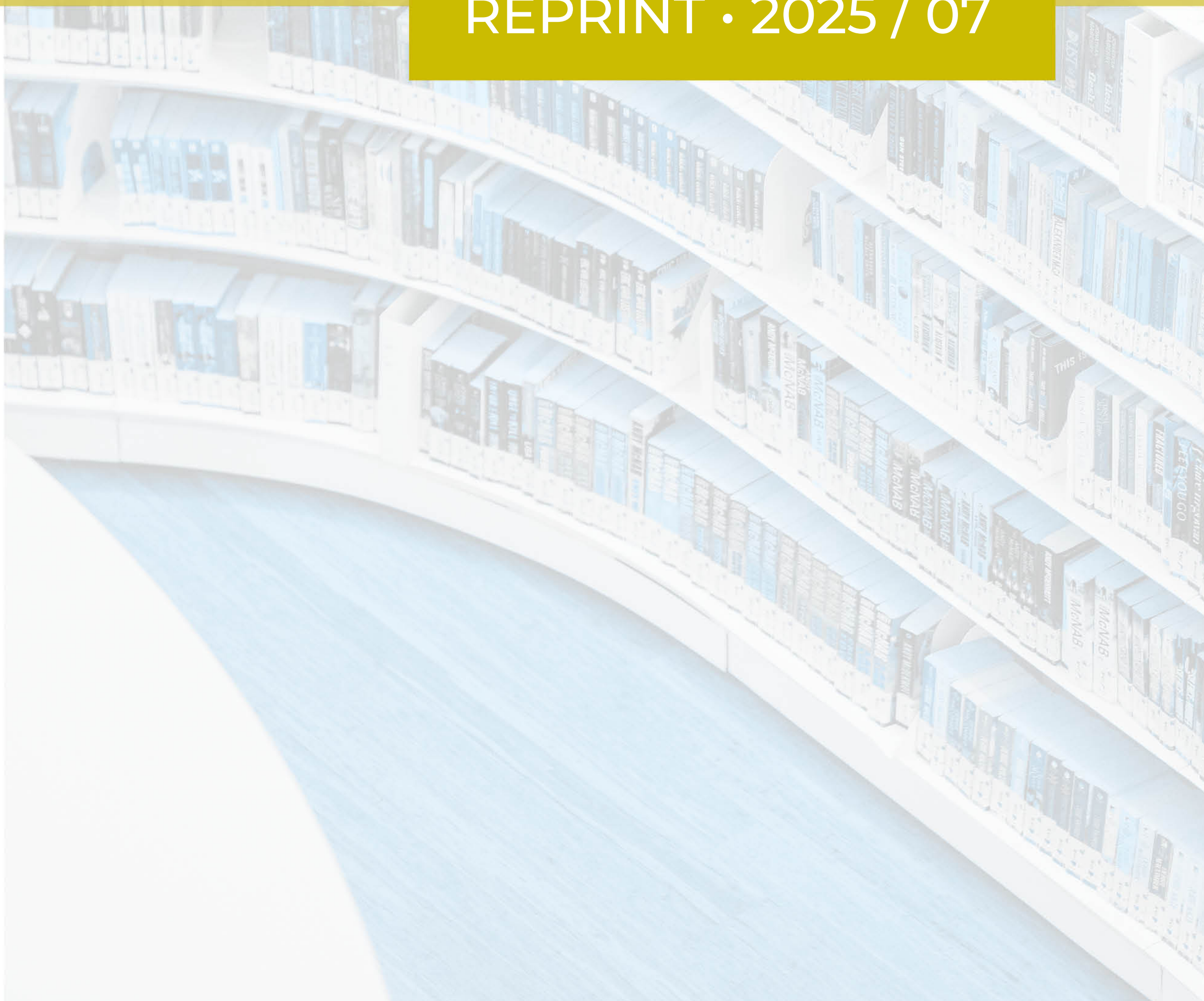
UNCOVERING THE PROFILE OF PASSIVE EXCHANGE-TRADED FUND RETAIL INVESTORS

Catherine D'Hondt, Mikael Petitjean,
Younes Elhichou Elmaya

REPRINT • 2025 / 07



UCLouvain

LFIN | LIDAM 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

Uncovering the profile of passive exchange-traded fund retail investors*

Catherine D'Hondt[†], Mikael Petitjean[‡], and Younes Elhichou Elmaya[§]

Abstract

We profile retail investors in passive exchange-traded funds (P-ETFs) using both trading records and survey-based data on a large pool of individuals over nearly a decade. There is strong evidence that retail investors who trade P-ETFs have a distinct profile. Both the probability and the magnitude of P-ETF investing are driven by higher education and longer trading experience. P-ETF investors self-report higher financial literacy, a lower return objective, lower risk tolerance, and a longer investment horizon. These investors are less overconfident, and their stock holdings show less home bias. Furthermore, the more active P-ETF users hold more concentrated stock portfolios and display lower stock portfolio turnover, pointing to a substitution effect between stocks and P-ETFs.

Résumé

Nous dressons le profil des investisseurs individuels qui négocient en bourse des trackers passifs (P-ETF). Nous utilisons les données d'enquête et les registres de transactions de tous les clients actifs sur une plateforme de courtage en ligne au cours d'une période d'environ 10 ans. Il apparaît clairement que ces investisseurs ont un profil distinct. La probabilité et l'ampleur des investissements dans les P-ETF sont liées à un niveau d'éducation plus élevé et à une plus longue expérience de négociation. Les investisseurs en P-ETF déclarent avoir des connaissances financières plus élevées, un objectif de rendement moins élevé, une tolérance au risque plus faible et un horizon d'investissement plus long. L'excès de confiance est

* The authors are grateful to the brokerage house for providing the data. They thank the editor and two anonymous referees for insightful comments that helped improve the paper. They also thank M-H. Broihanne, R. De Winne, G. Hubner as well as participants at the 6th ISCEF 2020, European Financial Management Association Annual conference (EFMA 2021), 13th RGS Doctoral Conference in Economics, 38th International conference of the French Finance Association (AFFI 2022) and participants at Louvain Finance seminars. Any 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 Dexia Group.

† LFIN/LIDAM (UCLouvain). Address: Chaussée de Binche 151, 7000 Mons, Belgium. Email: catherine.dhondt@uclouvain.be

‡ LFIN/LIDAM (UCLouvain, Belgium) & LARSH (IAE Valenciennes, Université Polytechnique Hauts-de-France). Address: Les Tertiales, Rue des Cent Têtes, 59313 Valenciennes Cedex 9, France. Corresponding author. Email: mikael.petitjean@uphf.fr

§ Dexia Group. Address: Tour Bastion, Place du Champs de Mars 5, 1050 Brussels, Belgium. Email: younes.elhichou@dexia.com

moins présent chez eux et les actions qu'ils détiennent sont moins orientés vers le marché domestique. En outre, les utilisateurs de P-ETF les plus actifs détiennent des portefeuilles d'actions plus concentrés et affichent un taux de rotation plus faible dans leurs portefeuilles d'actions, ce qui suggère l'existence d'un effet de substitution entre actions et P-ETF.

1. Introduction

Passive exchange-traded funds (P-ETFs), designed to replicate the performance of an index, have emerged as a highly appealing innovation in the finance industry. Their allure stems from their cost-effectiveness, transparency, and tax efficiency, setting them apart from active ETFs, which have historically struggled to surpass the benchmark after accounting for fees. P-ETFs have amassed an impressive $3 trillion in assets over the past decade, dwarfing the relatively meager $200 billion drawn by active ETFs. As pointed out by Cheng et al. (2018), ETFs have been heralded as "the harbingers of a new era of low-cost/low-risk investment opportunities available to the general public". Nowadays retail investors can indeed trade a wide range of well-diversified, highly liquid, low-cost funds, allowing them to apply the central tenets of modern portfolio theory.[1]

Research on the profile of retail investors who adhere to Warren Buffett's advice[2] and include P-ETFs in their portfolios is scarce. To the best of our knowledge, the primary reference in this area is Bhattacharya et al. (2017) but their focus mainly revolves around the performance of retail investors who use P-ETFs. They find that, compared to non-users, P-ETF users do not demonstrate improved portfolio performance. Regarding their specific profile, Bhattacharya et al. (2017) report only a few characteristics: P-ETF users tend to be younger, wealthier (in terms of both portfolio value and overall wealth), and have shorter relationships with brokerage firms.

1 Understanding the factors impacting the performance of passive and active investment vehicles is crucial for retail investors. Many studies have shown the dominance of ETFs over mutual funds. For instance, Prather et al. (2009) find that ETFs outperform mutual funds when considering total annual costs. Similarly, Farinella and Kubicki (2018) show that P-ETFs exhibit statistically higher annual returns and lower annual fees compared to mutual funds.

2 Warren Buffett strongly advocates for P-ETFs, particularly for retail investors with limited resources in terms of capital, time, or skills (see, e.g., Schroeder, 2008).

Considering a wider range of investment products, Davydov et al. (2017) provide evidence suggesting that younger investors have relatively higher skills in using exchange-traded products (i.e., ETFs but also exchange-traded commodities and exchange-traded notes), although they tend to underperform those who invest in mutual funds. Focusing exclusively on ETF users, Enete et al. (2018) report that these investors display greater subjective financial knowledge compared to non-users. However, these authors do not distinguish between passive and active ETFs, including leveraged, inverse, or even inverse-leveraged ETFs that strongly deviate from any passive strategies. According to D'Hondt et al. (2020), such leveraged products are typically used by overconfident gamblers with a propensity for highrisk investments, even if inverse-leveraged ETFs may serve as hedging instruments for some highly loss-averse investors.

Building upon the aforementioned research, we focus on retail investors who invest in PETFs,[3] referred to as P-ETF users, to sketch their profile and identify their main motivations. By doing so, we add to the above literature by determining a comprehensive profile of these P-ETF users relying on both field data (i.e., actual trading accounts) and survey-based data (i.e., information collected by the brokerage house within its *Know Your Customer* process under MiFID[4] regulation). To the best of our knowledge, this combination of field and survey data in sketching the profile of P-ETF investors is unique and allows us to cover a very large set of individual characteristics.[5]

Our dataset comes from a large Belgian online brokerage house and encompasses the entire population of its clients who traded in the post-MiFID period (i.e., after November 2007) and took the MiFID-related Appropriateness test (A-test).[6] Our population includes 26,140 retail investors spanning nearly a decade from January 2003 to March 2012. Bhattacharya et al. (2017) ignore survey data and do not have access to the entire population.[7] We therefore ensure an accurate representation of the

3 This means our study excludes any active and/or leveraged ETFs from consideration.

4 MiFID refers to the Markets in Financial Instruments Directive. This piece of European regulation came into force in late 2007. More information about MiFID requirements is provided in Section 3.

5 D'Hondt et al. (2020) also use a blend of trading and survey-based data to outline the profile of ETF users. However, these authors concentrate on retail investors who engage in trading leveraged exchangetraded products.

6 As explained in detail in Section 3, investors who ask for execution under MiFID have to complete the A-test when they want to trade "complex" instruments. In adherence to the precautionary principle, securities that exhibit variability in returns and possess characteristics that might not be easily understood by the average investor are classified as complex instruments.

7 They deal with a sample of 5,869 German investors from August 2005 to March 2010.

brokerage house's client population post-MiFID implementation since our analysis is free from several common types of bias: there is no omission bias, as we do not disregarded survey data; no convenience sampling or pre-selection bias, given the absence of illogical exclusion[8]; no survivorship bias, because our dataset accounts for clients who have ceased trading; and no non-response bias, as all clients must respond to the MiFID questionnaire.

In addition to detailed trading records, we have sociodemographic and survey-based characteristics for each investor. The latter were directly collected by the brokerage house within its MiFID questionnaires. Such rich data enable us to consider four sets of factors likely to influence P-ETF investing: (1) sociodemographics (including gender, age, education, and spoken language), (2) subjective individual characteristics (such as financial literacy, return objective, risk tolerance, investment time horizon, and wealth and financial situation), (3) objective measures of trading activity (such as experience, trade duration, portfolio size, trading intensity and underperformance), and (4) behavioral variables (including proxies for overconfidence, narrow framing, underdiversification, and home bias). Relying on this very large set of covariates, we mitigate the risk of omission bias and are able to combine the statements made by retail investors with their actual behavior.

Using the data at hand, we further contribute to the literature by accounting for two stylized features in the trading data that have been disregarded in the literature, i.e., the excess zeros[9] and the overdispersion in the distribution of the non-zero counts[10]. We deal with these two

8 Following Glaser and Weber (2009) and Bellofatto et al. (2018), the only filter that we apply is the regular 5-month trading experience requirement to disregard irrelevant, very short-lived investors.

9 The excess of zeros in a dataset, often encountered in count data, implies that there are more observations with a value of zero than what standard count models (like Poisson or standard negative binomial models) would predict. In statistics and regression analysis, an excess of zeros can lead to several issues. First, traditional count data models like Poisson or negative binomial regression assume that the number of occurrences of an event follows a specific distribution throughout the dataset. These models might not fit well if there's an excess of zeros, as they do not account for the dual process generating the zeros and the counts. Second, when standard models are used in the presence of excess zeros, the parameter estimates and their standard errors can be biased. This can lead to incorrect inferences about the relationship between the predictors and the outcome variable.

10 Overdispersion is particularly relevant in models that assume a Poisson or binomial distribution, where the mean and variance are typically expected to be equal or proportionally related. When overdispersion occurs, it implies that there is more variability in the data than the model accounts for, indicating the presence of additional, unmodeled sources of variation. In the context of regression analysis, overdispersion can have several implications. Overdispersion suggests that the basic assumptions of the model (e.g., Poisson or binomial distribution for the response variable) are not fully appropriate for the data. This can lead to incorrect conclusions about the significance of predictors or the model's overall fit. Standard errors of estimates may be underestimated in the presence of overdispersion, leading to overly optimistic p-values. This can increase the risk of Type I errors (falsely declaring a result significant). Addressing overdispersion requires adjusting the model or using alternative statistical methods. This is what we do by using a quasiPoisson or negative binomial regression instead of a standard Poisson regression.

features in a zero-inflated negative binomial (ZINB) model where the decision to invest in P-ETFs and the positive counts of trades in P-ETFs are generated by separate processes, explaining at once the *likelihood* of trading P-ETFs and the *magnitude* of P-ETF trading. This model is therefore suited to handle both the excess zeros (i.e., no P-ETF trades for non-users) and overdispersion in the non-zero counts (i.e., small or large P-ETF trades among users).[11] To the best of our knowledge, this kind of model has never been used to examine retail investing in investment funds. Binary logistic models (Bailey et al., 2011; Bhattacharya et al., 2017) and probit models (Müller and Weber, 2010) have usually been applied.

Our key findings can be summarized as follows. Compared to non-users, P-ETF users are more likely to have a university degree, self-report higher financial literacy and declare a longer investment horizon. They also exhibit longer trading experience, hold more diversified stock portfolios, and display stronger stock trading intensity. By contrast, investors who declare a higher return objective, self-report higher risk tolerance, and those who are more subject to overconfidence and home bias, are less likely to invest in P-ETFs.

When focusing exclusively on P-ETF users, the number of P-ETF trades is positively associated with age, higher education, higher financial literacy, longer investment horizon, and lower return objective. The number of P-ETFs trades is also positively related to the length of trading experience. Nevertheless, our findings provide evidence of a trade-off between the magnitude of P-ETF trades and the trading activity on stocks. Investors who actively trade P-ETFs tend to hold more concentrated stock portfolio (i.e., smaller portfolio size), and they display longer trade duration on stocks and lower stock portfolio turnover. These results are consistent with investors making fewer adjustments to their stock holdings when they actively trade P-ETFs. This empirical evidence suggests a substitution effect between stocks and P-ETFs as the number of trades in P-ETFs increases. Finally, we also notice that stock holdings of P-ETFs users are less subject to home bias as their trading activity in P-ETFs increases.

11 The ZINB model has gained popularity in both biomedical research (Cho et al., 2023) and various other fields, including finance. It has been applied to study trade duration (Blasques et al., 2023), insurance claims (Yip and Yau, 2005), and mergers and acquisitions (Ho et al., 2009), among other topics.

The rest of the paper is organized as follows. Section 2 reviews the relevant literature and presents our hypotheses regarding the main drivers of retail investment in P-ETFs. Section 3 describes our data and provides descriptive statistics for our sample of investors. Our empirical results are reported in Section 4. Section 5 concludes.

2. Literature and hypotheses

The objective of this paper is to identify the profile of retail P-ETF investors and their main motivations. For that purpose, we consider a blend of sociodemographic indicators, individual characteristics, trade- and portfolio- based variables, and proxies for common behavioral biases. The detailed description of these variables is provided in Section 3.2. These four sets of variables are used to capture the investor's level of financial sophistication as well as to what extent he or she is subject to common behavioral biases. A large body of literature has shown that these two aspects significantly impact retail investors' investment behavior and wealth and, to a larger extent, social welfare. It seems therefore natural to build upon financial sophistication and common behavioral biases to formulate testable hypotheses about the profile of P-ETF users and the magnitude of their trading activity in P-ETFs.

2.1. Financial sophistication

Financial sophistication is a broad concept that has been widely investigated in past research (e.g., Lusardi et al., 2009; Calvet et al., 2009; Van Rooij et al., 2011; Lusardi and Mitchell, 2014; Lusardi et al., 2017; Lindblom et al., 2018; Titman et al., 2022). From a theoretical standpoint, this notion entails a mixture of knowledge, skills, experience, and the ability to make sound investment decisions. Practically speaking, financial sophistication is usually assessed using many indicators such as the level of financial literacy, the length of trading experience or the number of trades, the level of education, the financial situation and wealth, and the investment horizon.

Financial literacy refers to the ability to process economic information and make sound decisions about financial planning, wealth accumulation, debt, and pensions (Lusardi and Mitchell, 2014). A large body of evidence shows that retail investors greatly benefit from financial literacy. Higher financial literacy is associated with better day-to-day financial management abilities (e.g., Hilgert et al., 2003) and successful retirement planning (e.g., Lusardi and Mitchell, 2007). Higher financial literacy helps reduce the negative effect of behavioral biases on wealth (e.g., Baker and Nofsinger, 2002) and the lack of diversification (e.g., Gaudecker, 2015). There is also empirical evidence showing that investors with lower financial literacy achieve poor portfolio performance (e.g., Gaudecker, 2015; Bellofatto et al., 2018; Bianchi, 2018).

Among retail investors, the level of financial literacy is usually estimated with two approaches. When based on a set of questions addressing the basics at the root of saving and investment decisions (such as compounding, inflation, and diversification), financial literacy measurement is labelled as *objective* because actual knowledge is revealed through correct answers. By contrast, measures of *subjective* financial literacy rely on questions asking individuals to self-assess their financial knowledge and expertise (using a preset scale). Although correlation between objective and subjective literacy should not be taken for granted (Lusardi and Mitchell, 2014), prior research points to a positive relationship between subjective literacy and the actual level of knowledge (e.g., Dorn and Huberman, 2005; Van Rooij et al., 2011; Babiarz and Robb, 2014).[12] According to Courchane and Zorn (2005), subjective financial literacy is one of the most significant determinant of the way people invest. In the same vein, Bellofatto et al. (2018) report that retail investors with higher subjective financial literacy seem to invest in a smarter way, even after controlling for gender, age, portfolio value, trading experience and education. These investors are also less subject to the disposition effect and achieve better performance.

Financial literacy is also known as driver of awareness of investment alternatives to common stocks. Using survey data, Müller and Weber (2010) provide evidence of a positive relationship between financial literacy and the probability of investing passively. Bellofatto et al. (2018) show that investors with higher financial literacy are more likely to invest in funds; they tend to concentrate their portfolios on a small set of stocks and achieve diversification through investment funds holding. Consistent findings are reported in Enete et al. (2018) who show that ETF users have

12 Put differently, retail investors tend to have a relatively reliable perception of their level of knowledge, even if exceptions are still possible (Lusardi and Mitchell, 2011).

better financial knowledge than non-users. According to Hsiao and Tsai (2018), individuals with higher financial literacy are also more likely to trade derivatives.

As mentioned earlier, experience is another facet of sophistication that is key in the decision-making process of investing. We know from past research that trading experience mitigates the reluctance of retail investors to realize losses (e.g., Feng and Seasholes, 2005) and, to a larger extent, positively influences overall investor behavior (e.g., Lyons et al., 2006). Retail investors are indeed able to learn from their past experience, as documented in Nicolosi et al. (2009) and Koestner et al. (2017). Focusing on fund investments, Wang (2011) shows that experience, in addition to financial knowledge and income, is a significant factor impacting the way younger individuals invest in mutual funds. According to Bailey et al. (2011), sophisticated investors, defined as better informed, older, with higher income and more experience, "make good use" of mutual funds since they hold a high proportion of funds for long periods of time, avoid high expense funds, and achieve relatively good performance.

Based on the aforementioned literature on financial sophistication, we formulate our first two hypotheses to be tested in Section 4 as follows:

H1. The more financially sophisticated retail investors are, the more likely they are to invest in P-ETFs.

H1*. Among the users of P-ETFs, the more financially sophisticated users trade P-ETFs more often.

2.2. Behavioral biases

Behavioral biases usually lead to costly investment mistakes among retail investors (e.g., Singh, 2010; Ramiah et al., 2015; Koestner et al., 2017). Although investors lacking financial sophistication are often those more subject to behavioral biases, the consideration of both concepts in empirical studies remains insightful. When data is available, it is advisable not to overlook the distinction between these concepts for valid reasons. For instance, highly financially sophisticated individuals may display overconfidence in their own abilities and knowledge, resulting in excessive risk-taking or excessively optimistic investment decisions. Consequently, it is crucial to carefully consider the distinct impact of financial sophistication and behavioral biases when profiling retail investors in P-ETFs.

Underdiversification is one of the most common investment mistakes made by retail investors (Koestner et al., 2017). In most empirical studies, retail investors hold typically only a few stocks in their portfolios, often between 3 and 7 stocks depending on the sample (e.g., Kelly, 1995; Polkovnichenko, 2005; Kumar and Lee, 2006; Korniotis and Kumar, 2013; Leal et al., 2017; Bellofatto et al., 2018). This lack of diversification negatively affects the investors' financial well-being in general (Barber and Odean, 2000) and depends on the degree of sophistication (Goetzmann and Kumar, 2008). P-ETFs being low-cost[13]/low-risk investment opportunities easily available to retail investors (Cheng et al., 2018), P-ETF users are arguably expected to be more aware of the benefits of diversification than non-users. Hence, we expect the P-ETFs users to hold more diversified stock portfolios than non-users. On the one hand, the same reasoning might also apply to the subsample of P-ETF users: among investors who trade P-ETFs, those whose stock portfolio is more diversified are expected to trade P-ETFs more extensively. On the other hand, we cannot rule out the possibility that (some) P-ETFs users see P-ETFs as better cost-effective investment alternatives to common stocks to diversify their holdings. In this case where P-ETFs users might prefer P-ETFs to stocks, we should observe a positive association between more concentrated stock portfolios and the number of P-ETF trades.

A particular case of underdiversification is the home bias, i.e., the situation in which investors overweight domestic stocks within their portfolio because these stocks look more familiar—geographically, culturally or socially. This term "home bias" was initially used by French and Poterba (1991) to refer to the tendency for investors to invest the majority of their portfolio in domestic stocks, ignoring the benefits of diversifying into foreign stocks. This concept has however been broadened out over the years. Retail investors are known for having preferences for companies with geographical proximity (e.g., Bailey et al., 2011; Boyle et al., 2012; Massa and Simonov, 2006). Both cultural proximity and language play a role in that matter as shown in Grinblatt and Keloharju (2001). Building on this stream of literature, we expect a negative relationship between retail P-ETF investment (or its magnitude) and home bias: when investors show a marked preference for specific domestic stocks, it arguably

13 Although higher transaction costs on investment alternatives to common stocks were arguably mentioned in the past as one main reason for underdiversification (Evans and Archer, 1968), this explanation is no longer relevant since the emergence of ETFs and index funds has led to a drastic reduction in explicit transaction costs (Domian et al., 2007).

follows that their inclination and level of investment in broad-based ETFs would diminish.

Individuals generally suffer from overconfidence[14] (e.g., Fischoff, 1982; Moore and Healy, 2008; De Bondt and Thaler, 1995; Camerer and Lovallo, 1999) and retail investors are not an exception to the rule. The detrimental impact of overconfidence on trading behavior and portfolio performance has been widely documented in the literature (e.g., Barber and Odean, 2000, 2001; Dorn and Huberman, 2005; Glaser and Weber, 2009). Overconfidence especially leads to overtrading (Odean, 1999) and riskier portfolios (Carpentier and Suret, 2011). Considering mutual funds, Bailey et al. (2011) find that overconfident investors are more likely to select higher expense funds and avoid index funds. Moreover, when overconfident investors buy mutual funds, they trade them frequently and prefer active funds with high expense ratios rather than passive index funds with low expense ratios, such as ETFs. Consistent findings are reported in Davydov et al. (2017). Hence, we expect that P-ETFs are not targeted by overconfident investors.

Narrow framing is another common behavioral bias among retail investors. In short, this bias refers to their tendency to evaluate and select assets in isolation (Barberis et al., 2006). Its existence is related to the broader concept of mental accounting developed by Thaler (1980) to describe the process by which people keep goals and move towards those goals separately from each other.[15] Because of narrow framing, retail investors tend to not consider their portfolio as a whole, contrary to the assumptions of Markowitz's portfolio choice theory. Narrow framing is even accentuated when portfolios contain a very low number of different stocks (Kumar and Lim, 2008). According to Kumar (2009), there is a positive relationship between uncertainty and narrow framing; i.e., investors are more likely to adopt focused and narrower decision frames in more uncertain environments. The degree of narrow framing is, however, negatively related to investor financial sophistication and trading experience (Liu et al., 2010). Focusing on mutual funds, Bailey et al. (2011) find that investors who exhibit higher scores of narrow framing are less

14 Overconfidence is observed when people's subjective confidence in their own ability is greater than their actual performance. In short, this concept is often related to (a) an overestimation of one's skills, (b) a belief that one is better than the median person, and (c) overprecision in one's belief (often labelled miscalibration).

15 According to Baker and Nofsinger (2002), this process helps people with their self-control, since keeping things separately allows better measurement of the progress of each goal.

likely to invest in both equity mutual funds and index funds. This evidence suggests that investors who suffer from narrow framing might have less willingness to invest in P-ETFs.

Building on past empirical findings about the four above-mentioned behavioral biases (i.e., underdiversification, overconfidence, narrow framing and home bias), we posit our last two hypotheses as follows:

H2. Retail investors who are more subject to behavioral biases are less likely to invest in P-ETFs.

H2*. Among the users of P-ETFs, those who are more subject to behavioral biases trade P-ETFs less often.

3. Data and sample

3.1. Data

The data come from a large Belgian online brokerage house and cover the trading accounts of 26,140 retail investors over the period from January 2003 to March 2012. For each investor, we have detailed trading records (ISIN code, timestamp, trade direction, executed quantity, trade price, explicit transaction costs, etc.). For P-ETFs, we count 22,376 trades involving 3,484 retail investors (i.e., 13.33% of the retail investors in the sample). These P-ETF users trade 313 different P-ETFs, among which 196 are equities-linked and 72 commoditieslinked.[16]

In addition to the trading records, we have individual data that are either sociodemographic (age, gender, education, and spoken language) or survey-based. The descriptive statistics for the sociodemographic indicators are provided in Section 3.2.1. The survey data provide subjective investor characteristics[17] collected by the brokerage house within the context of the European MiFID regulation.[18] In short, MiFID distinguishes among the execution of orders, provision of financial advice, and supply of

16 The list of the ten most-traded P-ETFs in the sample is available in Table 7 in appendix.

17 These survey data are reported online by each investor. Hence, answers are self-reported decisions that investors make on their own, without any interaction with a financial intermediary. Since most of these data imply self-assessment, we consider them subjective individual characteristics.

18 MiFID is a key piece of the regulation of financial markets in the European Union 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 but several reviews have been recently implemented. For more 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).

portfolio management services. Under MiFID requirements, brokerage firms and banks must collect specific information about their retail clients' needs and preferences using the so-called MiFID questionnaires (also referred to as "MiFID tests"). Investors who ask for execution only have to complete the appropriateness test (A-test hereafter) when they want to trade "complex" instruments.[19] The aim of this 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 have to complete the suitability test (S-test hereafter). Unlike the A-test, the S-test also includes assessment of investment objectives and financial situation.[20] Some detailed items of the S-test are provided in Table 10 (in appendix) as examples.

For the purpose of this study, we distinguish between "subjective" variables, which are based on the aforementioned A-test and S-test, and "objective" variables, which characterize the actual trading activity of retail investors. In addition, we build several behavioral variables, i.e., proxies used to capture overconfidence, narrow framing, lack of diversification, and home bias. The descriptive statistics about the subjective variables are provided in Section 3.2.2. The objective variables are presented in Section 3.2.3, while the behavioral variables are described in Section 3.2.4.

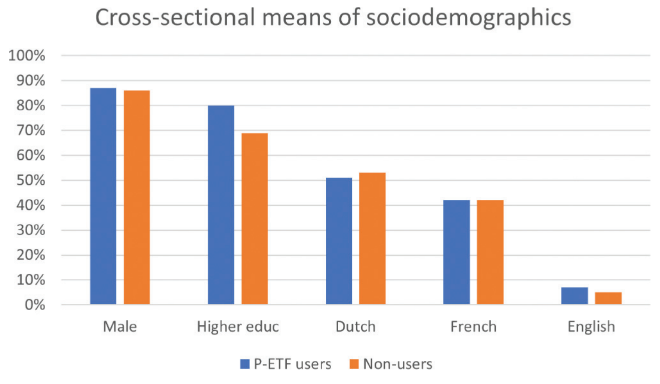
3.2. Descriptive statistics

Following Bhattacharya et al. (2017), we classify as P-ETF users any investors who traded a P-ETF *at least once* over the whole sample period (i.e., January 2003-March 2012). To drop outliers, we use trading experience as filter. Specifically, we exclude investors whose trading experience length is shorter than five months to disregard very short-lived investors as in Glaser and Weber (2009) and Bellofatto et al. (2018).

19 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, most brokers and banks tend to consider almost all financial securities with random returns as complex instruments.

20 Such items are usually covered in investment policy statements (IPSs) used in portfolio management delegation. According to Boone and Lubitz (2004), the main elements in an IPS are goals and objectives, fears and concerns, investment time frame, expected mortality, retirement time frame, shorter-term financial needs, risk tolerance attitudes, and risk capacity. In our data, the S-test covers most of these elements.

Figure 1: Histogram of cross-sectional means for sociodemographic indicators
This figure shows the distribution of investors (in %) depending on gender, higher education and spoken language. For each variable, the left-hand bar refers to P-ETF users while the right-hand bar refers to non-users. 'Higher educ' refers to investors with a university degree or equivalent.



3.2.1. Sociodemographic variables

The set of sociodemographic indicators includes age, gender, education and spoken language.[21] We use dummy variables for gender (Gender=1 for men) and education (Higher education=1 if the investor holds a university degree or equivalent). Regarding language, we opt for three different dummy variables: Dutch-speaking (=1 for investors who selected Dutch), French-speaking (=1 for investors who selected French), and English-speaking (=1 for investors who chose English). All these binary variables will be used as usual control variables in the ZINB regression model in Section 4.

Figure 1 provides the histogram of cross-sectional means for these sociodemographic variables with a distinction between P-ETF users and non-users. Table 1 reports the corresponding descriptive statistics. P-ETF users are on average somewhat older than non-users (50.32 for P-ETF users versus 47.91 for non-users). The proportion of men is high and similar in the two subsamples, i.e., 87% among P-ETF users versus 86% among non-users. Although statistically significant at the 5% level, this

21 Belgium has three official languages: French, Dutch and German. French and Dutch are spoken the most. Nevertheless, retail investors had to choose from the three languages available on the online trading platform: French, Dutch or English.

Table 1: Cross-sectional statistics for sociodemographic indicators

Metrics		P-ETF users Mean	P-ETF users Std Dev	P-ETF users N	Non-users Mean	Non-users Std Dev	Non-users N	Diff
Age (2012)	in years	50.32	13.40	3,484	47.91	13.24	22,656	2.41***
Gender	dummy=1 if male	0.87	0.33	3,484	0.86	0.35	22,656	0.01**
Higher education	dummy=1 if university	0.80	0.41	3,484	0.69	0.46	22,656	0.11***
Dutch-speaking	dummy=1 if Dutch	0.51	0.50	3,484	0.53	0.50	22,656	−0.02**
French-speaking	dummy=1 if French	0.42	0.49	3,484	0.42	0.49	22,656	0.00
English-speaking	dummy=1 if English	0.07	0.26	3,484	0.05	0.21	22,656	0.02***

This table reports the cross-sectional means, standard deviations, mean differences, as well as the number of investors among the subsamples of P-ETF users and non-users. For each investor, 'Age' is determined as the difference between 2012 and the year of birth. 'Gender' is a dummy variable equal to one for men. 'Higher education' is a dummy variable set to one for investors with a university degree or equivalent. 'Dutch-speaking', 'French-speaking' and 'English-speaking' are three additional dummies based on the language selected by the investor on the online trading platform. *, **, and *** indicate that means significantly differ at the level of 10%, 5% or 1%, respectively.

Table 2: Statistics about MiFID questionnaires

	Number of investors	A-test	S-test
P-ETF users	3,484	3,484	2,142
	15%	100%	61%
Non-users	22,656	22,656	10,696
	85%	100%	47%
Total	26,140	100%	49%

This table reports basic statistics about the number of P-ETF users and nonusers who completed both the A-test and the S-test.

difference is negligible. Regarding education, the proportion of investors who hold a university degree is significantly higher among the P-ETF users than among non-users (80% versus 69%). Regarding spoken language, the proportion of Dutch-speaking investors is the highest in both subsamples (53% and 51%). The proportion of French-speaking investors is exactly the same (42%) in both subsamples, while the proportion of English-speaking investors is slightly higher among the P-ETF users (7% versus 5%).

3.2.2. Subjective variables

Since MiFID questionnaires carry relevant and valuable information (e.g., Mazzoli and Marinelli, 2014; D'Hondt and Roger, 2017; Broihanne and Orkut, 2018; Bellofatto et al., 2018; D'Hondt et al., 2021; Broihanne and Orkut, 2021; Orkut, 2021), we use the two MiFID questionnaires (i.e., A-test and S-test) to define several subjective variables. As mentioned in Section 2, subjective measures of financial literacy are particularly helpful in profiling investors (e.g., Courchane and Zorn, 2005; Xiao et al., 2011; Allgood and Walstad, 2016; Anderson et al., 2017; Lind et al., 2020).

Table 2 provides the number of investors who completed both the A-test and the S-test, with a distinction between P-ETF users and non-users. All the investors in the sample filled in the A-test,[22] but only 49% of them also completed the S-test. Note that the brokerage house that

22 The brokerage house that provided us with the data implemented the A-test for an exhaustive list of instruments, including shares traded on a non-European market or on a European non-regulated market (such as Multilateral Trading Facilities under the MiFID typology). Because of this precautionary principle, all the retail investors who traded in the post-MiFID period (i.e., after November 2007) took the A-test.

provided us with the data did not offer portfolio management delegation services over the period under scrutiny. The motive for completing the S-test was receiving free access to a financial advice tool on stocks. This tool, directly available on the broker's trading platform, did not provide personalized advice but financial information on stocks, analysts' reports, and professional recommendations. In our sample, 61% of P-ETF users took the S-test while only 47% of non-users did so.

Using the items available in the A-test, we summarize information into four subjective variables, which allows us to attribute a specific score to each investor for each of them. We follow the same approach with the different items available in the S-test and summarize information into five additional subjective variables of interest. The cross-sectional statistics are reported in Table 3, in Panel A for the variables built upon the A-test and in Panel B for those based on the S-test.

Panel A of Table 3 clearly shows that P-ETF users display a higher financial literacy than non-users. The average score of P-ETF users regarding the 'Knowledge of financial markets' is 1.72, while the corresponding average for non-users is only 1.29. Similarly, when focusing on "complex" instruments, P-ETF users declared a relatively higher level of knowledge (7.22 vs 5.32), self-reported more experience (5.96 vs 4.33) and stated being more aware of the risk of incurring losses (7.67 vs 6.96). All these differences are statistically significant at the 1% level.

Consistent results emerge in Panel B of Table 3. Among the investors who completed the S-test, P-ETF users declared on average a higher financial literacy than non-users (5.01 vs 4.41). Moreover, P-ETF users exhibit on average a slightly higher risk tolerance than non-users (3.99 vs 3.91). P-ETF users also display a higher average score for both variables 'Investment time horizon' and 'Wealth and financial situation'. This means that P-ETF users are, on average, wealthier and declare longer investment horizons than non-users. Again, all these differences are significant at the 1% level. The only exception is the average level of return objective that does not statistically differ between the two subsamples of investors.

3.2.3. Objective variables

Using the trading records, we first rebuild the end-of-month individual portfolios and then compute a set of objective variables aiming at characterizing the actual trading activity on stocks. Building on past empirical

Table 3: Cross-sectional statistics for subjective variables

Metrics		P-ETF users			Non-users			
		Mean	Std Dev	N	Mean	Std Dev	N	Diff
Panel A: A-test	(Score Min-Max)							
Knowledge of financial markets	(score 0-3)	1.72	0.89	3,484	1.29	0.97	22,656	0.43***
Knowledge of "complex" instruments	(score 0-18)	7.22	5.78	3,484	5.32	5.39	22,656	1.89***
Experience with "complex" instruments	(score 0-18)	5.96	5.69	3,484	4.33	5.16	22,656	1.63***
Awareness of losses with "complex" instruments	(score 0-9)	7.67	2.74	3,484	6.96	3.43	22,656	0.71***
Panel B: S-test	(Score Min-Max)							
Risk tolerance	(score 1-5)	3.99	0.91	2,142	3.91	0.95	10,696	0.08***
Financial markets and products knowledge	(score 0-7)	5.01	1.35	2,142	4.41	1.43	10,696	0.60***
Investment time horizon	(score 2-11)	8.24	2.44	2,142	7.51	2.67	10,696	0.73***
Wealth and financial situation	(score 0-17)	9.09	3.58	2,142	8.07	3.63	10,696	1.02***
Return objective	(score 1-5)	3.25	0.99	2,142	3.26	1.07	10,696	−0.02

This table provides cross-sectional means and standard deviations, as well as the number of investors and mean differences, with a distinction between P-ETF users and non-users. Panel A refers to the A-test, while Panel B refers to the S-test. 'Knowledge of financial markets' refers to one specific question where investors have to self-assess their knowledge of financial markets on a 4-level scale (wherein Level 0 is associated with a basic knowledge and Level 3 refers to an experienced investor "who manages any aspect of the financial markets"). For 'Knowledge of "complex" instruments', 'Experience with "complex" instruments' and 'Awareness of losses with "complex" instruments', the score elicits the investor's self-assessment regarding his/her own abilities, past experience and skills. For example, a score of 18 (out of 18) for 'Knowledge of "complex" instruments' means that the investor considers himself/herself as an expert in the field. In the opposite, a score of 0 indicates that the investor declared no knowledge about "complex" instruments. For the variables 'Risk tolerance', 'Financial markets and products knowledge', 'Investment time horizon' and 'Wealth and financial situation', the score reflects the investor's direct self-assessment on the corresponding item. 'Return objective' refers to the annual yield investors expect on their investments, selected on a 5-level scale wherein level 1 corresponds to "a yield without any risk of capital loss" and level 5 to "the inflation rate + more than 12% per year". *,**, and *** indicate that means significantly differ at the level of 10%, 5% or 1%, respectively.

research, we compute the following variables for each investor: (1) number of stock trades over the whole sample period; (2) trading experience in months, expressed as the number of months between the first and last trades on stocks (e.g., Nicolosi et al., 2009; Bellofatto et al., 2018); (3) number of different stocks traded (NDST) over the entire sample period (e.g., Feng and Seasholes, 2005; Koestner et al., 2017); (4) trade duration computed as the average number of days between two consecutive trades on stocks (e.g., D'Hondt and Roger, 2017); (5) average number of stocks held in end-of-month portfolios (e.g., Bellofatto et al., 2018; D'Hondt et al., 2021); (6) stock trading intensity (STI), which is a dummy variable equal to one if the investor's monthly portfolio turnover[23] is in the upper quartile of the sample and equals zero otherwise; and (7) underperformance on stocks (UPS), which is equal to one when the investor's net excess Sharpe ratio on stocks is in the lower quartile of the sample and equals zero otherwise.[24]

Table 4 shows that both the 'Number of stock trades' and the 'Number of different stocks traded' exhibit a higher mean (almost double) for P-ETF users than for non-users, i.e., 133.6 and 33.87 for P-ETF users versus 71.94 and 17.27 for non-users. This reveals that P-ETF users not only trade much more on stocks but also trade a larger universe of stocks. PETF users also exhibit a longer trading experience than non-users (56 versus 46 months). In addition, P-ETF users display on average a shorter trade duration on stocks (57 days for P-ETF users vs 91 days for non-users), suggesting that P-ETF users trade stocks more frequently. Furthermore, the average P-ETF user holds an eight-stock portfolio, while the average non-user holds a five-stock portfolio. Interestingly, the proportion of P-ETF users with intense trading activity on stocks (STI) is lower than that of non-users. Similarly, the proportion of underperformers on stocks (UPS) is approximately 5 percentage points lower among P-ETF users than among non-users. All these differences are highly significant (both economically and statistically) and overall indicate that P-ETF users are more experienced traders on stocks than non-users.

23 Following the approach of Hoffmann et al. (2013) and Bellofatto et al. (2018), the portfolio turnover is measured 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 end of this month. Barber and Odean (2001) and Koestner et al. (2017) also use turnover to measure what they call overtrading.

24 As in Bellofatto et al. (2018), we calculate the net Sharpe ratio in excess of the Sharpe ratio of the market. This performance measure is computed using the Eurostoxx 600 index as the market portfolio and the 12-month Belgian T-bill rate as the risk-free rate.

Table 4: Cross-sectional statistics for objective variables

Metrics		P-ETF users Mean	P-ETF users Std Dev	P-ETF users N	Non-users Mean	Non-users Std Dev	Non-users N	Diff
Number of stock trades		133.6	295.5	3,484	71.94	161.8	22,656	61.67***
Trading experience	(in months)	56.43	27.36	3,484	46.43	26.63	22,656	10.00***
Number of different stocks traded (NDST)		33.87	38.38	3,484	17.27	21.09	22,656	16.60***
Trade duration	(in days)	57.07	96.40	3,484	91.12	158.5	22,656	−34.05***
Number of stocks		8.28	8.40	3,484	4.84	5.30	22,656	3.44***
Stock trading intensity (STI)	Dummy=1 if turnover >= Q3	0.220	0.414	3,484	0.245	0.430	22,656	−0.025***
Underperformance on stocks (UPS)	Dummy=1 if Net excess Sharpe ratio <= Q1	0.203	0.402	3,484	0.257	0.437	22,656	−0.054***

This table reports cross-sectional means and standard deviations, as well as number of investors and mean differences, with a distinction between P-ETF users and non-users. 'Number of stock trades' is the total number of trades executed on stocks. 'Trading experience' is computed as the difference between the date of the last trade and the date of the first trade on stocks, expressed in number of months. 'Number of different stocks traded' (NDST) is the number of different stocks traded. 'Trade duration' is computed as the average number of days between two consecutive trades on stocks. 'Number of stocks' is the average number of stocks held in the end-of-month portfolio. 'Stock trading intensity' (STI) is a dummy variable set to one if the investor's monthly portfolio turnover (i.e., the average of the absolute values 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) is in the upper quartile of the sample and set to zero otherwise. 'Underperformance on stocks' (UPS) is dummy variable equal to one when the investor's net excess Sharpe ratio on stocks is in the lower quartile of the sample and equal to zero otherwise. All these variables are defined over the whole sample period (i.e., January 2003–March 2012). *, **, and *** indicate that means significantly differ at the level of 10%, 5% or 1%, respectively.

3.2.4. Behavioral variables

Using the trading records at hand, we build a fourth set of variables made of proxies for the behavioral biases reviewed in Section 2, namely overconfidence, narrow framing, underdiversification, and home bias.

Combining the two aforementioned objective variables measuring stock trading intensity (STI) and stock underperformance (UPS), we create the proxy for overconfidence as in Odean (1999), Barber and Odean (2001) and Bailey et al. (2011). This proxy is a dummy variable equal to one when the investor simultaneously trades stocks intensively and underperforms (i.e., when the two dummies STI and UPS are set to one) and zero otherwise.

To measure narrow framing, we adopt the approach of Kumar and Lim (2008) and Bailey et al. (2011) by using the degree of clustering in the investors' trades.[25] For each investor i, we compute trade clustering (TC_i) as 1–*(number of trading days$_i$/number of trades$_i$)*. However, because TC_i is correlated with portfolio size, number of stocks, and trading frequency, we opt for the adjusted measure based on the peer group (ATC_i),[26] which is computed as TC_i– *mean trade clustering of the peer group*. A positive (negative) value for ATC_i indicates that the trades of investor i are more (less) clustered than those of their counterparts who have the same number of stocks, the same trading frequency and a similar portfolio size. Based on ATC_i, we build a dummy variable to capture narrow framing, where the latter is equal to one when the ATC_i is less than zero and zero otherwise.[27]

We use two proxies for underdiversification. The first proxy, called 'Trading underdiversification', is based on the number of different stocks traded during the whole period. It reflects how broad the investor's universe of stocks is (Bellofatto et al., 2018). More specifically, we consider that an investor suffers from a lack of diversification when his/her set of different stocks traded is in the lower quartile of the sample. The second proxy for underdiversification aims at capturing lack of diversification at

25 These authors especially identify whether an investor executes trades separately (i.e., one trade at a time) or executes multiple trades simultaneously.

26 A peer group is made of investors who are in the same quartile in terms of portfolio size, trading frequency, and number of stocks. An investor can belong to one of the sixty-four groups.

27 Kumar and Lim (2008) state that "a low *TC* measure for an investor indicates that her trades are temporally separated and thus the degree of narrow framing is likely to be higher. In particular, the tradeclustering measure is zero for investors who execute each trade on a separate day. These investors are more likely to adopt narrow decision frames in their investment choices".

the stock portfolio level. To build it, we rely on the monthly average Herfindahl-Hirschman index (HHI hereafter) that is equal to the sum of the squared stock portfolio weights (e.g., Goetzmann and Kumar, 2008; Koestner et al., 2017). HHI ranges from 0 for well-diversified portfolios to 1 for portfolios including only one stock. Our second proxy for underdiversification, called 'Portfolio underdiversification', is set to 1 when the investor's monthly HHI falls in the upper quartile of the sample and zero otherwise. The correlation between these two proxies for underdiversification is equal to 0.41 (significant at the 1% level), suggesting they are complementary and not substitutes.

Regarding home bias, we take into account the small size of Belgium (and of the Belgian stock market) and we compute for each investor the 'Home bias ratio', which captures the concentration of his/her trades on the Belgian stock market and on those of neighbouring countries (France, the Netherlands, Germany and Luxembourg).[28] As in Bailey et al. (2011), we compute this ratio over the full sample period as the aggregate number of trades on companies headquartered in the home and neighbouring countries divided by the total number of trades. By construction, this ratio ranges between 0 and 1, and is equal to one when all trades are executed on stocks issued in these five countries.

Table 5 reports cross-sectional statistics for our behavioral variables, with a distinction between P-ETF users and non-users. Overall, P-ETF users appear less subject to behavioral biases. The proportion of overconfident investors is higher among non-users than among PETF users (10.5% versus 7.2%). This is consistent with Table 4, wherein both stock trading intensity (STI) and underperformance on stocks (UPS) prevail more among non-users (25% and 26% versus 22% and 20%). Next, the two proxies for underdiversification indicate that P-ETF users suffer less from it than non-users. The proportion of P-ETF users (non-users) whose set of stocks traded belongs to the lower quartile is approximately 7% (17%). Similarly, the proportion of P-ETF users (non-users) whose monthly HHI is in the upper quartile is approximately 14% (28%). In both cases, the difference is substantial. Regarding home bias, the average ratio is slightly higher among non-users than among P-ETF users (0.750

28 To exit Belgium and reach a neighbouring country, the *maximum* distance a citizen would need to travel is approximately 200 kilometers. It is therefore not surprising that many Belgians feel a sense of familiarity and comfort when going in any of these neighbouring countries. Some citizens even feel more at home when going aboard than in the other linguistic regions within their home country.

Table 5: Descriptive statistics for behavioral variables

		P-ETF users			Non-users			
Metrics		Mean	Std Dev	N	Mean	Std Dev	N	Diff
Overconfidence	Dummy=1 if STI=1 & UPS=1	0.072	0.258	3,484	0.105	0.306	22,656	−0.033***
Narrow framing	Dummy =1 if ATC<= 0	0.561	0.496	3,484	0.574	0.494	22,656	−0.014
Trade underdiversification	Dummy=1 if NDST <= Q1	0.069	0.254	3,484	0.166	0.373	22,656	−0.097***
Portfolio underdiversification	Dummy=1 if HHI >= Q3	0.137	0.344	3,484	0.276	0.447	22,656	−0.139***
Home bias ratio	(score between 0 & 1)	0.725	0.254	3,484	0.751	0.297	22,656	−0.026***

This table reports cross-sectional means and standard deviations, as well as the number of investors and mean differences, with a distinction between P-ETF users and non-users. 'Overconfidence' is a dummy variable set to one if the investor displays both overtrading (i.e., when 'Stock trading intensity' (STI) is set to one) and underperformance (i.e., when 'Underperformance on stocks' (UPS) is set to one) and zero otherwise. 'Narrow framing' is a dummy variable set to one if the investor's adjusted trade clustering (defined as trade clustering minus the mean of trade clustering of the peer group) is less than zero and zero otherwise. 'Trade underdiversification' is a dummy variable set to one if the investor's number of different traded stocks is in the lower quartile of the sample and zero otherwise. 'Portfolio underdiversification' is a dummy variable set to one if the investor's monthly HHI (defined as the sum of the squared stock portfolio weights) is in the upper quartile of the sample and zero otherwise. 'Home bias ratio' is computed for any investor as the number of trades executed on stocks issued in the home country (Belgium) or in neighbouring countries (France, the Netherlands, Germany, and Luxembourg) divided by the total number of trades. *,**, and *** indicate that means statistically differ at the level of 10%, 5% or 1%, respectively.

versus 0.725). This suggests that the trades executed by non-users are more concentrated locally than those completed by P-ETF users. All these differences are statistically significant at the 1% level. The only exception is narrow framing, for which no significant difference is found between the two subsamples.

4. Empirical results

To sketch the profile of P-ETF users who took the two MiFID tests, we opt for a zero-inflated negative binomial (ZINB) model (Lambert, 1992). Such a model allows us to not only determine the probability for an investor to invest in P-ETFs but also explain the magnitude of his/her trades on these financial instruments. This "two-step-in-one" approach is fresh, compared to the models used in past research on retail trading in investment funds. Bailey et al. (2011) rely on a binary logistic model to examine whether a combination of behavioral factors impact the use of mutual funds by individuals. Bhattacharya et al. (2017) also uses a logit model to study the drivers of retail participation in P-ETFs, while Müller and Weber (2010) prefers a probit model to investigate the drivers of ETF/index fund awareness and choice.

The dependent variable in our ZINB model is the total number of trades on P-ETFs for each investor.[29] Basically, the idea behind this ZINB regression model is that the decision to invest in P-ETFs and the positive counts of trades in P-ETFs are generated by separate processes. This feature is especially appropriate to deal *at once* with both the group of investors who have never invested in P-ETFs (i.e., non-users) and the group of investors who traded P-ETFs to either a small or a large extent (i.e., P-ETF users). In other words, the ZINB model is particularly suited for our count data that exhibit excess zeros and overdispersion, as illustrated in Figure 2.

Building on Cameron and Trivedi (2013), the ZINB model can be specified as:

29 Since our dependent variable can take integer values (count data), Poisson regression models might have been an appropriate alternative. However, there is a dominance of zeros in our data since approximately 83% of the retail investors in our sample did not trade any P-ETFs. The ZINB model is actually a modified Poisson regression model used to tackle two common issues in Poisson regressions for count data (e.g., Ridout et al., 1998; Sheu et al., 2004; Burger et al., 2009). The first issue is overdispersion, which means that the sample variance is larger than the sample mean (while the variance of the Poisson distribution must be equal to its mean). The second issue is the excess of zeros, i.e., the situation in which there is an excessive number of observations equal to 0 (Greene, 2000).

$$P[y=j]=\begin{cases}\pi+(1-\pi)f_2(0), & \text{if } j=0\\ (1-\pi)f_2(j), & \text{if } j>0,\end{cases}$$

where the base count density is the negative binomial distribution $f_2(y)$,[30] y is the number of P-ETF trades, and j is any nonnegative value. To inflate the probability of a zero, we add a separate component, $\pi = f_1(0)$ such that the probability π depends on a set of regressors in a binary logistic model.[31] The ZINB model can be viewed as a weighted mixture of two components, $\pi = f_1(0)$ and $f_2(y = 0)$, used to predict the probability for $j = 0$, i.e., the probability of belonging to the group of non-users. The ZINB model is reduced to $(1-\pi)f_2(j)$ if the number of P-ETF trades is strictly positive, characterizing the probability of trading P-ETFs at least once.

Table 6 presents the results for three specifications of the above ZINB model, with a distinction between the logit ($\pi = f_1(0)$) and the negative binomial regression ($f_2(y)$) components. These results are based on 12,838 cross-sectional observations, i.e., 12,838 retail investors who completed the S-test (see Table 2). Panel A refers to the sociodemographic indicators that are included in the three specifications (Model 1, Model 2, and Model 3). Panel B refers to the subjective variables (from the S-test)[32] that are also used in the three models. Panel C is devoted to the set of objective variables related to trading added in both Model 2 and Model 3. Panel D refers to our proxies for common behavioral biases included in the last version of the model (Model 3). Sociodemographic indicators in Panel A are usual control variables. Panels B and C include variables of interest to test our hypotheses H1/H1* about financial sophistication, while variables in Panel D are useful to test our hypotheses H2/H2* about being subject to common behavioral biases. Table 9 reporting the Variance Inflation Factors (VIF) is available in appendix and shows that multicolinearity is not an issue in the three models.

A key point regarding the interpretation of the ZINB models is that the signs of the coefficient estimates in the binary logistic component are often in the opposite direction to the signs of the coefficient estimates in the negative

30 $f_2(y) = P(y|x) = \frac{\Gamma(y_i+\theta)}{\Gamma(\theta)y_i!}\mu_i^{\theta}(1-\mu_i)^{y_i}$, where $\theta > 0$, $y_i = 0, 1, 2, \ldots$, $\mu_i = \frac{\theta}{(\theta+\lambda_i)}$, $E(y_i) = \lambda_i$ and $Var(y_i) = \lambda_i(1 + \alpha\lambda_i)$.

31 $\pi = f_1(0) = P(y = 0 \mid x) = G(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k) = G(\beta_0 + \mathbf{x}\beta_k)$ with $G(.)$ being the logistic cumulative distribution function, where $G(z) \in]0, 1[$ for all $z \in \Re$, and $G(z) = exp(z)/[1 + exp(z)]$.

32 Similar results are obtained when considering the subjective variables built upon the A-test. As explained in Section 3.1, the A-test focuses on financial knowledge and experience, while the S-test also cover investment objectives and financial situation.

Figure 2: Histogram of investors as a function of P-ETF trades

This figure shows the distribution of investors depending on the total number of P-ETF trades. The horizontal axis refers to the number of P-ETF trades, while the vertical axis refers to the number of investors. The left-hand distribution includes both P-ETF users and non-users. The right-hand distribution includes only P-ETF users.

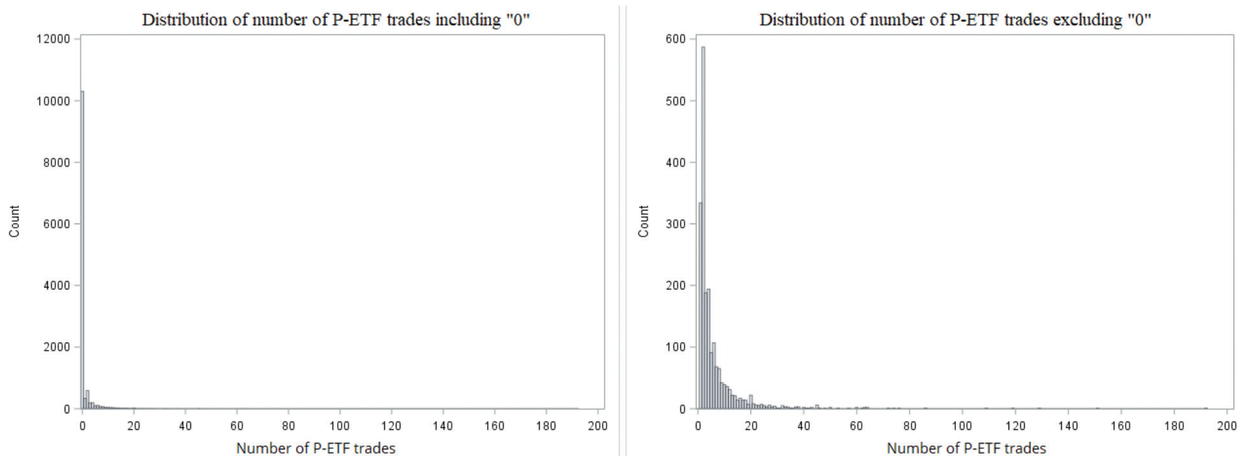


Table 6: Results of the ZINB model

	Model 1				Model 2				Model 3			
	Logit		Negative Binomial		Logit		Negative Binomial		Logit		Negative Binomial	
Parameter	Coefficient	OR	Coefficient	IRR	Coefficient	OR	Coefficient	IRR	Coefficient	OR	Coefficient	IRR
Dependent variable Number of P-ETF trades												
Independent variables												
Panel A : Sociodemographic variables												
Gender male (Dummy)	−0.001	0.999	0.059	1.061	−0.004	0.996	−0.027	0.973	0.006	1.006	−0.021	0.979
Age (in Years)	−0.009***	0.992	0.021***	1.021	0.001	1.001	0.011***	1.011	0.001	1.001	0.011***	1.011
Higher education (Dummy)	−0.313***	0.731	0.118	1.126	−0.259***	0.772	0.173**	1.188	−0.258***	0.772	0.162**	1.175
Dutch-speaking (Dummy)	0.055	1.056	−0.118	0.888	0.120*	1.127	−0.018	0.982	0.129*	1.138	−0.011	0.989
English-speaking (Dummy)	−0.442**	0.643	0.224	1.252	−0.426***	0.653	0.453***	1.572	−0.384**	0.681	0.376***	1.457
Panel B : Subjective variables from S-test												
Financial markets and products knowledge (score 0-7)	−0.287***	0.751	0.151***	1.163	−0.244***	0.784	0.121***	1.128	−0.242***	0.785	0.115***	1.122
Return objective (score 1-5)	0.066*	1.068	−0.140***	0.870	0.084**	1.088	−0.148***	0.863	0.098***	1.103	−0.140***	0.869
Risk tolerance (score 1-5)	0.168***	1.182	0.088**	1.092	0.144***	1.154	0.053	1.054	0.138***	1.148	0.044	1.045
Investment time horizon (score 2-11)	−0.078***	0.925	0.004	1.004	−0.055***	0.947	0.032**	1.033	−0.053***	0.949	0.036**	1.036
Wealth and financial situation (score 0-17)	−0.026**	0.974	−0.001	0.999	−0.008	0.992	0.005	1.005	−0.008	0.992	0.002	1.002
Panel C : Objective variables from trading												
Log(Trading experience)					−0.233***	0.792	0.259***	1.295	−0.212***	0.809	0.341***	1.407
Log(Trade duration)					0.020	1.021	−0.574***	0.563	0.034	1.035	−0.604***	0.547
Log(Number of stocks)					−0.516***	0.597	−0.486***	0.615	−0.487***	0.614	−0.463***	0.630
Stock trading intensity (STI)					−0.143	0.867	−0.355***	0.701	−0.235**	0.790	−0.420***	0.657
Underperformance on stocks (UPS)					0.045	1.046	0.078	1.081	−0.091	0.913	−0.008	0.993

Table 6 (Continued)**:** Results of the ZINB model

	Model 1				Model 2				Model 3			
	Logit		Negative Binomial		Logit		Negative Binomial		Logit		Negative Binomial	
Parameter	Coefficient	OR	Coefficient	IRR	Coefficient	OR	Coefficient	IRR	Coefficient	OR	Coefficient	IRR
Panel D : Behavioral variables												
Overconfidence									0.392**	1.479	0.257	1.293
Narrow framing									−0.039	0.962	−0.015	0.985
Trade underdiversification									0.020	1.020	0.700***	2.014
Portfolio underdiversification									0.140	1.150	−0.031	0.969
Home bias ratio									0.297**	1.345	−0.308**	0.735
α (dispersion)	3.32***	27.51			2.06***	7.81			2.00***	7.389		
N	12,838				12,838				12,838			
%(Number of P-ETF trades >0)	16.68%				16.68%				16.86%			
Log likelihood	−11,237				−10,929				−10,906			
AIC	22,520				21,924				21,897			
BIC	22,691				22,270				22,218			

This table reports the results of three ZINB regression models, with a distinction between the logistic and negative binomial components. The dependent variable is the number of P-ETF trades. Panel A refers to sociodemographic indicators. 'Gender' is a dummy variable equal to one for men. 'Age' is defined as the difference between 2012 and the investor's year of birth. 'Higher Education' is a dummy variable set to one for investors with a university degree (or equivalent). 'Dutch-speaking' and 'English-speaking' are dummies based on the language selected by the investor on the online trading platform. Panel B includes subjective variables (from the S-test). For 'Financial markets and products knowledge', 'Risk tolerance', 'Investment time horizon' and 'Wealth and financial situation', the score reflects the investor's direct self-assessment regarding the item. 'Return objective' refers to the annual yield investors expect on their investments, selected on a 5-level scale, wherein level 1 corresponds to 'a yield without any risk of capital loss' and level 5 to 'the inflation rate + more than 12% per year'. Panel C includes objective variables characterizing trading activity. 'Trading experience' is computed as the difference between the date of the last trade and the date of the first trade, expressed in number of months. 'Trade duration' is computed as the average number of days between two consecutive trades on stocks. 'Number of stocks' is the average number of stocks held in the end-of-month portfolio. 'Stock trading intensity' (STI) is a dummy variable set to one if the investor's monthly portfolio turnover (i.e., the average of the absolute values 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) is in the upper quartile of the sample and zero otherwise. 'Underperformance on stocks' (UPS) is dummy variable equal to one when the investor's net excess Sharpe ratio on stocks is in the lower quartile of the sample and zero otherwise. Panel D refers to behavioral variables. 'Overconfidence' is a dummy variable set to one if the investor displays both overtrading (STI=1) and underperformance on stocks (UPS=1) and zero otherwise. 'Narrow framing' is a dummy variable set to one if the investor's adjusted trade clustering (defined as trade clustering minus the mean trade clustering of the peer group) is less than zero and zero otherwise. 'Trade underdiversification' is a dummy variable set to one if the investor's number of different traded stocks is in the lower quartile of the sample and zero otherwise. 'Portfolio underdiversification' is a dummy variable set to one if the investor's monthly HHI (defined as the sum of the squared stock portfolio weights) is in the upper quartile of the sample and zero otherwise. 'Home bias ratio' is computed as the number of trades executed on stocks issued in the home country (Belgium) or in neighbouring countries (France, the Netherlands, Germany, and Luxembourg) divided by the total number of trades. 'IRR' is the incidence rate ratio, which is equal to the exponential of the corresponding coefficient estimate in the negative binomial component of the model. 'OR' is the odds ratio, which is equal to the exponential of the corresponding coefficient estimate in the logistic component of the model. *,**, and *** indicate that the coefficient estimate is statistically significant at 10%, 5% or 1%, respectively.

binomial component (Long et al., 2006). In our specific case, the binary logistic component of the model predicts the probability of *not* investing in P-ETFs; i.e., a negative coefficient estimate implies therefore a higher probability of being a P-ETF user. By contrast, the negative binomial component of the model predicts the number of P-ETF trades; i.e., a positive coefficient estimate indicates a higher magnitude of P-ETF trading.

4.1. The probability of trading P-ETFs

The results of estimating sequentially the three model specifications are provided in Table 6. For the sake of brevity, we focus on Model 3 that includes all the variables necessary to test our hypotheses H1 and H2.[33]

Regarding the sociodemographic indicators (Panel A), 'Higher education' is positively related to the probability of trading P-ETFs since the negative coefficient estimate indicates a lower probability of *not* investing in P-ETFs. More precisely, the odds of *not* investing in P-ETFs decrease by a meaningful factor of 0.772 for retail investors who hold a university degree. Put differently, this means that, All being else equal, the odds of *not* investing in PETFs decrease by 22.8% for retail investors with a university degree.[34] Language also seems to play some role in the sample: compared to French-speaking investors, English-speaking investors are more likely to trade P-ETFs[35] while Dutch-speaking investors are somewhat less likely to do so.

In Panel B, four subjective variables (out of five) are significantly associated with the probability of investing in P-ETFs (at the 1% level). On the one hand, 'Financial markets and products knowledge' and 'Investment time horizon' exhibit negative coefficient estimates, meaning that investors with higher subjective financial literacy and investors who declare longer investment horizons are more likely to use P-ETFs. Specifically, the odds of *not* investing in P-ETFs decrease by 21.5% (5%) for retail investors who report higher financial literacy (longer investment horizons). These findings allow us to validate H1 about the positive impact of financial sophistication

33 The overdispersion parameter α is statistically significant at the 1% level for the three ZINB models, which provides strong support for a negative binomial model (instead of a Poisson model). Regarding the goodness of fit, Model 3 is the best since the log likelihood increases from Model 1 to Model 3. Model 3 also displays the lowest AIC and BIC values.

34 If retail investors without university degree are used as the reference group, it implies that the odds of *not* investing in P-ETFs increase by 29.5% (=1/0.772 -1) for these investors compared to those who hold a university degree.

35 All being else equal, the odds of *not* investing in P-ETFs decrease by 31.9% for English-speaking investors, compared to French-speaking investors. In contrast, the odds of *not* investing in P-ETFs increase by 46.8% (=1/0.681 -1) for French-speaking retail investors, compared to English-speaking investors.

on the probability of using P-ETFs. Note that these results are consistent with those documented in Enete et al. (2018) and Bellofatto et al. (2018) about the positive link between subjective financial literacy and the probability of trading investment funds. On the other hand, Panel B reports positive coefficient estimates for both return objective and risk tolerance. Considering the odds ratios, the probability of *not* investing in P-ETFs increases by about 10% (15%) for investors who declare higher return objectives[36] (higher risk tolerance[37]).

In Panel C of Table 6, the coefficient estimate of 'Trading experience'[38] is significant (at the 1% level) and exhibits a negative sign. All being else equal, the odds of *not* investing in P-ETFs decrease by 19.1% when the length of trading experience doubles. This further validates H1 on the positive relationship between financial sophistication (here measured by actual experience) and the probability of investing in P-ETFs. Next, two other objective measures of trading activity are also significant. First, the 'Number of stocks' held in the end-of-month portfolio is positively associated with the probability of investing in P-ETFs. Its odds ratio is equal to 0.614, indicating that the odds of *not* investing in P-ETFs decrease when portfolio size increases. Second, 'Stock trading intensity' is also positively related to the probability of being a P-ETF user. The odds of *not* investing in P-ETFs for investors who trades stocks intensively decrease by a factor of 0.79. Considering that stock portfolio size and trading intensity may be indicative of financial sophistication, these findings are consistent with H1.

Results reported in Panel D of Table 6 are insightful about the influence of behavioral biases on the probability of trading P-ETFs (H2). Among the four biases under scrutiny, two of them exhibit significant coefficient estimates (at the 5% level). First, overconfidence emerges as a strong determinant of the probability of investing in P-ETFs. Specifically, when retail investors are overconfident, their odds of *not* investing

36 All being else equal, retail investors who targeted higher return objectives (than the lowest level in the scale) by one score unit have a 1.103 times higher chance of *not* investing in P-ETFs than those who selected the lowest level.

37 Retail investors who are less risk-tolerant might choose to invest more in diversified products such as P-ETFs because they can reduce the overall risk and volatility of their portfolios by diversifying away specific risk. By spreading investments across different assets, P-ETFs can help mitigate the impact of negative events affecting a specific investment. They lower their exposure to any single investment and minimize the potential losses associated with concentrated holdings. Investors who are less risk-tolerant might also be more inclined to prioritize capital preservation over aggressive bets, leading them to prefer diversified products over individual stocks.

38 Based on Glaser and Weber (2009), we use the natural logarithm of the trading experience, the trade duration, the average number of stocks held in a portfolio, and the turnover, since all these variables are positively skewed.

in P-ETFs increase by 47.9%, which is significant. This finding is in line with Bailey et al. (2011) who report that even though overconfident investors have higher allocations in mutual funds, they have a smaller proportion of their portfolio invested in index funds because they focus more on actively managed funds. This leads us to validate H2 for overconfidence. Second, home bias is another determinant of the probability of investing in P-ETFs. Looking at the odds ratio, when retail investors are subject to home bias, their odds of *not* investing in P-ETFs increase by 34.5%. Therefore, H2 is also validated for home bias. Regarding narrow framing and underdiversification, the results are inconclusive.

To sum up, Table 6 provides empirical evidence allowing us to fully validate H1 and validate H2 for overconfidence and home bias. Hence, retail investors who trade P-ETFs have a distinct profile. Compared to non-users, P-ETF users are more likely to have a university degree, and they declare higher financial literacy, a longer investment horizon, a lower return objective and lower risk tolerance. They also have longer trading experience, hold more diversified stock portfolios, and display stronger stock trading intensity. These investors are also less subject to overconfidence and home bias.

4.2. The magnitude of P-ETF trades

Unlike the logistic component of the ZINB model that is estimated on the full sample of retail investors, it is worth recalling that the negative binomial component is estimated on the subsample of P-ETF users only. The latter represents 13.33% of the retail investors in the sample (see Section 3). The results for the negative binomial component of the ZINB model in Table 6 are relevant to identify the factors that truly drive the magnitude of P-ETF investment (among the P-ETF users) and test our hypotheses H1* and H2*.

Among the sociodemographic indicators (see Panel A), three variables emerge as significant drivers of the number of P-ETF trades. First, the level of education is positively related to the magnitude of P-ETF investment. Considering the incidence rate ratio (IRR),[39] P-ETF users who hold

39 The odds ratio is only appropriate when the dependent variable is a dummy variable. The incidence rate ratio (IRR) is the equivalent measure in the negative binomial regression where the dependent variable takes any nonnegative values. The IRR corresponds to the exponential of the coefficient estimate and measures the estimated percentage change in the number of P-ETF trades for each unit variation in the explanatory variable. For example, IRR =2 means that the estimated number of the dependent variable doubles for each unit increase in the explanatory variable.

a university degree execute on average 17.5% more P-ETF trades than those with a lower level of education. Second, language is another significant determinant of the magnitude of P-ETF trading. Specifically, the English-speaking P-ETF users execute on average 45.7% more trades on P-ETFs than their French-speaking counterparts. Next, to some extent, age is another significant driver of the number of P-ETF trades. When age increases by one year, P-ETF trades are expected to increase by 1.01%.

When focusing on Panel B, the magnitude of P-ETF investment is explained by three subjective variables. Note that these individual characteristics have also been previously identified as significant determinants of the probability of trading P-ETFs. First, subjective financial literacy is positively related to the number of P-ETF trades. More precisely, the number of P-ETF trades increases by 12.2% when financial literacy increases by one level (on the scale), which is substantial in economic terms. Second, the number of P-ETF trades is positively associated with longer investment horizons: the number of P-ETF trades increases by 3.6% when investment horizon increases by one level (on the scale). These two findings allow us to validate H1* about the positive impact of financial sophistication on the magnitude of P-ETF trading. Next, P-ETF users who declare higher return objectives execute less trades on P-ETFs. The number of P-ETF trades decreases by a sizeable factor of 0.869 (i.e., −13.1%) when the return objective increases by one level (on the scale).

Panel C in Table 6 provides evidence allowing us to further validate H1* using actual trading experience. P-ETF users who have more experience execute more P-ETF trades. Specifically, the coefficient estimate indicates that the number of P-ETF trades increases by 40.7% when trading experience (in months) doubles. By contrast, trade duration on stocks, portfolio size (i.e., the number of stocks in the end-of-month portfolio), and stock trading intensity are negatively related to the magnitude of P-ETF investment. In other words, PETF users execute more trades on P-ETFs when they hold more concentrated stock portfolios and do not trade stocks intensively. This suggests a trade-off between the magnitude of PETF investing and the magnitude of stock trading. P-ETF users who trade P-ETFs more often not only hold a lower number of stocks in their portfolio but also rebalance their stock holdings less frequently. This evidence points to a substitution effect between stocks and P-ETFs.

Regarding our hypothesis H2*, Panel D of Table 6 provides findings showing that both underdiversification and home bias are significantly related to the number of P-ETF trades. On the one hand, consistent with the abovementioned substitution effect between stocks and P-ETFs, underdiversification (based on the universe of stocks traded) is positively associated with the magnitude of P-ETF investing. All being else equal, the number of P-ETF trades increases by 101.4% for P-ETF users who concentrate their trading on a few stocks (compared to their counterparts who trade a broader universe of stocks). This evidence clearly reinforces the idea that some P-ETF users prefer P-ETFs over stocks or use P-ETFs as substitutes for stocks. This result is also consistent with Bellofatto et al. (2018) who find that investors with higher financial literacy tend to concentrate their stock portfolio more extensively and achieve diversification through investment fund holdings. This leads us to reject H2* for underdiversification. On the other hand, Panel D reveals that home bias is negatively related to the number of P-ETF trades (at the 5% level). More specifically, there is a decline in trading activity for P-ETFs as the home bias ratio increases. This indicates that investors with a stronger tendency towards home bias engage in fewer P-ETF transactions compared to those less influenced by this bias. H2* is therefore validated for home bias. There is no significant evidence in Table 6 supporting H2* for either overconfidence or narrow framing. Overconfidence does not explain the magnitude of P-ETF investment, although it negatively affects the probability of trading P-ETFs.

To summarize, the number of P-ETF trades is positively associated with age, higher education, higher financial literacy, longer investment horizon, and lower return objective. The number of P-ETFs trades is also positively related to the length of trading experience. Of particular interest, our findings provide evidence of a trade-off between the magnitude of P-ETF trades and the trading activity on stocks since investors who actively trade P-ETFs tend to hold more concentrated stock portfolios (i.e., smaller portfolio size), and they display longer trade duration on stocks and lower stock portfolio turnover. We also find that stock holdings of P-ETFs users are less subject to home bias as their trading activity in P-ETFs increases.

4.3. Robustness check

Table 6 reports the results of estimating the models using the entire sample period January 2003-March 2012. However, since the MiFID regulation came into force in late 2007, investors who started trading before MiFID enforcement have accumulated some trading experience before completing the MiFID questionnaires. This experience might have influenced their answers as well as their behavior in the post-MiFID period, i.e., from January 2008 to March 2012. As a robustness check, we run the same three models on a subsample of 12,838 investors, excluding all investors who started trading before January 2008. The purpose is to check whether the information content of the MiFID survey data (captured by subjective variables in Panel B) remains unchanged. The results are available in Table 8 in appendix. Overall, the main findings are similar and we notice only a few changes. In Model 3, two variables are no longer significant in the logistic component, i.e., 'Dutch-speaking' and 'Log(Trading experience)'. For the latter, the filter that was applied to define this new sample has mechanically reduced heterogeneity in terms of trading experience; this is likely to explain why the length of experience is no longer significant for the probability of trading P-ETFs in Table 8. Regarding the negative binomial component of Model 3, three variables are no longer significant, namely 'Higher education', 'English-speaking' and 'Stock trading intensity'. By contrast, the coefficient estimates in both components of Model 3 are still significant for our subjective variables (in Panel B) and our behavioral variables (in Panel D). This confirms the robustness of our main results.

5. Conclusion

In this study, we focus on retail investors who invest in P-ETFs, referred to as P-ETF users, to sketch their profile and identify their main motivations. By doing so, we add to the literature by determining a comprehensive profile of these P-ETF users relying on both actual trading accounts and survey-based data. To the best of our knowledge, this combination of field and survey data in profiling P-ETF investors is unique and allows a large coverage of individual characteristics. Building on past research, we formulate testable hypotheses to examine the role of financial sophistication and common behavioral biases in retail P-ETF trading. To further contribute to the literature, we opt for a zero-inflated negative binomial (ZINB) regression model, which allows us to explain at once the *likelihood* of trading P-ETFs and the *magnitude* of P-ETF trading. This model is also well-suited for our count data that exhibit excess zeros and overdispersion.

We find strong evidence that retail investors who trade P-ETFs have a distinct profile. Compared to non-users, P-ETF users are more sophisticated investors. They are more likely to have a university degree, and they declare higher financial literacy, a longer investment horizon, a lower return objective and lower risk tolerance. They also have longer trading experience, hold more diversified stock portfolios, and display stronger stock trading intensity. These investors are also less subject to overconfidence and home bias.

We identify similar drivers and additional determinants when we zoom in on the subsample of P-ETF users to analyse the extent to which they trade P-ETFs. Higher education, higher subjective financial literacy, longer investment horizons, lower return objectives, and longer trading experience are all factors that significantly boost the magnitude of P-ETF investment. In addition, the number of P-ETF trades is positively associated with age and negatively related to home bias. Of particular interest, our findings provide evidence of a trade-off between the magnitude of P-ETF trades and the trading activity on stocks. Retail investors who trade P-ETFs more intensively hold a lower number of stocks in their end-of-month portfolio and rebalance their stock holdings less extensively. This particular result points to a substitution effect between stocks and P-ETFs as the trading in P-ETFs increases. We might conjecture that the most active P-ETF users apply a core-satellite portfolio strategy, in which P-ETFs are the core and stocks represent the satellite. Further research is obviously needed to better identify such a strategy as well as its performance.

As a potential limitation of our empirical work, we note that our sample period includes the 2008 financial crisis. Some retail investors in our sample might have been hit very hard by this crisis, which might have impacted their subjective financial literacy and risk tolerance, as well as their return objectives. More broadly, since our MiFID data were collected by the brokerage house through a one-shot questionnaire, information self-reported once by investors is therefore considered stable over the whole period. This is another shortcoming of this study since these individual characteristics might evolve over time.

This paper paves the way for further research on retail investing in P-ETFs. First, future research should examine whether our main findings can be generalized across time, especially in noncrisis periods. Investigating whether there is substantial variation in the behaviors of P-ETF users would be of particular interest. Future research could analyze how P-ETF investors with different levels of activity on P-ETFs respond to stressful market conditions, or to upward versus downward market trends.

6. Appendix

Table 7: The ten most traded P-ETFs in the sample

ETF	Underlying	Number trades	Volume trades
LYXOR ETF BEL 20	Equity	2,348	11,646,158.95
GOLD BULLION SEC	Commodities	1,607	31,304,732.32
US NAT GAS FD ETF	Commodities	1,015	5,315,626.30
U.S. OIL FUND ETF	Commodities	925	8,357,962.81
ETF BRA IBOVESPA LYX	Equity	890	5,724,898.89
LYXOR ETF CHINA ENT.	Equity	809	11,547,219.10
LYXOR ETF CAC 40	Equity	784	12,794,247.82
LYXOR ETF DJ ES 50	Equity	738	11,870,854.82
ETFS NATURAL GAS	Commodities	599	4,219,940.55
LYXOR ETF MSCI INDIA	Equity	582	5,449,180.23
Other (303 passive ETF)	–	12,079	92,588,179.85
Total		22,376	200,819,001.6

This table reports the underlying assets, the number of trades and the volume of trades (in EUR) of the 10 most traded P-ETFs in the sample.

Table 8: Results of the ZINB model for the subsample of retail investors trading from January 2008

	Model 1				Model 2				Model 3			
	Logit		Negative Binomial		Logit		Negative Binomial		Logit		Negative Binomial	
Parameter	Coefficient	OR	Coefficient	IRR	Coefficient	OR	Coefficient	IRR	Coefficient	OR	Coefficient	IRR
Dependent variable Number of P-ETF trades												
Independent variables												
Panel A : Sociodemographic variables												
Gender male (Dummy)	0.029	1.030	0.139	1.149	0.009	1.009	0.080	1.083	0.014	1.014	0.148	1.159
Age (in Years)	−0.003	0.997	0.018***	1.018	0.004	1.004	0.014***	1.014	0.004	1.004	0.014***	1.014
Higher education (Dummy)	−0.398***	0.671	0.108	1.114	−0.298**	0.742	0.238	1.269	−0.317**	0.728	0.169	1.184
Dutch-speaking (Dummy)	0.032	1.033	0.008	1.008	0.080	1.083	0.019	1.019	0.071	1.074	−0.002	0.998
English-speaking (Dummy)	−0.605**	0.546	0.219	1.245	−0.619**	0.539	0.392	1.480	−0.450*	0.638	0.326	1.385
Panel B : Subjective variables (S-test)												
Financial markets and products knowledge (score 0-7)	−0.166***	0.847	0.182***	1.205	−0.170***	0.843	0.164***	1.178	−0.153***	0.858	0.160***	1.174
Return objective (score 1-5)	0.059	1.060	−0.239***	0.787	0.085	1.088	−0.244***	0.783	0.129**	1.137	−0.215***	0.807
Risk tolerance (score 1-5)	0.181***	1.198	0.186***	1.205	0.182***	1.200	0.138**	1.148	0.166**	1.180	0.121*	1.129
Investment time horizon (score 2-11)	−0.097***	0.908	0.023	1.024	−0.070***	0.932	0.057**	1.058	−0.062**	0.940	0.078***	1.081
Wealth and financial situation (score 0-17)	−0.032*	0.968	−0.002	0.999	−0.020	0.981	−0.005	0.995	−0.017	0.983	−0.009	0.991
Panel C : Objective variables												
Log(Trading experience)					−0.172	0.842	0.278**	1.320	−0.161	0.852	0.412***	1.510
Log(Trade duration)					−0.086	0.917	−0.481***	0.618	−0.092	0.912	−0.530***	0.589
Log(Number of stocks)					−0.617***	0.539	−0.500***	0.607	−0.545***	0.580	−0.424***	0.655
Stock trading intensity (STI)					−0.153	0.859	−0.118	0.889	−0.354*	0.702	−0.188	0.829
Underperformance on stocks (UPS)					−0.072	0.930	0.061	1.062	−0.296**	0.744	−0.067	0.936

Table 8 (Continued)**:** Results of the ZINB model for the subsample of retail investors trading from January 2008

	Model 1				Model 2				Model 3			
	Logit		Negative Binomial		Logit		Negative Binomial		Logit		Negative Binomial	
Parameter	Coefficient	OR	Coefficient	IRR	Coefficient	OR	Coefficient	IRR	Coefficient	OR	Coefficient	IRR
Panel D : Behavioral variables												
Overconfidence									0.537**	1.711	0.279	1.322
Narrow framing									0.080	1.083	0.066	1.069
Trade underdiversification									0.061	1.063	0.824***	2.279
Portfolio underdiversification									0.276	1.318	−0.012	0.988
home bias ratio									0.862***	2.367	−0.391**	0.676
α (dispersion)	2.78***	16.039			1.86***	6.45			1.70***	5.437		
N	4,762				4,762				4,762			
%(Number of P-ETF trades >0)	13.02%				13.02%				13.02%			
Log likelihood	−3,338				−3,255				−3,227			
AIC	6,723				6,576				6,541			
BIC	6,871				6,790				6,819			

This table reports the results of three ZINB regression models, with a distinction between the logistic and negative binomial components. The dependent variable is the number of P-ETF trades. Panel A refers to sociodemographic indicators. 'Gender' is a dummy variable equal to one for men. 'Age' is defined as the difference between 2012 and the investor's year of birth. 'Higher Education' is a dummy variable set to one for investors with a university degree (or equivalent). 'Dutch-speaking' and 'English-speaking' are dummies based on the language selected by the investor on the online trading platform. Panel B includes subjective variables (from the S-test). For 'Financial markets and products knowledge', 'Risk tolerance', 'Investment time horizon' and 'Wealth and financial situation', the score reflects the investor's direct self-assessment regarding the item. 'Return objective' refers to the annual yield investors expect on their investments, selected on a 5-level scale, wherein level 1 corresponds to 'a yield without any risk of capital loss' and level 5 to 'the inflation rate + more than 12% per year'. Panel C includes objective variables characterizing trading activity. 'Trading experience' is computed as the difference between the date of the last trade and the date of the first trade, expressed in number of months. 'Trade duration' is computed as the average number of days between two consecutive trades on stocks. 'Number of stocks' is the average number of stocks held in the end-of-month portfolio. 'Stock trading intensity' (STI) is a dummy variable set to one if the investor's monthly portfolio turnover (i.e., the average of the absolute values 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) is in the upper quartile of the sample and zero otherwise. 'Underperformance on stocks' (UPS) is dummy variable equal to one when the investor's net excess Sharpe ratio on stocks is in the lower quartile of the sample and zero otherwise. Panel D refers to behavioral variables. 'Overconfidence' is a dummy variable set to one if the investor displays both overtrading (STI=1) and underperformance on stocks (UPS=1) and zero otherwise. 'Narrow framing' is a dummy variable set to one if the investor's adjusted trade clustering (defined as trade clustering minus the mean trade clustering of the peer group) is less than zero and zero otherwise. 'Trade underdiversification' is a dummy variable set to one if the investor's number of different traded stocks is in the lower quartile of the sample and zero otherwise. 'Portfolio underdiversification' is a dummy variable set to one if the investor's monthly HHI (defined as the sum of the squared stock portfolio weights) is in the upper quartile of the sample and zero otherwise. 'Home bias ratio' is computed as the number of trades executed on stocks issued in the home country (Belgium) or in neighbouring countries (France, the Netherlands, Germany, and Luxembourg) divided by the total number of trades. 'IRR' is the incidence rate ratio, which is equal to the exponential of the corresponding coefficient estimate in the negative binomial component of the model. 'OR' is the odds ratio, which is equal to the exponential of the corresponding coefficient estimate in the logistic component of the model. *,**, and *** indicate that the coefficient estimate is statistically significant at 10%, 5% or 1%, respectively.

Table 9: Variance Inflation Factors

	Model 1	Model 2	Model 3
Panel A : Sociodemographic variables			
Gender male (Dummy)	1.0143	1.0209	1.0220
Age (in Years)	1.1920	1.3026	1.3072
Higher education (Dummy)	1.0742	1.0821	1.0835
Dutch-speaking (Dummy)	1.0943	1.0985	1.1035
English-speaking (Dummy)	1.0664	1.0691	1.0886
Panel B : Subjective variables (S-test)			
Financial markets and products knowledge (score 0-7)	1.4463	1.4840	1.4896
Return objective (score 1-5)	1.2171	1.2467	1.2648
Risk tolerance (score 1-5)	1.3291	1.3339	1.3371
Investment time horizon (score 2-11)	1.3437	1.3787	1.3791
Wealth and financial situation (score 0-17)	1.3416	1.3514	1.3526
Panel C : Objective variables			
Log(Trading experience)		1.2780	1.4498
Log(Trade duration)		2.2111	2.6477
Log(Number of stocks)		2.4597	2.8688
Stock trading intensity (STI)		2.0361	2.6421
Underperformance on stocks (UPS)		1.0993	1.6080
Panel D : Behavioral variables			
Overconfidence			2.6477
Narrow framing			1.1343
Trade underdiversification			1.6483
Portfolio underdiversification			1.4329
Home bias ratio			1.0566

This table reports the Variance Inflation Factors (VIF) when estimating OLS regression models. The dependent variable is the number of P-ETF trades. Panel A refers to sociodemographic indicators. 'Gender' is a dummy variable equal to one for men. 'Age' is defined as the difference between 2012 and the investor's year of birth. 'Higher Education' is a dummy variable set to one for investors with a university degree (or equivalent). 'Dutch-speaking' and 'English-speaking' are dummies based on the language selected by the investor on the online trading platform. Panel B includes subjective variables (from the S-test). For 'Financial markets and products knowledge', 'Risk tolerance', 'Investment time horizon' and 'Wealth and financial situation', the score reflects the investor's direct self-assessment regarding the item. 'Return objective' refers to the annual yield investors expect on their investments, selected on a 5-level scale, wherein level 1 corresponds to 'a yield without any risk of capital loss' and level 5 to 'the inflation rate + more than 12% per year'. Panel C includes objective variables characterizing trading activity. 'Trading experience' is computed as the difference between the date of the last trade and the date of the first trade, expressed in number of months. 'Trade duration' is computed as the average number of days between two consecutive trades on stocks. 'Number of stocks' is the average number of stocks held in the end-of-month portfolio. 'Stock trading intensity' (STI) is a dummy variable set to one if the investor's monthly portfolio turnover (i.e., the average of the absolute values 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) is in the upper quartile of the sample and zero otherwise. 'Underperformance on stocks' (UPS) is dummy variable equal to one when the investor's net excess Sharpe ratio on stocks is in the lower quartile of the sample and zero otherwise. Panel D refers to behavioral variables. 'Overconfidence' is a dummy variable set to one if the investor displays both overtrading (STI=1) and underperformance on stocks (UPS=1) and zero otherwise. 'Narrow framing' is a dummy variable set to one if the investor's adjusted trade clustering (defined as trade clustering minus the mean trade clustering of the peer group) is less than zero and zero otherwise. 'Trade underdiversification' is a dummy variable set to one if the investor's number of different traded stocks is in the lower quartile of the sample and zero otherwise. 'Portfolio underdiversification' is a dummy variable set to one if the investor's monthly HHI (defined as the sum of the squared stock portfolio weights) is in the upper quartile of the sample and zero otherwise. 'Home bias ratio' is computed as the number of trades executed on stocks issued in the home country (Belgium) or in neighbouring countries (France, the Netherlands, Germany, and Luxembourg) divided by the total number of trades.

Table 10: 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%?	
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.

References

Allgood, S. and W. B. Walstad (2016). The effects of perceived and actual financial literacy on financial behaviors. *Economic Inquiry 54*(1), 675–697.

Anderson, A., F. Baker, and D. T. Robinson (2017). Precautionary savings, retirement planning and misperceptions of financial literacy. *Journal of Financial Economics 126*(2), 383–398.

Babiarz, P. and C. A. Robb (2014). Financial literacy and emergency saving. *Journal of Family and Economic Issues 35*(1), 40–50.

Bailey, W., A. Kumar, and D. Ng (2011). Behavioral biases of mutual fund investors. *Journal of Financial Economics 102*(1), 1–27.

Baker, H. K. and J. R. Nofsinger (2002). Psychological biases of investors. *Financial Services Review 11*(2), 97.

Barber, B. M. and T. Odean (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The Journal of Finance 55*(2), 773–806.

Barber, B. M. and T. Odean (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics 116*(1), 261–292.

Barberis, N., M. Huang, and R. H. Thaler (2006). Individual preferences, monetary gambles, and stock market participation: A case for narrow framing. *American Economic Review 96*(4), 1069–1090.

Bellofatto, A., C. D'Hondt, and R. De Winne (2018). Subjective financial literacy and retail investors' behavior. *Journal of Banking & Finance 92*, 168–181.

Bhattacharya, U., B. Loos, S. Meyer, and A. Hackethal (2017). Abusing ETFs. *Review of Finance 21*(3), 1217–1250.

Bianchi, M. (2018). Financial literacy and portfolio dynamics. *The Journal of Finance 73*(2), 831–859.

Blasques, F., V. Holý, and P. Tomanová (2023). Zero-inflated autoregressive conditional duration model for discrete trade durations with excessive zeros. *Studies in Nonlinear Dynamics & Econometrics, Forthcoming*.

Boone, N. M. and L. S. Lubitz (2004). *Creating an Investment Policy Statement: Guidelines & Templates*. AdvisorPress.

Boyle, P., L. Garlappi, R. Uppal, and T. Wang (2012). Keynes meets markowitz: The trade-off between familiarity and diversification. *Management Science 58*(2), 253–272.

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

Broihanne, M.-H. and H. Orkut (2018). Investment goals and mental accounting in French retail clients. *Finance 39*(1), 107–144.

Burger, M., F. Van Oort, and G.-J. Linders (2009). On the specification of the gravity model of trade: zeros, excess zeros and zero-inflated estimation. *Spatial Economic Analysis 4*(2), 167–190.

Calvet, L. E., J. Y. Campbell, and P. Sodini (2009). Measuring the financial sophistication of households. *American Economic Review 99*(2), 393–98.

Camerer, C. and D. Lovallo (1999). Overconfidence and excess entry: An experimental approach. *American Economic Review 89*(1), 306–318.

Cameron, A. C. and P. K. Trivedi (2013). *Regression analysis of count data*, Volume 53. Cambridge university press.

Carpentier, C. and J. Suret (2011). La rationalité des décisions des investisseurs. Technical report, Rapport préparé pour le compte de l'Autorité des marchés financiers et du Ministère des finances du Québec.

Cheng, S., M. Massa, and H. Zhang (2018). The Unexpected Activeness of Passive Investors: A Worldwide Analysis of ETFs. *The Review of Asset Pricing Studies 9*(2), 296–355.

Cho, H., C. Liu, J. S. Preisser, and D. Wu (2023). A bivariate zero-inflated negative binomial model and its applications to biomedical settings. *Statistical Methods in Medical Research 32*(7), 1300–1317.

Courchane, M. and P. Zorn (2005). Consumer literacy and credit worthiness. In *Federal Reserve System Conference, Promises and Pitfalls: as consumer options multiply, who is being served and at what cost*.

Davydov, D., O. Florestedt, J. Peltomäki, and M. Schön (2017). Portfolio performance across genders and generations: The role of financial innovation. *International Review of Financial Analysis 50*, 44–51.

De Bondt, W. F. and R. H. Thaler (1995). Financial decision-making in markets and firms: A behavioral perspective. *Handbooks in Operations Research and Management Science 9*, 385–410.

D'Hondt, C., R. De Winne, and M. Merli (2021). Do retail investors bite off more than they can chew? A close look at their return objectives. *Journal of Economic Behavior and Organization 188*, 879–902.

D'Hondt, C., R. McGowan, and P. Roger (2020). Trading leveraged Exchange-Traded Products is hazardous to your weath. *The Quarterly Review of Economics and Finance 80*, 287–302.

D'Hondt, C. and P. Roger (2017). Investor sentiment and stock return predictability: The power of ignorance. *Finance 38*(2), 7–37.

Domian, D. L., D. A. Louton, and M. D. Racine (2007). Diversification in portfolios of individual stocks: 100 stocks are not enough. *Financial Review 42*(4), 557–570.

Dorn, D. and G. Huberman (2005). Talk and Action: What Individual Investors Say and What They Do. *Review of Finance 9*(4), 437–481.

Enete, S., M. Reiter, W. Usrey, A. Scott, and M. Seay (2018). Characteristics of etf owners: Exploring the role of investor knowledge, fee aversion, and financial advice seeking. Technical report, Working Paper, 2019 Academic Research Colloquium for Financial Planning and Related Disciplines.

Evans, J. L. and S. H. Archer (1968). Diversification and the reduction of dispersion: an empirical analysis. *The Journal of Finance 23*(5), 761–767.

Farinella, J. and R. Kubicki (2018). The performance of exchange traded funds and mutual funds. *Journal of Accounting & Finance 18*(4), 1–55.

Feng, L. and M. S. Seasholes (2005). Do investor sophistication and trading experience eliminate behavioral biases in financial markets? *Review of Finance 9*(3), 305–351.

Fischoff, B. (1982). For those condemned to study the past: Heuristics and biases in hindsight. In D. Kahneman, P. Slovic, and A. Tversky (Eds.), *Judgment Under Uncertainty: Heuristics and Biases*, Chapter 23, pp. 335–351. Cambridge University Press Cambridge.

French, K. R. and J. M. Poterba (1991). Investor diversification and international equity markets. *American Economic Review 81*(2), 222–226.

Gaudecker, H.-M. V. (2015). How does household portfolio diversification vary with financial literacy and financial advice? *The Journal of Finance 70*(2), 489–507.

Glaser, M. and M. Weber (2009). Which past returns affect trading volume? *Journal of Financial Markets 12*(1), 1–31.

Goetzmann, W. N. and A. Kumar (2008). Equity portfolio diversification. *Review of Finance 12*(3), 433–463.

Greene, W. H. (2000). *Econometric analysis (International edition)*. Pearson US Imports & PHIPEs.

Grinblatt, M. and M. Keloharju (2001). How distance, language, and culture influence stockholdings and trades. *The Journal of Finance 56*(3), 1053–1073.

Hilgert, M. A., J. M. Hogarth, and S. G. Beverly (2003). Household financial management: The connection between knowledge and behavior. *Fed. Res. Bull. 89*, 309.

Ho, W. Y., P. Wang, and J. D. Alba (2009). Merger and acquisition fdi, relative wealth and relative access to bank credit: Evidence from a bivariate zero-inflated count model. *International Review of Economics & Finance 18*(1), 26–30.

Hoffmann, A. O., T. Post, and J. M. Pennings (2013). Individual investor perceptions and behavior during the financial crisis. *Journal of Banking & Finance 37*(1), 60–74.

Hsiao, Y.-J. and W.-C. Tsai (2018). Individual investor trading and return patterns around earnings announcements. *Journal of Banking and Finance 88*, 15–29.

Kelly, M. (1995). All their eggs in one basket: Portfolio diversification of us households. *Journal of Economic Behavior & Organization 27*(1), 87–96.

Koestner, M., B. Loos, S. Meyer, and A. Hackethal (2017). Do individual investors learn from their mistakes? *Journal of Business Economics 87*(5), 669–703.

Korniotis, G. M. and A. Kumar (2013). Do portfolio distortions reflect superior information or psychological biases? *Journal of Financial and Quantitative Analysis 48*, 1–45.

Kumar, A. (2009). Who gambles in the stock market? *The Journal of Finance 64*(4), 1889–1933.

Kumar, A. and C. Lee (2006). Retail investors sentiment and returns comovements. *The Journal of Finance 61*, 2451–2486.

Kumar, A. and S. S. Lim (2008). How do decision frames influence the stock investment choices of individual investors. *Management Science 54*(6), 1052–1064.

Lambert, D. (1992). Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics 34*(1), 1–14.

Leal, C. C., G. Loureiro, and M. J. R. Armada (2017). Selling winners, buying losers: Mental decision rules of individual investors on their holdings. *European Financial Management 24*, 362–386.

Lind, T., A. Ahmed, K. Skagerlund, C. Strömbäck, D. Västfjäll, and G. Tinghög (2020). Competence, confidence, and gender: The role of objective and subjective financial knowledge in household finance. *Journal of Family and Economic Issues 41*(4), 626–638.

Lindblom, T., T. Mavruk, and S. Sjögren (2018). East or west, home is best: The birthplace bias of individual investors. *Journal of Banking & Finance 92*, 323–339.

Liu, Y.-J., M.-C. Wang, and L. Zhao (2010). Narrow framing: Professions, sophistication, and experience. *Journal of Futures Markets: Futures, Options, and Other Derivative Products 30*(3), 203–229.

Long, S. J., J. S. Long, and J. Freese (2006). *Regression models for categorical dependent variables using Stata*. Stata press.

Lusardi, A., P.-C. Michaud, and O. S. Mitchell (2017). Optimal financial knowledge and wealth inequality. *Journal of Political Economy 125*(2), 431–477.

Lusardi, A. and O. Mitchell (2007). Financial literacy and retirement planning: New evidence from the rand american life panel. Technical report, Michigan Retirement Research Center Research Paper No. WP 2007-157.

Lusardi, A. and O. Mitchell (2011). Financial literacy around the world: an overview. *Journal of Pension Economics & Finance 10*(4), 497–508.

Lusardi, A. and O. Mitchell (2014). The economic importance of financial literacy: Theory and evidence. *Journal of Economic Literature 52*, 5–44.

Lusardi, A., O. Mitchell, and V. Curto (2009). Financial literacy and financial sophistication among older americans. Technical Report No. w15469, National Bureau of Economic Research.

Lyons, A. C., L. Palmer, K. S. Jayaratne, and E. Scherpf (2006). Are we making the grade? a national overview of financial education and program evaluation. *Journal of Consumer Affairs 40*(2), 208–235.

Massa, M. and A. Simonov (2006). Hedging, familiarity and portfolio choice. *The Review of Financial Studies 19*(2), 633–685.

Mazzoli, C. and N. Marinelli (2014). Determinants of risk-suitable investment portfolios: Evidence from a sample of italian householders. *International Journal of Finance and Banking 1*(1), 13–28.

Moore, D. A. and P. J. Healy (2008). The trouble with overconfidence. *Psychological review 115*(2), 502.

Müller, S. and M. Weber (2010). Financial literacy and mutual fund investments: who buys actively managed funds? *Schmalenbach Business Review 62*(2), 126–153.

Nicolosi, G., L. Peng, and N. Zhu (2009). Do individual investors learn from their trading experience? *Journal of Financial Markets 12*(2), 317–336.

Odean, T. (1999). Do investors trade too much? *American Economic Review 89*(5), 1279– 1298.

Orkut, H. (2021). Foreign stock investment and sophistication of French retail investors. *Finance 42*, 39–80.

Polkovnichenko, V. (2005). Household portfolio diversification: A case for rank-dependent preferences. *The Review of Financial Studies 18*(4), 1467–1502.

Prather, L. J., T.-H. Chu, M. I. Mazumder, and J. C. Topuz (2009). Index funds or ETFs: the case of the S&;P 500 for individual investors. *Financial Services Review 18*(3), 213.

Ramiah, V., X. Xu, and I. A. Moosa (2015). Neoclassical finance, behavioral finance and noise traders: A review and assessment of the literature. *International Review of Financial Analysis 41*, 89–100.

Ridout, M., C. G. Demétrio, and J. Hinde (1998). Models for count data with many zeros. In *Proceedings of the XIXth international biometric conference*, pp. 179–192. International Biometric Society Invited Papers. Cape Town, South Africa.

Schroeder, A. (2008). *The snowball: Warren Buffett and the business of life*. AdvisorPress.

Sheu, M.-l., T.-w. Hu, T. E. Keeler, M. Ong, and H.-Y. Sung (2004). The effect of a major cigarette price change on smoking behavior in california: a zero-inflated negative binomial model. *Health Economics 13*(8), 781–791.

Singh, R. (2010). Behavioral finance studies: Emergence and developments. *Journal of Contemporary Management Research 4*(2), 1–9.

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization 1*(1), 39–60.

Titman, S., C. Wei, and B. Zhao (2022). Corporate actions and the manipulation of retail investors in china: An analysis of stock splits. *Journal of Financial Economics 145*(3), 762–787.

Van Rooij, M., A. Lusardi, and R. Alessie (2011). Financial literacy and stock market participation. *Journal of Financial Economics 101*(2), 449–472.

Wang, A. (2011). Younger generations' investing behaviors in mutual funds: Does gender matter? *The Journal of Wealth Management 13*(4), 13.

Xiao, J. J., C. Tang, J. Serido, and S. Shim (2011). Antecedents and consequences of risky credit behavior among college students: Application and extension of the theory of planned behavior. *Journal of Public Policy & Marketing 30*(2), 239–245.

Yip, K. C. and K. K. Yau (2005). On modeling claim frequency data in general insurance with extra zeros. *Insurance: Mathematics and Economics 36*(2), 153–163.