 75% of both non S-investors and S-investors exhibit a trading experience longer than 24 months (the minimum required by the filter). Regarding trading activity, the typical non S-investor executed 71 (59) trades, for a corresponding trading volume of €684,555 (€568,221) in the pre-MiFID (post-MiFID) period. For S-investors, the average number of trades (85 and 78) and the trading volume (€740,565 and €643,168) are slightly higher in both periods. For the two types of investors, we notice a decrease in trading activity in the post-MiFID period.

[14] As mentioned earlier, some detailed items of the *S*-test are provided in Table 8 (in appendix) as examples.

The stock portfolio-based variables reveal considerable heterogeneity in both behavior and performance across investors. The typical non S-investor holds a four-stock (six-stock) portfolio in the pre-MiFID (post-MiFID) period, compared to a six-stock (ten-stock) portfolio for the typical S-investor. Such underdiversification is consistent with past empirical evidence on retail investors (e.g., Kumar and Lee (2006) and Polkovnichenko (2010) for the US and Broihanne et al. (2016) for France). For non S-investors, the average end-of-month portfolio value is approximately €52,102 in the pre-MiFID period and €67,152 in the post-MiFID period (with a median of €11,026 and €10,749, respectively). S-investors appear somewhat wealthier in each period, with an average portfolio value of €61,264 or €93,357 (and a median of €16,128 or €21,026). Portfolios over the January 2008–March 2012 period are on average more diversified (from 4 to 6 stocks for non S-investors and from 6 to 10 stocks for S-investors) and exhibit higher monetary values, especially for S-investors.

Consistent with the presence of some highly active investors who hold very small portfolios, the average monthly portfolio turnover[15] is high, i.e., approximately 13.02 (18.21) for non S-investors and 5.52 (6.14) for S-investors in the pre-MiFID (post-MiFID period). The corresponding medians are much lower (i.e., 0.28 and 0.04 for non S-investors and 0.30 and 0.06 for S-investors) and show that both types of investors seem to have reduced their trading intensity in the second period.

Since one of our hypotheses addresses the propensity of investors to invest in lottery-like stocks, Table 1 also reports statistics about the lottery-like stock ratio. This ratio is defined as the number of lottery-like stocks held in the end-of-month portfolio divided by the total number of stocks in this end-of-month portfolio. To compute that ratio, we build on Kumar (2009) to flag lottery-like stocks at the monthly frequency.[16] Accordingly, the stocks identified as lottery-like are those in the lowest 50^{th} stock price percentile, the highest 50^{th} volatility percentile, and the highest 50^{th} skewness percentile. Volatility and skewness are computed using daily returns over the last 6 months. For the sake of simplification, we use total volatility and total skewness since Kumar (2009) reports that both idiosyncratic and total measures provide equivalent results. The typical non S-investor displays a lottery-like stock ratio of 6.75% and 6.72% in the pre-MiFID and post-MiFID period, respectively. This ratio is slightly lower for the typical S-investors in both periods, i.e., 5.41% and 6.08%, respectively. In comparison, the proportion of lottery-like stocks in retail portfolios is equal to 3.74% in the sample of Kumar (2009).

Regarding stock portfolio performance, Table 1 shows that the typical non S-investor earned a monthly realized return of 0.07% or −0.19% net of explicit transaction costs in the pre-MiFID period. However, the typical non S-investor's gross

[15] The monthly portfolio turnover is calculated as the sum of all the purchases and sales in a particular month, divided by the average of the portfolio values at the beginning and the end of this particular month. Since the average portfolio value can sometimes be very small, extreme values for turnover are likely. Hence, the median turnover is more representative. The average turnover has to be interpreted with caution because it can be very high without having any economic meaning. In Sect. 5, portfolio turnovers are winsorized with a 1% cutoff.

[16] This means that a given stock may be considered a lottery-like stock for a specific month and not necessarily over the full sample period.

Table 1 Descriptive statistics for trade-based variables and stock portfolio-based variables before and after January 1, 2008

	Non S-investors (N=10,963)				S-investors (N=10,270)			
	Mean	Median	Q1	Q3	Mean	Median	Q1	Q3
Panel A: BEFORE JANUARY 1, 2008								
Number of trades	71	25	9	67	85	36	14	90
Trading volume on stocks (in €)	684,555	87,870	21,346	365,448	740,565	124,982	37,383	441,063
Trading experience (in months)	39	39	24	53	37	36	24	52
Portfolio size (in # stocks)	4	3	1	5	6	4	2	7
Portfolio value (in €)	52,102	11,026	3990	31,402	61,264	16,128	5981	43,764
Portfolio turnover	13.02	0.28	0.15	0.58	5.52	0.30	0.17	0.55
Lottery-like stock ratio (%)	6.75	4.10	0.82	9.69	5.41	3.39	1.00	7.46
Gross return (%)	0.07	0.22	−0.94	1.08	0.42	0.48	−0.39	1.16
Net return (%)	−0.19	0.01	−1.18	0.88	0.16	0.26	−0.67	0.97
Volatility	0.18	0.07	0.05	0.12	0.14	0.06	0.04	0.10
Panel B: AFTER JANUARY 1, 2008								
Number of trades	59	13	4	45	78	27	9	73
Trading volume on stocks (in €)	568,221	46,501	9442	212,309	643,168	89,844	23,261	321,048
Trading experience (in months)	49	51	51	51	50	51	51	51
Portfolio size (in # stocks)	6	3	2	7	10	6	3	13
Portfolio value (in €)	67,152	10,749	3070	35,425	93,357	21,026	6681	58,916
Portfolio turnover	18.21	0.04	0.01	0.15	6.14	0.06	0.02	0.18
Lottery-like stock ratio (%)	6.72	5.39	2.29	9.31	6.08	5.11	2.94	7.84
Gross return (%)	−1.15	−0.80	−1.80	−0.09	−0.98	−0.73	−1.44	−0.21
Net return (%)	−1.23	−0.87	−1.89	−0.15	−1.07	−0.80	−1.54	−0.28
Volatility	0.24	0.09	0.06	0.14	0.16	0.08	0.06	0.11

This table reports the cross-sectional mean, median, and lower and upper quartiles for trade-based variables and stock portfolio-based variables computed on a per investor basis over the January 2003–December 2007 period (Panel A) and the January 2008–March 2012 period (Panel B), with a distinction between non S-investors and S-investors. The sample is composed of 10,963 non S-investors (who only took the *A*-test) and 10,270 S-investors (who completed both the *A*-test and the *S*-test). 'Number of trades' is the total number of trades executed on stocks. 'Trading volume on stocks' is the total monetary trading volume (purchases plus sales). 'Trading experience' refers to the number of months between the last and first trade. 'Portfolio size' is the average number of stocks in the end-of-month portfolio. 'Portfolio value' is the average end-of-month portfolio value. 'Portfolio turnover' refers to monthly portfolio turnover, calculated as the sum of all the purchases and sales in a particular month, divided by the average of the portfolio values at the beginning and the end of this particular month. 'Lottery-like stock ratio' is defined as the number of lottery-like stocks held in the end-of-month portfolio divided by the total number of stocks in the end-of-month portfolio. Using daily returns over the last 6 months, the stocks identified as lottery-like are those in the lowest 50th stock price percentile, the highest 50th total volatility percentile, and the highest 50th total skewness percentile. 'Gross return' is the geometric average of monthly portfolio gross returns. 'Net return' is the geometric average of monthly portfolio net returns. 'Volatility' is the standard deviation of monthly portfolio gross returns and net realized returns both become negative in the post-MiFID period, i.e., −1.15% and −1.23%. The same phenomenon is observed for the typical S-investor who achieved a monthly realized return of 0.42% (−0.98%) or 0.16% (−1.07%) after transaction costs in the pre-MiFID (post-MiFID) period. Regardless of investor type, even the returns in the upper quartile are negative over the January 2008–March 2012 period. The monthly volatility of returns is approximately 0.18 for the typical non S-investor and 0.14 for the typical S-investor in the pre-MiFID period, while it increases in the post-MiFID period to 0.24 and 0.16, respectively. Such results showing negative average returns and higher volatility in the post-MiFID period are not surprising since this period includes the 2008 global financial crisis. These findings overall confirm the detrimental impact of this crisis on investor stock portfolio performance.

3.3 Individual characteristics of S-investors and non S-investors

In addition to trading records, we have investors' sociodemographic indicators and MiFID survey data. Since each investor in our sample completed the *A*-test, the latter provides us with a set of individual characteristics available for every investor. The *A*-test under scrutiny is an online survey consisting of categorical questions for which the investor has to select a single answer.[17] Answers are self-reported decisions that investors make on their own, without any interaction with a financial intermediary. Since the questions imply a self-assessment of financial knowledge and experience, these survey data deliver subjective individual variables for which statistics are reported in Panel A of Table 2.

Approximately 14% of the S-investors and 16% of the non S-investors selected the highest level of financial knowledge. In contrast, 10.96% of S-investors and 17.60% of non S-investors chose the lowest level, thereby declaring a basic financial knowledge. The remaining investors in each group, who represent a clear majority, are spread across the two intermediate levels. Regarding the experience in complex instruments, approximately 25% of both S-investors and non S-investors evaluated their experience as good (the highest level on the scale). Moreover, 24.83% of S-investors and 30.10% of non S-investors declared no experience (lowest level), while the others positioned themselves in between (average experience). Additionally, more than 50% of the investors in both groups declared having already invested at least once in a complex instrument.

Table 2 also provides statistics about the sociodemographic variables in Panel B. As for education, 72.17% of non S-investors and 75.16% of S-investors have a university degree or equivalent. Only less than 7% of non S-investors and 5% of S-investors have no degree, which indicates that most of the investors are educated. In both groups, there is also a large majority of males. Regarding the spoken language, slightly more than 55% of the investors selected Dutch as main language,

[17] Some questions are grouped together when they relate to the same topic. For example, the *A*-test contains distinct questions regarding the investor's knowledge about options, future contracts, warrants, structured products, etc. For convenience, we group these questions together and create a single item related to the knowledge of complex instruments.

while English was rarely chosen. Regarding the professional status, almost 17% of non S-investors and 19% of S-investors declared executive responsibilities. Regarding age (unreported in Table 2), the typical non S-investor and S-investor is 50.51 and 52.49 years old, respectively.

Most of the differences reported in the last column of Table 2 are statistically significant at the 1% level, suggesting that S-investors and non S-investors differ on a large set of individual characteristics. At first glance, S-investors appear more sophisticated than non S-investors, i.e., they tend to report a higher financial literacy and more experience, and they also count a higher proportion of highly educated and executive investors. To properly capture the impact of the information tool (and the related financial knowledge acquisition) on investors in the post-MiFID period, we need to compare S-investors with truly comparable non S-investors.

4 Methodology

We apply a quasi-natural experiment methodology (PSM-DiD) to chech whether S-investors with access to the information tool in the post-MiFID period behave differently from their matched counterparts (who neglected the tool).

The DiD methodology is frequently used to study the impact of a given treatment on a "treatment group" versus a "control group" in a natural experiment. It allows calculating the effect of a treatment (i.e., an independent variable) on an outcome (i.e., a dependent variable) by comparing the average change over time in the outcome variable for the treatment group to the average change over time in this outcome variable for the control group (Bertrand et al. 2004). However, the DiD methodology assumes that the treated and control groups are observationally equivalent pre-treatment. Statistics reported in Sect. 3.3 show that S- and non S-investors differ on a large set of individual characteristics. Comparing their trading behavior might therefore be misleading since any differences in trading behavior might be due to other investor-related effects that are correlated with the inclination toward financial knowledge acquisition. As stated in Stuart (2010), when estimating causal effects using observational data, it is desirable to replicate experiments as closely as possible by obtaining treated and control groups with similar distributions of covariates. To do so, we opt for PSM that is generally used before DiD when the assumption of the treated and untreated groups being observationally equivalent pre-treatmentis is violated. It entails matching individuals in the treated and control groups with the same observed characteristics.

PSM has been frequently used in the literature on retail investor behavior (e.g., Gerhardt and Hackethal (2009), Kramer (2012), D'Hondt and Roger (2017)). In our case, PSM enables us to identify non S-investors matched to S-investors and select truly comparable investors, that is, investors who mainly differ in their appetite for financial knowledge acquisition. The details of our PSM randomization is presented in Sect. 4.1 and allows to test our first hypothesis (H1).

Using the PSM results, we apply the DiD methodology wherein the *S*-test is the treatment (since taking the *S*-test gives access to the information tool), S-investors constitute our treatment group, and matched non S-investors are our control group.

Table 2 Individual characteristics of S-investors and non S-investors

	Non S-investors		S-investors		
	N	Freq.(%)	N	Freq. (%)	Diff. (%)
Panel A: Subjective individual variables					
Subjective knowledge of financial markets					
Level 0	1929	17.60	1126	10.96	6.63***
Level 1	3229	29.45	2794	27.21	2.25***
Level 2	4041	36.86	4867	47.39	−10.53***
Level 3	1764	16.09	1483	14.44	1.65***
Subjective experience in complex instruments					
Level 0	3300	30.10	2550	24.83	5.27***
Level 1	4869	44.41	5168	50.32	−5.91***
Level 2	2794	25.49	2552	24.85	0.64
Investment in complex instruments					
No	5186	47.30	4237	41.26	6.05***
Yes	5777	52.70	6033	58.74	−6.05***
Panel B: Sociodemographic variables					
Level of education					
Level 0	714	6.51	435	4.24	2.28
Level 1	2337	21.32	2116	20.60	0.71
Level 2	7912	72.17	7719	75.16	−2.99***
Gender					
Female	1477	13.47	886	8.63	4.85***
Male	9486	86.53	9384	91.37	−4.85***
Language					
French speaker	3957	36.09	4090	39.82	−3.73***
Dutch speaker	6389	58.28	5812	56.59	1.69***
English speaker	617	5.63	368	3.58	2.04***
Professional status					
Executive	1815	16.56	1926	18.75	−2.20***
Other	9148	83.44	8344	81.25	2.20***
N	10,963		10,270		

This table reports the number of investors and the empirical frequencies for individual characteristics, with a distinction between non S-investors and S-investors. The sample is composed of 10,963 non S-investors (who only took the A-test) and 10,270 S-investors (who completed both the A-test and the S-test). Panel A refers to individual characteristics that are survey data based on the MiFID Appropriateness questionnaire, while Panel B contains sociodemographic indicators. For 'Subjective knowledge of financial markets', level 0 is associated with basic knowledge while level 3 refers to an experienced investor who manages all aspects of financial markets. For 'Subjective experience in complex instruments', levels 0, 1 and 2 correspond to no experience, average experience and good experience, respectively, in options, structured products, Forex and futures contracts. 'Investment in complex instruments' is yes if the investor reports having already invested in options, structured products, Forex or futures contracts and no otherwise. 'Level of education' is based on a three-level scale, wherein level 0 refers to no qualification, level 1 to secondary school/high school qualification, and level 3 to university degree or equivalent. 'Gender' distinguishes between female and male. 'Language' is French speaker if the investor selected French as main spoken language, Dutch speaker if the investor selected Dutch as main spoken language, and English speaker if the investor selected English as main spoken language. 'Professional status' is executive if the investor claimed executive responsibilities and other in all other cases. The last

Table 2 (continued)

column reports the difference between the non S-investors and S-investors and its significance. ***, **, and * indicate significance at 1%, 5%, and 10%, respectively

Specifically, we consider investors' behavior through DiD regressions including five different dependent variables, which are presented in Sect. 4.2.

4.1 PSM randomization

Since we consider the enforcement of MiFID as an exogenous shock in our DiD approach, we need to build two highly comparable groups of treated and control investors who should have homogeneous expectations and be exposed to similar environments in the post-MiFID period. According to Stuart (2010), nearest-neighbor matching is one of the most common and easy-to-implement matching methods. In its simplest form, 1:1 nearest-neighbor matching selects the closest control individual to each treated individual *i*. The distance between two individuals is based on their respective propensity scores, defined as the probability of receiving the treatment given the observed covariates (Rosenbaum and Rubin 1983). The propensity score is advantageous because it facilitates the construction of matched sets with similar distributions of covariates, without requiring close or exact matches on any of the individual variables (Stuart 2010).

We compute the propensity scores using a logit model wherein the dependent variable, Y_i, is a binary variable that equals 1 if investor *i* completed the *S*-test (to have access to the information tool) and 0 otherwise. The probability of showing an inclination toward financial knowledge acquisition is conditioned on a set of regressors, *X*, and is given by $\text{Prob}[Y = 1|X] = \wedge(x\prime b)$, where $\wedge(.)$ is the logistic cumulative distribution function. When selecting the regressors, Stuart (2010) recommends including variables associated with both the treatment (i.e., taking the *S*-test) and the outcome (i.e., the trading behavior). Building on past research, we use a large set of regressors as discussed below.

First, as explained in Sect. 2, financial literacy is crucial in this paper since we posit in H1 that the decision to take the *S*-test (and acquire financial knowledge) is positively related to financial literacy and education. Building on Lusardi et al. (2017), we argue that the *S*-test is likely to attract more sophisticated investors because the tradeoff between the costs and benefits of financial knowledge acquisition should be lower for them. In other words, financial knowledge acquisition should attract investors who are the most able to effectively use it, that is, investors with higher financial literacy.

Beyond Lusardi et al. (2017)'s work, financial literacy is well known to be associated with both the demand for financial advice (e.g., Hackethal et al. 2012; Collins 2012; Georgarakos and Inderst 2014; Calcagno and Monticone 2015) and trading behavior (e.g., Van Rooij et al. 2011; Balloch et al. 2014; Hsiao and Tsai 2018;

Bellofatto et al. 2018).[18] In addition, education and other sociodemographic indicators are also known to affect the propensity for financial knowledge acquisition (e.g., Bluethgen et al. 2008; Haslem 2008; Chalmers and Reuter 2010; Gerhardt and Hackethal 2009; Kramer 2012). Regarding trading behavior, past research has shown that women hold less risky assets than men (e.g., Bernasek and Shwiff 2001; Dwyer et al. 2002; Agnew et al. 2003; Charness and Gneezy 2012) and that age has an impact on the mix of risky assets, although it is not clear whether young people invest more (Ackert et al. 2002) or less in stocks (Shum and Faig 2006) than their older counterparts. There is also evidence that native language and culture impact stock holdings and trades (e.g., Grinblatt and Keloharju 2001). Investor trading experience is also known to affect behavior (e.g., Glaser and Weber 2009; Nicolosi et al. 2009; Koestner et al. 2017) and portfolio size (e.g., Glaser 2003; Vissing-Jorgensen 2003; Dorn and Huberman 2005; Guiso and Jappelli 2008; Goetzmann and Kumar 2008; Abreu and Mendes 2012).

Building on the above literature and the data at hand, we opt for the following logit model:

$$\begin{gathered} Y_i = \alpha + \beta_1 \mathrm{FL1}_i + \beta_2 \mathrm{FL2}_i + \beta_3 \mathrm{FL3}_i + \beta_4 \mathrm{ECI1}_i + \beta_5 \mathrm{ECI2}_i + \beta_6 \mathrm{ICI}_i \\ + \beta_7 \mathrm{LE1}_i + \beta_8 \mathrm{LE2}_i + \beta_9 \mathrm{Male}_i + \beta_{10} \mathrm{FS}_i + \beta_{11} \mathrm{ES}_i + \beta_{12} \mathrm{Exec}_i + \beta_{13} \mathrm{Age}_i \\ + \beta_{14} \mathrm{LOG}(\mathrm{MPV}_i) + \beta_{15} \mathrm{Exp}_i \end{gathered} \tag{1}$$

where the set of regressors consists of an intercept, an initial set of subjective individual characteristics (listed in Panel A of Table 2), a second set of sociodemographic variables (see Panel B of Table 2) including age, and a third set composed of two additional variables intended to control for actual past trading behavior (i.e., before the completion of the MiFID questionnaires and having access to the information tool). In Eq. 1, $\mathrm{FL1}_i$, $\mathrm{FL2}_i$, and $\mathrm{FL3}_i$ are $N-1$ dummies for the level of subjective financial literacy selected by investor i; $\mathrm{ECI1}_i$ and $\mathrm{ECI2}_i$ are $N-1$ dummies for the level of subjective experience in complex instruments reported by investor i; ICI_i is another dummy set to one when investor i declares having already traded complex instruments; $\mathrm{LE1}_i$ and $\mathrm{LE2}_i$ are $N-1$ dummies for the level of education of investor i; Male_i is a dummy set to one when investor i is male; FS_i and ES_i are $N-1$ dummies for spoken language, with FS_i being set to one when investor i selected French and ES_i being set to one when investor i selected English; Exec_i is a dummy set to one when investor i declared executive responsibilities; Age_i is the age of investor i at the end of the year 2012; $\mathrm{LOG}(\mathrm{MPV}_i)$ is the log of investor i's average end-of-month portfolio value computed over the January 2003–December 2007 period; and Exp_i is the number of months in which investor i's stock portfolio is not empty over the January 2003–December 2007 period.

Using the propensity scores estimated with Eq. 1, we opt for matching with replacement to associate each S-investor to the closest non S-investor. Table 3 reports the results of the before-match model, i.e., when estimating the logit model on the full sample of investors. Focusing on Panel A devoted to subjective individual

[18] Financially literate households are known to invest more in stocks (e.g., Van Rooij et al. 2011; Balloch et al. 2014) and in derivatives (Hsiao and Tsai 2018).

variables, the results clearly indicate a positive relationship between the demand for the information tool and subjective financial literacy, whatever the level of knowledge. This evidence provides support for our first hypothesis stating that the probability of taking the *S*-test (and acquiring financial knowledge) is higher for investors with higher financial literacy. Investors who have already invested in complex instruments are also more likely to ask for the information tool. By contrast, investors who perceive themselves as highly experienced in complex instruments are significantly less likely to do so. This last result might be due to the distinction between self-assessed knowledge and experience, which can be viewed as two distinct components of financial literacy. Overall, the findings in Panel A lead us to validate our first hypothesis and confirm that the *S*-test is taken by investors with higher financial literacy. These findings do not support prior evidence that subjective financial literacy is negatively correlated with the demand for financial advice (e.g., Hung and Yoong 2014; Georgarakos and Inderst 2014; Calcagno and Monticone 2015). In particular, Calcagno and Monticone (2015) provide evidence that while subjective financial literacy is negatively correlated with the demand for financial advice, objective financial literacy exhibits the opposite relationship. In our sample, we do not have objective measures of financial literacy, i.e., based on a set of questions addressing the fundamental concepts at the roots of saving and investment decisions (such as compounding, inflation, and diversification). Hence, we cannot rule out the possibility that, for some investors, objective financial literacy differs from subjective financial literacy.

Panel B of Table 3 shows that education is also positively related to the demand for the information tool (and financial knowledge acquisition). Such a relationship is consistent with the aforementioned result about financial literacy, as well as with the findings of Hung and Yoong (2014), Collins (2012), Hackethal et al. (2012), and Hoechle et al. (2017) showing that the likelihood of asking for financial advice increases with the level of education.

Our results regarding gender and age reveal that being a male and older significantly increases the inclination toward financial knowledge acquisition. Regarding spoken language, we find significant coefficient estimates revealing that, compared to Dutch-speaking investors, French speakers are more likely to ask for the information tool while English speakers are less likely to do so. Next, unlike Bluethgen et al. (2008) and Hackethal et al. (2012), who do not observe any effect of executive responsibilities on the demand for financial advice, we find that investors who report executive responsibilities are more likely to ask for the information tool.

The last two variables in Panel C of Table 3, which capture some aspects of actual past behavior, display coefficient estimates of opposite signs. While investors with higher portfolio value are more likely to ask for the information tool, investors with longer trading experience are less likely to do so. Although the negative impact of trading experience on the likelihood of asking for the information tool is not in line with Gerhardt and Hackethal (2009), who find that investors' ex ante trading experience is positively related to the decision to refer to a financial advisor, it is consistent with that found for the subjective assessment of experience in Panel A.

Using the propensity scores estimated with the logit model, we match S-investors with non S-investors to build a sample of homogeneous investors on the whole set of

Table 3 Results of the logit model for propensity score matching

Independent variables	Before-match Estimates	After-match Estimates
Intercept	−2.535***	0.059
Panel A: Subjective individual variables		
Subjective knowledge of financial markets 1 (FL1)	0.262***	−0.034
Subjective knowledge of financial markets 2 (FL2)	0.531***	−0.014
Subjective knowledge of financial markets 3 (FL3)	0.231***	−0.052
Subjective experience in complex instruments 1 (ECI1)	0.017	0.030
Subjective experience in complex instruments 2 (ECI2)	−0.170***	0.036
Investment in complex instruments (ICI)	0.175***	−0.028
Panel B: Sociodemographic variables		
Level of education 1 (LE1)	0.181**	−0.106
Level of education 2 (LE2)	0.221***	−0.079
Male	0.551***	0.022
French speaker (FS)	0.068**	0.012
English speaker (ES)	−0.435***	−0.010
Executive (Exec)	0.108***	0.035
Age	0.009***	0.002
Panel C: Trading variables		
Portfolio value (LOG(MPV))	0.130***	−0.010
Trading experience (Exp)	−0.009***	0.001
Pseudo R^2	0.043	0.000
N	21,233	20,540
S-investors	10,270	10,270
Non S-investors	10,963	10,270

This table reports the parameter estimates for a logit model wherein the dependent variable, Y_i, is a binary variable that equals 1 if investor i completed the *S*-test and 0 otherwise. Based on Eq. 1, the results are reported before-match (column 2) and after-match (column 3). The set of independent variables consists of an intercept, an initial set of subjective individual characteristics (in Panel A), a second set of sociodemographic variables (in Panel B), and a third set composed of two additional variables intended to control for actual trading behavior before the completion of the MiFID questionnaires (in Panel C). $FL1_i$, $FL2_i$, and $FL3_i$ are $N-1$ dummies for the level of financial literacy selected by investor i; $ECI1_i$ and $ECI2_i$ are $N-1$ dummies for the level of self-assessed experience in complex instruments reported by investor i; ICI_i is another dummy set to one when investor i declares having already traded complex instruments; $LE1_i$ and $LE2_i$ are $N-1$ dummies for the level of education of investor i; $Male_i$ is a dummy set to one when investor i is male; FS_i and ES_i are $N-1$ dummies for the spoken language, with FS_i being set to one when investor i selected French and ES_i being set to one when investor i selected English; $Exec_i$ is a dummy set to one when investor i declared executive responsibilities; Age_i is the age of investor i at the end of 2012; $LOG(MPV_i)$ is the log of investor i's average end-of-month portfolio value computed over the January 2003–December 2007 period; and Exp_i is the number of months when investor i's stock portfolio is not empty over the January 2003–December 2007 period. ***, **, and * indicate significance at 1%, 5%, and 10%, respectively covariates. We ultimately have 10,270 S-investors and 10,270 matched non S-investors (with 4663 duplicates).[19] Building on Fang et al. (2014) and Brogaard et al. (2017), we conduct several diagnostic tests to assess the quality of this matching. This quality is crucial for the validity of the DiD analysis that critically depends on the assumption that the underlying trends in the outcome variable are the same for both treatment and control groups. First, we compare the before-match logit model with the after-match logit model whose results are also reported in Table 3. As expected, when estimating the after-match model, non S-investors and S-investors no longer differ on the large set of explanatory variables. The Pseudo R^2 also drops to zero in the after-match model. For the second diagnostic test, we rely on Table 4 which reports the distribution of estimated propensity scores for the treatment group, control group, and the difference in estimated propensity scores postmatching. Propensity scores of both groups line up closely and the difference is trivial. The final diagnostic test consists in examining the univariate comparisons between the two groups' characteristics and their corresponding *p*-values provided in Table 5. All the *p*-values are well above 10%, meaning there are no statistical differences between the control group and the treatment group.

4.2 DiD regressions

We estimate the following panel data regression model:

$$\begin{aligned} Y_{i,t} =& a_0 + a_1.\text{AFTER}_t + a_2.\text{TREAT}_i + a_3.\text{TREAT} \times \text{AFTER}_{i,t} \\ &+ \delta.X_i + \beta_1.\text{Rm}_t + \beta_2.\text{Crisis}_t + \epsilon_{i,t} \end{aligned} \tag{2}$$

where $Y_{i,t}$ characterizes investor i's behavior in month t; AFTER_t is a dummy variable set to one when month t is in the post-MiFID period; TREAT_i is a dummy variable set to one when investor i completed the *S*-test; and $\text{TREAT} \times \text{AFTER}_{i,t}$ is the interaction of these two dummies (which is equal to 1 for S-investors in the January 2008–March 2012 period). With this specification, the coefficient a_2 captures the pre-existing differences in trading behavior between the two groups of investors, while a_1 captures any aggregate temporal factors that affect trading behavior in the same manner across the two types of investors. Our main variable of interest reflecting the impact of the *S*-test on investors' behavior in the post-MiDID period is then TREAT × AFTER. We add a set of covariates (X_i) in the regression to generate more precise estimates.[20] Importantly, we use the same regressors as those used in PSM (see Sect. 4.1) to generate highly comparable control and treated groups before observing the changes in the outcome variable ($Y_{i,t}$). Finally, we also add two

[19] Since we allow replacement, we actually have 5607 matched non S-investors, some of them being matched to several S-investors.

[20] According to Angrist and Pischke (2008), the inclusion of covariates generates smaller residual variance since "a regression along this point is the result that even in a scenario with no omitted variable bias, the long regression generates more precise estimates of the coefficients on the variables included in the short regression whenever these variables have some predictive power for outcomes because these covariates lead to a smaller residual variance" (Chapter 3, p. 62).

time-varying control variables in Eq. 2, namely, the market return in month t (Rm_t) and a dummy variable for the financial crisis period ($Crisis_t$) set to one for months from January 2008 to June 2009. As a proxy for the market portfolio, we use the STOXX Europe 600 index. When estimating this panel data regression model, we cluster standard errors by month to address potential issues due to cross-sectional correlation (Seasholes and Zhu 2010).

The choice for $Y_{i,t}$, the dependent variable in Eq. 2, is driven by the hypotheses presented in Sect. 2.[21] We first use stock portfolio net returns (Model 1) to check whether S-investors succeed in achieving higher returns, as postulated in H2. Second, we use the size (in number of stocks) of end-of-month portfolios as an indicator of diversification (Model 2). This allows us to check the conjecture made in H3 that S-investors hold better diversified stock portfolios. Next, we consider both the number of stock trades (Model 3) and portfolio turnover (Model 4) to test whether S-investors increase their trading intensity as H4 posits. To check H5 regarding the propensity of S-investors to invest less in lottery-like stocks, we consider the lottery-like stock ratio (Model 5) as defined in Sect. 3.2. Furthermore, to check our conjecture that more sophisticated S-investors benefit the most from financial knowledge acquisition and therefore make better trading decisions than less sophisticated S-investors (see H6), we replicate all the aforementioned models separately on the two subsets of investors: the first one contains more sophisticated investors, i.e., those with a "Dynamic" or an "Aggressive" investment profile, while the other one includes less sophisticated investors, i.e., those with a "Conservative" or a "Neutral" investment profile. The results of all these models are reported in Sect. 5.

5 Empirical results

A key assumption for the DiD estimation to be valid is the parallel trends hypothesis, i.e., the treatment and control groups are supposed to follow parallel trends in the absence of the treatment. To check this assumption, we plot the cross-sectional monthly averages of each of the five outcome variables under scrutiny over the whole pre-MiFID period. Trends are parallel over the last 2 years before treatment (see Figs. 2, 3, 4, 5 and 6 available in the appendix). Other events may have resulted in deviations before 2006 since diverging latent trends are noticeable for some of the outcome variables between 2003 and 2005. We therefore consider the pre-treatment period from January 2006 to December 2007 when estimating our models to ensure

[21] We did not consider the effect of financial knowledge acquisition on the home/local bias because Belgian retail investors tend to be well diversified internationally (e.g., D'Hondt et al. 2020) and we found no theoretical background to hypothesize that the use of the information tool would lead S-investors to exhibit a lower home/local bias. In our sample, only 34% of the total trades are executed on Belgian stocks. Regarding the disposition effect (DE) that is another widespread bias among retail investors, it has been disregarded because it is not possible to calculate a satisfactory DE measure per investor and per month. Indeed, Table 9 (in the appendix) shows that, on average, non S-investors (S-investors) executed a total of 121 (158) trades over the entire sample period, which corresponds to approximately 1.09 (1.42) trade per month. We are grateful to an anonymous reviewer for raising these points.

Table 4 Estimated propensity score distributions

Group	N	Mean	Minimum	Median	Maximum	STD
Treatment	10,270	0.0239	−1.8627	0.0374	1.5963	0.4143
Control	10,270	0.0238	−1.8108	0.0374	1.3401	0.4141
Difference	–	0.0001	−0.0519	0.0000	0.2562	0.0002

This table reports the statistical distribution of estimated propensity scores for the treatment group (i.e., S-investors), control group (i.e., matched non S-investors), and the difference in estimated propensity scores postmatching

that our DiD setting is valid.[22] In Table 6, we provide the DiD test results and the mean DiD estimators for the five outcome variables. The results of the five corresponding specifications of our PSM-DiD regression model including control variables (see Eq. 2) are reported in Table 7. These findings allow us to validate or refute our main hypotheses (H2–H6).

Table 6 shows that, for each outcome variable, the gap between the change observed in the treatment group and the change in the control group is significant at the 1%. This indicates that taking the *S*-test and having access the information tool significantly impact the trading behavior of S-investors. In particular, the mean DiD estimators reveal for S-investors a decrease in portfolio net returns and an increase in portfolio size, number of trades, portfolio turnover, and lottery-like stock ratio. This evidence suggests mixed effects of financial knowledge acquisition on the trading behavior of S-investors in the post-MiFID period.

Panel A of Table 7 reports the results of estimating our PSM-DiD regressions, which mostly confirm the previous findings. When considering net portfolio returns in Model 1, the coefficient estimate of TREAT_i shows that S-investors achieve higher net returns than matched non S-investors. The coefficient estimate of AFTER_t does not significantly differ from zero, indicating that net returns have not really improved over time. However, the coefficient estimate of $\mathrm{TREAT} \times \mathrm{AFTER}_{i,t}$ is negative and significant at the 10% level, meaning that the net returns earned by S-investors slightly deteriorate in the post-MiFID period. This result prevents us from validating H2 and is instead consistent with the empirical evidence provided by Guiso and Jappelli (2006). These authors report a negative relationship between financial knowledge acquisition and returns, which they related to overconfidence. Unsurprisingly, net portfolio returns are strongly positively related to market returns. The coefficient estimate of the crisis dummy is also positive at the 10% level, suggesting that net portfolio returns tend to be slightly higher during the 2008 crisis period.

Regarding Model 2, the coefficient estimate of TREAT_i reveals that the average number of stocks held in the portfolio is significantly higher among S-investors. The coefficient estimate of AFTER_t is also positive at the 1% level, meaning that both

[22] When estimating our DiD regressions using the entire sample period from 2003 to 2012, results reveal overall similar patterns.

Table 5 Differences in variables

	Control group	Treatment group	Difference	p-value
Panel A: Subjective individual variables				
Subjective knowledge of financial markets 1 (FL1)	0.2755	0.2721	0.0034	0.5839
Subjective knowledge of financial markets 2 (FL2)	0.4691	0.4739	−0.0048	0.4934
Subjective knowledge of financial markets 3 (FL3)	0.1488	0.1444	0.0044	0.3747
Subjective experience in complex instruments 1 (ECI1)	0.5011	0.5032	−0.0021	0.7588
Subjective experience in complex instruments 2 (ECI2)	0.2504	0.2485	0.0019	0.7471
Investment in complex instruments (ICI)	0.5917	0.5874	0.0043	0.5326
Panel B: Sociodemographic variables				
Level of education 1 (LE1)	0.2112	0.2060	0.0052	0.3628
Level of education 2 (LE2)	0.7500	0.7516	−0.0016	0.7839
Male	0.9121	0.9137	−0.0016	0.6740
French speaker (FS)	0.3944	0.3982	−0.0038	0.5683
English speaker (ES)	0.0365	0.0358	0.0007	0.7937
Executive (Exec)	0.1827	0.1875	−0.0048	0.3691
Age	52.3040	52.4882	−0.1842	0.3102
Panel C: Trading variables				
Portfolio value (LOG(MPV))	9.7002	9.6835	0.0167	0.4503
Trading experience (Exp)	37.2713	37.3962	−0.1249	0.5604
N	10,270	10,270		

This table reports the univariate comparisons between the treatment and control groups' characteristics and their corresponding p-values. Panel A refers to subjective individual characteristics, Panel B to sociodemographic variables, and Panel C to variables based on actual trading behavior before the completion of the MiFID questionnaires. $FL1_i$, $FL2_i$, and $FL3_i$ are $N-1$ dummies for the level of financial literacy selected by investor i; $ECI1_i$ and $ECI2_i$ are $N-1$ dummies for the level of self-assessed experience in complex instruments reported by investor i; ICI_i is another dummy set to one when investor i declares having already traded complex instruments; $LE1_i$ and $LE2_i$ are $N-1$ dummies for the level of education of investor i; $Male_i$ is a dummy set to one when investor i is male; FS_i and ES_i are $N-1$ dummies for the spoken language, with FS_i being set to one when investor i selected French and ES_i being set to one when investor i selected English; $Exec_i$ is a dummy set to one when investor i declared executive responsibilities; Age_i is the age of investor i at the end of 2012; LOG(MPV_i) is the log of investor i's average end-of-month portfolio value computed over the January 2003–December 2007 period; and Exp_i is the number of months when investor i's stock portfolio is not empty over the January 2003–December 2007 period

S-investors and matched non S-investors have improved their portfolio diversification over time. However, the coefficient estimate of $TREAT \times AFTER_{i,t}$ is positive and significant at the 1% level, which indicates that portfolio diversification has improved more for S-investors. This leads us to validate H3. Regarding the control variables, only the coefficient estimate of the crisis dummy is highly significant and reveals that the 2008 crisis period was detrimental to portfolio diversification.

Table 6 DiD test results

	Treatment group (S-investors)		Control group (Matched non S-investors)			
	Before	After	Before	After	Mean DiD Estimator	p-value
Net return	0.0003	−0.0051	−0.0015	−0.0052	−0.0018	< 0.0001
Portfolio size	6.6650	9.7140	5.1946	6.9436	1.3000	< 0.0001
Number of trades	2.0720	1.4283	1.8269	1.0999	0.0834	< 0.0001
Portfolio turnover	0.3027	0.1693	0.2919	0.1518	0.0067	0.0020
Lottery-like stock ratio	0.0470	0.0598	0.0562	0.0657	0.0032	< 0.0001

This table reports the DiD test results for the five outcome variables under scrutiny: investor portfolio net return, portfolio size, number of stock trades, portfolio turnover, and lottery-like stock ratio. Portfolio returns are computed using an approximation of the modified Dietz method as follows: $R_t = \frac{\mathrm{EMV}_t - \mathrm{EMV}_{t-1} - P_t + S_t}{\mathrm{EMV}_{t-1} + P_t}$ where R_t is the portfolio gross return for month t, EMV_t is the end-of-month portfolio value in month t, and P_t and S_t are the aggregate monetary value of all purchases and sales executed during month t. For net returns, transaction costs paid during month t are subtracted from the aggregate monetary value (in the numerator). Portfolio size is defined as the number of stocks in the end-of-month portfolio. Number of trades refers to the total number of trades executed on stocks per month. Portfolio turnover is the aggregate monetary value of all purchases and sales executed over the month divided by the average of the portfolio values at the beginning and the end of this particular month. Lottery-like stock ratio is defined as the number of lottery-like stocks held in the end-of-month portfolio divided by the total number of stocks in the end-of-month portfolio. Using daily returns over the last 6 months, the stocks identified as lottery-like are those in the lowest 50$^{\text{th}}$ stock price percentile, the highest 50$^{\text{th}}$ total volatility percentile, and the highest 50$^{\text{th}}$ total skewness percentile. "Before" refers to the pre-MiFID period (i.e., January 2006-December 2007) while "After" refers to the post-MiFID period (i.e., January 2008-March 2012). Both net returns and portfolio turnover are winsorized with a 1% cutoff

Both Model 3 and Model 4 deliver insights related to H4. The results of Model 3 show that S-investors trade significantly more than their matched counterparts (i.e., the coefficient estimate of TREAT_i is positive and significant at the 1% level). Interestingly, while the coefficient estimate of AFTER_t is significantly negative, the one of $\mathrm{TREAT} \times \mathrm{AFTER}_{i,t}$ is positive and strongly significant. Although trading activity (in number of trades) has decreased since MiFID enforcement, S-investors have been trading more than their matched counterparts. Such results suggest that taking the *S*-test (and acquiring financial knowledge) leads to an increase in trading intensity because of overconfidence as posited in H4. Regarding the control variables, only the crisis dummy exhibits a significant coefficient estimate (at the 5% level), revealing that trading activity was higher during the 2008 financial crisis.

Portfolio turnover in Model 4 offers a more comprehensive picture of trading intensity. Like in Model 3, the coefficient estimate of TREAT_i is positive and significant at the 1% level. This means that portfolio turnover is significantly higher among S-investors than among their matched counterparts. Consistent with what is observed for the number of trades, portfolio turnover has significantly

Table 7 PSM-DiD results

	Net return (Model 1)	Portfolio size (Model 2)	Number of trades (Model 3)	Portfolio turnover (Model 4)	Lottery-like stock ratio (Model 5)
Panel A: All S-investors and matched non S-investors					
Constant	−0.0158***	−23.9636***	−5.6000***	0.1699***	0.0718***
$AFTER_t$	0.0032	2.1007***	−0.8420***	−0.1621***	−0.0004
$TREAT_i$	0.0017***	1.5081***	0.2577***	0.0113***	−0.0090***
$TREAT \times AFTER_{i,t}$	−0.0018*	1.3000***	0.0833***	0.0067**	0.0032**
Rm_t	0.9467***	−0.3352	−1.3108	−0.0748	0.1145
$Crisis_t$	0.0113*	−1.0078***	0.2820**	0.0598***	0.0320*
X_i					
N	1,472,655	1,540,500	1,540,500	1,540,500	1,540,500
R^2	0.2337	0.1883	0.0453	0.0157	0.0105
Panel B: More sophisticated S-investors and their matched non S-investors					
Constant	−0.0151***	−24.7562***	−5.6906***	0.2008***	0.0735***
$AFTER_t$	0.0029	2.1748***	−0.8327***	−0.1629***	−0.0004
$TREAT_i$	0.0016**	1.6290***	0.1865***	−0.0030	−0.0093***
$TREAT \times AFTER_{i,t}$	−0.0014	1.5093***	0.1473***	0.0171***	0.0027*
Rm_t	0.9445***	−0.3588	−1.4219	−0.0799	0.1127
$Crisis_t$	0.0113*	−1.0761***	0.2630**	0.0577***	0.0325**
X_i					
N	1,196,849	1,251,900	1,251,900	1,251,900	1,251,900
R^2	0.2351	0.1873	0.0450	0.0147	0.0108
Panel C: Less sophisticated S-investors and their matched non S-investors					
Constant	−0.0176***	−19.6240***	−5.1643***	0.0543***	0.0702***
$AFTER_t$	0.0042	1.7791***	−0.8825***	−0.1587***	−0.0002
$TREAT_i$	0.0022**	0.7861***	0.5744***	0.0807***	−0.0080***
$TREAT \times AFTER_{i,t}$	−0.0032**	0.3923***	−0.1941***	−0.0383***	0.0054***
Rm_t	0.9561***	−0.2330	−0.8289	−0.0528	0.1227
$Crisis_t$	0.0114*	−0.7112***	0.3644***	0.0687***	0.0300*
X_i					
N	275,806	288,600	288,600	288,600	288,600
R^2	0.2281	0.2039	0.0492	0.0223	0.0104

This table reports the results for Eq.(2) where the dependent variable ($Y_{i,t}$) is investor i's portfolio net return (Model 1), portfolio size (Model 2), number of stock trades (Model 3), portfolio turnover (Model 4), or lottery-like stock ratio (Model 5) in month t. Portfolio returns are computed using an approximation of the modified Dietz method as follows: $R_t = \frac{\mathrm{EMV}_t - \mathrm{EMV}_{t-1} - P_t + S_t}{\mathrm{EMV}_{t-1} + P_t}$ where R_t is the portfolio gross return for month t, EMV_t is the end-of-month portfolio value in month t, and P_t and S_t are the aggregate monetary value of all purchases and sales executed during month t. For net returns, transaction costs paid during month t are subtracted from the aggregate monetary value (in the numerator). Portfolio size is defined as the number of stocks in the end-of-month portfolio. Number of trades refers to the total number of trades executed on stocks per month. Portfolio turnover is the aggregate monetary value of all purchases and sales executed over the month divided by the average of the portfolio values at the beginning and the end of this particular month. Lottery-like stock ratio is defined as the number of lottery-like stocks held in the end-of-month portfolio divided by the total number of stocks in the end-of-month port-

Table 7 (continued)

folio. Using daily returns over the last 6 months, the stocks identified as lottery-like are those in the lowest 50^{th} stock price percentile, the highest 50^{th} total volatility percentile, and the highest 50^{th} total skewness percentile. In the set of regressors, AFTER_t is a dummy variable set to one when month t is in the post-MiFID period, TREAT_i is a dummy variable set to one when investor i completed the *S*-test, and $\text{TREAT} \times \text{AFTER}_{i,t}$ is the interaction of these two dummies, Rm_t is the market return in month t, and Crisis_t is a dummy variable for the financial crisis period set to one for months from January 2008 to June 2009. X_i is the set of covariates used for PSM (see Eq. 1). N gives the number of observations. ***, **, and * indicate significance at 1%, 5%, and 10%, respectively. Both net returns and portfolio turnover are winsorized with a 1% cutoff. Standard errors are clustered by month

increased for S-investors in the post-MiFID period, despite an overall decrease of it over time (as revealed by the coefficient estimate of AFTER_i). This finding further validates H4 regarding reinforcement in beliefs and/or ongoing access to financial information bolstering overconfidence. Such a result is in line with the overconfident online investors in Barber and Odean (2001b) and Barber and Odean (2002) or with the framework of automated advice used by Hackethal et al. (2012). Also consistent with what is found in Model 3, the coefficient estimate of the crisis dummy shows that portfolio turnover is significantly higher during the crisis period. Retail investors traded more intensively during that highly volatile period, which is consistent with the findings of Hoffmann et al. (2013) showing that individual investors continue to trade actively during the 2008 global financial crisis and do not de-risk their investment portfolios.

The last model in Table 7 (Model 5) provides the results about the investors' propensity to hold lottery-like stocks. The coefficient estimate of TREAT_i is negative at the 1% level, meaning that S-investors display a lower lottery-like stock ratio than their matched counterparts. However, the coefficient estimate of $\text{TREAT} \times \text{AFTER}_{i,t}$ indicates that their propensity to hold lottery-like stocks in their portfolio has significantly increased since MiFID enforcement. This result does not provide support for H5. On the contrary, although S-investors are less likely to invest in lottery-like stocks than their matched counterparts, the information tool has deteriorated their behavior on that particular aspect (i.e., the proportion of lottery-like stocks in their portfolio is higher). It is also noteworthy that the crisis dummy exhibits a positive coefficient estimate. Although its significance is weak, this finding is consistent with Kumar (2009), who reports that the trading activity on lottery-like stocks tends to increase during crisis periods.

To check our last hypothesis (H6) that more sophisticated S-investors are expected to benefit the most from the information tool, we need to turn to Panels B and C in Table 7 reporting the results of estimating our PSM-DiD regressions on the subset of more sophisticated investors (Panel B) and the subset of less sophisticated investors (Panel C).[23] Both provide additional insights into how S-investors changed their behavior depending on their investment profile. In Model 1, both coefficient

[23] Univariate comparisons across S-investors depending on the investment profile are available in Table 14 in the appendix.

estimates of $\text{TREAT} \times \text{AFTER}_{i,t}$ are negative: −0.0014 but not significant at any conventional levels in Panel B and −0.0032 significant at the 5% level in Panel C. This reveals that the decrease in net stock portfolio returns in the post-MiFID period is significantly lower only for less sophisticated S-investors, which provides support for H6. In Model 2, both coefficient estimates of $\text{TREAT} \times \text{AFTER}_{i,t}$ are positive (i.e., 1.5093 and 0.3923, respectively) and significant at the 1%. The higher marginal effect in Panel B indicates however that the increase in portfolio diversification in the post-MiFID period looks higher among more sophisticated S-investors. Such a result also provides support for H6.

Interestingly, Model 3 reveals that the number of trades has significantly increased for more sophisticated S-investors, while it has significantly decreased for less sophisticated investors. In both cases, the significance level is at 1%. Using portfolio turnover as the dependent variable, Model 4 tells the same story since it displays a positive coefficient estimate for $\text{TREAT} \times \text{AFTER}_{i,t}$ in Panel B (i.e., 0.0171 significant at the 1% level) but a negative coefficient estimate in Panel C (i.e., −0.0383 significant at the 1% level). This reveals that portfolio turnover has increased (decreased) among more (less) sophisticated S-investors, which does not provide support for H6. By contrast, we observe a distinctive phenomenon in Model 5, for which we find positive and significant coefficient estimates for $\text{TREAT} \times \text{AFTER}_{i,t}$ in both panels (i.e., 0.0027 significant at the 10% level in Panel B and 0.0054 significant at the 5% level in Panel C). Nevertheless, the higher marginal effect in Panel C indicates that the propensity to hold lottery-like stocks has significantly increased more among less sophisticated S-investors than among more sophisticated S-investors. This finding would rather provide support for H6.

Overall, the results in Panels B and C provide mixed evidence inconsistent with more sophisticated S-investors benefiting the most from the information tool. Although these investors do not display a decrease in stock portfolio net returns, exhibit a higher increase in diversification and a lower increase in the propensity to hold lottery-like stocks, they also display a significant increase in trading activity (measured in number of trades or in portfolio turnover). Such conflicting findings do not validate H6. However, they at least show that the investment profile does matter to explain heterogeneity in behavior across S-investors.

6 Robustness checks

Since we do not have the accurate date at which each S-investor took the *S*-test (and was consequently granted access to the information tool), we assume that investors did it in early 2008, i.e., before trading in the post-MiFID period. Although this assumption looks reasonable because investors were all already active before January 2008, we conduct a first robustness check and replicate our main analysis by dropping either the first six months of the year 2008 or the entire year 2008. The results available in Table 10 in the appendix are overall similar to the ones reported above.

As a second robustness check, we replicate our main analysis on two alternative samples to examine whether our findings depend on sample selection. We consider both a less restrictive filter, i.e., a holding period of at least 6 months for stock portfolios in both the pre- and post-MiFID periods, and a more restrictive filter, i.e., a holding period of at least 24 months for stock portfolios in both the pre- and post-MiFID periods. The results are available in the appendix, in Tables 11 and 12 for PSM and in Table 13 for the PSM-DiD regressions. Our main findings are confirmed on the two alternative samples.

7 Discussion

The *S*-test is used in this paper as a proxy for the investor's appetite for financial knowledge acquisition, and the investors who decide to take this test, that is, S-investors, are expected to acquire financial knowledge, mainly through ongoing financial information available in the online tool. However, our empirical work has some limitations because we do not know whether S-investors use the information tool at their disposal or how they use it. Despite this shortcoming, our empirical findings show that (1) the demand for the information tool is positively related to both financial literacy and education in our broad sample, and (2) they are significant changes in the behavior of S-investors. Such results showing significant shifts for S-investors suggest that these investors did actually acquire financial knowledge. Indeed, if S-investors had not used the information tool or had not acquired financial knowledge, no significant shifts in their behavior would have been observed and no difference between treatment and control groups would have been identified.

Nevertheless, we cannot rule out that some investors possibly chose to take the *S*-test for purposes other than financial knowledge acquisition. For instance, investors could have taken the *S*-test to join a kind of "privileged customer" category of the brokerage house. Kreis and Mafael (2014) show that providing customers with loyalty programs can be an effective tool since these programs are not merely something that adds to the value of a product or service but also something that creates value per se. Kang et al. (2015) and Brashear-Alejandro et al. (2016) also show that loyalty programs are important customer relationship management tools. In line with this idea, we argue that the brokerage house that provided us with the data has established the information tool as its own specialized portal, and according to Claessens et al. (2002), portals are recent delivery channels for financial services that attempt to process and personalize information to capture consumers.

Another alternative is that some investors considered taking the *S*-test as a default option since the access to the information tool is free of charge and the additional time required to complete the online survey is not particularly long. New technology adopters are also more likely to make that choice. In this regard, Lee et al. (2002) conclude that positive perceptions of technology-based service innovations by consumers are crucial to computer banking adoption behavior. More broadly, such motivations could derive from a strong advantage of online recommendations compared to more traditional human advice. As mentioned in Horn et al. (2020), robo-advice provides increased flexibility regarding the time of the consultation and lower consultation costs. The online information tool under scrutiny in this paper is not a robo-advisor (as explained in the introduction), but it entails the aforementioned advantages that are likely to be attractive to retail investors.[24]

The consumers' perception of the legal protection provided by financial services might also play a role in the decision to take the *S*-test. This idea is developed in Georgarakos and Inderst (2014), who report evidence that financial advice matters differently to two types of households. For households whose perceived own financial capability is low, trust in financial advice is important. By contrast, what matters the most to more educated households is the perception of their legal rights as consumers of financial services. Regardless of the true motivation behind the decision to take the *S*-test, it should be associated with a positive perceived value among customers. Hence, if retail investors associate value with the *S*-test, they are likely to consider it an investment and should consequently expect to earn some benefits from it.

Another limitation that cannot be ruled out is that S-investors may also acquire financial knowledge without using the information tool. For instance, they might rely on other or complementary sources of information. Such a phenomenon would actually not contradict our main findings. On the contrary, it would reinforce our results showing that financial literacy, the demand for advice and financial knowledge acquisition are complementary.

Regarding the investors who disregarded the *S*-test (i.e., non S-investors), we ignore whether their decision to forgo the *S*-test is because they do not need financial information or because they already have access to other sources of financial information. At first glance, if non S-investors had an appetite for financial knowledge acquisition in general, they would be very unlikely to neglect the free access to an information tool that is easily available on their broker's trading platform. Another reason for non S-investors to reject this information tool might be that although they have a need for financial knowledge, they prefer human advice (i.e., provided by a human financial advisor). Finally, since the information tool does not provide personalized advice, non S-investors might simply not be concerned about general information on stocks, analysts' reports, and professional recommendations.

8 Conclusion

In this paper, we use rich field data to examine whether retail investors who asked for access to an online information tool (and are supposed to have acquired financial knowledge) differ in their trading behavior from those who neglected it on purpose. This tool is not a robo-advisor, but it is similar to one. On the one side, investors

[24] Recently, Oehler et al. (2021) used a survey to identify the determinants of the use of robo-advisors. These authors show that, aside from classical determinants (such as the willingness to take risk), personal characteristics (such as locus of control) influence this decision. This finding is consistent with those reported in Tauni et al. (2015) showing that investors' personality traits (such as extraversion, conscientiousness and openness) moderate the relationship between financial knowledge acquisition and asset allocation decisions.

who are interested in this information tool have to complete a profile survey (i.e., the MiFID *S*-test). They are subsequently assigned a given investment profile. On the other side, the online tool provides investors with general information on stocks, analysts' reports, and professional recommendations, but it does not deliver personalized advice (as a traditional financial advisor would do). Taking the *S*-test arguably reveals the investor's appetite for financial knowledge acquisition.

To investigate the impact of the information tool (and financial knowledge acquisition) on investor trading behavior, we mainly rely on the theoretical model of Lusardi et al. (2017) with endogenous financial information acquisition to formulate several hypotheses, and we adopt a PSM-DiD approach. First, we use propensity score matching to pair investors with access to the information tool (i.e., S-investors) with comparable counterparts who neglected it on purpose (i.e., matched non S-investors). This matching is based on a large set of covariates including variables characterizing investor trading behavior before having access to the information tool. The results show that subjective financial literacy and education are both positively related to financial knowledge acquisition, as suggested in Lusardi et al. (2017)'s model. We also find that male investors, older investors, French-speaking investors, investors who declare executive responsibilities, and wealthier investors are more likely to take the *S*-test to have access to the information tool. By contrast, investors with longer trading experience are less likely to do so.

Next, we run PSM-DiD regressions on a homogeneous subset of 10,270 S-investors and 10,270 matched non S-investors. We provide evidence that financial knowledge acquisition has mixed effects on retail trading behavior: better portfolio diversification but higher trading intensity and lower net returns. However, when discriminating between investment profiles to proxy for investor sophistication, we do not find that more sophisticated investors benefited the most from the information tool: although these investors do not display a decrease in stock portfolio net returns, exhibit a higher increase in diversification and a lower increase in the propensity to hold lottery-like stocks, they also display a higher increase in both number of trades and portfolio turnover. Such findings reveal however that the investment profile does matter to explain heterogeneity in behavior across S-investors.

To a broader extent, our results suggest that the trading behavior of investors is consistent with their attitude toward financial knowledge acquisition in the MiFID tests. Indeed, MiFID questionnaires can help identify investors with an inclination toward financial knowledge acquisition, which is useful for financial practitioners. In the same vein, building on the harmonized typology of financial advisors introduced by MiFID II (2014/65/EU), Migliavacca (2020) shows that financial advisors who are independent have an educational role for bank retail clients. Specifically, she finds that the lower the selling incentive of financial advisors, the higher their ability to improve the financial literacy of their clients. Interestingly, this effect is also driven by retail clients' "willingness to learn", that is, their self-assessment of some personal interest toward economics and finance. Since this willingness to learn appears very close to the appetite for financial information (proxied by taking the *S*-test in this paper), our results add to the stream of papers showing the usefulness of MiFID questionnaires (e.g., Mazzoli and Marinelli 2014; D'Hondt and Roger 2017; Broihanne and Orkut 2018; Bellofatto et al. 2018; D'Hondt et al. 2021a;

Broihanne and Orkut 2021; Orkut 2021). Such questionnaires that are at the core of the *Know Your Customer* process for investment firms provide reliable information about the actual behavior of retail investors. Such data are highly informative and therefore valuable for investment firms, researchers, and regulators.

Despite a few limitations discussed in Sect. 7, this paper paves the way for further research on endogenous financial knowledge acquisition and trading behavior. Future research might collect field data that allow researchers to test whether the demand for human versus online financial advice differs across retail investors. Another promising avenue for future research would be to use data enabling any measurement of the extent to which investors use the information tool at their disposal (e.g., connection time, time spent in the tool). Any measurement of how they actually use it (e.g., do investors look for information about the stocks they already own, or do they look for investment opportunities?) would also be of great interest.

Appendix

See Figs. 1, 2, 3, 4, 5 and 6 and Tables 8, 9, 10, 11, 12, 13 and 14

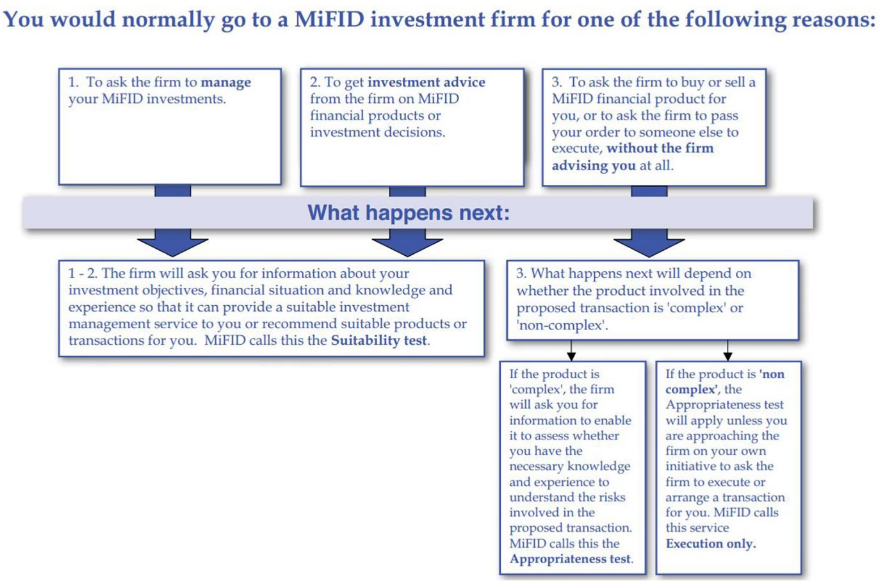

Fig. 1 MiFID services. This figure exhibits the three types of services in the MiFID regulation and information retail investors need to provide accordingly. Source: Committee of European Securities Regulators (2008)

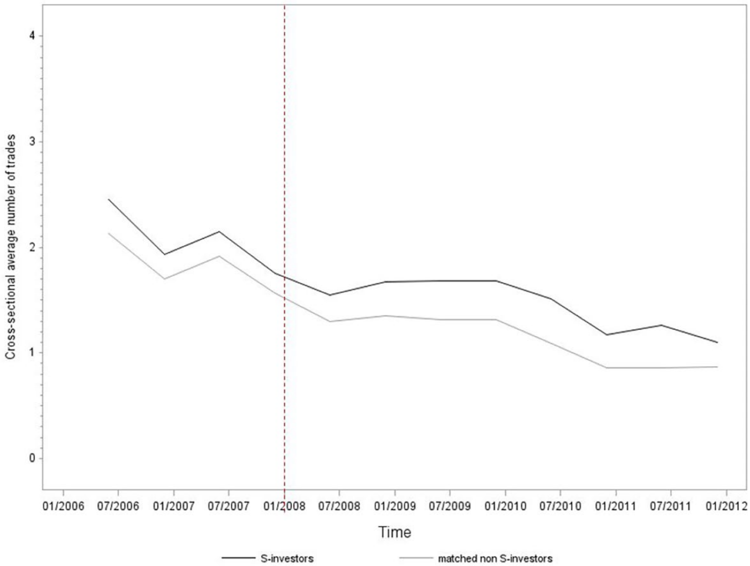

Fig. 2 Monthly average number of trades over the period 2006–2011. This figure displays the cross-sectional monthly average number of trades per investor over the 6-year period from 2006 to 2011 (using half-yearly observations). The black line refers to S-investors while the gray line refers to matched non S-investors

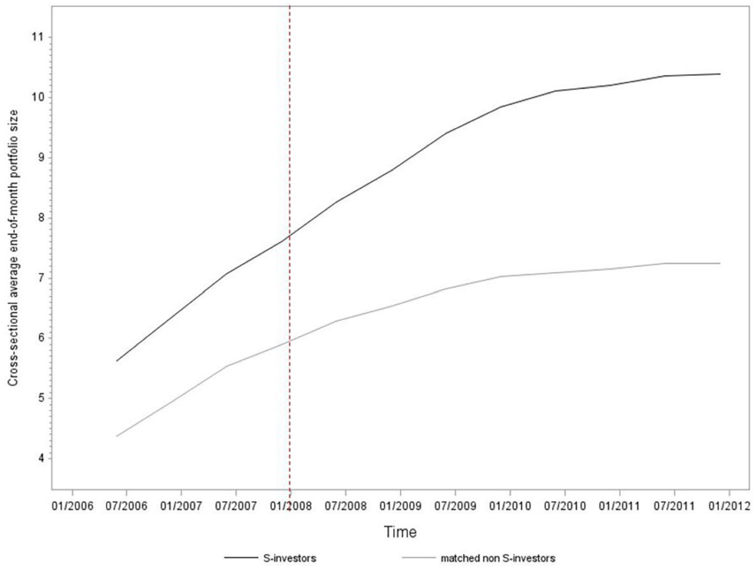

Fig. 3 Monthly average portfolio size over the period 2006–2011. This figure displays the cross-sectional monthly average investor end-of-month portfolio size (in number of stocks) over the 6-year period from 2006 to 2011 (using half-yearly observations). The black line refers to S-investors while the gray line refers to matched non S-investors

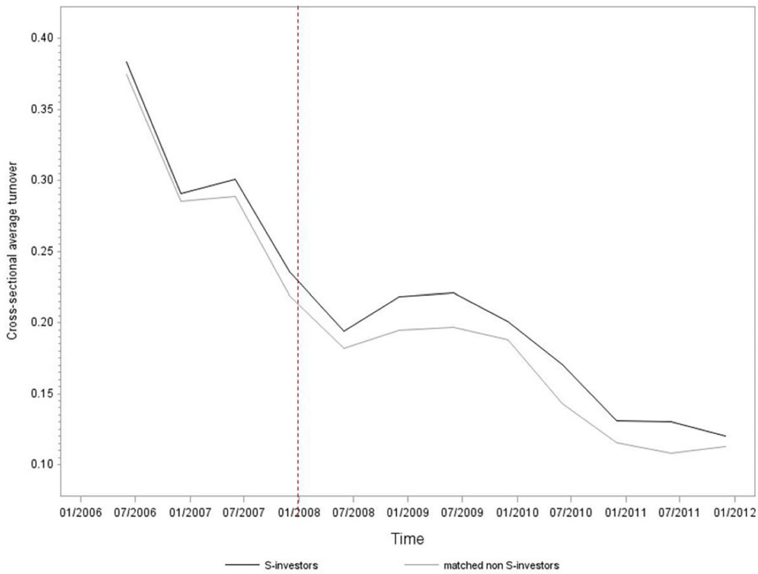

Fig. 4 Monthly average portfolio turnover over the period 2006–2011. This figure displays the cross-sectional monthly average investor portfolio turnover over the 6-year period from 2006 to 2011 (using half-yearly observations). The black line refers to S-investors while the gray line refers to matched non S-investors

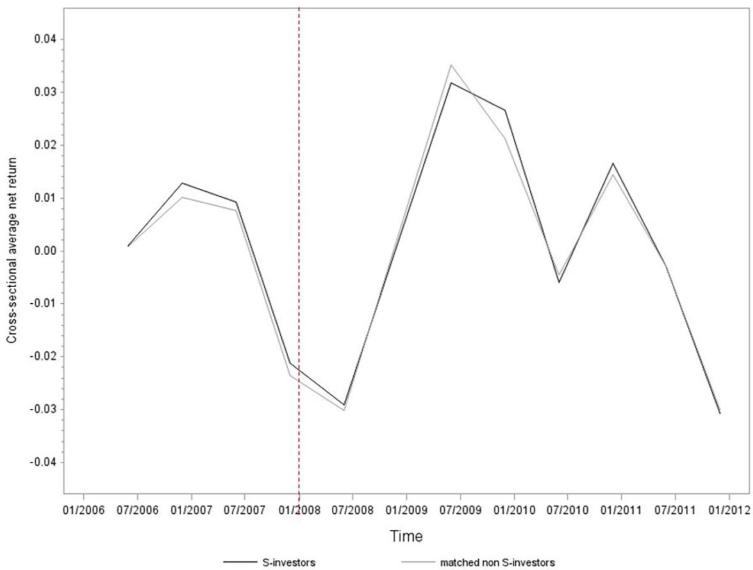

Fig. 5 Monthly average net return over the period 2006–2011. This figure displays the cross-sectional monthly average investor net return over the 6-year period from 2006 to 2011 (using half-yearly observations). The black line refers to S-investors while the gray line refers to matched non S-investors

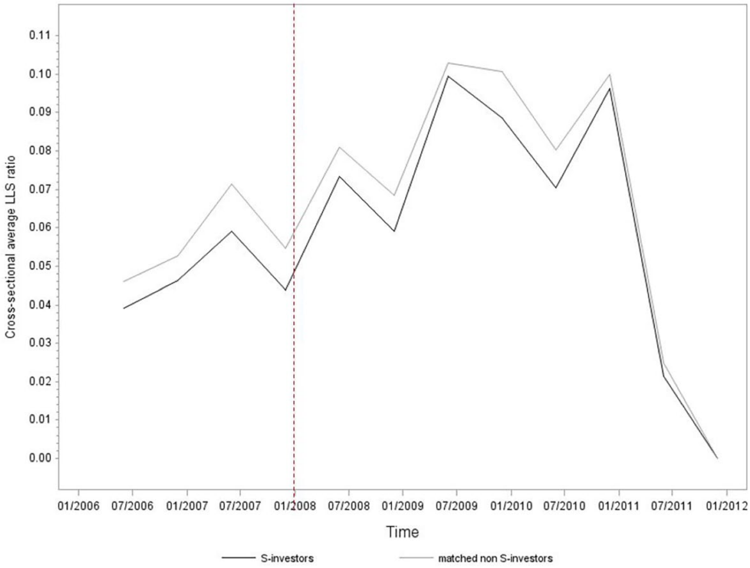

Fig. 6 Monthly average lottery-like stock ratio over the period 2006–2011. This figure displays the cross-sectional monthly average investor lottery-like stock ratio over the 6-year period from 2006 to 2011 (using half-yearly observations). The black line refers to S-investors while the gray line refers to matched non S-investors

Table 8 Items from the MiFID Suitability questionnaire – Examples

(Return objective) Taking into account the current interest rates, what is the yield you expect of your investments?	
Level 1	A yield without any risk of capital loss
Level 2	The inflation rate increased with 2 to 4% per year
Level 3	The inflation rate increased with 5 to 7% per year
Level 4	The inflation rate increased with 8 to 12% per year
Level 5	The inflation rate increased with more than 12% per year
(Financial literacy) What is your knowledge of the financial markets?	
Level 1	I know very little about it and I am not really interested in it.
Level 2	I am not familiar with investments, but I am interested in it.
Level 3	I have sufficient experience to acknowledge the importance of risk diversification.
Level 4	I have a good knowledge of the financial markets.
	I am aware that the financial markets can strongly fluctuate, that sector and asset categories have different characteristics regarding revenue, growth and risk profile.
Level 5	I consider myself as an experienced investor who thoroughly masters all the aspects of the financial markets.
(Risk tolerance) What would you do if you noticed that 6 months after having invested, this investment (in line with the market situation), shows a drop of 20%?	

Table 8 (continued)

Level 1	A total disaster! The security of my investments is my most important criterion. I do not want to take any risks!
Level 2	You will sell at a loss and transfer the money in safer investments.
Level 3	You are anxious but will not move until the situation improves.
Level 4	This was a calculated risk. You leave your investments as they are and will wait until the situation improves.
Level 5	You take advantage of the market situation (low prices) to invest even more.
	By doing so you reduce your average buy price.

This table reports the details on some questions in the MiFID Suitability test. These questions aim to elicit the investor's return objective, financial literacy, and risk tolerance

Table 9 Descriptive statistics for the entire sample period

	Non S-investors				S-investors			
	Mean	Median	Q1	Q3	Mean	Median	Q1	Q3
Panel A: Trade-based variables								
Number of trades	121	41	15	114	158	70	29	165
Trading volume on stocks (in €)	1,157,044	156,948	37,788	613,485	1,343,778	252,523	81,075	844,679
Trading experience (in months)	91	95	76	106	90	91	75	105
Panel B: Stock portfolio-based variables								
Portfolio size (in # stocks)	5	3	2	6	8	6	3	10
Portfolio value (in €)	60,004	11,904	4096	34,929	78,139	20,130	7275	53,192
Portfolio turnover	15.68	0.16	0.08	0.36	6.68	0.17	0.10	0.36
Lottery-like stock ratio (%)	6.82	5.40	2.72	9.28	5.89	4.79	2.81	7.61
Gross return (%)	−0.63	−0.40	−1.25	0.21	−0.39	−0.25	−0.86	0.24
Net return (%)	−0.78	−0.52	−1.42	0.09	−0.55	−0.37	−1.02	0.12
Volatility	0.25	0.09	0.07	0.15	0.17	0.08	0.06	0.12

This table reports the cross-sectional mean, median, and lower and upper quartiles for trade-based variables in Panel A and stock portfolio-based variables in Panel B. They are computed on a per investor basis over the January 2003–March 2012 period, with a distinction between non S-investors and S-investors. The sample counts 10,963 non S-investors (who only took the *A*-test) and 10,270 S-investors (who completed both the *A*-test and the *S*-test). 'Number of trades' is the total number of trades executed on stocks. 'Trading volume on stocks' is the total monetary trading volume (purchases plus sales). 'Trading experience' refers to the number of months between the last and first trade. 'Portfolio size' is the average number of stocks in the end-of-month portfolios. 'Portfolio value' is the average end-of-month portfolio value. 'Portfolio turnover' refers to monthly portfolio turnover, calculated as the sum of all the purchases and sales in a particular month, divided by the average of the portfolio values at the beginning and the end of this particular month. 'Lottery-like stock ratio' is defined as the number of lottery-like stocks held in the end-of-month portfolio divided by the total number of stocks in the end-of-month portfolio. Using daily returns over the last 6 months, the stocks identified as lottery-like are those in the lowest 50th stock price percentile, the highest 50th total volatility percentile, and the highest 50th total skewness percentile. 'Gross return' is the geometric average of monthly portfolio gross returns. 'Net return' is the geometric average of monthly portfolio net returns. 'Volatility' is the standard deviation of monthly portfolio gross returns

Table 10 PSM-DiD results for robustness check 1

	Net return	Portfolio size	Number of trades	Portfolio turnover	Lottery-like stock ratio
	(Model 1)	(Model 2)	(Model 3)	(Model 4)	(Model 5)
Panel A: Without the first six months of the year 2008					
Constant	−0.0150***	−23.9921***	−5.6130***	0.1685***	0.0707***
$AFTER_t$	0.0032	2.0490***	−0.8474***	−0.1625***	−0.0005
$TREAT_i$	0.0017**	1.5077***	0.2575***	0.0114***	−0.0090***
$TREAT \times AFTER_{i,t}$	−0.0019**	1.4043***	0.0935***	0.0074**	0.0035**
Rm_t	0.9545***	−0.6498	−1.0761	−0.0318	0.1008
$Crisis_t$	0.0126	−0.8123***	0.3224**	0.0681***	0.0330
X_i					
N	1,352,881	1,417,260	1,417,260	1,417,260	1,417,260
R^2	0.2272	0.1873	0.0454	0.0167	0.0094
Panel B: Without the entire year 2008					
Constant	−0.0128**	−24.0244***	−5.6574***	0.1626***	0.0722***
$AFTER_t$	0.0032	2.0020***	−0.8503***	−0.1624***	−0.0008
$TREAT_i$	0.0017**	1.5071***	0.2574***	0.0114***	−0.0090***
$TREAT \times AFTER_{i,t}$	−0.0015	1.5008***	0.0960***	0.0067**	0.0041**
Rm_t	0.8371***	−1.4988	0.0254	0.1260	0.1248
$Crisis_t$	0.0298***	−0.5647***	0.3518**	0.0702***	0.0482**
X_i					
N	1,232,100	1,294,020	1,294,020	1,294,020	1,294,020
R^2	0.1810	0.1872	0.0473	0.0184	0.0132

This table reports the results for Eq. (2) where the dependent variable ($Y_{i,t}$) is investor i's portfolio net return (Model 1), portfolio size (Model 2), number of stock trades (Model 3), portfolio turnover (Model 4), or lottery-like stock ratio (Model 5) in month t. Portfolio returns are computed using an approximation of the modified Dietz method as follows: $R_t = \frac{EMV_t - EMV_{t-1} - P_t + S_t}{EMV_{t-1} + P_t}$ where R_t is the portfolio gross return for month t, EMV_t is the end-of-month portfolio value in month t, and P_t and S_t are the aggregate monetary value of all purchases and sales executed during month t. For net returns, transaction costs paid during month t are subtracted from the aggregate monetary value (in the numerator). Portfolio size is defined as the number of stocks in the end-of-month portfolio. Number of trades refers to the total number of trades executed on stocks per month. Portfolio turnover is the aggregate monetary value of all purchases and sales executed over the month divided by the average of the portfolio values at the beginning and the end of this particular month. Lottery-like stock ratio is defined as the number of lottery-like stocks held in the end-of-month portfolio divided by the total number of stocks in the end-of-month portfolio. Using daily returns over the last 6 months, the stocks identified as lottery-like are those in the lowest 50th stock price percentile, the highest 50th total volatility percentile, and the highest 50th total skewness percentile. In the set of regressors, $AFTER_t$ is a dummy variable set to one when month t is in the post-MiFID period, $TREAT_i$ is a dummy variable set to one when investor i completed the S-test, and $TREAT \times AFTER_{i,t}$ is the interaction of these two dummies, Rm_t is the market return in month t, and $Crisis_t$ is a dummy variable for the financial crisis period set to one for months from January 2008 to June 2009. X_i is the set of covariates used for PSM (see Eq. 1). N gives the number of observations. ***, **, and * indicate significance at 1%, 5%, and 10%, respectively. Both net returns and portfolio turnover are winsorized with a 1% cutoff. Standard errors are clustered by month

Table 11 Results of the logit model for propensity score matching – sample based on a filter of 6 months of nonempty portfolios before and after January 1, 2008

Independent variables	Before-match Estimates	After-match Estimates
Intercept	−2.555***	0.068
Panel A: Subjective individual variables		
Subjective knowledge of financial markets 1 (FL1)	0.249***	−0.020
Subjective knowledge of financial markets 2 (FL2)	0.505***	−0.028
Subjective knowledge of financial markets 3 (FL3)	0.202***	−0.098
Subjective experience in complex instruments 1 (ECI1)	0.030	0.023
Subjective experience in complex instruments 2 (ECI2)	−0.204***	−0.020
Investment in complex instruments (ICI)	0.200***	0.012
Panel B: Sociodemographic variables		
Level of education 1 (LE1)	0.184***	−0.165**
Level of education 2 (LE2)	0.236***	−0.130*
Male	0.579***	−0.015
French speaker (FS)	0.075***	−0.006
English speaker (ES)	−0.430***	−0.005
Executive (Exec)	0.116***	0.032
Age	0.009***	−0.001
Panel C: Trading variables		
Portfolio value (LOG(MPV))	0.124***	0.010
Trading experience (Exp)	−0.007***	0.002
Pseudo R^2	0.041	0.000
N	23,912	22,972
S-investors	11,486	11,486
Non S-investors	12,426	11,486

This table reports the parameter estimates for a logit model wherein the dependent variable, Y_i, is a binary variable that equals 1 if investor i completed the S-test and 0 otherwise. Based on Eq. 1, the results are reported before-match (column 2) and after-match (column 3). The set of independent variables consists of an intercept, an initial set of subjective individual characteristics (in Panel A), a second set of sociodemographic variables (in Panel B), and a third set composed of two additional variables intended to control for actual trading behavior before the completion of the MiFID questionnaires (in Panel C). $FL1_i$, $FL2_i$, and $FL3_i$ are $N-1$ dummies for the level of financial literacy selected by investor i; $ECI1_i$ and $ECI2_i$ are $N-1$ dummies for the level of self-assessed experience in complex instruments reported by investor i; ICI_i is another dummy set to one when investor i declares having already traded complex instruments; $LE1_i$ and $LE2_i$ are $N-1$ dummies for the level of education of investor i; $Male_i$ is a dummy set to one when investor i is male; FS_i and ES_i are $N-1$ dummies for the spoken language, with FS_i being set to one when investor i selected French and ES_i being set to one when investor i selected English; $Exec_i$ is a dummy set to one when investor i declared executive responsibilities; Age_i is the age of investor i at the end of 2012; $LOG(MPV_i)$ is the log of investor i's average end-of-month portfolio value computed over the January 2003–December 2007 period; and Exp_i is the number of months when investor i's stock portfolio is not empty over the January 2003–December 2007 period. ***, **, and * indicate significance at 1%, 5%, and 10%, respectively

Table 12 Results of the logit model for propensity score matching – sample based on a filter of 24 months of nonempty portfolios before and after January 1, 2008

Independent variables	Before-match Estimates	After-match Estimates
Intercept	−2.661***	0.111
Panel A: Subjective individual variables		
Subjective knowledge of financial markets 1 (FL1)	0.245***	0.020
Subjective knowledge of financial markets 2 (FL2)	0.516***	0.019
Subjective knowledge of financial markets 3 (FL3)	0.224***	0.046
Subjective experience in complex instruments 1 (ECI1)	0.019	−0.002
Subjective experience in complex instruments 2 (ECI2)	−0.179***	−0.053
Investment in complex instruments (ICI)	0.187***	0.035
Panel B: Sociodemographic variables		
Level of education 1 (LE1)	0.174**	0.105
Level of education 2 (LE2)	0.245***	−0.006
Male	0.580***	−0.003
French speaker (FS)	0.044	−0.002
English speaker (ES)	−0.440***	−0.035
Executive (Exec)	0.168***	−0.002
Age	0.009***	0.000
Panel C: Trading variables		
Portfolio value (LOG(MPV))	0.144***	−0.008
Trading experience (Exp)	−0.009***	−0.002
Pseudo R^2	0.045	0.000
N	16,005	15,314
S-investors	7657	7657
Non S-investors	8348	7657

This table reports the parameter estimates for a logit model wherein the dependent variable, Y_i, is a binary variable that equals 1 if investor i completed the S-test and 0 otherwise. Based on Eq. (1), the results are reported before-match (column 2) and after-match (column 3). The set of independent variables consists of an intercept, an initial set of subjective individual characteristics (in Panel A), a second set of sociodemographic variables (in Panel B), and a third set composed of two additional variables intended to control for actual trading behavior before the completion of the MiFID questionnaires (in Panel C). $FL1_i$, $FL2_i$, and $FL3_i$ are $N-1$ dummies for the level of financial literacy selected by investor i; $ECI1_i$ and $ECI2_i$ are $N-1$ dummies for the level of self-assessed experience in complex instruments reported by investor i; ICI_i is another dummy set to one when investor i declares having already traded complex instruments; $LE1_i$ and $LE2_i$ are $N-1$ dummies for the level of education of investor i; $Male_i$ is a dummy set to one when investor i is male; FS_i and ES_i are $N-1$ dummies for the spoken language, with FS_i being set to one when investor i selected French and ES_i being set to one when investor i selected English; $Exec_i$ is a dummy set to one when investor i declared executive responsibilities; Age_i is the age of investor i at the end of 2012; $LOG(MPV_i)$ is the log of investor i's average end-of-month portfolio value computed over the January 2003–December 2007 period; and Exp_i is the number of months when investor i's stock portfolio is not empty over the January 2003–December 2007 period. ***, **, and * indicate significance at 1%, 5%, and 10%, respectively

Table 13 PSM-DiD results for robustness check 2

	Net return (Model 1)	Portfolio size (Model 2)	Number of trades (Model 3)	Portfolio turnover (Model 4)	Lottery-like stock ratio (Model 5)
Panel A: Investors with a least 6 months of nonempty portfolios before and after January 1, 2008					
Constant	−0.0171***	−22.5141***	−9.0317***	0.1656***	0.0615***
$AFTER_t$	0.0033	2.3059***	−0.3344***	−0.1569***	0.0026
$TREAT_i$	0.0018***	1.3498***	0.1306***	0.0107***	−0.0090***
$TREAT \times AFTER_{i,t}$	−0.0017**	1.1943***	−0.3106***	0.0091***	0.0034**
Rm_t	0.9700***	−0.5472	−1.8126*	−0.0991	0.1215
$Crisis_t$	0.0121*	−1.0115***	0.2706**	0.0654***	0.0313*
X_i					
N	1,594,000	1,722,900	1,722,900	1,722,900	1,722,900
R^2	0.2369	0.1956	0.0188	0.0147	0.0107
Panel B: Investors with a least 24 months of nonempty portfolios before and after January 1, 2008					
Constant	−0.0124***	−26.5419***	−8.9421***	0.1027***	0.0695***
$AFTER_t$	0.0030	1.7039***	−0.7166***	−0.1496***	−0.0068
$TREAT_i$	0.0015*	1.7886***	0.3424***	0.0092***	−0.0099***
$TREAT \times AFTER_{i,t}$	−0.0018*	1.4461***	−0.1836***	0.0021	0.0053***
Rm_t	0.9180***	0.0297	−1.4333	−0.0784	0.1152
$Crisis_t$	0.0108*	−0.9941***	0.3252***	0.0609***	0.0309*
X_i					
N	1,133,411	1,148,550	1,148,550	1,148,550	1,148,550
R^2	0.2297	0.1966	0.0240	0.0135	0.0093

This table reports the results for Eq.(2) where the dependent variable ($Y_{i,t}$) is investor i's portfolio net return (Model 1), portfolio size (Model 2), number of stock trades (Model 3), portfolio turnover (Model 4), or lottery-like stock ratio (Model 5) in month t. Portfolio returns are computed using an approximation of the modified Dietz method as follows: $R_t = \frac{\text{EMV}_t - \text{EMV}_{t-1} - P_t + S_t}{EMV_{t-1} + P_t}$ where R_t is the portfolio gross return for month t, EMV_t is the end-of-month portfolio value in month t, and P_t and S_t are the aggregate monetary value of all purchases and sales executed during month t. For net returns, transaction costs paid during month t are subtracted from the aggregate monetary value (in the numerator). Portfolio size is defined as the number of stocks in the end-of-month portfolio. Number of trades refers to the total number of trades executed on stocks per month. Portfolio turnover is the aggregate monetary value of all purchases and sales executed over the month divided by the average of the portfolio values at the beginning and the end of this particular month. Lottery-like stock ratio is defined as the number of lottery-like stocks held in the end-of-month portfolio divided by the total number of stocks in the end-of-month portfolio. Using daily returns over the last 6 months, the stocks identified as lottery-like are those in the lowest 50^{th} stock price percentile, the highest 50^{th} total volatility percentile, and the highest 50^{th} total skewness percentile. In the set of regressors, AFTER_t is a dummy variable set to one when month t is in the post-MiFID period, TREAT_i is a dummy variable set to one when investor i completed the S-test, and $\text{TREAT} \times \text{AFTER}_{i,t}$ is the interaction of these two dummies, Rm_t is the market return in month t, and Crisis_t is a dummy variable for the financial crisis period set to one for months from January 2008 to June 2009. X_i is the set of covariates used for PSM (see Eq. 1). N gives the number of observations. ***, **, and * indicate significance at 1%, 5%, and 10%, respectively. Both net returns and portfolio turnover are winsorized with a 1% cutoff. Standard errors are clustered by month

Table 14 Comparison across S-investors depending on the investment profile

	Conservative or Neutral	Dynamic or Aggressive	
	Mean	Mean	Diff.
Net return	−0.0062	−0.0049	***
Portfolio size	7.53	10.22	***
Number of trades	1.24	1.47	***
Portfolio turnover	0.1906	0.1755	***
Lottery-like stock ratio	0.0622	0.0593	***
N	1924	8346	

This table reports the cross-sectional means for the five dependent variables used in Table 7. These means are based on individual end-of-month portfolios and/or monthly trading activity over the January 2008–March 2012 period, with a distinction between S-investors who are assigned a Conservative or Neutral profile and those who received a Dynamic or Aggressive profile. These investment profiles are defined by the brokerage house and depend on the score obtained by the investor on his/her *S*-test. 'Net return' refers to stock portfolio return that is computed using an approximation of the modified Dietz method as follows: $R_t = \frac{EMV_t - EMV_{t-1} - P_t + S_t}{EMV_{t-1} + P_t}$ where R_t is the portfolio gross return for month t, EMV_t is the end-of-month portfolio value in month t, and P_t and S_t are the aggregate monetary value of all purchases and sales executed during month t. For net returns, transaction costs paid during month t are subtracted from the aggregate monetary value (in the numerator). 'Portfolio size' is defined as the number of stocks in the end-of-month portfolio. 'Number of trades' refers to the total number of trades executed on stocks per month. 'Portfolio turnover' is the aggregate monetary value of all purchases and sales executed over the month divided by the average of the portfolio values at the beginning and the end of this particular month. 'Lottery-like stock ratio' is defined as the number of lottery-like stocks held in the end-of-month portfolio divided by the total number of stocks in the end-of-month portfolio. Using daily returns over the last 6 months, the stocks identified as lottery-like are those in the lowest 50^{th} stock price percentile, the highest 50^{th} total volatility percentile, and the highest 50^{th} total skewness percentile. Both net returns and portfolio turnover are winsorized with a 1% cutoff. The last column reports whether the mean difference between the two groups of investors is statistically significant. *** indicates significance at 1%

Acknowledgements An earlier version of this paper was part of the PhD dissertation of A. Bellofatto when he was a research and teaching assistant at UCLouvain, LFIN/LIDAM. The authors are grateful to the online brokerage house for providing the data. They thank the editor and an anonymous reviewer for insightful comments that helped improve the paper. They also thank O. Corneille, W. De Bondt, R. De Winne, M. Merli and P. Roger for their comments. They also thank Ryan L. Davis for his useful discussion of the earlier version of the paper as well as the participants at the 66^{th} Annual Meeting of the Midwest Finance Association, at the 2017 Joint Conference of the Academy of Entrepreneurial Finance and the Academy of Behavioral Finance & Economics, and at the 34^{th} International Conference of the French Finance Association. Any remaining errors are the full responsibility of the authors. The comments, opinions, and views expressed in this paper do not necessarily reflect those of the managers and directors of Arvestar Asset Management.

Declarations

Conflict of interest Financial support was received from the Observatoire de l'Epargne Européenne. The authors have no Conflict of interest to declare.

References

Abreu, M., Mendes, V.: Information, overconfidence and trading: Do the sources of information matter? J. Econ. Psychol. **33**, 868–881 (2012)

Ackert, L.F., Church, B.K., Englis, B.: The asset allocation decision and investor heterogeneity: a puzzle? J. Econ. Behav. Organ. **47**, 423–433 (2002)

Agnew, J., Balduzzi, P., Sunden, A.: Portfolio choice and trading in a large 401 (k) plan. Am. Econ. Rev. **93**, 193–215 (2003)

Andries, M., Haddad, V.: Information aversion. J. Political Econ. **128**, 1901–1939 (2020)

Angrist, J., Pischke, J.S.: Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press, Princeton (2008)

Argentesi, E., Lütkepohl, H., Motta, M.: Acquisition of information and share prices: an empirical investigation of cognitive dissonance. German Econ. Rev. **11**, 381–396 (2010)

Balloch, A., Nicolae, A., Philip, D.: Stock market literacy, trust, and participation. Rev. Finance **19**, 1925–1963 (2014)

Barber, B.M., Odean, T.: Boys will be boys: Gender, overconfidence, and common stock investment. Q. J. Econ. **116**, 261–292 (2001)

Barber, B.M., Odean, T.: The internet and the investor. J. Econ. Perspect. **15**, 41–54 (2001)

Barber, B.M., Odean, T.: Online investors: Do the slow die first? Rev. Financ. Stud. **15**, 455–488 (2002)

Barthel, A.C., Lei, S.: Investment in financial literacy and financial advice-seeking: Substitutes or complements? Q. Rev. Econ. Finance **81**, 385–396 (2021)

Bellofatto, A., D'Hondt, C., De Winne, R.: Subjective financial literacy and retail investors' behavior. J. Bank Finance **92**, 168–181 (2018)

Bernasek, A., Shwiff, S.: Gender, risk, and retirement. J. Econ. Issues **35**, 345–356 (2001)

Bertrand, M., Duflo, E., Mullainathan, S.: How much should we trust differences-in-differences estimates? Q. J. Econ. **119**, 249–275 (2004)

Bhattacharya, U., Hackethal, A., Kaesler, S., Loos, B., Meyer, S.: Is unbiased financial advice to retail investors sufficient? Answers from a large field study. Rev. Financ. Stud. **25**, 975–1032 (2012)

Bhushan, R.: Firm characteristics and analyst following. J. Account. Econ. **11**, 255–274 (1989)

Bluethgen, R., Gintschel, A., Hackethal, A., Mueller, A.: Financial advice and individual investors' portfolios. Available at SSRN 968197 (2008)

Brashear-Alejandro, T., Kang, J., Groza, M.D.: Leveraging loyalty programs to build customer-company identification. J. Bus. Res. **69**, 1190–1198 (2016)

Brogaard, J., Li, D., Xia, Y.: Stock liquidity and default risk. J. Financ. Econ. **124**, 486–502 (2017)

Broihanne, M., Orkut, H.: MiFID questionnaire answers and stock market participation. Rev. Econ. **7**, 1164–1191 (2021)

Broihanne, M.H., Merli, M., Roger, P.: Diversification, gambling and market forces. Rev. Quantit. Finance Account. **47**, 129–157 (2016)

Broihanne, M.H., Orkut, H.: Investment goals and mental accounting in French retail clients. Finance **39**, 107–144 (2018)

Calcagno, R., Monticone, C.: Financial literacy and the demand for financial advice. J. Bank. Finance **50**, 363–380 (2015)

Calvet, L.E., Campbell, J.Y., Sodini, P.: Measuring the financial sophistication of households. Am. Econ. Rev. **99**, 393–98 (2009)

Chalmers, J., Reuter, J.: What is the impact of financial advisors on retirement portfolio choices and outcomes? National Bureau of Economic Research (NB10-05) (2010)

Charness, G., Gneezy, U.: Strong evidence for gender differences in risk taking. J. Econ. Behav. Organ. **83**, 50–58 (2012)

Claessens, S., Glaessner, T., Klingebiel, D.: Electronic finance: reshaping the financial landscape around the world. J. Financ. Serv. Res **22**, 29–61 (2002)

Collins, J.M.: Financial advice: A substitute for financial literacy? Financ. Serv. Rev. **21**, 307–322 (2012)
Committee of European securities regulators.: a consumer's guide to MiFID: Investing in financial products (2008)
Desagre, C., D'Hondt, C.: Googlization and retail trading activity. J. Behav. Exp. Finance **29**, 100453 (2021)
D'Hondt, C., De Winne, R., Ghysels, E., Raymond, S.: Artificial intelligence alter egos: Who might benefit from robo-investing? J. Empir. Finance **59**, 278–299 (2020)
D'Hondt, C., De Winne, R., Merli, M.: Do retail investors bite off more than they can chew? A close look at their return objectives. J. Econ. Behav. Organ. **188**, 879–902 (2021a)
D'Hondt, C., McGowan, R., Roger, P.: Trading leveraged exchange-traded products is hazardous to your wealth. Q. Rev. Econ. Finance **80**, 287–302 (2021b)
D'Hondt, C., Roger, P.: Investor sentiment and stock return predictability: the power of ignorance. Finance **38**, 7–37 (2017)
Dorn, D., Huberman, G.: Talk and action: What individual investors say and what they do. Rev. Finance **9**, 437–481 (2005)
Dwyer, P.D., Gilkeson, J.H., List, J.A.: Gender differences in revealed risk taking: evidence from mutual fund investors. Econ. Lett. **76**, 151–158 (2002)
Fang, V., Tian, X., Tice, S.: Does stock liquidity enhance or impede firm innovation? J. Finance **69**, 2085–2125 (2014)
Festinger, L.: A Theory of Cognitive Dissonance. Stanford University Press, Redwood (1957)
Gaudecker, H.M.V.: How does household portfolio diversification vary with financial literacy and financial advice? J. Finance **70**, 489–507 (2015)
Georgarakos, D., Inderst, R.: Financial advice and stock market participation. Working Paper No 1296, European Central Bank (2014)
Gerhardt, R., Hackethal, A.: The influence of financial advisors on household portfolios: A study on private investors switching to financial advice. Available at SSRN 1343607 (2009)
Glaser, M.: Online broker investors: Demographic information, investment strategy, portfolio positions, and trading activity. Available at SSRN 975985 (2003)
Glaser, M., Weber, M.: Which past returns affect trading volume? J. Financ. Mark. **12**, 1–31 (2009)
Goetzmann, W.N., Kumar, A.: Equity portfolio diversification. Rev. Finance **12**, 433–463 (2008)
Grinblatt, M., Keloharju, M.: How distance, language, and culture influence stockholdings and trades. J. Finance **56**, 1053–1073 (2001)
Grossman, S.J.: On the efficiency of competitive stock markets where trades have diverse information. J. Finance **31**, 573–585 (1976)
Grossman, S.J., Stiglitz, J.E.: On the impossibility of informationally efficient markets. Am. Econ. Rev. **70**, 393–408 (1980)
Guiso, L., Jappelli, T.: Information acquisition and portfolio performance. CSEF Working Paper No. 167.r (2006)
Guiso, L., Jappelli, T.: Financial literacy and portfolio diversification. Working Paper, European University Institute (2008)
Hackethal, A., Haliassos, M., Jappelli, T.: Financial advisors: A case of babysitters? J. Bank. Finance **36**, 509–524 (2012)
Haslem, J.A.: Why do mutual fund investors employ financial advisors? J. Invest. **17**, 91–94 (2008)
Hoechle, D., Ruenzi, S., Schaub, N., Schmid, M.: The impact of financial advice on trade performance and behavioral biases. Rev. Finance **21**, 871–910 (2017)
Hoffmann, A.O., Post, T., Pennings, J.M.: Individual investor perceptions and behavior during the financial crisis. J. Bank. Finance **37**, 60–74 (2013)
Horn, M., Oehler, A., Wendt, S.: Fintech for consumers and retail investors: opportunities and risks of digital payment and investment services. In: Walker, T., Gramlich, D., Bitar, M., Fardnia, P. (eds.) Ecological, Societal, and Technological Risks and the Financial Sector. Palgrave Macmillan, London (2020)
Hsiao, Y.J., Tsai, W.C.: Financial literacy and participation in the derivatives markets. J. Bank. Finance **88**, 15–29 (2018)
Hung, A., Yoong, J.: Asking for help: survey and experimental evidence on financial advice and behavior change. In: Mitchell, Olivia S., Smetters, Kent (eds.) The Market for Retirement Financial Advice. Published to Oxford Scholarship Online, Oxford (2014)

Jappelli, T., Padula, M.: Investment in financial literacy and saving decisions. J. Bank. Finance **37**, 2779–2792 (2013)
Kang, J., Alejandro, T.B., Groza, M.D.: Customer-company identification and the effectiveness of loyalty programs. J. Bus. Res. **68**, 464–471 (2015)
Koestner, M., Loos, B., Meyer, S., Hackethal, A.: Do individual investors learn from their mistakes? J. Bus. Econ. **87**, 669–703 (2017)
Kramer, M.M.: Financial advice and individual investor portfolio performance. Financ. Manag. **41**, 395–428 (2012)
Kreis, H., Mafael, A.: The influence of customer loyalty program design on the relationship between customer motives and value perception. J. Retail. Consum. Serv. **21**, 590–600 (2014)
Kumar, A.: Who gambles in the stock market? J. Finance **64**, 1889–1933 (2009)
Kumar, A., Lee, C.: Retail investors sentiment and returns comovements. J. Finance **61**, 2451–2486 (2006)
Lee, E.J., Eastwood, D.B., Lee, J.: A sample selection model of consumer adoption of computer banking. J. Financ. Serv. Res. **26**, 263–275 (2002)
Lusardi, A., Michaud, P.C., Mitchell, O.S.: Optimal financial knowledge and wealth inequality. J. Political Econ. **125**, 431–477 (2017)
Mazzoli, C., Marinelli, N.: Determinants of risk-suitable investment portfolios: evidence from a sample of Italian householders. Int. J. Finance Bank. **1**, 13–28 (2014)
Migliavacca, M.: Keep your customer knowledgeable: financial advisors as educators. Eur. J. Finance **26**, 402–419 (2020)
Nicolosi, G., Peng, L., Zhu, N.: Do individual investors learn from their trading experience? J. Financ. Mark. **12**, 317–336 (2009)
Oehler, A., Horn, M., Wendt, S.: Investor characteristics and their impact on the decision to use a robo-advisor. J. Financ. Serv. Res. **62**(1), 91–125 (2021)
Orkut, H.: Foreign stock investment and sophistication of French retail investors. Finance **42**, 39–80 (2021)
Polkovnichenko, V.: Individual investor portfolios. In: Kent Baker, H., Nofsinger, J.R. (eds.) Behavioral Finance: Investors, Corporations, and Markets, pp. 539–557. Wiley, New York (2010)
Rosenbaum, P.R., Rubin, D.B.: The central role of the propensity score in observational studies for causal effects. Biometrika **70**, 41–55 (1983)
Seasholes, M.S., Zhu, N.: Individual investors and local bias. J. Finance **65**, 1987–2010 (2010)
Shestopaloff, Y., Shestopaloff, A.: A hierarchy of methods for calculating rates of return. J. Perform. Measurement **12**, 39–52 (2007)
Shum, P., Faig, M.: What explains household stock holdings? J. Bank. Finance **30**, 2579–2597 (2006)
Stuart, E.A.: Matching methods for causal inference: a review and a look forward. Statist. Sci. **25**, 1–21 (2010)
Tauni, M.Z., Fang, H.X., Yousaf, S., et al.: The influence of investor personality traits on information acquisition and trading behavior: evidence from Chinese futures exchange. Personal. Individ. Differ. **87**, 248–255 (2015)
Van Rooij, M., Lusardi, A., Alessie, R.: Financial literacy and stock market participation. J. Financ. Econ. **101**, 449–472 (2011)
Vissing-Jorgensen, A.: Perspectives on behavioral finance: Does "irrationality" disappear with wealth? Evidence from expectations and actions. In: Gertler, M., Rogoff, K. (eds.) NBER Macroeconomics Annual 2003, pp. 139–208. The MIT Press, Cambridge (2003)
Volpe, R.P., Kotel, J.E., Chen, H.: A survey of investment literacy among online investors. Financ. Couns. Plan. **13**, 1–13 (2002)

Anthony Bellofatto is Risk & ESG Expert Lead at Arvestar Asset Management in Belgium. This paper is part of research work he carried out during his PhD at the Louvain Finance research center (LIDAM) and the Louvain School of Management, UCLouvain. His doctoral research focused on behavioral finance and aimed to better understand how individual characteristics (i.e., sociodemographic indicators and information reported in the MiFID questionnaires) help explain heterogeneity in behavior across online retail investors.

Marie-Hélène Broihanne is a Full Professor of Finance at the University of Strasbourg. She is deputy head of the LaRGE Research Center at EM Strasbourg. Her main research interests are related to behavioral finance. Her research work aims to better understand the behavior of retail investors and their decision-making under uncertainty. Her recent publications focus on the MiFID questionnaires used by banks in the EU. Her papers are published in top journals, such as Journal of Banking & Finance, Journal of Risk and Uncertainty, Theory and Decision, Finance Research Letters, Review of Quantitative Finance and Accounting, and Journal of Comparative Psychology.

Catherine D'Hondt is a Full Professor of Finance at the Louvain School of Management, UCLouvain, in Belgium. She is affiliated with the Louvain Finance research center (LIDAM). Her main research interests are related to behavioral finance, decision-making, or market microstructure. Her research work aims to better understand the behavior of online retail investors. Her papers are published in top journals, such as Review of Finance, Journal of Economic Behavior & Organization, Journal of Banking & Finance, Finance Research Letters, Journal of Behavioral and Experimental Finance, Journal of Gambling studies, Psychonomic Bulletin & Review, and Personality and Individual Differences.