What drives retail investment in passive exchange-traded funds *

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

April 28, 2022

Abstract

We profile retail investors in passive exchange-traded funds (P-ETFs) using both trading records and survey-based data provided by a major brokerage firm for a very large pool of clients over nearly a decade. There is overall strong evidence that retail investors who trade P-ETFs are more sophisticated than those who do not. Both the probability and magnitude of retail P-ETF investing are driven by higher education and longer trading experience. Retail P-ETF investors self-report higher financial literacy, a longer investment horizon, and a lower target return. These investors are less overconfident, and their investment is less biased locally. The more active retail P-ETF users hold a lower number of stocks and modify the composition of their stock portfolio less extensively, pointing to a substitution effect between stocks and P-ETFs.

*The authors are grateful to the brokerage house for providing the data. We would like to thank participants at the 6th ISCEF 2020, European Financial Management Association Annual conference (2021), 13th RGS Doctoral Conference in Economics and LFIN seminar participants at UCLouvain. Any errors are the full responsibility of the authors.

†Louvain School of Management (LFIN-LIDAM, UCLouvain).

‡Louvain School of Management (LFIN-LIDAM, UCLouvain).

§IÉSEG School of Management, Univ. Lille, CNRS, UMR 9221 - LEM - Lille Economie Management; Louvain School of Management (LFIN-LIDAM, UCLouvain). Corresponding author. Email: mikael.petitjean@uclouvain.be - Address: Chaussée de Binche 151, 7000 Mons, Belgium.

1 Introduction

Passive exchange-traded funds (P-ETFs) seek to mimic the performance of the stock market or a subset of it. They have become one of the most attractive innovations in finance, mostly because they are cheaper, less opaque, and more tax efficient when compared to active ETFs which have historically failed to outperform the benchmark net of fees. Passive ETFs have attracted nearly $3 trillion in assets over the last 10 years, while active ETFs have drawn only $200 billion. Even Warren Buffett, who is cited in Schroeder (2008), p. 685, strongly recommend P-ETFs to retail investors who have by definition limited resources, i.e. limited capital, time, or skills: 'Stocks are the things to own over time. Productivity will increase and stocks will increase with it. There are only a few things you can do wrong. One is to buy or sell at the wrong time. Paying high fees is the other way to get killed. The best way to avoid both of these is to buy a low-cost index fund, and buy it over time. Be greedy when others are fearful, and fearful when others are greedy, but don't think you can outsmart the market [...] Very few people should be active investors'.

Very little is known about the profile of retail investors who follow Warren Buffet's advice and hold P-ETFs in their portfolio. To the best of our knowledge, Bhattacharya et al. (2017) is the only major reference on this topic, but they essentially focus on whether P-ETF retail investors display better performance. Compared to non-users, they find that P-ETF users do not improve their portfolio performance. Regarding the ETF users' profile more specifically, they document that P-ETF users are younger, wealthier (in terms of both portfolio value and overall wealth), and have a shorter relationship with brokerage.[1]

Despite these primary findings, no previous study has relied on both trading and survey-based data to establish the drivers of P-ETF investing and determine more comprehensively

[1]Other studies consider the population of retail investors who hold a very heterogeneous set of products. For example, Davydov et al. (2017) consider retail investors who invest in exchange-traded products (ETPs) which include ETFs, ETCs (exchange-traded commodities), and ETNs (exchange-traded notes). They provide evidence that younger investors may be relatively more skilful users of ETPs. They also find that ETP investors perform worse than those who invest in mutual funds. Enete et al. (2018) focus on ETF users only. They find that ETF users have more subjective financial knowledge than non-users. They nevertheless do not distinguish between passive and active ETFS, such as leveraged, inverse, or inverse leveraged ETFs. These active ETFs have a multiplier different from 1 and are not passive. D'Hondt et al. (2021) show that that the typical user of leveraged ETPs looks like an overconfident gambler willing to take a high risk, while inverse leveraged ETPs are used by some highly loss averse investors for hedging purposes. In our study, we exclude them.

the profile of retail investors who hold P-ETFs.[2] Our data set includes the trading records of 26,140 retail investors at a major brokerage firm from January 2003 to March 2012. It also contains matched survey-based information collected within the framework of the MiFID regulation in place in the European Union member states since November 2007.[3]

Contrary to Bhattacharya et al. (2017), we therefore use both types of data, which leads us to include four sets of factors potentially driving retail ETF investing, namely, sociodemographics (such as gender, age, education, and language), subjective individual characteristics (such as target return, risk tolerance, and financial literacy), objective measures of trading activity (such as experience and stock portfolio size), and behavioural variables (such as proxies for overconfidence, lack of diversification, and local bias).[4] We thereby minimize the risk of introducing an omission bias and are able to contrast what retail investors declare to what they actually do.

To further contribute to the literature, we opt for a zero-inflated negative binomial (ZINB) regression model. The benefit of using a ZINB model is twofold since it helps explain the *likelihood* of trading P-ETFs but also the *magnitude* of P-ETF trading. This approach allows us to go beyond what was done previously.

Our main findings may be summarized as follows. The retail investors who are more likely to invest in P-ETFs are those who are better educated, self-report both higher financial literacy and a longer investment horizon, exhibit longer trading experience, hold a better diversified stock portfolio, and trade stocks more intensively. In contrast, retail investors who declare a higher target return, who suffer more from either overconfidence or local bias, are less likely to invest in P-ETFs. When we only consider retail investors who use P-ETFs, we find that their trading activity in P-ETFs is boosted by higher education, better financial literacy, longer investment horizons, lower target returns, and stronger trading

[2]To the best of our knowledge, D'Hondt et al. (2021) is the only study that also relies on both trading and survey-based data to sketch the profile of ETF users. However, this study focuses on the typical retail investor who trades leveraged ETPs.

[3]MiFID stands for the Markets in Financial Instruments Directive. MiFID I (2004/39/EC) is known as the first version of this Directive, while a review of it was implemented in January 2018 (known as MiFID II (2014/65/UE)). 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`).

[4]Behavioural variables are based on transaction records as in Bailey et al. (2011) who investigate the effect of behavioral biases on the mutual fund choices of US individual investors.

experience. The most active investors in P-ETFs are also older and less subject to local bias. Furthermore, there is a trade-off between the magnitude of their investments in P-ETFs and the magnitude of their investments in stocks. Investors who trade P-ETFs more actively hold a lower number of stocks in portfolios and modify the composition of their stock portfolios less extensively (i.e., have a lower turnover). This result points to a substitution effect between stocks and P-ETFs as trading in P-ETFs increases.

This paper contributes to the ongoing debate over passive retail investing by providing strong evidence that retail investors who trade P-ETFs are more sophisticated than those who do not.

The rest of the paper is organized as follows. Section 2 reviews the relevant literature and presents our five testable hypotheses regarding the drivers of retail investment in P-ETFs. Section 3 describes our data and provides 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, considering sociodemographics, individual characteristics, trading activity, and behavioural biases.[5] A particularly interesting research question is whether P-ETF users differ in terms of financial literacy, financial experience, and behavioural biases. The reason why we focus on these three aspects is justified by the widely documented impact these factors have on both the investors' wealth and social welfare in general.

2.1 Financial literacy and retail investment in P-ETFs

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). There is ample evidence that financial literacy benefits investors. In summary, 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

[5] A detailed description of our variables is provided in Section 3.2.

Mitchell (2007)). It also helps reduce the negative effect of behavioural biases on wealth (e.g., Baker and Nofsinger (2002)) and the probability of underdiversification (e.g., Gaudecker (2015)). Empirical evidence also shows that less literate investors exhibit lower portfolio performance (Gaudecker (2015), Bellofatto et al. (2018), Bianchi (2018)).

The level of financial literacy of retail investors may usually be evaluated in two ways. When based on a set of questions addressing the basics at the root of saving and investment decisions,[6] measures of financial literacy are labelled as *objective* because actual knowledge is revealed through correct answers. By contrast, measures of *subjective* financial literacy rely on questions asking people to self-assess their financial knowledge and expertise. Although the correlation between objective and subjective literacy may not be taken for granted (Lusardi and Mitchell, 2014),[7] 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)). According to Courchane and Zorn (2005), subjective financial literacy is one of the most significant factors in determining the way people invest. In the same vein, Bellofatto et al. (2018) find that 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 less prone to the disposition effect and display higher performance.

Of particular interest for our research question, financial literacy is also a robust driver of awareness about investment alternatives to stocks. Using survey data, Müller and Weber (2010) provide evidence that there is a positive relationship between financial literacy and the probability of investing passively. Bellofatto et al. (2018) show that investors who report higher financial literacy are more likely to invest in funds; i.e., they tend to concentrate their portfolios on a small set of stocks and achieve diversification through investment funds holding. Enete et al. (2018) find that ETF users have more subjective or objective financial knowledge than non-users of ETFs.

Based on the aforementioned findings, we formulate our first two hypotheses to be tested with the ZINB model in Section 4.

[6]These basics include topics, such as compounding, inflation, and diversification.

[7]There might be exceptions to the rule (e.g., Lusardi and Mitchell (2011)), but individual investors tend to have a relatively reliable perception of their knowledge.

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

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

2.2 Financial experience and retail investment in P-ETFs

In addition to financial literacy, experience is another key element in the decision-making process of investing. For example, trading experience eliminates the reluctance of retail investors to realize losses (Feng and Seasholes, 2005) and positively influences overall investor behaviour (Lyons et al., 2006). Retail investors learn from their trading experience, as documented in Nicolosi et al. (2009) or Koestner et al. (2017).

Regarding investments in funds, Wang (2011) show that experience (in addition to knowledge and income) is an important factor impacting the way younger generations invest in mutual funds. Bailey et al. (2011) report that sophisticated investors, i.e., better informed, higher income, older, and more experienced investors, make good use of mutual funds. In particular, they hold a high proportion of funds for long periods of time, avoid high expense funds, and experience relatively good performance.

Based on prior evidence, we expect financial experience to be positively correlated with retail investment in P-ETFs. Hence, we formulate our next two hypotheses as follows.

H2. The more experienced retail investors are, the more likely they are to invest in P-ETFs.

H2*. Among the users of P-ETFs, the more experienced users trade P-ETFs more often.

2.3 Behavioral biases and retail investment in P-ETFs

Behavioral biases lead to costly investment mistakes among retail investors (e.g., Singh (2010), Ramiah et al. (2015), Koestner et al. (2017)). It is well established in the literature that individuals generally suffer from overconfidence[8] (e.g., Fischoff (1982), Moore and Healy

[8]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

(2008), De Bondt and Thaler (1995), and Camerer and Lovallo (1999)). The detrimental impact of overconfidence on trading behaviour and/or portfolio performance has been widely documented (among others, Barber and Odean (2000), Barber and Odean (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, which leads us to hypotheses H3 and H3*.

H3. Overconfident retail investors are less likely to invest in P-ETFs.

H3*. Among the users of P-ETFs, the more overconfident users trade P-ETFs less often.

Narrow framing is another recognized behavioural bias among retail investors. This bias refers to their tendency to evaluate and select assets in isolation (Barberis et al., 2006). In other words, contrary to the assumptions of Markowitz's portfolio choice theory, retail investors do not consider their portfolio as a whole. Kumar and Lim (2008) find that narrow framing is accentuated when portfolios contain a very low number of different stocks. Kumar (2009) report 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 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 likely to invest in both equity mutual funds and index funds. This might suggest that investors who suffer from narrow framing may have less willingness to invest in P-ETFs. Accordingly, we formulate the following two hypotheses.

H4. Retail investors who suffer more from narrow framing are less likely to invest in P-ETFs.

that one is better than the median person (i.e., better-than-average effect), and (c) overprecision in one's beliefs (sometimes labelled miscalibration).

H4*. Among the users of P-ETFs, those who suffer more from narrow framing trade P-ETFs less often.

Underdiversification is another well-documented investment mistake made by retail investors (Koestner et al., 2017). The majority of retail investors and households typically own only a few stocks in their portfolios (e.g., Kelly (1995); Polkovnichenko (2005)). For example, Kumar and Lee (2006) and Korniotis and Kumar (2013) document that a typical investor at a major U.S. discount brokerage house over the period 1991-1996 holds a four-stock portfolio (with a median of three). Consistently, Leal et al. (2017) reports a three-stock portfolio for a typical investor at a Portuguese brokerage firm over the 2003-2007 period, and Bellofatto et al. (2018) find a four-stock portfolio for the median investor at a Belgian brokerage house over the 2003-2012 period. This lack of diversification negatively affects financial well-being in general (e.g., Barber and Odean (2000)) and depends on the degree of investor sophistication (e.g., Goetzmann and Kumar (2008)).

Although transaction costs have usually been mentioned as one of the major drivers of underdiversification (Evans and Archer, 1968), they are no longer relevant because of the emergence of ETFs and index funds, which have allowed for a drastic reduction in explicit transaction costs (Domian et al., 2007). P-ETF users are arguably expected to be more aware of diversification principles than non-users of P-ETFs since P-ETFs are *de facto* inexpensive diversification tools at the disposal of investors. Hence, the non-users of P-ETFs are expected to hold less diversified stock portfolios than P-ETF users. The same reasoning might also apply to the subsample of P-ETF users: among P-ETF users, those whose stock portfolio is more underdiversified are expected to trade P-ETFs less extensively. Such expectations lead to hypotheses H5 and H5*:

H5. Retail investors whose stock portfolios are more underdiversified are less likely to invest in P-ETFs.

H5*. Among the users of P-ETFs, those who suffer more from underdiversification trade P-ETFs less often.

A particular case of underdiversification is the local bias, i.e., the situation in which investors overweight stocks from one geographical area within their portfolio because these stocks look more familiar—geographically, culturally or socially. In addition to the home bias initially pointed out by French and Poterba (1991), investors have preferences for companies with geographical or professional proximity (e.g., Bailey et al. (2011), Boyle et al. (2012), Massa and Simonov (2006)). Moreover, both cultural distance and language play a role, 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 local bias. Therefore, we posit our last two hypotheses as follows.

H6. Retail investors who suffer more from local bias are less likely to invest in P-ETFs.

H6*. Among the users of P-ETFs, those who suffer more from the local bias trade P-ETFs less often.

3 Data and sample

3.1 Data

Our data come from an online Belgian brokerage house and cover the trading accounts of 26,140 retail investors over the period January 2003–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). These P-ETF users trade 313 different P-ETFs, among which 196 are equities-linked and 72 commodities-linked.[9]

In addition to the trading data, we matched individual data that were either sociodemographic (age, gender, education, and spoken language) or survey-based. The descriptive statistics of the sociodemographics are provided in Subsection 3.2.1. The survey-based data comprise subjective investor characteristics[10] collected by the brokerage house within the context of the European MiFID regulation. In short, this piece of regulation has made it compulsory for investment firms to collect specific information about their retail clients'

[9]The list of the ten most-traded P-ETFs in the sample is given in Table 7 in appendix.

[10]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 characteristics.

needs and preferences. Accordingly, investment firms operating in the EU are obliged to submit questionnaires (which are then referred to as "MiFID tests") to their clients. Specifically, two tests are used depending on the client's request: (1) the Appropriateness test (hereafter, the A-test) and (2) the Suitability test (hereafter, the S-test). Investors who ask for financial advice and/or portfolio management services have to fulfil the S-test, which aims at assessing their level of knowledge and experience, their investment objectives and their financial situation.[11] Investors who ask only for order execution (without advice or portfolio management services) have to fulfil the A-test when trading "complex"instruments. This lighter test aims to ensure that the investor has the necessary experience and knowledge to understand the risks involved in these "complex"financial instruments.

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 behavioural variables, i.e., proxies used to characterize overconfidence, narrow framing, lack of diversification, and local bias. The descriptive statistics about our subjective variables are provided in Subsection 3.2.2. The objective variables are presented in Subsection 3.2.3, while the behavioural variables are described in Subsection 3.2.4.

3.2 Descriptive statistics

As in Bhattacharya et al. (2017) and Bailey et al. (2011), we classify as a P-ETF user any investor who traded a P-ETF at least once over the whole sample period (i.e., 2003-2012). To drop outliers, we use trading experience as filter and we exclude investors with less than 5-month trading experience to disregard very short-lived investors as in Glaser and Weber (2009) or Bellofatto et al. (2018).

[11] Such items are usually covered in investment policy statements (IPSs) used in portfolio management delegation. There are some points according to Boone and Lubitz (2004) that are primordial for an IPS: 1) client goals and objectives, 2) client fears and concerns, 3) investment time frame, 4) expected mortality, 5) retirement time frame, 6) shorter-term financial needs, 7) risk tolerance attitudes, and 8) risk capacity. In our data, the S-test covers almost all those points.

3.2.1 Sociodemographic variables

The set of sociodemographics includes age, gender, education and spoken language.[12] We build dummy variables for gender (a dummy set to 1 for men), education (a dummy set to 1 if the investor holds a university degree or equivalent) and language (one dummy set to 1 for each language). All these variables will be used as usual control variables in the ZINB regression model in Section 4.

Table 1 reports the summary statistics for these sociodemographics, with a distinction between P-ETF users and non-users. 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 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 speakers is the highest in both subsamples (53% and 51%). The proportion of French speakers is exactly the same (42%) in both subsamples, while the proportion of English speakers is slightly higher among the P-ETF users (7% versus 5%).

Table 1: Descriptive statistics for sociodemographics

		P-ETF users			Non-users			
Metrics		Mean	Std Dev	N	Mean	Std Dev	N	Diff
Age (2012)	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 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 if the investor is a male. '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 the means significantly differ at the level of 10%, 5% or 1%, respectively.

[12] Belgium has three official languages: French, Dutch and German. French and Dutch are spoken the most. On the online trading platform, investors can choose from three available languages: French, Dutch and English.

3.2.2 Subjective variables

Table 2 gives the number of investors who completed the A-test and the S-test, with a distinction between P-ETF users and non-users. All the investors filled in the A-test,[13] but only 49.11% of them also completed the S-test. Note that the brokerage house did not offer portfolio management services over the sample period. Hence, the investors completed the S-test to be granted free access to a financial advice tool on stocks. This online tool provided investors with general information on stocks, analysts' reports, and professional recommendations. A total of 61.48% of the investors who took the S-test are P-ETF users.

Table 2: Number of completed A-tests and S-tests

	Number of investors	S-test	A-test
P-ETF users	3,484	2,142	3,484
		61.48%	100.0%
Non-users	22,656	10,696	22,656
		47.21%	100.0%
Total	26,140	49.11%	100.0%

This table reports the number of P-ETF users and non-users as well as the number and percentage of completed A-tests and S-tests, depending on whether the investor is a P-ETF user or not.

Using the A-test data, we summarize the available items into four variables, which allows us to attribute for each of them a final score to each investor. We follow the same approach with the S-test data and gather the available information into five variables. Table 3 displays the resulting subjective variables in Panel A for the A-test and in Panel B for the S-test.

[13] 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.

Table 3: Descriptive statistics for subjective variables

Metrics		P-ETF users			Non-users			Diff
		Mean	Std Dev	N	Mean	Std Dev	N	
Panel A: Appropriateness 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 the losses with “complex” instruments	(score:0-9)	7.67	2.74	3,484	6.96	3.43	22,656	0.71***
Panel B: Suitability 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***
Target return	(score 1-5)	3.30	1.01	2,142	3.26	1.07	10,696	0.08*

This table provides the cross-sectional means, standard deviations, mean differences, as well as the number of investors among the 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 scale of four levels: Level 0 is associated with a basic knowledge, while 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 the losses with “complex” instruments’, the score elicits the investor’s self-assessment regarding his/her own abilities and skills. For example, a score of 18 out of 18 for the item ’Knowledge of “complex” instruments’ means that the investor considers himself/herself as an expert in the field. Conversely, a score of 0 (out of 18) indicates that the investor declares having no knowledge about “complex” instruments. For ’Risk tolerance’, ’Financial markets and products knowledge’, ’Investment time horizon’ and ’Wealth and financial situation’, the score reflects the investor’s direct self-assessment regarding the item. ’Target return’ refers to the annual yield investors expect on their investments, selected on scale of 5 levels, of which 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 the means significantly differ at the level of 10%, 5% or 1%, respectively.

Panel A of Table 3 shows that P-ETF users clearly display a higher financial literacy than non-users. This holds for both the knowledge of financial markets in general and the knowledge, experience and awareness about “complex” instruments in particular. The mean score of P-ETF users regarding the ’Knowledge of financial markets’ is 1.72, while the corresponding mean for non-users is only 1.29. Similarly, when focusing on “complex” instruments, P-ETF users (1) exhibit a relatively higher level of knowledge (7.22 vs 5.32); (2) have more experience (5.96 vs 4.33); and (3) are more aware of the risk of incurring losses (7.67 vs 6.96). All these differences are statistically significant at the 1% level.[14]

Consistent with Panel A, Panel B of Table 3 reveals that P-ETF users display better 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 have a higher average score for both the items ’Investment time horizon’ and ’Wealth and financial situation’. This means that P-ETF users are, on average, wealthier and have a longer investment horizon than non-users. Again, all these differences are statistically significant

[14]For subjective variables, we opt for Wilcoxon signed-rank tests, which are an appropriate nonparametric alternative to the t-test.

at the 1% level. Nevertheless, the level of target return appears to differ much less between the two subsamples.

3.2.3 Objective variables

Using the trading data, 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 prior empirical 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; (6) stock trading intensity (STI), which is a dummy variable equal to one if the investor's monthly portfolio turnover[15] 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.[16]

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 only 71.94 and 17.27 for non-users. This reveals that P-ETF users not only trade twice as many stocks but also trade on a set of different stocks that is twice as large as that of non-users. P-ETF users also exhibit a longer trading experience than non-users (56 versus 46 months). In addition, P-ETF users display a shorter trade duration on average (57 days for P-ETF users vs 91 days for non-users), suggesting that P-ETF users trade stocks more frequently. Furthermore, the typical P-ETF user holds an

[15]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.

[16]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.

eight-stock portfolio, while the typical non-user holds a five-stock portfolio. Interestingly, the proportion of P-ETF users with intense trading activity on stocks is lower than that of non-users. Similarly, the proportion of underperformers 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.

Table 4: Descriptive statistics for objective variables

		P-ETF users			Non-users			
Metrics		Mean	Std Dev	N	Mean	Std Dev	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	8.37	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 the cross-sectional means, standard deviations, mean differences, and number of investors, 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 last and the first trading dates, 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

Since it is well established in the literature that individuals tend to suffer from several behavioural biases, we compute additional variables in order to capture some of these biases.

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

Second, we adopt the approach of Kumar and Lim (2008) and Bailey et al. (2011) by using the degree of clustering in the investors' trades to measure narrow framing.[17] For

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

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 use the adjusted measure based on the peer group (ATC_i),[18] and this measure 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 equal to zero otherwise.[19]

Third, we use two proxies to assess underdiversification. On the one hand, trading underdiversification is based on the number of different stocks traded during the whole period, which 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. On the other hand, we use the monthly average Herfindahl-Hirschman index (HHI hereafter) to build our second proxy aiming at capturing portfolio underdiversification. Building on Goetzmann and Kumar (2008) and Koestner et al. (2017), this index is equal to the sum of the squared stock portfolio weights. It ranges from 0 for well-diversified portfolios to 1 for portfolios including only one stock. Our second proxy is set to 1 when the investor's monthly HHI falls in the upper quartile of the sample and is set to zero otherwise.

Finally, we also compute for each investor the local bias ratio, which captures the concentration of his/her trades on the Belgian stock market and on those of neighbouring countries. 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 (Belgium, France, Netherlands, Germany and Luxembourg) divided by the total number of trades. By construction, this ratio ranges between 0 and 1 and is equal to one when all

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

[19] Kumar and Lim (2008) states 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 trade-clustering 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".

trades are executed on stocks issued in these five countries.

Table 5 reports the cross-sectional statistics for the above behavioural variables. Overall, P-ETF users look less prone to behavioural biases. The proportion of overconfident investors is higher among non-users than among P-ETF users (10.5% versus 7.2%). This is consistent with Table 4, wherein both stock trading intensity and underperformance prevail more among non-users (25% and 26% versus 22% and 20%). For the lack of diversification, the two proxies indicate that P-ETF users suffer less from this bias 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%). For the local bias, the average ratio is slightly higher among non-users than among P-ETF users (0.750 versus 0.725). This suggests that the trades executed by non-users are more locally concentrated 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.

Table 5: Descriptive statistics for the behavioural 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	0.334	3,484	0.276	0.447	22,656	-0.139***
Local bias ratio	(score between 0 & 1)	0.725	0.254	3,484	0.751	0.297	22,656	0.026***

This table reports the cross-sectional means, standard deviations, mean differences, as well as the number of investors, with a distinction between P-ETF users and non-users. 'Overconfidence' is a dummy variable set to one if the investor displays both overtrading ('Stock trading intensity' (STI) is set to one) and underperformance ('Underperformance on stocks' (UPS) set to one), and is set to 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 is set to 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 and set to 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 and set to zero otherwise. 'Local bias ratio' is computed for any investor as the number of trades executed on stocks issued in the home (Belgium) and in neighbouring countries (France, 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.

4 Empirical results

To identify the key drivers of the retail P-ETF investors' profile, 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 retail investment in these instruments. This "two-step-in-one" approach is innovative compared with the approaches used in the prior empirical research on investment funds. Bailey et al. (2011) opt for a binary logistic model to examine whether a combination of behavioural 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 passive 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 ZNIB model is the total number of trades on P-ETFs for each investor.[20] Hence, the basic 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 is appropriate for our retail investor sample that includes two distinct groups: the group of investors who do not invest in P-ETFs (non-users) and the group of investors who invest in P-ETFs to a small or large extent (P-ETF users).

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

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

where the base count density is the negative binomial distribution $f_2(y)$,[21] 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.[22] 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

[20]Since our dependent variable can take integer values (count data), Poisson regression models could be considered adequate. However, there is a dominance of zeros in our data. Approximately 83% of the retail investors in our database have no trade on P-ETFs, which would make the adoption of a Poisson regression inconsistent. The ZINB model is 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 in the sample, 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 related to an excess of zeros, i.e., the situation in which there is an excessive number of observations equal to 0 (Greene, 2000).

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

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

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 observations, i.e., 12,838 retail investors. Panel A refers to the sociodemographics 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)[23] that are also used in the three models. Among them, we are especially interested in the variable "Financial markets and products knowledge' since it is directly related to hypotheses H1/H1* about financial literacy. Panel C is devoted to a set of objective variables added in both Model 2 and Model 3. These variables are useful to test hypotheses H2/H2* about financial experience. Panel D refers to proxies for different behavioural biases available in Model 3. In particular, these variables allow us to test H3/H3* about overconfidence, H4/H4* about narrow framing, H5/H5* about underdiversification, and H6/H6* about local bias.

A key point regarding the interpretation of the ZINB models is that the signs of the coefficient estimates from the binary logistic equation are often in the opposite direction to the signs of the coefficient estimates in the count (negative binomial) equation (Long et al., 2006). In our specific case, the logistic process predicts the probability of not investing in P-ETFs; i.e., a negative coefficient estimate implies a higher probability of being a P-ETF investor. In contrast, the negative binomial process predicts the number of P-ETF trades; i.e., a positive coefficient estimate indicates a higher magnitude of P-ETF investment.

4.1 The probability of retail investing in P-ETFs

Table 6 provides results that are consistent across the three model specifications. We therefore focus on the significant predictors in the logit component of Model 3.[24]

[23]Whether the A-test is instead considered in the regression analysis does not change the empirical conclusions regarding the tested hypotheses. The advantage of the S-test is the wider coverage in terms of the risk-return profile of the investor. The A-test strictly focuses on financial knowledge and awareness.

[24]The overdispersion parameter α (2.00) 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, Table 6 shows that Model 3 is improved, with the log likelihood increasing (by 345 between

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 (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
Target return (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												
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
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
Local bias ratio									0.297**	1.345	-0.308**	0.735
α (dispersion)	3.32***	27.51			2.05***	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				22,170				22,218			

This table reports the results of three ZINB regressions, with a distinction between the logististic and negative binomial components. The dependent variable is the number of P-ETF trades. Panel A refers to the sociodemographics as the usual control variables. 'Gender' is a dummy variable equal to one if the investor is a male. 'Age' is determined 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 both dummies based on the language selected by the investor on the online trading platform. Panel B includes the 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. 'Target return' refers to the annual yield investors expect on their investments, selected on scale of 5 levels, of which 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 the objective variables characterizing trading activity. 'Trading experience' is computed as the difference between the last and the first trading dates, 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 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. Panel D refers to behavioural variables. 'Overconfidence' is a dummy variable set to one if the investor displays both overtrading (STI is set to one) and underperformance (UPS set to one) and is set to 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 set to 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 and set to 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 and set to zero otherwise. 'Local bias ratio' is computed for any investor as the number of trades executed on stocks issued in the home (Belgium) and in neighbouring countries (France, 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 coefficient estimate. 'OR' is the odds ratio, which is equal to the exponential of the coefficient estimate. *,**, and *** indicate that the coefficient estimate is statistically significant at 10%, 5% or 1%, respectively.

Regarding the sociodemographics in Panel A, the two dummies 'Higher education' and 'English-speaking' significantly impact the probability of investing in P-ETFs. On the one hand, retail investors who hold a university degree are more likely to invest in P-ETFs. More precisely, the odds of *not* investing in P-ETFs decrease by a meaningful factor of 0.772 for a highly educated retail investor compared to investors who are not highly educated. All else equal, this means the odds of *not* investing in P-ETFs decrease by 22.8% for retail investors with a university degree.[25] On the other hand, the odds ratio is even lower (at 0.681) when we consider the marginal effect of the language selected on the trading platform. All else equal, the odds of *not* investing in P-ETFs decrease by 31.9% for English-speaking investors compared to French-speaking investors.[26]

In Panel B, four subjective variables (out of five) significantly impact the probability of investing in P-ETFs. Focusing first on the variable "Financial markets and products knowledge', we validate **H1** on the positive impact of financial literacy since its coefficient estimate is negative and statistically significant at 1%. The corresponding odds ratio indicates that the odds of *not* investing in P-ETFs decrease by 21.5% for retail investors who self-report high financial knowledge. Three other subjective variables also exhibit insightful coefficients. A higher 'Target return' by one score unit leads to a higher probability of *not* investing in P-ETFs.[27] In other words, the higher the return targeted by retail investors is, the less likely they are to invest in P-ETFs. Regarding 'Risk tolerance', we observe that the more risk tolerant retail investors are, the less likely they are to invest in P-ETFs. All else equal, retail investors who display higher risk tolerance by one score unit have a 1.148 times higher chance of *not* investing in P-ETFs than those who do not. Interestingly, a one-unit increase in the score of 'Investment time horizon' decreases the odds of *not* investing in P-ETFs by a factor of 0.949. We conclude from Panel B that the retail investors who are more likely to invest in P-ETFs are those who are more financially literate, target lower returns, are less

Models 1 and 3). Model 3 also displays the lowest AIC and BIC values.

[25] 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 retail investors compared to retail investors who hold a university degree.

[26] 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.

[27] All else equal, this means that retail investors who target higher returns by one score unit have a 1.103 times higher chance of *not* investing in P-ETFs than those who do not.

risk tolerant, and declare a longer investment horizon, all else equal.

In Panel C of Table 6, the coefficient estimate of 'Trading experience' (in log)[28] is significant and exhibits a negative sign. We therefore validate **H2** on the positive relationship between financial experience and the probability of investing in P-ETFs.[29] Two other objective measures of trading activity are also significant. All else equal, the 'Number of stocks' held in the end-of-month portfolio is positively associated with the probability of investing in P-ETFs, since the odds ratio is equal to 0.614. This is also the case for the variable 'Stock trading intensity', which means that retail investors who trade stocks intensively are more likely to invest in P-ETFs. The odds of *not* investing in P-ETFs for an investor who trades stocks intensively decrease by a factor of 0.79. In summary, we learn from Panel C that the retail investors who are more likely to invest in P-ETFs are those who are more experienced, hold a larger number of different stocks in their portfolio, and trade these stocks more intensively (ceteris paribus).

For the influence of behavioural biases, Panel D in Table 6 shows that two of them, 'overconfidence' and 'local bias', emerge as significant drivers of the probability of investing in P-ETFs. Both negatively affect the chance for a retail investor to trade P-ETFs. We therefore validate **H3** and **H6**. Specifically, when retail investors are overconfident, their odds of *not* investing in P-ETFs increase by 47.9%. 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 equity portfolio invested in index funds because they focus more on actively managed funds. Regarding the local bias, when retail investors invest (exclusively) locally, their odds of *not* investing in P-ETFs also increase by a slightly lower percentage equal to 34.5%. We do not find any supportive evidence regarding **H4** and **H5**. Hence, we conclude from Panel D that the retail investors who are less likely to invest in P-ETFs are those who are overconfident and invest locally.

[28]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.

[29]All else equal, this means that the odds of *not* investing in P-ETFs decrease by 19.1% when the retail investors' trading experience doubles.

4.2 The magnitude of P-ETF investment

Unlike the ZINB model's logit component that is estimated on the full sample of retail investors (including those who do not trade P-ETFs), the negative binomial component is estimated on the subsample of P-ETF users. The latter represents 13.33% of the retail investors in the entire sample. The purpose here is to identify the factors that truly drive the magnitude of P-ETF investment among retail investors who trade P-ETFs.

In Panel A of Table 6, higher education and the use of English are again significant drivers. All else equal, both are positively correlated with the magnitude of P-ETF investment. Their respective coefficient estimates are positive, and their incidence rate ratios (IRRs) are greater than 1.[30] Specifically, P-ETF users who hold a university degree complete 17.5% more P-ETF transactions on average than those who do not hold such a degree. Similarly, the P-ETF users who selected English on the trading platform execute on average 45.7% more trades on P-ETFs than the French-speaking P-ETF users. Interestingly, investor age also determines the number of P-ETF trades to a lesser extent.[31] We conclude from Panel A that the magnitude of P-ETF investment is stronger when the retail investors are highly educated, English-speaking, and older.

When focusing on Panel B, we observe that the magnitude of P-ETF investment is impacted by the same three subjective variables that drive the probability of trading P-ETFs. We first validate **H1*** since the coefficient estimate of the "Financial markets and products knowledge' variable is positive and significant. The number of P-ETF trades increases by 12.2% for each score unit increase in financial literacy, which is economically substantial. Second, the number of P-ETF trades slightly increases when retail investors have longer investment time horizons: for each score unit increase in the variable 'Investment time horizon', the number of P-ETF trades increases by 3.6%. Next, P-ETF users who target higher returns complete a lower number of P-ETF trades. For each score unit increase in the

[30]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.

[31]When age increases by 1 year, P-ETF trades are expected to increase by 1.01%.

variable ’Target return’, the number of P-ETF trades decreases by a sizeable factor of 0.869 (i.e., -13.1%). From Panel B, we conclude that P-ETF users who self-report higher financial literacy, longer investment horizons, and lower target returns trade P-ETFs more often.

Panel C in Table 6 reveals that P-ETF users who are more experienced trade P-ETFs more frequently, which validates **H2***. Specifically, the coefficient estimate indicates that the number of P-ETF trades increases by 40.7% when trading experience (in months) doubles. In addition, three other drivers negatively affect the magnitude of P-ETF investment, namely, the trade duration on stocks, the number of stocks in the end-of-month portfolio, and the stock trading intensity. When P-ETF users hold a low number of stocks that they do not trade in quick succession in a stock portfolio whose turnover is low, they are more likely to trade P-ETFs more intensively. This suggests a trade-off between the magnitude of P-ETF investment and the magnitude of stock investing. P-ETF users who trade P-ETFs more often not only hold a lower number of stocks but also modify the composition of their stock portfolio less extensively. This points to a substitution effect between stocks and P-ETFs.

For the behavioural variables available in Panel D of Table 6, the local bias is the only one that negatively impacts both the probability and the magnitude of investing in P-ETFs. More specifically, trading activity in P-ETFs is 26.5% (i.e., 1 minus 0.735) lower for P-ETF users who suffer more from local bias than for those who suffer less. **H6*** is therefore validated. By contrast, when considering trading underdiversification, we reject **H5***, albeit partially. All else equal, the number of P-ETF trades increases by 101.4% for P-ETF users who concentrate their stock trading on a few stocks (compared to those P-EFT users who do not). This is totally consistent with the findings of Panel C regarding the substitution effect between stocks and P-ETFs. This result is also in line 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. For the remaining hypotheses, there is no evidence to support **H3*** and **H4***. Of particular interest, overconfidence does not appear to drive the magnitude of P-ETF investment, although it affects the probability of trading P-ETFs.

4.3 Robustness check

All the models were estimated using the full sample of retail investors and the entire 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 influence 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 ran the same three models on a subsample excluding all investors who started trading before January 2008. The purpose was to check whether the information content of the MiFID survey data (captured in Panel B) remained unchanged. Overall, the main findings are similar and we notice only a few changes. In Model 3, two variables lose their significance in the Logit part, i.e., 'Dutch-speaking' and 'log(Trading experience)', while three variables are no more significant in the negative binomial part of Model 3, i.e., 'Higher education', 'English-speaking' and 'Stock trading intensity'. These empirical results are available in Table 8 in appendix.

5 Conclusion

To our knowledge, this study is the first of its kind to profile retail investors by using both trading and survey-based data on a large sample over nearly a decade. Such a data set allows us to include a large set of sociodemographic factors, subjective characteristics, trading activity measures, and behavioural variables. We build on the literature and formulate several hypotheses to test whether retail investors who invest in P-ETFs differ from those who do not. To further contribute to the prior empirical research on investment funds, we opt for a zero-inflated negative binomial (ZINB) regression model to identify the main drivers of retail investment in P-ETFs. The advantage of this approach is twofold since it enables us to explain the likelihood of retail investors trading P-ETFs as well as the magnitude of the retail investment in P-ETFs.

There is overall strong evidence that retail P-ETF investment is made by more sophisticated investors. When focusing on the likelihood of trading P-ETFs, we find that retail investors who are better educated, self-report both higher financial literacy and a longer investment horizon, exhibit longer trading experience, hold a better diversified stock portfolio, and trade stocks more intensively are more likely to invest in P-ETFs. By contrast, retail investors who declare higher target returns and suffer more from either overconfidence or local bias are less likely to invest in P-ETFs.

We identify similar drivers when we zoom in on the subsample of P-ETF users to analyse the extent to which they trade these products. Higher education, the use of English, higher financial literacy, longer investment horizons, lower target returns, and stronger financial experience are all factors that boost the magnitude of P-ETF investment. In addition, the most active investors in P-ETFs are older and less subject to local bias. Nevertheless, overconfidence does not appear to drive the magnitude of P-ETF investment, although it affects the probability of trading P-ETFs.

Finally, a trade-off seems to exist among P-ETF users between the magnitude of their P-ETF investment and the magnitude of their stock investment. Retail investors who trade P-ETFs more intensively hold a lower number of stocks in their end-of-month portfolio and modify the composition of their stock portfolio 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.

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 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	0.186***	1.205	-0.170***	0.843	0.164***	1.178	-0.153***	0.858	0.160***	1.174
Target return (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
Panel D : Behavioral variables												
Overconfidence									0.532**	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
Local 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%				13%				13%			
Log likelihood	-3,338.36				-3,255.04				-3,227			
AIC	6,722.71				6,576,09				6,540.95			
BIC	6,871.48				6,789,54				6,819.10			

This table reports the results of three ZINB regressions, with a distinction between the logististic and negative binomial components. The dependent variable is the number of P-ETF trades. Panel A refers to sociodemographics as usual control variables. 'Gender' is a dummy variable equal to one if the investor is a male. 'Age' is determined 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 both 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. 'Target return' refers to the annual yield investors expect on their investments, selected on scale of 5 levels, 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 last and the first trading dates, 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 is set to one) and underperformance (UPS 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 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, and zero otherwise. 'Portfolio underdiversification' is a dummy variable set to one if the investor's monthly HHI (defined as the sum of squared stock portfolio weights) is in the upper quartile, and zero otherwise. 'Local bias ratio' is computed for any investor as the number of trades executed on stocks issued in home (Belgium) and neighboring countries (France, 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 coefficient estimate. 'OR' is the odds ratio, which is equal to the exponential of the coefficient estimate. *,**,*** indicate that the coefficient estimate is statistically significant at 10%, 5% or 1%, respectively.

References

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.

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.

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.

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.

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. McGowan, and P. Roger (2021). Trading leveraged exchange-traded products is hazardous to your wealth. *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.

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. *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.

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.

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. *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.

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. 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.

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.

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.

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

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.

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.