Financial Literacy and Multi-Asset Portfolio Diversification*

Rudy De Winne† Aiste Petkeviciute‡

January 30, 2021

Abstract

This paper investigates the relationship between financial literacy and portfolio diversification. We rely on a dataset including both the trading records of more than 24,000 retail investors and their answers to questionnaires that provide information regarding financial literacy and other personal characteristics. Since the trading records are not limited to stocks, we assess portfolio diversification based on different asset classes (bonds, funds, options, warrants) and show that this leads to a significant diversification improvement. We find a strong and robust relationship between financial literacy and portfolio diversification, regardless of the financial literacy score and the diversification measure that are used. The relevance of relying on a data-consuming approach for accurate diversification measurement is also discussed and seems not crucial.

Keywords: Portfolio diversification, Financial literacy, Individual investors, Multi-asset class portfolios

1 Introduction

Portfolio diversification is used to protect the investor against the unsystematic risks of the market. It helps reduce the investment risk without compromising the expected return. This rather intuitive concept of "not putting all of one's eggs in the same basket" was first formally introduced by Markowitz (1952) in the Modern Portfolio Theory (MPT). He defined portfolio diversification as a strategy that aims to combine portfolio assets with returns that are less than perfectly correlated in order to lower the variance of the portfolio for a given level of portfolio return. Since the introduction of MPT, the diversification benefits have been uniformly recognized. However, we notice that portfolio underdiversification still ranks among the most harmful mistakes investors make (Von Gaudecker (2015)).

We investigate the relationship between investors' financial literacy and portfolio diversification. Previous studies that have examined this link used mostly stock portfolios and concluded that

*The authors are grateful to the online brokerage house for providing the data. Any errors are the full responsibility of the authors. Declarations of interest: none.

†Louvain Finance (LIDAM), UCLouvain, Belgium

‡Corresponding author; Louvain Finance (LIDAM), UCLouvain, Mailing address: Chaussée de Binche 151, 7000 Mons, Belgium – Email: aiste.petkeviciute@uclouvain.be

investors hold very concentrated portfolios. However, Von Gaudecker (2015) and Calvet et al. (2007) find that investors' diversification scores improve when taking stocks and mutual funds into account. Using a larger universe of asset classes including bonds, funds, options, stocks and warrants, we analyze how portfolio diversification evolves when retail investors are able to trade a wider range of assets. For our analysis, we rely on a dataset containing the trading records of more than 24,000 investors over nearly ten years as well as their answers to MiFID questionnaires, including specific questions about financial literacy. This combination allows to compute different diversification measures at monthly intervals and to analyze how they are affected by scores of financial literacy built from the investors' answers to the questionnaire.

We compare the results given by the most popular portfolio diversification measures (i.e. number of positions in a given portfolio, a diversification index derived from the Herfindahl-Hirschman Index, and the Shannon's measure of Entropy) and check whether some information can be gained from the use of more sophisticated diversification measures. The latter are weight-based and require historical data to value the portfolio positions at regular time intervals (e.g., at the end of each month). The availability of such data and the problems that arise in combining them with investors' portfolio positions, especially for some asset classes, make it useful to check whether this heavy and uninspiring work is really worth it and whether there is an added value in terms of diversification assessment. Therefore, each sophisticated diversification measure is computed based on monthly investors' portfolio positions estimated using three different approaches: the price-free, the internal and the external approaches. First, the price-free approach evaluates investors' portfolio positions using the quantity of each asset only (basically, we assume that each asset unit is worth 1 €). Second, the internal approach consists in valuing the asset position held by a given investor at the most recent trading price for that asset, regardless of the investor who actually carried out this transaction. Finally, we evaluate investors' portfolios using historical asset prices downloaded from an external database, when available, and an internal price otherwise.

We contribute to the literature concerning individual investors' portfolio diversification in three ways. First, we consider an investor's aggregate portfolio including different asset classes (beyond stocks and sometimes funds) and check whether diversification is substantially improved. Second, we analyze the relationship between financial literacy and investor diversification. Finally, we provide some analysis regarding the accuracy of simple diversification measures compared with those that require price time series to value the portfolio positions at regular time intervals and to compute the corresponding weights.

Our results are interesting in different respects. First, we find that assessing diversification in a multiple asset classes framework results in a significant improvement of investors' portfolio diversification scores. Second, we find that financial literacy is strongly related to the level of investors' portfolio diversification. This result is robust to the choice of diversification measures and to the approach used to evaluate investors' portfolio positions. The robustness of the results to an approach which is less data-consuming is important. This could indeed encourage many researchers to extend their analyzes to other asset classes for which the problem of external data is more acute.

The remainder of the paper is structured as follows. Section 2 reviews the relevant literature and presents an overview of portfolio diversification measures. Section 3 describes our sample and provides some descriptive statistics. Section 4 presents the results. Section 5 concludes.

2 Literature review

2.1 Portfolio diversification and investor financial literacy

Since the Modern Portfolio Theory of Markowitz (1952), diversification is considered as a key point for portfolio performance. While financial theory tells how portfolios should be built, many papers show that most retail investors depart from this in a large extent and actually hold poorly diversified portfolios. This potentially leads to huge losses that are highly avoidable (Friend & Blume (1975), Kelly (1995), Goetzmann & Kumar (2004), Von Gaudecker (2015), Calvet et al. (2009)). More specifically, Florentsen et al. (2019) find that the aggregate cost of underdiversification of Danish investors equals 3.1%, meaning that the investors could obtain 3.1% higher returns yearly on average. They also report that 66% of their sample investors own 1 stock only while this proportion is estimated at around 30% of the investors by Goetzmann et al. (2005). However, many investors hold positions in different assets (Calvet et al. (2007), Goetzmann & Kumar (2004)). It may therefore be questionable to assess portfolio diversification on a stock-only basis, as in Friend & Blume (1975), Kelly (1995), Evans & Archer (1968), Florentsen et al. (2019) or Goetzmann et al. (2005). Calvet et al. (2007) found that 76% of the investors owned mutual funds and that owning cash and mutual funds can limit the losses from severely underdiversified stock-portfolios. They also insist on the significant improvement of portfolio diversification once the information about mutual funds is included. Von Gaudecker (2015) observed that while stock-portfolios display a strong underdiversification, mutual fund portfolios show a much better diversification level. The author finds that many households reduce the riskiness of their portfolios by investing in mutual funds. A huge amount of studies assess portfolio diversification by counting the number of positions held by investors (Abreu & Mendes (2010); Goetzmann & Kumar (2004); DeMiguel et al. (2009); Tu & Zhou (2011)). Doing so, they implicitly assume that investors' portfolios are equally-weighted (each position is assigned with an equal weight) or that portfolio weights are not relevant for diversification quality assessment. Authors find that the average number of positions held in a portfolio fluctuate: US investors hold 4 stocks on average (Goetzmann & Kumar (2004)), Portuguese investors hold 2.6 positions on average (Abreu & Mendes (2010)) and Danish investors hold 2 stocks on average (Florentsen et al. (2019)). More descriptive statistics regarding samples from our literature review can be found in the Table 1.

Beyond the number of assets held in a portfolio, diversification quality also depends on the weights of these positions. This led several researchers to use a Herfindahl–Hirschman Index (HHI) or a modified version of this index for diversification measurement (Dorn & Huberman (2005), Hoffmann et al. (2010), Koestner et al. (2017), Bellofatto et al. (2018)). The use of this index for diversification measurement relies on the idea that, given the very large number of positions in the market portfolio, the weight of each position is close to zero. Thus, the smaller the portfolio weights are, the closer they are to those of the market portfolio and the more diversified

Table 1: Sample descriptions from the literature

Author	Country	Universe of assets	Mean number of assets	Number of investors
Florentsen et al. (2019)	Denmark	Stocks	2 stocks	1,000
Hoffmann et al. (2010)	The Netherlands	Stocks	6.57 stocks	5,500
Barber & Odean (2000)	US	Stocks	4.3 stocks	66 465
Goetzmann & Kumar (2008)	US	Stocks	4.71 stocks	62 387
Von Gaudecker (2015)	The Netherlands	Stocks, funds	3.34 assets	1,604
Koestner et al. (2017)	Germany	Stocks, funds	6.71 assets	19,487
Abreu & Mendes (2010)	Portugal	Stocks	2.6 assets	1,559

Notes: The table illustrates some sample characteristics in the reviewed articles about portfolio diversification of individual investors. Column 1 indicates the authors of the article. Column 2 shows the country of the sample used in their analysis. Column 3 indicates the assets that are considered in the analysis. Column 4 indicates the mean number of assets held in investors' portfolios. Column 5 indicates the size of the analyzed sample. The results of mean and median number of assets in the last line (Abreu & Mendes (2010)) refer to mean/median values of assets/securities/issuers rather than *only* assets.

the portfolio is.

However, some researchers find that even when examining diversification of portfolios including stocks and funds (Koestner et al. (2017)), some of them remain poorly diversified. This raises natural questions about retail investors' financial literacy and about how it may affect portfolio diversification and performance. However, this relationship is not easy to analyze since it is rather uncommon to have data covering both portfolio holdings and financial literacy. Portfolio holdings are often built from trading records while financial literacy is generally assessed through a questionnaire. Since *financial literacy* is not available in most datasets, a lot of researchers use proxies for financial literacy, such as education and wealth (Calvet et al. (2007)), Volpe et al. (2002)), that are more often available.

Volpe et al. (2002) find that investors' scores regarding financial literacy are higher for men and that financial literacy scores improve with age, income level and education level. Bianchi (2018) confirms Volpe et al.'s (2002) results and note that higher scores of financial literacy are associated with the investors being male, not married and that financial literacy increases with investor's education level, income and wealth.

Regarding the impact of financial literacy on the diversification ability of investors, Guiso & Jappelli (2008) use questionnaires to assess financial literacy and to gather information about portfolio composition. Based on this, they find that investors' financial literacy is strongly correlated with portfolio diversification. Abreu & Mendes (2010) observe that individual investors' capability to diversify their portfolios improves with education level, income level, investment experience, specific knowledge of financial markets and financial instruments as well as it depends on the information sources used to gather the financial knowledge. Calvet et al. (2009) also provide similar results. They find that higher education, wealth and owning real estate are associated with Swedish investors underdiversifying their portfolios less. Goetzmann & Kumar (2008) also find that diversification level increases with age, income, wealth, education and financial sophistication. They also note that investors with a greater propensity to invest in foreign equities or mutual funds exhibit a better level of diversification regarding their domestic portfolios too. Florentsen et al. (2019), as Goetzmann & Kumar (2008), confirm that low income, low wealth and low education are related to lower diversification scores.

2.2 Diversification measures

Financial literature about retail investors' diversification rely on different diversification measures. The most commonly used measures are the number of portfolio positions, the Herfindahl-Hirschmann Index (HHI) or its variants and Shannon's measure of entropy (SE).

The easiest measure which is commonly used to define portfolio diversification (Goetzmann & Kumar (2008), Barber & Odean (2000), Abreu & Mendes (2010)), is simply counting the number of monthly/daily positions in investors' portfolios. However, this measure assumes that all assets have equal weights within a given portfolio therefore often overstates the level of diversification (Friend & Blume (1975)).

The second diversification measure we use is a variation of the Herfindahl-Hirschmann Index (HHI) initially used to evaluate the *economic concentration* in an industry or market. The HHI is computed by first squaring the market share of each firm i in a given market m, which is referred to as *weight* and then by summing these weights. The HHI is calculated based on the following formula:

$$HHI = \sum_{i=1}^{N}(\omega_i - \frac{1}{N})^2 = \sum_{i=1}^{N}(\omega_i - \omega_m)^2 \tag{1}$$

where ω_i is the portfolio weight of a firm i, N is the number of firms in the market. The bigger the value of HHI is, the bigger is the market concentration.

Since diversification can be considered as the opposite of concentration, some researchers (Koestner et al. (2017), Goetzmann & Kumar (2008)) use the HHI measure in the field of finance as a measure of portfolio concentration. Given that the market portfolio is supposed to include a very large number of positions, N is very high and ω_m is close to zero. Consequently, HHI measure applied to portfolios can be approximated to the sum of squared weights of each position i in the portfolio:

$$HHI = \sum_{i=1}^{N}(\omega_i - \omega_m)^2 \approx \sum_{i=1}^{N}\omega_i^2 \tag{2}$$

It is popular not only because its computation is relatively easy but also because it explains the degree of diversification of an unevenly distributed portfolio at least as well as the number of securities in a portfolio does so for an evenly distributed portfolios Woerheide & Persson (1992)). However, HHI in Equation 2 measure portfolio concentration. Since we are interested in the level of portfolio diversification, which is opposite to concentration, we use a variation of HHI score, presented by Woerheide & Persson (1992). It is calculated in the following way:

$$DI = 1 - HHI = 1 - \sum_{i=1}^{N} \omega_i^2 \tag{3}$$

The DI takes values between 0 (least diversified) and 1 (most diversified), which makes its interpretation more intuitive than using the original HHI scoring to talk about the level of portfolio diversification.

We also use another variation of DI, which is introduced by Dorn & Huberman (2005). The authors adopted a technique to in order to better represent the already-diversified nature of funds. Every position that an investor holds in mutual funds is substituted by a 100 equally-weighted non-overlapping positions. Dorn & Huberman (2005) used the original HHI score, which means the lower is the value taken by the index, the better is the level of the diversification. In this case, once again, to ease the interpretation accordingly to what we want to measure, we compute a modified DI (M_DI). M_DI takes values from 0 to 1 - indicating increasing levels of portfolio diversification.

The last measure we are using to evaluate the level of portfolio diversification of individual investors is called Shannon's entropy (SE) and is computed in the following way:

$$SE(\omega) = -\sum_{i=1}^{N} \omega_i log \omega_i \tag{4}$$

where $\omega = (\omega_1, ..., \omega_N)$ is a portfolio allocation among N risky assets, $\sum_{i=1}^{N} \omega_i = 1$ and $\omega_i \geq 0$ for each i=1, 2, ..., N.

Shannon's entropy comes from the information theory where it is used to measure the disorder (Bera & Park (2008)). This index has an advantage of not having an upper-bound regarding the values it can take. Other indices, such as HHI, that are constrained between 0 and 1 use up 3/4 of the possible values to explain the level of diversification of portfolios containing 1-5 stocks (Woerheide & Persson (1992)).

3 Data, sample and descriptive statistics

3.1 Data and sample

Our analysis relies on an extensive dataset provided by an online Belgian brokerage house combining the trading records of 78,243 retail investors over a period of 9.25 years (January 2003 - March 2012) and individual characteristics that are either sociodemographic or survey-based. The latter have been collected within the context of MiFID European regulation[1] that came into effect in November 2007 in all the countries of the European Union. The goal of the directive is to offer a high degree of protection to individual investors regarding different financial instruments as well as to improve the competitiveness of the financial markets. One of the consequences of

[1] 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)).

the MiFID directive is that the investment banks have to implement two assessment tests to be filled in individually by their clients. The Suitability test (hereafter S-test) is designed to help the investment bank match the investment objectives of the investors (given their income level, household situation, investment horizon) and evaluate their experience and financial knowledge while the Appropriateness test (hereafter A-test) aims to capture investors' understanding and experience regarding financial instruments that are classified as complex[2] (e.g. when the risks of the returns are hard to understand).

The trading records contain information such as an International Security Identification Number (ISIN) of the traded asset (bond, fund, stock, option or warrant), trading price, executed quantity, time-stamp, associated transaction costs, the currency an asset is traded in, whether the trade is a sell or a buy, a code for the market where a trade was completed in, and an anonymous investor's ID code, which allows us to match it with the individual characteristics. Among these characteristics, we have investors' answers to the Suitability Test, their answers to the Appropriateness Test as well as their personal characteristics such as year of birth, gender and education level.

For the purpose of our study, we need both the trading records and the individual characteristics, including the answers to several MiFID questions. Therefore we select the investors who have completed Both the Appropriateness and the Suitability tests. This results in a sample of 24,513 investors. We then construct their monthly portfolios using the trading records. To assess the weight of any position in a given portfolio at regular intervals (day, week or month), we must value each position in that portfolio, which requires market prices available from an external database. We outsource to Thomson Reuters Eikon and Eurofidai databases to find this information. Columns 4 and 5 of Table 2 describe the availability of daily prices that match our dataset in the Thomson Reuters Eikon and Eurofidai databases. We believe that the Column 5 is more accurately describing the actual situation because it takes into account the popularity of each financial asset and accounts for each *investor-monthly position* combination in our dataset. Column 4 assigns the same weight to an asset and does not make a distinction between an asset held by multiple of investors during several years and an asset held by 1 investor during 1 month. Therefore, the Column 5 shows that we possess daily prices mostly for funds, stocks and bonds. Considering the sample as a whole, as we see in the last row of the Table 2, our daily prices cover more than 3/4 of the monthly positions that investors hold.

3.2 Descriptive statistics of the sample

Over our sample period, the investors have placed 3,156,542 trades. On average, an investor trades 13 times a year across the classes of financial assets. If we look only at stocks, this number falls to 10 times a year. These trades cover five classes of financial assets with their respective class and trading frequency described in Table 2. Stocks are the most traded class of assets and make up a majority of trades in our sample (74%). The least traded asset-class is bonds, which makes up only 0.2% of all the trades.

We also observe that an average investor has 1.89 asset classes represented in his/her portfolio,

[2] Among the financial asset classes considered in this, only options and warrants fall under the category of *complex instruments*.

Table 2: Distribution of trades across asset classes and availability of daily prices

Financial asset class	N° of trades	% of all trades	% monthly prices (1)	% monthly prices (2)
Stocks	2,339,809	74%	65.5%	88.5%
Warrants	318,379	10%	11%	5%
Funds	342,120	10.8%	82%	94%
Options	150,155	5%	0.02%	0.01%
Bonds	6,097	0.2%	66.1%	40.2%
Total	3,156,542	100%	59%	76%

Notes: The table presents a list of financial asset classes analyzed in our sample (Column 1), the number of trades placed per financial asset class (Column 2), what percentage it represents given all the trades placed by the investors (Column 3). Column 4 shows the availability of monthly prices per asset class if we disregard the frequency at which the asset is held by the investors, i.e. if an asset is held by half of the investors or by only one investor - it appears to have the same weight in this calculation. The percentage is computed by disregarding the weight of each asset (how long it was held, how many investors chose it). Column 5 exhibits the availability of monthly prices per class of asset and takes into account how many investors hold a position in a given asset and for how long. The percentage is computed by taking into account all *investor-monthly position* combinations regarding every traded asset in our sample.

a median investor holds 2 asset classes and only 41% of investors trade stocks-only. This already suggests that excluding other-than-stock types of assets might provide misleading results regarding their level of portfolio diversification.

We investigate the ensemble of trades to find that only 29.40% of the trades concern Belgian (domestic) financial assets meaning that investors in our sample tend to diversify internationally. However, when the distinction is made between European and non-European assets, we notice that a majority (81.57%) of investors feel more comfortable investing closer to home - in Europe. Overall, 59.88% of the trading activity were purchases and 40.12% were sales.

Table 3: The proportion at which a given number of monthly positions is held by the investors monthly (at the multi-class level as well as at the individual level to each asset class)

Number of monthly positions	Aggregate level	Stocks	Funds	Bonds	Warrants	Options
1	15.8%	18.7%	36.0%	57.3%	38.2%	27.1%
2	12.0%	13.6%	16.3%	18.3%	14.9%	14.2%
3	9.4%	10.2%	10.3%	9.3%	9.3%	9.7%
<5	44.7%	50.8%	70.5%	90.3%	68.2%	57.5 %
<10	68.9%	75.1%	91.0%	97.2%	85.2%	74.7%
10 - 20	15.9%	16.4%	5.4%	1.7%	7.7%	10.6%
>20	11.2%	7.1%	1.4%	0.4%	5.0%	12.4%
>50	1.7%	0.6%	0.1%	0.0%	0.7%	4.2%
Participation rate	100%	94.5%	21.7%	6.7%	10.9%	8.2%

Notes: This table shows how many positions investors monthly hold in their portfolios. Column 1 indicates the number of positions held (multi-asset or regarding separate asset types). Column 2 presents the percentage of investors-monthly portfolios that hold 1 position, 2 positions and so on. The following columns show how many stocks, funds, bonds, warrants and options respectively investors hold in their monthly portfolios.

Using the history of trades of 24,513 investors during the period of 2003 January - 2012 March, we construct their monthly portfolios. An investor' monthly portfolio reflects the non negative positions in each asset that (s)he holds at the end of the month. The aggregate number of monthly positions, summarized in Table 3 according to different asset classes, held by investors equals to 1,364,097.

Looking at the number of positions in investors' portfolios, we notice that less than 20% of investors in our sample hold one position only. These results contrast findings by Florentsen et al. (2019), Goetzmann et al. (2005) and show a higher level of portfolio diversification in our sample. By examining each asset-class separately, we observe that the mode for the number of positions held in bonds, in warrants or in funds is 1 position. The most popular asset-class by far to hold less than 5 monthly positions in is bonds. Investors that have a higher number of monthly positions (10-20) in one asset-class, tend to prefer stocks. After that, we notice that bigger portfolios (more than 20, more than 50 monthly positions) in one asset-class only are mostly dominated by options. Stocks, however, remain the most-popular asset-class given that 94.5% of the investors hold stocks at one moment or another during our analyzed period.

Table 4: Details about investors' net income, investments and savings

Income (€)	% of our sample	Saved income (%)	% of our sample	Invested amount (€)	% of our sample
less than 20,000	14%	less than 10%	25%	less than 20,000	31%
20,000 - 40,000	37%	10 - 20 %	30%	20,000 - 50,000	19%
40,000 - 75,000	33%	20 - 30 %	22%	50,000 - 250,000	30%
75,000 - 150,000	13%	30 - 40 %	12%	250,000 - 1M	16%
more than 150,000	3%	more than 40%	11%	more than 1M	4%

Notes: The table presents information about investors' net income, amount of money that is already invested and their savings. Columns 1 and 2 show the net income classes and indicates the percentage of investors belonging to each of them. Columns 3 and 4 present categories for saved net income expressed in percentage of their annual net income and the fraction of investors that belong to each category. Columns 5 and 6 display already invested amounts in the financial assets and the percentage of investors that invested in them.

Table 4 contains information about investors' income and its distribution between investments and savings. Most investors (70%) earn between 20,000 and 75,000 euros per year. 55% of the investors are dedicating less than 20% of their yearly income to their savings. We also notice that the amounts they decide to invest are mostly selected to be below 20,000 euros per year or between 50,000 and 250,000 euros per year.

Our sample is composed of 10% female and 90% male investors. Regarding their education level, 73% of the sample have a university diploma. The average investor is 56 years old[3]. Majority of the investors in our sample (76%) are real estate owners.

Since Belgium is a multilingual country, the investors could place their trades in different languages: Dutch, English and French. Our dataset allows us to tell that the majority of the investors are originally from the Flemish side of Belgium - 54% of traders used Dutch language for their trading activity. 42% of the investors can be identified as French-speaking and 4% chose English as their investment language.

Table 5 reports some descriptive statistics about investors according to the asset classes they

[3] Age is calculated in 2008 using the year of birth

Table 5: investor profile and trading across asset-classes

	Stock trader	Fund trader	Option trader	Warrant trader	Bond trader	All traders
N° of traders	23,646	8,370	2,171	4,009	1,794	24,513
% of men	90 %	90 %	94 %	92 %	90 %	90%
University education	73 %	79 %	79 %	79 %	81 %	73%
Real estate owners	76 %	80 %	80 %	78 %	84 %	77%
Low income	51 %	47 %	44 %	47 %	45 %	51%
Long investment horizon	67 %	72 %	76 %	72 %	72 %	67%
Average duration	19	24	29	16	48	21
Median duration	10	18	19	3	45	10

Notes: This table describes investors' personal characteristics according to the asset-class that they trade. Row 1 describes the five asset classes an investor can trade in - stocks, funds, options, warrants and bonds. The list of personal characteristics is shown in the Column 1 - what is the percentage of men in the sample, how many investors have a university degree, how many of investors own real estate, how many belong to a low income category (earns less than 40 000 € per year), how many have a longer investment horizon, and the typical duration of a position (expressed in months) with respect to each asset class.
Rows indicates the number of investors that held each of the assets for at least a month, the percentage of men per class of asset they trade, the fraction of traders population that has a university degree, whether they own real estate, how many of the investors belong to a low income class and whether they have a long investment horizon. Last two rows indicate the average and the median number of months a position is held in each of the asset classes.

trade. Across asset-classes we see that the investors are predominantly male and highly educated. Around 80% of investors in every group have already invested in real estate, the biggest fraction of them trading bonds. The highest percentage of people earning a low income (less than 40,000 € per year) is being observed amongst stock traders. The investors trading other-than-stock asset classes belong mostly to medium or high income classes. Most of the investors have longer term goals regarding their investments: they want to invest for a period longer than 3 years. Last two rows of Table 5 show that investors holding assets in their portfolios the longest are the bond traders. On average, they hold an asset during 4 years. Fund and option traders hold their assets on average two times shorter than bond traders and the shortest period (one year on average) of asset-holding is observed among warrant and stock traders.

3.3 Financial literacy of investors

Among investors' personal characteristics, we are particularly interested in financial literacy. So we compute five different diversification scores based on the investors' answers to the A-test. The computed scores aim to understand not only investors knowledge and experience in the financial markets in general but also to get a sense of a more detailed knowledge investors possess about the so-called *complex instruments* - shares traded on a non-European market or on a European non-regulated market, non UCITS funds, warrants, turbos, options, structured products, Forex (spots, options and forward outrights), futures and CFDs (Contracts For Difference). These scores provide a more diverse and detailed view of the level of investors financial literacy. Table 6 presents the financial literacy scores we computed as well as some descriptive statistics regarding the total points achieved by the investors.

Table 6: Description of financial literacy measures

Financial literacy measures	Description	Results of investors scores	
$Score\_FL$	Investors evaluate their knowledge according to each of the *complex instruments*.	<50% points 79%	>50% points 22%
$Score\_Exp1$	Investors tell whether they have already invested in each of the *complex instruments*.	<50% points 87%	>50% points 13%
$Score\_Exp2$	Investors evaluate their experience related to each of the *complex instruments*.	<50% points 77%	>50% points 23%
$Score\_Exp3$	Investors indicate how many transactions per year on average they place of each of the *complex instruments*.	<12 trans. 62%	>12 trans. 38%
$Knowledge\_FM$	Investors indicate their knowledge of financial markets in the A-test.	Limited 53%	Advanced 47%
$Trading\ experience$	We calculate the number of sales an investor executed within given asset classes during the whole trading period.	<50 trans. 81%	>50 trans 19%

Notes: Table presents the financial literacy scores we use in our analysis. Column 1 shows the name of the financial literacy measure, Column 2 describes the measure and Column 3 presents some descriptive statistics regarding the financial literacy scores of the investors in the sample. First 4 rows of financial literacy measures ($Score\_FL$, $Score\_Exp1$, $Score\_Exp2$, $Score\_Exp3$) are computed given the investors answers about *complex* instruments traded in the financial markets. This category of instruments includes shares traded on a non-European market or on a European non-regulated market, non UCITS funds, warrants, turbos, options, structured products, Forex (spots, options and forward outrights), futures and CFDs (Contracts For Difference). The following financial literacy measure ($Knowledge\_FM$) refers to a simpler question regarding investors knowledge of financial markets in the A-test. All the above mentioned financial literacy measures are self-declared. The last financial literacy proxy called $Trading\ experience$ gives a number of sales per investor at the end of each month. The end-of-month value includes the sales made during the month as well as the sales executed during all the previous trading months (50 trades is reported as a mean-value, therefore, serves as a benchmark).

Table 6 shows that most investors (79%) admit not having a deep knowledge about *complex* financial instruments. Even a larger share of investors (87%) have invested in less than a half of 9 types of suggested financial instruments that are categorized as *complex*. Investors also claim to have limited experience investing in this type of financial instruments (77%) and that they place less than 12 transactions per year (62%) in *complex* financial instruments. This result is confirmed by the objective measure of financial literacy (*Trading experience*) which shows that the vast majority of investors (81%) have placed less than 50 transactions during the whole trading period. However, the tendency changes when the investors are asked to evaluate their knowledge regarding financial markets - given both financial literacy tests, roughly half of the investors report having advanced level of knowledge (meaning that they have a good level of knowledge about the different growth, revenue and the risk profile regarding different asset classes and/or sectors).

Table 7 reports correlation coefficients between financial literacy scores. We observe that *Score_ FL*, which represents how investors evaluate their own knowledge level about *complex* finan-

Table 7: Spearman correlations between financial literacy scores

	$Score\_FL$	$Score\_Exp1$	$Score\_Exp2$	$Score\_Exp3$	$Knowledge\_FM$
$Score\_FL$		0.750***	0.887***	0.486***	0.603***
$Score\_Exp1$	0.750***		0.798***	0.527***	0.504***
$Score\_Exp2$	0.887***	0.798***		0.511***	0.552***
$Score\_Exp3$	0.486***	0.527***	0.511***		0.399***
$Knowledge\_FM$	0.603***	0.504***	0.552***	0.399***	
$Trading\_experience$	0.616***	0.510***	0.564***	0.426***	0.110***

Notes: This table presents correlation coefficients between different financial scores that we use in our analysis. *Score_FL* represents how investors evaluate their own knowledge regarding *complex* financial instruments. *Score_Exp1* indicates whether investors have already invested in *complex* financial instruments. *Score_Exp2* represents investors' self-evaluation of their experience regarding their trading experience of *complex* financial instruments. *Score_Exp3* reports how many trades of *complex* financial instruments on average an investor pleases per year. *Knowledge_FM* stands for investors' answers regarding their level of financial market knowledge in the A-test. *Trading experience* represents the cumulated number of sales an investor has made. Since this value is evolving with time, for this correlation table we use the value measured in 2008 (when the MiFID questionnaires to declare the other, self-reported financial literacy scores were filled in), or the last available value.

cial instruments, is highly correlated with their past investments in these instruments (*Score_Exp1*) and even more correlated with investors' self-assessment regarding experience (*Score_Exp2*). The better an investor perceives her own past investment experience, the higher is the reported knowledge about considered financial instruments.

Additionally, if an investor has already invested in *complex* financial instruments (*Score_Exp1*), she is likely to be positive about her investment experience (*Score_Exp2*). What seems to be odd is that the average number of transactions an investor has placed regarding *complex* financial instruments (*Score_Exp3*) is only weakly related to her knowledge of financial markets, as reported in the A-test (*Knowledge_PM*). However, we see that the tendency is similar when we look at the objective financial literacy measure (*Trading experience*). It is also only weekly correlated with the investor's knowledge of financial markets while the correlation coefficients with other financial literacy measures display stronger relationships.

3.4 Diversification measures

First, we compute the simplest diversification measures mentioned in the literature, i.e. the number of monthly positions held in a portfolio as well as the number of positions in a particular asset class held. We also compute the number of monthly positions held regarding other types of assets (bonds, funds, options and warrants).

We also compute three variants for each of the other diversification measures presented in Subsection 2.2. Each variant is denoted by a subscript a, b or c according to the data-availability level. First, measures denoted by the subscript a (DI_a, DI_a^s, M_DI_a) are computed usinga price-free approach. It means that we use only the quantity of each asset rather than the price to estimate the value of each portfolio. In this step, it is as if we assume that every asset unit in a portfolio is worth 1 €. Second, diversification measures with a b-subscript (DI_b, DI_b^s, M_DI_b) are computed based on the evaluation of monthly investors' portfolios given the internal price (available in our primary sample, without outsourcing to the external databases) for each asset. This approach consists in valuing the asset position held by a given investor at the most recent trading price for that asset, regardless of the investor who actually carried out this transaction. For months where for a given asset no internal price is available, we use the last trade-price from our sample and we assume that it remains constant until the following trade. Third, portfolio diversification measures denoted with a c-subscript (DI_c, DI_c^s, M_DI_c) are computed using monthly investors' portfolios which values are estimated using the last price of the month when available and at last internal trade price otherwise.

Table 8 provides some statistics about the alternative diversification measures. Panel A shows that, despite the high investor propensity to invest in stocks, the average number of monthly positions (all asset classes together) and the average number of stock positions differs considerably. This suggests that using stock-only portfolios does not show a full picture regarding the diversification level of investors. We also observe that the median is systematically lower than the mean for any asset class. This suggests that a handful of large investors have a high number of monthly positions in their portfolios.

From Panels B, C and D of Table 8, we observe that, for each DI variant, the mean is lower than the median, suggesting a distribution skewed to the left. This seems to contrast with the results for the number of positions but it could also be explained by some naive diversification strategy for many investors who tend to assign an equal weight to each portfolio position. This could also indicate that the number of monthly positions in a portfolio depends on investor wealth rather than on any diversification strategy. From Panels B to E, we also notice that portfolio diversification measures based on quantities only (a subscript) are lower than the other two on average. This could be in line with investors trying to dedicate the same part of their investment across their portfolio positions.

Since the c-subscript measures rely on more accurate information, we can reasonably assume that it is more representative of the actual portfolio diversification. Consequently, these results seem to show that ignoring the price component in the weight computation is harmful and that, without historical prices allowing to price the different positions, the use of the purchase price of the position could be a rather good solution. Focusing on Panels B (stock-only portfolios) and D (multi-asset portfolios with an emphasis on funds), we observe very similar results. For both stocks and funds, a rather complete daily price history is available (see Table 2) but the similar results between b- and c-subscript measures do not show an advantage for the most data-consuming measure.

Table 9 contains the Spearman correlations between the alternative diversification measures and their various ways of computing them (based on monthly portfolios evaluated using the quantity owned of each asset, without using the information related to prices (a), based on monthly portfolios evaluated at the trading prices of each asset and assuming the prices stay constant between trades (b), based on monthly portfolios evaluated at asset's daily price and substituted by a trade price when the daily price is not available (c)).

Since all measures are supposed to assess portfolio diversification, all these correlations are

Table 8:
Descriptive statistics about the diversification measures

Diversification measure	Mean	Q1	Median	Q3
Panel A				
Number of positions	9.64	2.00	5.00	12.00
Number of stocks	7.09	2.00	4.00	9.00
Number of funds	0.88	0.00	0.00	0.00
Number of bonds	0.16	0.00	0.00	0.00
Number of options	0.93	0.00	0.00	0.00
Number of warrants	0.60	0.00	0.00	0.00
Panel B				
DI_a^s	0.48	0.23	0.55	0.73
DI_b^s	0.56	0.41	0.66	0.81
DI_c^s	0.56			
$DI_a^s - DI_b^s$	0.080***			
$DI_b^s - DI_c^s$	-0.008***			
$DI_a^s - DI_c^s$	0.076***			
Panel C				
DI_a	0.50	0.28	0.57	0.74
DI_b	0.56	0.42	0.66	0.81
DI_c	0.56	0.40	0.66	0.80
Panel D				
M_DI_a	0.50	0.28	0.57	0.74
M_DI_b	0.55	0.37	0.65	0.80
M_DI_c	0.55	0.36	0.65	0.80
Panel E				
SE_a	1.02	0.51	1.02	1.53
SE_b	1.17	0.64	1.15	1.74
SE_c	1.16	0.62	1.14	1.73

Notes: The table reports the mean, median, first and third quartiles for each diversification measure. Panel A presents the mean, median first and third quartiles regarding the number of monthly positions that investors hold within the 5 asset-classes as well as at the aggregate level. Panel B shows DI diversification scores that are computed based on stock-only portfolios. Panel C presents DI diversification scores computed taking into account all asset-classes (stocks, bonds, funds, warrants and options) in the portfolios. Panel D also considers multi-asset classes but substitutes every position held in funds by a 100 equally-weighted non overlapping positions. Each of Panels B-D contain three variants of an DI diversification measure, computed according to Equation 3. Panel E presents the mean, median, first and third quartiles regarding Shannon's Entropy, which is computed according to Equation 4. Each Panel within Panels from B to E contains one diversification measure, computed using different information. Each subscript a refers to a measure computed using investors' monthly portfolios valued using only the quantity of each asset (without taking the price into account). Subscript b indicates a measure computed using investor monthly portfolios where every position of each asset is estimated at the most recent internal trading price for that asset, regardless of the investor who actually carried out this transaction. Subscript c refers to diversification measures that were computed using daily external prices (and a last available internal trading price, when an external daily external price was not available) to calculate portfolio value.

positive and significant. The lowest correlations are observed between each of the counting measures (N^s and N^a) and the other measures. All the DI and SE are highly correlated ($\rho > 0.75$), suggesting that whatever the diversification measure used, the portfolios will be ranked similarly in terms of diversification.

For each measure, let's compare the correlations between pairs for variants represented by a a, b or c-subscript. We observe that the b-subscript measure is systematically more correlated than the

Table 9: Spearman correlation coefficients between different diversification measures

	N^s	N^a	DI_a	DI_b	DI_c	DI^s_a	DI^s_b	DI^s_c	M_DI_a	M_DI_b	M_DI_c	SE_a	SE_b	SE_c
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
N^s		0.555***	0.367***	0.390***	0.385***	0.403***	0.421***	0.420***	0.344***	0.362***	0.358***	0.450***	0.488***	0.482***
N^a	0.555***		0.614***	0.621***	0.614***	0.611***	0.636***	0.633***	0.599***	0.570***	0.571***	0.772***	0.796***	0.791***
DI_a	0.367***	0.614***		0.829***	0.836***	0.945***	0.816***	0.822***	0.936***	0.767***	0.777***	0.955***	0.815***	0.821***
DI_b	0.390***	0.621***	0.829***		0.969***	0.804***	0.924***	0.909***	0.784***	0.917***	0.898***	0.818***	0.944***	0.923***
DI_c	0.385***	0.614***	0.836***	0.969***		0.806***	0.905***	0.921***	0.802***	0.901***	0.929***	0.824***	0.918***	0.945***
DI^s_a	0.403***	0.611***	0.945***	0.804***	0.806***		0.872***	0.877***	0.889***	0.755***	0.759***	0.911***	0.797***	0.799***
DI^s_b	0.421***	0.636***	0.816***	0.924***	0.905***	0.872***		0.985***	0.766***	0.867***	0.852***	0.812***	0.888***	0.874***
DI^s_c	0.420***	0.633***	0.822***	0.909***	0.921***	0.877***	0.985***		0.771***	0.853***	0.867***	0.817***	0.876***	0.886***
M_DI_a	0.344***	0.599***	0.936***	0.784***	0.802***	0.889***	0.766***	0.771***		0.804***	0.823***	0.894***	0.768***	0.785***
M_DI_b	0.362***	0.570***	0.767***	0.917***	0.901***	0.755***	0.867***	0.853***	0.804***		0.969***	0.754***	0.861***	0.851***
M_DI_c	0.358***	0.571***	0.777***	0.898***	0.929***	0.759***	0.852***	0.867***	0.823***	0.969***		0.763***	0.845***	0.874***
SE_a	0.450***	0.772***	0.955***	0.818***	0.824***	0.911***	0.812***	0.817***	0.894***	0.754***	0.763***		0.884***	0.890***
SE_b	0.488***	0.796***	0.815***	0.944***	0.918***	0.797***	0.888***	0.876***	0.768***	0.861***	0.845***	0.884***		0.979***
SE_c	0.482***	0.791***	0.821***	0.923***	0.945***	0.799***	0.874***	0.886***	0.785***	0.851***	0.874***	0.890***	0.979***	

Notes: The table presents Spearman correlation coefficients between every type of diversification measures computed in our analysis. The columns are divided into five sections, every section contains the same but differently computed diversification measure. The difference in the ways the measures are computed reflects different ways of evaluating monthly investor portfolios and is denoted by indices a, b and c. Each subscript a means that monthly investor portfolios are evaluated using only the quantity in each asset rather than the prices. Subscript b indicates a measure computed using investor monthly portfolios where every position of each asset is estimated at the most recent internal trading price for that asset, regardless of the investor who actually carried out this transaction. The subscript c indicates that the portfolios are evaluated using external daily prices from Thomson Reuters Eikon. When information about the external daily prices is not available to compute the diversification measures with the subscript c, the internal trade price is used. Columns (1)-(2) contain correlation measures between simple diversification measures (N^s which is the number of stock-monthly positions in a portfolio and N^a which is the number of monthly positions in a portfolio) and other diversification measures. Last four column sections, Columns (3)-(11) contain different modifications of Diversification Index (DI), computed using the Equation 3. which takes values between 0 and 1 (1 meaning the portfolio is the most diversified). Columns (3)-(5) refer to DI and include all types of monthly positions analyzed in our study - stocks, bonds, funds, options and warrants. Columns (6)-(8) present results of DI when it is computed using stock-only portfolios. Columns (9)-(11) exhibit modified DI correlation coefficients. The modification consists of substituting every fund by a 100 equally-weighted and non-overlapping positions, as suggested by Dorn & Huberman (2005). Columns (12)-(14) present three versions of Shannon's Entropy (SE) correlation coefficients. Significance levels are *p<0.1, **p<0.05, and ***p<0.01.

a-subscript measure with the c-subscript . This confirms the comments we built from Table 8: b- and c-subscript measures are quite similar in terms of diversification measurement. From all this, we can conclude that assuming that the value of the portfolio positions remains constant over time is not a big issue and that we do not necessarily need historical prices for all the assets included in a portfolio. A deeper analysis will be necessary to check whether this holds for portfolios that exhibit a higher proportion of more volatile assets (e.g., options and warrants).

We complete our analysis by providing information whether these diversification measures can tell us anything about the quality of the diversification. In order to do so, we calculate the standard deviation of portfolio returns per investor. We use the overall duration of each portfolio's existence and the average per investor value of the diversification scores. Table 10 displays the Spearman correlation coefficients between the standard deviation of net returns (denoted $S.d$) and the average scores of diversification measures (denoted $\overline{N^s}$, $\overline{N^a}$, and so on). It shows that the higher the standard deviation is, the lower the diversification scores are, which makes us conclude that high diversification scores capture also the riskiness of the portfolios.

Table 10: Spearman correlation coefficients between different average values of diversification measures and the standard deviation of net returns

	$S.d.$	$\overline{N^s}$	$\overline{N^a}$	$\overline{DI_c^s}$	$\overline{DI_c}$	$\overline{M\_DI_c}$	$\overline{SE_c}$
$S.d.$		-0.017**	-0.023***	-0.037***	-0.038***	-0.036***	-0.033***
$\overline{N^s}$	-0.017**		0.599***	0.483***	0.431***	0.401***	0.529***
$\overline{N^a}$	-0.023***	0.599***		0.633***	0.658***	0.616***	0.809***
$\overline{DI_c^s}$	-0.037***	0.483***	0.633***		0.880***	0.828***	0.858***
$\overline{DI_c}$	-0.038***	0.431***	0.658***	0.880***		0.932***	0.955***
$\overline{M\_DI_c}$	-0.036***	0.401***	0.616***	0.828***	0.932***		0.887***
$\overline{SE_c}$	-0.033***	0.529***	0.809***	0.858***	0.955***	0.887***	

Notes: The table presents Spearman correlation coefficients between standard deviation of net returns ($S.d.$) and the average values of portfolio diversification measures (remaining columns). The standard deviation is calculated taking into account the total amount of monthly net returns during which each investor holds a portfolio. The remaining columns show the mean value of different diversification scores for each investor - $\overline{N^s}$ represents the average number of stocks held in an investor's portfolio, $\overline{N^a}$ shows the average number of assets held in an investor's portfolio, $\overline{DI_c^s}$ shows the average values per investor of a Diversification Index, computed based on stock-only portfolios and using monthly-asset prices when available and an internal price otherwise. $\overline{DI_c}$ represents the average values per investor of a Diversification Index, computed including all asset-classes as well as evaluating the portfolios using the monthly-price when available and an internal price otherwise. $\overline{M\_DI_c}$ shows the average values per investor of a *Modified* Diversification Index, which is computed including all asset-classes and using monthly asset prices when available and an internal price otherwise. The difference between the modified version of Diversification Index (M_DI_c) and the general version (DI_c) is that every position in funds is substituted by a 100 of equally-weighted non-overlapping positions, to underline the intrinsically better level of diversification of funds. SE_c represents the average values of Shannon's Entropy diversification measure computed including all asset-classes and using monthly-asset prices to evaluate the portfolios, when available and an internal price otherwise. Significance levels are *p<0.1, **p<0.05, and ***p<0.01.

4 Diversification and individual characteristics

This section is devoted to the multivariate analysis of the relationship between financial literacy and portfolio diversification. We just described some empirical results related to different diversification measures. As mentioned in Subsection 3.3, our dataset also allows to identify different financial literacy measures. We use OLS regression models with different diversification measures as dependent variables and use different financial literacy scores as dependent variables to check whether the relationships are sensitive to the score used. In addition, we use some usual control variables such as gender, age, and others. Table 11 reports the estimates of six models that differ only by a dependent variable which are the diversification measures we use. They are the following: (1) average number of monthly stock positions ($\overline{N^s}$), (2) average number of monthly positions ($\overline{N^a}$), (3) and (4) two average DI measures computed based on stock-only portfolios in the models ($\overline{DI^s_a}$, $\overline{DI^s_c}$) and (5) and (6) two average DI measures computed using multi-asset portfolios in the models ($\overline{DI_a}$, $\overline{DI_c}$). Subscript a in the Models (3) and (5) means that the diversification measures are computed using monthly portfolios that are valued using only the quantity of each asset (without taking the price into account). Subscript c in Models (4) and (6) refers to diversification measures based on monthly portfolios that are evaluated using the external daily price of each asset, when it is available, and the last available internal trading price otherwise. Higher values of DI measures represent a higher level of portfolio diversification.

Based on the results in Table 11, we observe that portfolio diversification significantly increases with age[4], education level[5], income[6]. The positive effect of Age on portfolio diversification could be explained by accumulated trading experience, as confirmed by the findings of Nicolosi et al. (2009). Regarding the income dummies used to identify $Medium$ and $High$ income investors, our results are in line with Calvet et al. (2007) who find that wealthier investors invest more efficiently. The higher the income of an investor is, the better she diversifies her portfolio. As for gender, the parameter estimate is positive and significant for the measure $\overline{N^s}$ (Models 1), which means that men hold on average a higher number of stocks in their portfolios. However, when diversification is measured by the average number of positions in any asset class, or an average of a DI measure (Models 2-6), the diversification scores do not seem to be affected by gender.

The parameter estimates for financial literacy ($Score\_FL$) is positive and highly significant for most of the diversification measures. This means that financial literacy positively affects portfolio diversification. However, we notice that this is not the case for Models 3-4, where DI is measured using stock-only portfolios. This could indicate that more financially literate investors dedicate a lower fraction of their portfolios to stocks and prefer to hold multi-asset portfolios or that the restriction of the portfolio to stock positions only could lead to unreliable results.

As mentioned in Subsection 3.3, financial literacy is assessed based on a questionnaire filled in by the investors. We build different scores related to financial literacy based on their answers to a set of questions included in the A-test. These questions aim at assessing level of knowledge

[4] Our dataset contains the year of birth. For each investor, *Age* is computed based on year 2012.

[5] *Education* is a dummy variable. It is set to one when the investor has a university or an equivalent degree, and to zero otherwise.

[6] This refers to the annual net income. We define *Low income* as less than 40,000€, *Medium income* as between 40,000 and 75,000€ and *High income* as above 75,000€.

regarding *complex* financial instruments. This category of instruments includes shares traded on a non-European market or on a European non-regulated market, non UCITS funds, warrants, turbos, options, structured products, Forex (spots, options and forward outrights), futures and CFDs (Contracts For Difference). Tables 11 and 12 use one of these scores (*Score_FL*) as a proxy for financial literacy. *Score_FL* indicates how an investor perceives her own level of knowledge concerning *complex* financial instruments.

Given the multiplicity of financial literacy scores available (Table 6), we implement a robustness check by substituting a financial literacy measure by another. Let's remember that *Score_Exp1* indicates whether an investor has already invested in each of the categories of *complex* financial instruments and that *Score_Exp2* provides information about how an investor evaluates her own experience regarding each product belonging to a *complex* financial instrument category. *Score_Exp3* indicates, on average, how many trades an investor placed yearly in those *complex* financial instruments. *Knowledge_FM* illustrates investors' level of expertise regarding financial markets. Finally, *Trading experience* constitutes an objective financial literacy measure - it indicates the cumulated number of asset-sales. Given that these measures are highly correlated (see Table 7), we expect to find the same results. However, our results differ slightly. Table 15 report the parameter estimates for identical models where only the financial literacy measure has been replaced and some small differences in the results were observed. Table 15 indicates a negative impact of financial literacy scores to the average diversification measure (DI) computed using stock-only portfolios. The results are exactly the same (therefore not repeated) when we use $Score\_Exp2$ as a proxy for investor financial literacy. Once again, the results, as regarding the Table 11, suggest that the restriction of the portfolio to stock positions only could lead to unreliable results. Table 13, where an objective financial literacy score $Trading\ experience$ is used, displays a positive influence of investor financial literacy to the average portfolio diversification scores. The results are identical (therefore not repeated) when we use $Score\_Exp3$ and $Knowledge\_FM$ to indicate the level of investors financial literacy. The increase in financial knowledge is associated with a better portfolio diversification capability. Tables 13 and 14 complete the analysis provided by Tables 11 and 12 by using monthly, not average, diversification scores and by including not only a different financial literacy measure that is objective ($Trading\ experience$) but also independent variables such as $Hold\ funds$ which indicates whether a given investor has funds in her monthly portfolio and $Portfolio\ value$, which gives the value of a monthly investor's portfolio in millions of euros. Tables 13 and 14 show us that in addition of diversification scores increasing with age, education and income level, they also increase when investors hold funds and invest in multiple asset-classes.

Table 12 reports results for similar OLS regression models as those in Table 11, but with different dependent variables. We use the same independent variables in order to explain the rest of diversification measures that we computed in our analysis - Modified DI measure ($\overline{M\_DI_a}$, $\overline{M\_DI_c}$) and Shannon's Entropy $\overline{SE_a}$, $\overline{SE_c}$). Models (1) and (2) from Table 12 are identical to Models (5) and (6) from Table 11. Again, Table 12 shows that diversification increases with age. University graduates diversify better than investors with a lower level of education. Income also has a positive influence on the diversification scores of the investors. Once again, we observe that regardless of the diversification measure, financial literacy scores matter - the higher the financial knowledge about *complex* financial instruments of the investors is, the better their diversification scores are. In order to check for the robustness, we compute the same regressions, that differ only by the used financial literacy measure (from the Table 6). We observe that our results are robust - regardless of a financial literacy measure, the results are equivalent to those displayed in Table 12 (therefore are not repeated).

Table 11:
Parameter Estimates from OLS regressions models for average diversification scores (measured by number of positions, DI^s and DI)

	Dependent variable:					
	$\overline{N^s}$	$\overline{N^a}$	$\overline{DI^s_a}$	$\overline{DI^s_c}$	$\overline{DI_a}$	$\overline{DI_c}$
	(1)	(2)	(3)	(4)	(5)	(6)
Gender	0.601**	0.228	0.004	0.007	0.001	0.002
	(0.239)	(0.159)	(0.006)	(0.006)	(0.005)	(0.006)
Age_2008	0.180***	0.143***	0.005***	0.005***	0.005***	0.005***
	(0.005)	(0.004)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Education	0.688***	0.436***	0.017***	0.013***	0.022***	0.016***
	(0.169)	(0.113)	(0.004)	(0.004)	(0.004)	(0.004)
Medium income	0.296*	0.221**	0.005	0.007	0.002	0.004
	(0.161)	(0.107)	(0.004)	(0.004)	(0.004)	(0.004)
High income	1.398***	1.148***	0.025***	0.024***	0.024***	0.020***
	(0.208)	(0.138)	(0.005)	(0.005)	(0.005)	(0.005)
Score_FL	2.866***	0.759***	−0.006	−0.005	0.018***	0.015***
	(0.146)	(0.097)	(0.004)	(0.004)	(0.003)	(0.004)
Intercept	−2.656***	−1.611***	0.157***	0.221***	0.220***	0.286***
	(0.370)	(0.247)	(0.009)	(0.010)	(0.008)	(0.009)
Observations	23,863	23,863	23,863	23,863	23,863	23,863
Adjusted R^2	0.065	0.071	0.068	0.061	0.062	0.055

Notes: This table reports OLS regression results for six different models that differ only by the dependent variable used. The dependent variables are the average number of stocks in an investor's portfolio ($\overline{N^s}$), the average number of monthly positions in an investor's portfolio ($\overline{N^a}$), the average DI measure computed using stock-only portfolios ($\overline{DI^s_a}$, $\overline{DI^s_c}$) and the average DI measure computed including all assets in the portfolios ($\overline{DI_a}$, $\overline{DI_c}$). Subscript a in Models (3) and (5) means that the monthly investors' portfolios are evaluated using only the quantity in each asset, without taking the price into account. Subscript c in Models (4) and (6) means that investors' monthly portfolios are computed using external daily prices of the assets, when available, and a last available internal trading price otherwise. In the list of dependent variables, *Gender* is a dummy set to 1 for males, *Age_2008* is the investor's age in 2008, *Education* is a dummy set to 1 for investors with a university degree, *Medium income* is a dummy set to 1 for investors that earn between 40,000 and 75,000 € per year, *High income* is a dummy set to 1 for investors earning above 75,000 € per year. The last dependent variable *Score_FL* is computed given the investors' answers to the A-test. It indicates how investors evaluate their knowledge regarding *complex* instruments traded in the financial markets. Significance levels are *p<0.1, **p<0.05, and ***p<0.01.

Table 12:
Parameter Estimates from OLS regressions models for diversification
(measured by DI, M_DI and SE)

	Dependent variable:					
	$\overline{DI_a}$	$\overline{DI_c}$	$\overline{M\_DI_a}$	$\overline{M\_DI_c}$	$\overline{SE_a}$	$\overline{SE_c}$
	(1)	(2)	(3)	(4)	(5)	(6)
Gender	0.001	0.002	0.002	0.006	0.011	0.012
	(0.005)	(0.006)	(0.005)	(0.006)	(0.013)	(0.014)
Age_2008	0.005***	0.005***	0.004***	0.004***	0.012***	0.013***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0003)	(0.0003)
Education	0.022***	0.016***	0.024***	0.014***	0.047***	0.035***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.009)	(0.010)
Medium income	0.002	0.004	0.001	0.002	0.006	0.012
	(0.004)	(0.004)	(0.004)	(0.004)	(0.009)	(0.010)
High income	0.024***	0.020***	0.018***	0.014***	0.069***	0.071***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.011)	(0.012)
Score_FL	0.018***	0.015***	0.017***	0.012***	0.053***	0.053***
	(0.003)	(0.004)	(0.003)	(0.004)	(0.008)	(0.009)
Intercept	0.220***	0.286***	0.235***	0.283***	0.304***	0.397***
	(0.008)	(0.009)	(0.008)	(0.009)	(0.020)	(0.022)
Observations	23,863	23,863	23,863	23,863	23,863	23,863
Adjusted R^2	0.062	0.055	0.055	0.049	0.076	0.072

Notes: This table reports OLS regression results for six different models that differ only by the dependent variable used. The dependent variables are the average DI measures computed including all types of assets in the portfolios ($\overline{DI_a}$, $\overline{DI_c}$), the average modified versions of DI measures where each position in funds is substituted by a 100 non-overlapping equally-weighted positions ($\overline{M\_DI_a}$, $\overline{M\_DI_c}$) and average Shannon's Entropy measures ($\overline{SE_a}$, $\overline{SE_c}$). Subscript a in the models (1), (3) and (5) means that the monthly investors' portfolios are evaluated using only the quantity of each asset rather than taking the price into account. Subscript c in Models (2), (4) and (6) means that investors' monthly portfolios are computed using external daily asset prices, when available, and a last available internal trading price otherwise. In the list of dependent variables, *Gender* is a dummy set to 1 for males, *Age_2008* is the investor's age in 2008, *Education* is a dummy set to 1 for investors with a university degree, *Medium income* is a dummy set to 1 for investors that earn between 40,000 and 75,000 € per year, *High income* is a dummy set to 1 for investors earning above 75,000 € per year. The following dependent variable *Score_FL* refers to the financial literacy measures computed given the investors' answers to the A-test. It reflects investors' answers about their knowledge regarding *complex* instruments traded in the financial markets. Significance levels are *p<0.1, **p<0.05, and ***p<0.01.

Table 13:
Parameter Estimates from OLS regressions models for diversification (measured by number of positions, DI^s and DI), using *Trading experience* as a proxy for financial literacy

	Dependent variable:					
	N^s	N^a	DI^s_a	DI^s_c	DI_a	DI_c
	(1)	(2)	(3)	(4)	(5)	(6)
Gender	−0.249	0.176	−0.006	−0.003	−0.002	−0.003
	(0.239)	(0.273)	(0.006)	(0.006)	(0.006)	(0.006)
Age	0.118***	0.121***	0.005***	0.005***	0.004***	0.004***
	(0.012)	(0.009)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Education	0.581***	0.980***	0.022***	0.014***	0.021***	0.012***
	(0.138)	(0.194)	(0.004)	(0.004)	(0.004)	(0.004)
Medium income	0.386***	0.623***	0.003	0.004	0.001	0.003
	(0.126)	(0.194)	(0.004)	(0.004)	(0.004)	(0.004)
High income	1.363***	1.819***	0.026***	0.025***	0.021***	0.018***
	(0.203)	(0.308)	(0.005)	(0.005)	(0.005)	(0.005)
Hold funds	2.529***	6.980***	0.066***	0.069***	0.096***	0.074***
	(0.167)	(0.222)	(0.004)	(0.004)	(0.004)	(0.004)
Portfolio value	−0.001	−0.004	−0.00003	−0.0001	−0.0001	−0.0002
	(0.013)	(0.020)	(0.0001)	(0.0002)	(0.0002)	(0.0003)
Trading experience	0.035***	0.063***	0.0003***	0.0003***	0.0003***	0.0003***
	(0.008)	(0.005)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Intercept	−0.999**	−1.492***	0.212***	0.287***	0.249***	0.321***
	(0.426)	(0.465)	(0.010)	(0.010)	(0.009)	(0.010)
Observations	1,364,097	1,364,097	1,305,971	1,305,971	1,364,097	1,364,097
Adjusted R^2	0.242	0.302	0.081	0.078	0.084	0.070

Notes: This table reports OLS regression results for six different models that differ only by the dependent variable used. The dependent variables are the DI measures computed including all types of assets in the portfolios (DI_a, DI_c), the modified versions of DI measures where each position in funds is substituted by a 100 non-overlapping equally-weighted positions (M_DI_a, M_DI_c) and Shannon's Entropy measures (SE_a, SE_c). Subscript a in the models (1), (3) and (5) means that the monthly investors' portfolios are evaluated using only the quantity of each asset rather than taking the price into account. Subscript c in Models (2), (4) and (6) means that investors' monthly portfolios are computed using external daily asset prices, when available, and a last available internal trading price otherwise. In the list of dependent variables, *Gender* is a dummy set to 1 for males, *Age* is the investor's age computed based on her birth year, *Education* is a dummy set to 1 for investors with a university degree, *Medium income* is a dummy set to 1 for investors that earn between 40,000 and 75,000 €per year, *High income* is a dummy set to 1 for investors earning above 75,000 €per year. *Hold funds* is a dummy variable set to 1 to indicate if an investor held funds in her portfolio during a given month. *Portfolio value* describes the value of a monthly portfolio, evaluated using external prices upon their availability and is expressed in millions of euros. *Trading experience* refers to the number of sales made by each investor within given asset classes. At the end of each month, the value of *Trading experience* reflects the number of sales an investor has made during the considered month as well as all the previous trading months. Significance levels are *p<0.1, **p<0.05, and ***p<0.01.

Table 14:
Parameter Estimates from OLS regressions models for diversification (measured by DI, M_DI and SE), using *Trading experience* as a proxy for financial literacy

	Dependent variable:					
	DI_a	DI_c	M_DI_a	M_DI_c	SE_a	SE_c
	(1)	(2)	(3)	(4)	(5)	(6)
Gender	−0.002	−0.003	−0.002	−0.0005	−0.001	−0.006
	(0.006)	(0.006)	(0.006)	(0.006)	(0.014)	(0.016)
Age	0.004***	0.004***	0.004***	0.004***	0.011***	0.012***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0005)	(0.001)
Education	0.021***	0.012***	0.020***	0.013***	0.043***	0.028***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.010)	(0.011)
Medium income	0.001	0.003	−0.001	0.002	0.006	0.013
	(0.004)	(0.004)	(0.004)	(0.004)	(0.009)	(0.010)
High income	0.021***	0.018***	0.013***	0.011**	0.068***	0.068***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.012)	(0.013)
Hold funds	0.096***	0.074***	0.116***	0.039***	0.267***	0.238***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.009)	(0.011)
Portfolio value	−0.0001	−0.0002	−0.0001	−0.0002	−0.0002	−0.0005
	(0.0002)	(0.0003)	(0.0002)	(0.0002)	(0.001)	(0.001)
Trading experience	0.0003***	0.0003***	0.0003***	0.0003***	0.001***	0.001***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0003)	(0.0003)
Intercept	0.249***	0.321***	0.267***	0.325***	0.340***	0.450***
	(0.009)	(0.010)	(0.009)	(0.010)	(0.024)	(0.027)
Observations	1,364,097	1,364,097	1,364,097	1,364,097	1,364,097	1,364,097
Adjusted R^2	0.084	0.070	0.084	0.056	0.124	0.115

Notes: This table reports OLS regression results for six different models that differ only by the dependent variable used. The dependent variables are the DI measures computed including all types of assets in the portfolios (DI_a, DI_c), the modified versions of DI measures where each position in funds is substituted by a 100 non-overlapping equally-weighted positions (M_DI_a, M_DI_c) and Shannon's Entropy measures (SE_a, SE_c). Subscript a in the models (1), (3) and (5) means that the monthly investors' portfolios are evaluated using only the quantity of each asset rather than taking the price into account. Subscript c in Models (2), (4) and (6) means that investors' monthly portfolios are computed using external daily asset prices, when available, and a last available internal trading price otherwise. In the list of dependent variables, *Gender* is a dummy set to 1 for males, *Age* is the investor's age computed based on her birth year, *Education* is a dummy set to 1 for investors with a university degree, *Medium income* is a dummy set to 1 for investors that earn between 40,000 and 75,000 €per year, *High income* is a dummy set to 1 for investors earning above 75,000 €per year. *Hold funds* is a dummy variable set to 1 to indicate if an investor held funds in her portfolio during a given month. *Portfolio value* describes the value of a monthly portfolio, evaluated using external prices upon their availability and is expressed in millions of euros. *Trading experience* refers to the number of sales made by each investor within given asset classes. At the end of each month, the value of *Trading experience* reflects the number of sales an investor has made during the considered month as well as all the previous trading months. Significance levels are *p<0.1, **p<0.05, and ***p<0.01.

5 Conclusion

It is well-known that portfolio diversification is key for investors but empirical evidence suggests individual investors tend to hold poorly diversified portfolios and that these mistakes may result in significant losses. Based on a dataset combining trading records for more than 24,000 retail investors over a ten-year period and answers to questionnaires these investors had to fill in within the context of MiFID European regulation, we examine the relationship between financial literacy and portfolio diversification.

Our paper differs from many past studies in different respects. First, we rely on financial literacy measures that have been gathered with this purpose rather than on general proxies, such as income or education, that are known to be correlated with financial literacy. Second, the computation of portfolio diversification measures are based on actual trading records rather than questionnaires submitted to investors to collect information about portfolio composition at a given time. Finally, the trading records include trades on bonds, mutual funds, options, stocks and warrants and are not limited to one or two of these asset classes. In other words, we rely on more specific proxies for financial literacy and on more accurate information about investors' portfolios and their evolution. These portfolios may include several asset classes.

We compute three main diversification measures: (1) the number of positions an investor holds in his/her portfolio, (2) a Diversification Index (DI), derived from the well-known Herfindhal-Hirschmann index and (3) the Shannon's Entropy measure (SE). In addition, several variants of these diversification measures are used to examine the effects of limiting portfolios to stock positions only or of enhancing the weight of funds in a portfolio.

Another problem we address is how sensitive are the results to the availability of data to value the portfolio positions at regular time intervals in order to get more or less accurate weights. For this, we first use historical data from external databases (Eurofidai and Eikon). This allows to have a precise, but time-consuming, approach to measure the weights for most of the assets held by the investors of our sample. Second, we consider a simpler approach - valuing any portfolio position based on the most recent trade price within our sample. Finally, we implement an overly simplified method that assigns the same value (i.e., 1€) for every asset unit held in a portfolio.

We contribute to the literature regarding portfolio diversification in three ways. First, based on accurate data of portfolios, some of which are multi-asset invested, we show that there is a significant improvement in investors' portfolio diversification when considering non-stocks positions. Second, regardless of the diversification measure and of the score to proxy financial literacy, we empirically evidence a positive and robust relationship between financial literacy and portfolio diversification. In other words: financial literacy matters. Third, our results are not very sensitive to the method we used to compute weights, whether it is time-consuming in gathering external data, approximative or simpler.

Our study shows the importance of financial literacy for individual investors' diversification capability, which is a useful result for policy makers that want to limit investors' losses due to portfolio underdiversification. This paper also gives important insights for future research by questioning the necessity of using an external database for historical prices to draw valid conclusions.

This paper is a first version. In the future, we plan to analyze more deeply the differences that exist between investors who hold stock-only portfolios and those who hold multi-asset portfolios as well as analyzing the specificity of the stock positions held by investors who are also invested in funds.

References

Abreu, M. & Mendes, V. (2010), 'Financial literacy and portfolio diversification', *Quantitative finance* **10**(5), 515–528.

Barber, B. M. & Odean, T. (2000), 'Trading is hazardous to your wealth: The common stock investment performance of individual investors', *The Journal of Finance* **55**(2), 773–806.

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

Bera, A. K. & Park, S. Y. (2008), 'Optimal portfolio diversification using the maximum entropy principle', *Econometric Reviews* **27**(4-6), 484–512.

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

Calvet, L. E., Campbell, J. Y. & Sodini, P. (2007), 'Down or out: Assessing the welfare costs of household investment mistakes', *Journal of Political Economy* **115**(5), 707–747.

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

DeMiguel, V., Garlappi, L. & Uppal, R. (2009), 'Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?', *The review of Financial studies* **22**(5), 1915–1953.

Dorn, D. & Huberman, G. (2005), 'Talk and action: What individual investors say and what they do', *Review of Finance* **9**(4), 437–481.

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

Florentsen, B., Nielsson, U., Raahauge, P. & Rangvid, J. (2019), 'The aggregate cost of equity underdiversification', *Financial Review* **54**(4), 833–856.

Friend, I. & Blume, M. E. (1975), 'The demand for risky assets', *The American Economic Review* pp. 900–922.

Goetzmann, W. N. & Kumar, A. (2004), 'Diversification decisions of individual investors and asset prices', *Yale School of Management Working Papers (Yale School of Management.)* .

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

Goetzmann, W. N., Kumar, A. et al. (2005), 'Why do individual investors hold under-diversified portfolios?'.

Guiso, L. & Jappelli, T. (2008), 'Financial literacy and portfolio diversification'.

Hoffmann, A. O., Shefrin, H. & Pennings, J. M. (2010), 'Behavioral portfolio analysis of individual investors', *Available at SSRN 1629786* .

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

Markowitz, H. (1952), 'The utility of wealth', *Journal of Political Economy* **60**(2), 151–158.

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

Tu, J. & Zhou, G. (2011), 'Markowitz meets talmud: A combination of sophisticated and naive diversification strategies', *Journal of Financial Economics* **99**(1), 204–215.

Volpe, R. P., Kotel, J. E. & Chen, H. (2002), 'A survey of investment literacy among online investors', *Journal of Financial Counseling and Planning* **13**(1), 1.

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

Woerheide, W. & Persson, D. (1992), 'An index of portfolio diversification', *Financial Services Review* **2**(2), 73–85.

Appendix

A Robustness check for results reported in Table 11 using alternative financial literacy scores

Table 15: Parameter Estimates from OLS regressions models for diversification (measured by number of positions, DI^s and DI); using an alternative financial literacy score *Score_Exp1*

	Dependent variable:					
	$\overline{N^s}$	$\overline{N^a}$	$\overline{DI^s_a}$	$\overline{DI^s_c}$	$\overline{DI_a}$	$\overline{DI_c}$
	(1)	(2)	(3)	(4)	(5)	(6)
Gender	0.646***	0.247	0.003	0.006	0.002	0.002
	(0.238)	(0.159)	(0.006)	(0.006)	(0.005)	(0.006)
Age	0.170***	0.141***	0.005***	0.005***	0.005***	0.005***
	(0.005)	(0.004)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Education	0.765***	0.473***	0.020***	0.016***	0.023***	0.017***
	(0.168)	(0.112)	(0.004)	(0.004)	(0.004)	(0.004)
Medium income	0.224	0.215**	0.006	0.007	0.002	0.004
	(0.161)	(0.108)	(0.004)	(0.004)	(0.004)	(0.004)
High income	1.333***	1.168***	0.028***	0.026***	0.025***	0.021***
	(0.207)	(0.138)	(0.005)	(0.005)	(0.005)	(0.005)
Score_Exp1	3.335***	0.675***	−0.014***	−0.012***	0.013***	0.011***
	(0.156)	(0.104)	(0.004)	(0.004)	(0.004)	(0.004)
Intercept	−2.116***	−1.454***	0.167***	0.233***	0.223***	0.289***
	(0.367)	(0.245)	(0.009)	(0.010)	(0.008)	(0.009)
Observations	23,863	23,863	23,299	23,299	23,863	23,863
Adjusted R^2	0.068	0.071	0.070	0.062	0.062	0.054

Notes: This table reports OLS regression results for six different models that differ only by the dependent variable used. The dependent variables are the average number of stocks in an investor's portfolio ($\overline{N^s}$), the average number of monthly positions in an investor's portfolio ($\overline{N^a}$), the average DI measure computed using stock-only portfolios ($\overline{DI^s_a}$, $\overline{DI^s_c}$) and the average DI measure computed including all assets in the portfolios ($\overline{DI_a}$, $\overline{DI_c}$). Subscript a in Models (3) and (5) means that the monthly investors' portfolios are evaluated using only the quantity of each asset rather than taking the price into account. Subscript c in Models (4) and (6) means that investors' monthly portfolios are computed using external daily prices of the assets, when available, and a last available internal trading price otherwise. In the list of dependent variables, *Gender* is a dummy set to 1 for males, *Age* is the investor's age in 2012, *Education* is a dummy set to 1 for investors with a university degree, *Medium income* is a dummy set to 1 for investors that earn between 40,000 and 75,000 € per year, *High income* is a dummy set to 1 for investors earning above 75,000 € per year. The last dependent variable *Score_Exp1* is computed given the investors' answers to the A-test. It indicates whether the investors have already invested in *complex* instruments traded in the financial markets. Significance levels are *p<0.1, **p<0.05, and ***p<0.01.