MiFID questionnaire answers and stock market participation

Marie-Hélène Broihanne and Hava Orkut

LaRGE, EM Strasbourg Business School

University of Strasbourg

Preliminary version: July 20, 2017

Abstract

The Markets in Financial Instruments Directive (MiFID - 2004/39/EC and 2014/65/UE) aims at strengthening the transparency, the efficiency of European financial markets but also the protection of investors. The last objective is reached by requiring investment service providers to address a questionnaire to their clients. In this paper, we combine MiFID questionnaire answers and banking records of more than 70,000 retail clients of a big European commercial bank. We demonstrate that *MiFID indicators*, i.e. self-declared attitudes towards risk and towards losses, provide an accurate explanation of stock investment decision, controlling variables commonly studied in the literature such as gender, age and income. We find that the probability of stockholding increases with risk tolerance and the willingness to hold further financial securities during the downturn. Besides, we show that MiFID indicators exhibit greater magnitude effects than those of classical variables. We also demonstrate that other specific variables available in our datasets, such as geographical variables, matrimonial regime and holding of other risky financial products are important drivers of stock market participation. Our results are consistent with prior studies on the usual determinants of stock market participation and are robust to robustness checks.

Keywords : Stock investment, MiFID questionnaire, attitudes towards risk, attitudes towards losses

JEL Classification : G02, G11, G28

1 Introduction

Understanding retail clients' investment behaviours has many implications for asset pricing under investment mistakes such as under-diversification[1] and non-participation[2] in the financial and insurance markets. Numerous behavioural finance empirical works have examined retail investors decisions and their consequences on asset prices (Hirshleifer, 2001, Brown and Cliff, 2005 and Baker and Wurgler, 2006, 2007). Specifically, investor sentiment, defined as optimism or pessimism about stock market (Baker and Wurgler, 2006), drives asset pricing and is positively correlated to assets mispricing (Brown and Cliff, 2005). The Capital Asset Pricing Model or CAPM (Sharpe, 1964, Lintner, 1965 and Mossin, 1966) assumes that investors are rational and risk averse, an hypothesis first formulated in the fundamental work of Markowitz (1959) on portfolio selection. Indeed, asset pricing theories rest on investors' preferences under uncertainty. In all models, risk aversion is taken into account in an utility function which is maximized under the Expected Utility Theory (von Neumann and Morgenstern, 1947). Therefore, in empirical works the use of a consistent methodology to assess attitudes towards risk is necessary for understanding retail clients' investment decision.

Attitudes towards risk have been commonly assessed either by using lottery and experiments that present simple choices (Holt and Laury, 2002 and Booij and Van de Kuilen, 2009) or willingness-to-pay surveys (Cummings et al., 1986 and Mitchell and Carson, 1989), i.e. *revealed preferences*, or by using secondary data reflecting actual investment decisions (Barber and Odean, 2001 and Hoffmann et al. 2015), i.e. *stated preferences*. In this paper, we employ a different approach in which the Markets in Financial Instruments Directive[3] (labelled MiFID henceforth) questionnaire answers are used to evaluate individual preferences. This directive aims at improving the competitiveness of European financial markets and at ensuring a harmonized protection to client according to their level of financial knowledge. MiFID requires investment service providers to get a deep knowledge of their clients thanks to a questionnaire, called *MiFID questionnaire*[4]. By using such a questionnaire, investment service providers have to offer advices and financial products that are perfectly suited to clients' situation. Studying the answers to MiFID questionnaire allows highlighting clients' declared characteristics, needs and preferences and also establishing their risk profile[5]. In a nutshell, MiFID

[1] In the US, we can quote Lease et al. (1974) and more recently Odean (1999), Mitton and Vorkink (2007), Kumar (2007) and Goetzmann and Kumar (2008).

[2] In the US, we can quote Mankiw and Zeldes (1991) and Poterba and Samwick (1995).

[3] Implemented in 2007, MiFID I (2004/39/EC) gathers 31 member states of the European Economic area (28 European member states and 3 other states: Iceland, Norway and Liechtenstein). It replaces the Investment Services Directive (ISD) adopted in 1993. From January 2018, MiFID II (2014/65/UE) will be implemented and then will replace the actual directive, MiFID I (2004/39/EC) that we consider in our paper. This new regulation aims at strengthening the transparency, the efficiency of financial markets but also the protection of investors. We have to precise that MiFID questionnaire is only imposed to the MiFID member states whereas it is not used in the US.

[4] Note that the directive does not impose a standard questionnaire to investment service providers. Hence, each bank is free to prepare and organize its own questionnaire.

[5] However, doing so requires that the questionnaire that was built by the institution explores directly clients' perceptions and attitudes. Questions are exploitable on a scientific point of view (and not only written in order to fit MiFID requirements).

questionnaire intends to collect the main variables that drive risk preferences in the literature such as socio-demographic, wealth and background risk characteristics.

Our paper combines two datasets, MiFID questionnaire answers and banking records, both being provided by a large European commercial bank. A few empirical works simultaneously analyze individuals' own preferences and real investment decisions (e.g. Dorn and Huberman, 2005 and Hoffmann et al., 2015). Our approach distinguishes from previous works as we use data collected under a regulatory approach. We specifically interest in the contribution of the use of MiFID questionnaire answers in explaining stock market participation.

Our paper contributes to the literature in behavioural finance for the following four reasons.

First of all, we complement prior behavioural finance works by showing how investors' profiles built under a regulatory approach which is aimed at investor protection, may explain investor behaviour. Specifically, we shed light the contribution of the use of MiFID questionnaire answers regarding attitudes towards risk and towards losses in explaining stock market participation. Although declared by retail clients, these attitudes are important determinants of stock market participation after controlling for other well-known variables such as individual characteristics or income. Under the MiFID requirements, as it is mandatory for investment service providers to address the questionnaire to their clients, we ask whether analysing collected data will be helpful at improving knowledge of investors' behaviour.

Second, our data are provided by a big European commercial bank. This allows us to include variables that have never been used in the literature, like matrimonial regime choice, or deepened, like geographical criteria. Indeed, we are the first to introduce the matrimonial regime which is linked both to marital status, thus a socio-demographic criterion, and to wealth, as it helps to structure patrimony allocation rules within spouses. From a geographical perspective, we also include the place of living that allows discriminating individuals living in the capital region from those living in province, and the country of birth. The impact of such geographical variables have never been tested in France, whereas, in the US, the difference in financial market participation between US immigrants and natives has been demonstrated in the literature (Osili and Paulson, 2004 and Chatterjee, 2009). Furthermore, we are also able to distinguish two other risky financial products which are equity-indexed annuities and retirement accounts.

Third, previous studies on French retail investors studied investors' behavioural biases such as the disposition effect (Boolell-Gunesh et al. 2009, 2012) or their market performance (Magron 2012, 2014) or the impact of retail investors on market volatility (Foucault et al., 2011). Stockholding by French retail investors has also been addressed by Arrondel et al. (2015) who demonstrate the positive impact of basic financial literacy on stock market participation. In their paper, data come from an individual investor's survey. In this paper, our measure of stock market participation is both

direct and indirect (via active "Equity Saving Plans" held by retail clients of the bank) as direct stockholding is infrequent in France (only 2.89% of retail clients directly hold stocks in our sample). However, our results may help governments and financial authorities to understand the drivers of stock market participation on a general point of view and to take policies in order to improve it.

Finally, our study is the first to combine declared data altogether with real data. Actually, preceding studies on the behaviour of retail investors (Bellofatto et al., 2014, Dorn and Huberman, 2005 and Hoffmann et al., 2015) generally use questionnaire or market data. Our first dataset (called dataset 1 hereafter) contains MiFID questionnaire answers and helps to understand retail clients' declared perceived risk and declared risk-taking behaviour. Our second dataset (called dataset 2 hereafter) contains information dealing with socio-demographic as well as wealth and patrimony data. Therefore, our approach allows confronting declaration which could be subjective (dataset 1) and reality which is objectively observable (dataset 2). Such a joint-analysis has never been carried out for studying stock market participation yet. Like Fan and Xiao (2006), we do not focus on quantitative differences in stock holding but only on the decision to hold. As our sample is representative of the whole population, our results would be helpful in understanding what drives stock holding decisions.

In this paper, we demonstrate that MiFID questionnaire answers regarding attitudes towards risk and losses enhance our understanding of stock investment decision. Specifically, we find that the probability of stockholding increases with risk tolerance and the willingness to hold further financial securities during the downturn. By focusing on socio-demographic indicators extracted from banking data, we also find that males, elderly, natives and those living in capital region are more likely to invest in stocks. We are the first to demonstrate that the choice of a specific matrimonial regime affects stock market participation. Indeed, retail clients opting for the separation of property regime are more likely to participate on stock market. Unlike retired retail clients, self-employed and those exercising no professional activity are more prone to invest in stock market than salaried ones. As for wealth and patrimony indicators, we find that the probability of investing in stocks increases with net monthly income and decreases as credit amount remaining to reimburse becomes higher. By focusing on retail clients' financial product holding, we find that the probability of investing in stocks increases if retail clients already hold other investment vehicles such as equity-indexed annuities and/or retirement accounts.

Our paper is organized as follows. Section 2 presents the related literature. Section 3 describes our datasets. Section 4 displays empirical results. Section 5 is dedicated to robustness checks. Section 6 concludes.

2 Related literature

In this section, we present results of prior studies for having a quick overview of the main determinants of stock market participation.

Previous works have identified gender differences in financial decision making. Indeed, women hold less risky assets (Riley and Chow, 1992, Hinz et al., 1997, Bernasek and Shwiff, 2001, Dwyer et al., 2002, Agnew et al., 2003 and Charness and Gneezy, 2012) and are less risk seeking[6] (Powell and Ansic, 1997, Jianakoplos and Bernasek, 1998, Booij and Van de Kuilen, 2009 and Booth and Nolen, 2012) than men. Jacobsen et al. (2014) point out that women have a more conservative approach since they consider financial markets as being riskier than men. Therefore, women are less likely to invest in stock market than men (van Rooij et al., 2011 and Almenberg and Dreber, 2015). According to Bajtelsmit and VanDerhei (1997) and Bajtelsmit et al. (1999), women allocate a smaller percentage of their financial assets to stocks than to bonds. Barber and Odean (2001) show that men trade stocks 45% more often than women on financial markets.

Age also influences individuals' investment decision[7]. Bodie and Crane (1997) find that the proportion of risky assets held by individuals decreases with age. Therefore, they complement the previous findings of Bakshi and Chen (1994), namely risk aversion increases with population' age. Besides, by working on asset allocation decisions of US households, Ackert et al. (2002) show that age has an impact on the mix of risky assets. Indeed, young households prefer investing more in stocks than in bonds. Combining cognitive aging and stock investment decisions among individual investors, Korniotis and Kumar (2011) show that older and experienced investors have a low tendency to hold risky portfolios and trade less frequently. According to Korniotis and Kumar (2011), introducing both age and experience allows distinguishing two confounding effects. Indeed, age refers to cognitive aging, i.e. the weakening of memory with age, whereas experience refers to accumulation of greater investment knowledge with age. In their study, Korniotis and Kumar (2011) demonstrate that the negative effects of age dominate the positive effects of experience, i.e. older investors have greater investment knowledge and financial experience but cognitive aging is responsible of their investment skills deterioration. However, the impact of age on individuals' investment decision is not always clear. Differences in findings may be attributed to the methodologies employed (laboratory experiments, surveys or portfolio holdings) and sample characteristics (households, investors). For example, Wang and Hanna (1997) show that the proportion of net wealth invested in risky assets increases with age. By focusing on a sample of faculty and staff working at a large university, Grable (2000) find that risk tolerance increases with age. By applying some sample selection criteria

[6] However, Grable and Joo (1999) and Hanna et al. (2001) do not find any link between gender and financial risk tolerance.

[7] However, Grable and Lytton (1998) do not find any link between age and risk tolerance.

dealing with financial net worth[8], Shum and Faig (2006) show that the decision to own stocks is positively correlated with age. Matching brokerage records and monthly survey data, Hoffmann et al. (2015) show that the probability of trading derivatives is higher among older and more experienced investors. Other studies indicate that the relationship between age and risk aversion is not linear. By deriving relative risk aversion indexes from actual asset allocations of the US population, Riley and Chow (1992) find a U-shaped relationship, i.e. risk aversion decreases with age then increases after 65 years old. In the same vein, by using psychometrically validated survey[9], Faff et al. (2008) show that young and older individuals are more risk tolerant compared to those who are middle aged.

In the US, the difference in financial market participation between US immigrants and natives has been demonstrated in the literature (Osili and Paulson, 2004 and Chatterjee, 2009). US immigrants hold less financial assets, such as stocks and mutual funds compared to natives. Chatterjee and Zahirovic-Herbert (2011) study the difference in homeownership between immigrants and natives. They find that natives are more likely to be owners compared to immigrants. However, by analysing immigrants who decide to become owners, they notice that they get a higher housing equity than natives. In line with their previous study, Chatterjee and Zahirovic-Herbert (2014) show that immigrants' financial asset ownership increases with the number of years of residence in the US. In a nutshell, studying the interaction between wealth or patrimony and geographic origin is important because immigrants are less wealthy than native-born households in the US (Love and Schmidt, 2015). To our knowledge, the country of birth has not been documented in Europe yet.

Concerning marital status, Agnew et al. (2003) analyze nearly 7,000 retirement accounts and find that stock allocation is higher among married investors than among their single counterparts. This finding is consistent with that of Grable (2000) who show that married individuals are more risk tolerant. By testing the hypothesis that marriage represents a sort of safe asset in portfolio framework, Bertocchi et al. (2011) show that married households are more likely to invest in risky assets than single ones. Nevertheless, Grable and Joo (2004) find the reverse evidence by analyzing a sample of faculty and staff from two large universities, namely risk tolerance is higher among single individuals[10]. Roszkowski et al. (1993) argue that single individuals are more risk tolerant than married ones since they are less faced to responsibilities, particularly with respect to dependents.

Occupational status has been mostly documented for understanding its impact on risk tolerance rather than on stock market participation. According to Maccrimmon and Wehrung (1986), individuals perceiving their incomes directly from their own activity are willing to take more risk compared

[8] In their sample, Shum and Faig (2006) only take into account US households having a financial net worth greater than or equal to 1,000$. They also require a positive total net worth and positive labour income in their sample.

[9] Faff et al. (2008) analyze the link between financial risk tolerance and risk aversion and show that they are strongly aligned for explaining decision-making under uncertainty.

[10] Grable and Joo (2004) indicate that women represent 55% of their sample.

to those having straight salary work or wage from an employer. They are then willing to choose riskier investments than salaried individuals do. Sung and Hanna (1996) argue that households with a self-employed head are more risk tolerant. Grable (2000) shows that risk tolerance is higher among professionally employed individuals perceiving a high level of income. Hence, professional occupations as well as education seem to drive individuals' attitudes towards risk. Recently, Agnew et al. (2003) show that stock allocations are higher among investors with more seniority on the job. As for portfolio composition, Fuertes et al. (2014) find that educated investors are more likely to hold better diversified equity portfolios.

As for income, Shum and Faig (2006) find that the decision to own stocks is positively associated to different measure of wealth such as financial net worth and labour income. Their findings are in line with that of Agnew et al. (2003), namely equity allocations are higher among investors with higher earnings. Besides, Barber and Odean (2001) indicate that individuals having a higher income have a higher tendency to accept market risk. Other studies underline that individuals having a higher income are more risk tolerant in comparison of those having a lower income (Cohn et al., 1975, Maccrimmon and Wehrung (1986), McInish et al., 1993, Grable, 2000, Bernheim et al., 2001, Hallahan et al., 2003 and Grable and Joo, 2004).

Concerning the indebtedness situation, Guiso et al. (1996), Fratantoni (1998) and Cardak and Wilkins (2009) show that credit-constrained households have a low tendency to hold risky assets. In a similar vein, investors who are less liquidity constrained are more likely to invest in stock market (Guiso and Sodini, 2013). Constantinides et al. (2002) show that borrowing constraint is responsible for the limited equity investment of young consumers. Furthermore, Thomas and Reza (2010) compare households with mortgage debt and those with no outstanding mortgage debt. They show that households with mortgage debt are less likely to own stocks and bonds. They underline that outstanding debt alone allows explaining households' asset market non-participation.

Individuals' risk perception and preferences are also known to be key drivers of investment decision making (Antonides and Van Der Sar, 1990). Financial risk tolerance has been shown to impact individuals' behaviour on financial markets. By using a survey, individuals' attitudes towards financial risk can be assessed and matched with brokerage records for studying simultaneously individuals' perception and their actual trading choices (Hoffmann et al., 2013, 2015). Indeed, investors with high levels of and upward revisions in risk tolerance have a high tendency to trade and hold riskier portfolios (Hoffmann et al., 2015). Focusing on the CAPM, Hariharan et al. (2000) work on a sample of investors nearing retirement, i.e. aged between 51-61, and find that risk tolerant investors invest less in risk-free assets. However, these authors show that risk tolerant investors nearing retirement do not reduce the quantity of bonds held for purchasing additional stocks.

Investment decision making is also influenced by individuals' sensitivity towards losses, i.e. loss

aversion. Specifically, loss aversion implies that individuals are more sensitive to losses than to the enthusiasm generated by gains. Indeed, loss aversion allows explaining low stock market participation. By using household survey data, Dimmock and Kouwenberg (2010) get loss aversion coefficient for each surveyed household and demonstrate that the probability of participating in equity markets decreases for households with high loss aversion. Furthermore, they show that, as loss aversion becomes higher, the probability of holding stocks decreases more than that of holding mutual funds. In an experiment framework, Schmidt and Traub (2002), Brooks and Zank (2005), Booij and Van de Kuilen (2009) and Rau (2014) show that women are more loss averse than men.

After looking at the related literature dealing with stock market participation, we focus on our sample characteristics (described throughout in Section 3), display our results (Section 4) and carry out robustness checks (Section 5) before concluding (Section 6).

3 Data

In order to assess MiFID questionnaire answers ability to explain stock investment decision, we use two datasets provided by a large European commercial bank. The first dataset (Dataset 1) contains MiFID questionnaire answers and the second dataset (Dataset 2) gathers banking records. In all, the two datasets are available for an important sample gathering more than 70,000 retail clients over the period 2007-2015. Specifically, we use data on retail clients of this commercial bank and not only on retail investors like in Barber and Odean (2000) and Bauer et al. (2009). Therefore, our sample differs from those obtained by a brokerage firm like in France (Boolell-Gunesh et al., 2009, 2012) or in other European countries (specifically in Belgium (Bellofatto et al., 2014), Germany (Weber et al., 2013), the Netherlands (Hoffmann et al., 2015) and the UK (Richards et al., 2011)) but also in the non-European countries such as in China (Chen et al., 2007 and Feng and Seasholes, 2005, 2008) and the US (Barber and Odean, 2001 and Korniotis and Kumar, 2011). We simultaneously analyze both datasets thanks to anonymity code attributed to each retail client.

In this paper, we use binary logit model in order to explain retail clients' stockholding. Our dependent variable is labelled "Stocks" and is coded 1 if the retail client has directly or indirectly held at least one stock in an "Equity Saving Plan" between 2007 and 2015 and 0 otherwise[11]. Like in Fan and Xiao (2006), we do not focus on quantitative differences of stockholding but only on the decision to participate on stock market. Thereby, our approach differs from that of Korniotis and Kumar (2011) who focus on the quantity of stocks held and also differs from that of Wachter and Yogo (2010) who study the portfolio share of financial wealth invested in stocks. This qualitative analysis of stocks is the most appropriate in our study.

[11]"Stocks" takes the value of one if the retail client holds at least one "active" equity saving plan, i.e. the quantity of stocks held differs from zero, and zero otherwise (if there is only cash in the saving plan for example).

In our sample, 11.05% of retail clients directly or indirectly invested in at least one stock throughout the period 2007-2015[12]. This proportion is in line with prior works that also underline low stock market participation in the UK (Attanasio et al., 2002), the US (Mankiw and Zeldes, 1991, Haliassos and Bertaut, 1995 and Poterba and Samwick, 1995) and France (Arrondel et al., 2015). According to the survey of Sofia of TNS Sofres, the number of individuals holding financial assets in France decreases from 20% to 11% and stockholding rate decreases from 15.9% to 8.1% between 2009 and 2015.

Several variables are available in our datasets and used to explain stock market participation. We classify them into 3 Panels (see Table 1): MiFID indicators (Panel A), socio-demographic indicators (Panel B) and wealth and patrimony indicators (Panel C). Variables belonging to Panel A are extracted from Dataset 1 and focus on MiFID questionnaire answers given by retail clients regarding attitudes towards risk and losses[13]. Variables of Panels B and C are extracted from Dataset 2 and relate to data recorded by the commercial bank. Both dependent and independent variables are presented in Table 1.

[12]The percentage of retail clients directly holding stocks is very low in our sample (1.69%).

[13]As the two datasets used in this study contains similar variables, among all the MiFID questions, we only focus on two questions referring to the attitudes towards risk and the attitudes towards losses.

Table 1 – Variable description

Variables	Definitions
Dependent variable	
Stocks	Dummy variable coded 1 if the retail client holds stock(s) in an "Equity saving plan" and 0 otherwise.
Independent variables	
	Panel A: **MiFID indicators**
Risk tolerance	Ordinal qualitative variable assessing retail clients' attitudes towards risk into 3 category variables: "Accepting", "SeekBetter" and "SeekHigh" (details are given in Table 2).
Losses	Ordinal qualitative variable assessing retail clients' attitudes towards losses into 4 category variables: "SellingAll", "SellingPart", "Waiting" and "Investing" (details are given in Table 2).
	Panel B: **Socio – demographic indicators**
Gender	Dummy variable coded 1 for males and 0 for females.
Age	Continuous variable referring to age of the retail clients as of 07/31/2015 (in years).
Native	Dummy variable coded 1 if the retail client is native of the country and 0 otherwise.
Paris	Dummy variable coded 1 if the retail client lives in and close to the capital (and biggest town) of the country and 0 otherwise.
Matrimonial	Dummy variable coded 1 if the retail client is subjected to the separation of property legal regime and 0 otherwise.
Occupations	Nominal qualitative variable referring to four professional category variables, "Self-employed", "Salaried", "Retired" and "No occupation", described below.
Self-employed	Dummy variable coded 1 if the retail client perceives directly his/her income from his/her own professional activity and 0 otherwise.
Salaried	Dummy variable coded 1 if the retail client has a wage or salary from an employer and 0 otherwise.
Retired	Dummy variable coded 1 if the retail client is retired and 0 otherwise.
No occupation	Dummy variable coded 1 for investors having no occupation (e.g. students and those having no professional activity) and 0 otherwise.
	Panel C: **Wealth and patrimony indicators**
Income	Ordinal qualitative variable referring to net monthly income categories expressed in euros: 0€; <1,500€; [1,500€; 3,000€[; [3,000€; 5,000€[; [5,000€; 10,000€]; >10,000€.
Credit	Ordinal qualitative variable referring to credit amount remaining to reimburse categories expressed in euros: 0€; <10,000€; [10,000€; 100,000€]; >100,000€
Annuities	Dummy variable coded 1 if the retail client holds an equity-indexed annuity and 0 otherwise.
Retirement	Dummy variable coded 1 if the retail client holds a retirement account and 0 otherwise.

Table 1 describes all variables. Independent variables are classified into three panels: Panel A: MiFID indicators; Panel B: Socio-demographic indicators and Panel C: Wealth and patrimony indicators.

Descriptive statistics are reported in Table 3.

The decision to hold stocks is usually found to depend on socio-demographics, wealth, background risk (human capital, housing credit, ...) and attitudes towards risk and towards losses. In Section 3.1, we present MiFID questionnaire answers regarding attitudes towards risk and losses represented by Panel A. Banking records represented by Panels B and C are exposed in Section 3.2 and Section 3.3 respectively.

3.1 Panel A: MiFID questionnaire answers

In this paper, we use the MiFID questionnaire provided by a big European commercial bank. This questionnaire satisfies both MiFID requirements and anti-money-laundering standards[14].

The MiFID questionnaire contains five sections. The first section focuses on retail clients' socio-demographic characteristics such as gender, age or marital status. The second section mainly deals with income perceived by retail clients. The third section is dedicated to the patrimony situation referring both to real estate and movable patrimonies. The fourth section concentrates on credit amount remaining to reimburse. The fifth section refers to retail clients' saving capacity. The last and biggest section is devoted to investment objectives and includes itself four subsections dealing with the main investment objectives, attitudes towards risk, experience with financial products and attitudes towards losses during a hypothetical downturn.

The questionnaire has been administered a maximum of three times to each retail client between 2007 and 2015. The first time the questionnaire has been administered to any client was when he/she subscribed any financial instrument after 2007. The second questionnaire has been administered to retail clients three years after the first one. The third questionnaire is administered to retail clients subscribing any financial instrument after having complemented the second one or three years after the second one. Answers provided by retail clients to whom the questionnaire has been administered at least twice are stable over time. For that reason, we only focus on the most recent MiFID questionnaire answers of these retail clients. Moreover, choosing to analyze the answers to the most recent questionnaire, whichever it is the first, second or third of any client, is in line with our objective. Actually, the number of unreported answers decreases between two successive questionnaires leading to an increase of the number of answers. Furthermore, the most recent MiFID questionnaire answers are always extracted at a date which is the closest one and prior from the date of banking data extraction, i.e. 07/31/2015.

In Panel A, we focus on retail clients' attitudes towards risk[15] (represented by variable "Risk

[14] Retail clients answer the questionnaire with a financial adviser of the bank. About 70% of them consider that this commercial bank is their principal bank, i.e. they use the current account opened within the bank on a daily basis.

[15] In the questionnaire, Likert scale has not been used for assessing individuals' risk attitudes. This unidimensional measure is used by Hoffmann et al. (2015) for recording answers dealing with return expectation, risk tolerance ans risk perception.

tolerance") and towards losses (represented by the variable "Losses"). We specify that all retail clients did not answer all MiFID questions. Attitudes towards risk and towards losses have not been reported by 7.63% and 7.26% of retail clients respectively. In Table 2, we present questions and recode proposals into numerical modalities[16] according to their natural ordering. Table 3 presents descriptive statistics of all variables.

Table 2 – MiFID questions regarding attitudes towards risk and losses

Variables	Questions	Modalities	Category variables	Proposals
Risk	As a general rule,	0	Accepting	Accepting lower remuneration by taking no risk on the invested capital.
tolerance	which assertion	1	SeekBetter	Seeking better remuneration by taking a capital risk.
	best describes you ?	2	SeekHigh	Seeking high performance by accepting a significant part of capital risk.
Losses	If in the coming months,	1	SellingAll	Selling all.
	your investments value	2	SellingPart	Selling a part of your portfolio.
	would decrease by 15%,	3	Waiting	Waiting until values increase.
	what would you do ?	4	Investing	Taking advantage of a lower price to invest again.

Table 2 presents the two questions dealing with attitudes towards risk and towards losses respectively. The first column indicates variable names attributed to MiFID questions which are reported in the second column. Numerical recoded modalities (third column) and category variables (fourth column) refer to proposals indicated in the MiFID questionnaire (the fifth column).

The variable "Risk tolerance" aims at assessing the level of financial risk that a retail client is willing to bear. Actually, by analysing the three proposals of the corresponding question, we notice that risk tolerance increases from the first to the third proposal. Indeed, the variable "Accepting" (coded 0) indicates that individuals are not tolerant towards risk at all. Risk tolerance appears in "SeekBetter" (coded 1) and becomes much higher in "SeekHigh" (coded 2)[17]. In Table 3, we notice that retail clients are not very risk tolerant. About 69% of retail clients are not risk tolerant at all and 29% of them declare to be able to take a capital risk. This high proportion may explain the low stock market participation that we identified previously. Only 1.75% of retail clients are capable of bearing a significant part of capital risk. Therefore, by focusing on these two latter proposals, about 31% of retail clients consider themselves risk tolerant. In the US, Hong et al. (2004) report similar result since they find that 32.53% of their sample are risk tolerant.

The variable "Losses" places retail clients in a hypothetical context since it consists in testing retail clients' attitudes when they face a loss. Actually, the bank here looks at the sensitivity towards losses, namely loss aversion. It is usually found that the sensitivity towards losses is higher than

[16]For *Risk tolerance*, the use of the modalities 0, 1 and 2 allows reflecting retail clients' risk tolerance level. As for *Losses*, we use the modalities 1, 2, 3 and 4 since each of them refer to a behaviour observed on financial markets.

[17]Our variable is more precise than that of Hong et al. (2004) since these authors use an indicator variable for knowing whether an individual is risk tolerant or not.

the enthusiasm generated by gains[18]. We notice that the loss is represented as a percentage and not by an amount expressed in euros. Thereby, retail clients may feel the same loss whatever their wealth situation. In this question, the percentage of loss has been freely chosen by the bank and reflects the volatility of the stock market which is around 15-20% on an annual basis[19]. For this variable, modalities are recoded in increasing order of final portfolio value. In Table 3, we find that retail clients are, on the average, more likely to wait until their investment values increase during the downturn. About 20% of retail clients prefer selling the entire or a part of their portfolio. There are only about 5% of retail clients who declare being interested by purchasing additional financial securities when they face a loss.

3.2 Banking records, Panel B: Socio-demographic indicators

In Table 3, descriptive statistics of Panel B indicate that our sample has a low proportion of men (51.24%). This specificity is seldom observed in European and American studies since samples in other academic studies are mainly (about 80%) composed of male retail investors, like France (Boolell-Gunesh et al., 2009), Belgium (Bellofatto et al., 2014), Germany (Weber and Welfens, 2007), the Netherlands (Bauer et al., 2009), the UK (Richards et al., 2011), Finland[20] (Grinblatt and Keloharju, 2009) and the US[21] (Barber and Odean, 2001). However, the gender parity of our sample is similar to those analyzed in East Asia (Feng and Seasholes, 2008) due to a high participation rate of Chinese women on financial markets compared to other countries (Chen et al., 2004 and Feng and Seasholes, 2005). As for retail investors' age, the average age is about 48 years old. Our result is close to that obtained by Feng and Seasholes (2005) in China (about 35 years) [22] and slightly lower than those obtained by van Rooij et al. (2011) in the Netherlands (about 51 years), by Dhar and Zhu, 2006 in the US (50 years) and Hallahan et al. (2004) in Australia (in their study, the largest number of individuals belongs to the age bracket 51-60 years).

Two geographical variables are analyzed in our study. Indeed, we distinguish retail clients who were born in the country ("Native") from those who were born in foreign countries. Besides, we distinguish retail clients living in Parisian region ("Paris") from those living in province. This latter variable allows us to test for the first time the impact of the biggest region of the country, in economic and size terms, on stock investment decision. About 85% of French retail investors were

[18] Tversky and Kahneman (1992) develop a fundamental theory, named Cumulative Prospect Theory, in order to illustrate the asymmetry existing between gains and losses. According to their experimental work, they estimate a loss aversion coefficient (λ) which allows quantifying the sensitivity towards losses. When λ is higher than 1, individuals exhibit loss aversion. They show that individuals are 2.25 times more sensitive towards losses than gains.

[19] In Belgium, a similar question is treated in the MiFID questionnaire studied by Bellofatto et al. (2014). Indeed, the percentage of loss has also been freely chosen by the online bank and represents 20% of the investments value.

[20] Working on a dataset from Finland, Grinblatt and Keloharju (2009) only focus on men enlisted into mandatory military service. Therefore, they do not consider women in their study.

[21] Lease et al. (1974) indicate that there are 80% of male investors in the 1960s.

[22] Feng and Seasholes (2005) work on a younger sample aged of 35 years old, on average. As for the US, Barber and Odean (2002) and Dhar and Zhu (2006) find respectively 52 and 50 years old for their samples.

born in France and 12% of them live in or close to the capital (and biggest town) of the country[23]. We can assume that natives are more likely to invest in stock market as demonstrated by Osili and Paulson (2004) and Chatterjee (2009). Demographic breakdown has also been mentioned by Tekce and Yilmaz (2015) for explaining Turkish retail investors' behaviour on stock market. They argue that about half of the individuals in their sample live in the most developed region containing the biggest Turkish city, namely Istanbul, and about 17% reside in the region where the capital, Ankara, takes place. In our case, the variable "Paris" allows capturing these two specificities. First, by studying the gross domestic product (GDP) per region, the Parisian region concentrates about one-third of national wealth[24]. Hence, this region displays the highest economic performance of the whole country and represents the most developed French region. Furthermore, it includes Paris which is both the capital and the largest French city. Geographic proximity has been documented by Grinblatt and Keloharju (2001) to be a stock holding determinant for firms that are located close to the place of living of investors. Therefore, we expect that retail clients living in or close to the capital are more likely to hold stock(s).

Matrimonial regime choice is also analyzed in our study. Complementing marital status, matrimonial regime helps at structuring patrimony allocation rules within spouses during the marriage but also after its breakdown (for divorced or widowed individuals). We are the first to test whether matrimonial regime choice has an impact on stock market participation or not. Among the different French matrimonial regimes[25], we particularly focus on the separator regime, also called "separation of property regime" (represented by the variable "Matrimonial" henceforth). As these terms suggest, this matrimonial regime implies that there is no joint-ownership between spouses. A married agreement and thus a notarial deed are required for applying it. Hence, each spouse is free to manage his or her own goods and is liable for any debt incurred by him or her before and after the marriage. The notion of "common goods" does not exist. All property is owned separately by each spouse. As a matter of fact, the separation of property regime allows individuals to behave as if they were single since any increase or decrease in wealth does not impact that of the other spouse. Furthermore, this

[23]Parisian region is also called *Île-de-France* and represents an administrative region of France. Our sample is quietly similar to the French demography. According to INSEE, 11.6% of the French population were born in a foreign country and about 18.8% of them live in Parisian region in 2014.

[24]INSEE displays this result in 2013.

[25]French matrimonial regimes are divided into two categories: community and separator regimes. Community regimes refer to the notion of common goods. Indeed, all goods acquired before but also after the marriage are considered like being common goods within spouses. These are universal community of property regime and community of movables and acquets regime (real property does not belong to the community in this latter regime). In France, the standard matrimonial regime is the community property regime. Fixed by default, this regime requires no formality and thus no married agreement contrary to the other community regimes. It implies that only goods acquired after the marriage are considered like being common goods. Each spouse keeps its own goods acquired before the marriage as well as those received by inheritance, gift or will. As for separator regime, we treat it throughout our paper. In all European countries, there is also legal marriage regime. In some countries, community regime is the default option (e.g. Belgium, Italy and Luxembourg) whereas in other countries separate regime is the default option (e.g. Germany, England, and Greece). In all countries, the legal matrimonial regime can be modified. As for United States, the legal matrimonial regime differs from one state to another one.

variable can be considered like being a proxy for patrimony protection desirability by an individual. About 10% of our sample is subject to the separation of property regime which is representative of the French population. Actually, according to INSEE[26], 6.1% of married individuals have chosen the separation of property regime in 1992. This rate increased to 10% in 2010.

As several professional categories are available in Dataset 2, we chose to gather and recode them as a 4 dummy variables[27] in Panel B: "Self-employed", "Salaried", "Retired" and "No occupation". The distinction between self-employed and salaried has already been used by Maccrimmon and Wehrung (1986) and Dorn and Sengmueller (2009)[28]. Actually, among individuals exercising a professional activity, this allows to distinguish individuals who are financially independent from an employer ("Self-employed") from those who are not ("Salaried"). We also distinguish retired individuals ("Retired") from those exercising no professional activity ("No occupation"). About half of retail clients have a salary from an employer ("Salaried") and about 13% of retail clients directly perceive their incomes from their professional activity ("Self-employed"). The proportion of self-employed individuals is similar to those indicated by Sung and Hanna (1996) and Dorn and Sengmueller (2009) and is higher than that reported by van Rooij et al. (2011). We also notice that 15.59% of the individuals in our sample are retired like van Rooij et al. (2011) who indicate that 18.4% of Dutch households are retired.

3.3 Banking records, Panel C: Wealth and patrimony indicators

Looking at wealth and patrimony indicators, we are particularly interested in the net monthly income[29] provided by the professional activity of retail clients and not on the whole wealth which also includes the real estate value[30] and financial assets. Actually, Kumar (2009) and Grinblatt et al. (2011) argue that distinguishing income from the other wealth components allows providing a detailed description of individuals' financial situation. Our approach differs then from that of Cho (2014) who assumes that wealth is only consisting of financial assets and housing wealth[31], Guiso et al. (2008) who combine financial wealth and income, Hong et al. (2004), who refer to the value

[26]INSEE (Institut National de la Statistique et des Etudes Economiques) designs the French national statistics bureau.

[27]For affecting an active socio-professional category (i.e. retired, students and those having no professional activity are excluded) into "Self-employed" or "Salaried", we have used retail clients' MiFID questionnaire answers as a question is dedicated to the source of income. A comparison can be made with the whole French population by using data provided by INSEE (*Institut National de la Statistique et des Etudes Economiques* in French) that designs the French national statistics bureau.

[28]Other studies use 3 categories, labeled professional, non-professional and retired or non-employed categories (Dhar and Zhu, 2006 and Goetzmann and Kumar, 2008) or distinguish finance-related jobs from other activities (Grinblatt and Keloharju, 2009 and Fuertes et al., 2014).

[29]In Panel C, only the variables "Income" and "Credit" are extracted from the MiFID questionnaire answers, i.e. Dataset 1, due to a high proportion of unreported banking data, i.e. Dataset 2. For these variables, the use of banking data would have significantly decreased the sample size by about 80%.

[30]In our banking records, homeownership is also available. However, this information is not used due to high proportion of missing data that would have significantly decreased the sample size by approximately 77%.

[31]Dorn and Sengmueller (2009) use the total net worth which includes all financial assets and real estate.

of all assets[32]. Retail clients perceive, on average, 2,418 euros per month[33]. This amount is close to that communicated by INSEE, which is around 2,202 euros, in 2013. By analyzing the different income brackets, we find that about 76% of retail clients earn less than 3,000 euros per month. The income bracket 1,500-3,000 euros, which is coded 2, concentrates the highest number of retail clients. Our sample is representative of the whole French population with regards to income (INSEE).

The credit amount remaining to reimburse is also taken into account in our analysis. An increase of the credit amount leads to a decrease of individuals' available wealth since it refers to overall indebtedness including consumer and real estate credits. An increase of income leads to a higher wealth level whereas an increase of credit amount exercises the reverse effect. Therefore, we are able to measure the impact of both effects on stock investment decision. Half of the retail clients in our sample still have credit reimburse[34]. Among them, we find that the credit bracket 10,000-100,000 euros concentrates the highest number of retail clients. In 2010, INSEE indicates that 47% of French household have still credit to repay and the amount remaining to reimburse is, on average, about 61,900€. Our result is consistent with French population since the average credit amount indicated by INSEE belongs to that credit bracket (coded 2). Our sample is then representative of the whole French population.

Other risky financial products are also included into our analysis. Like for stocks, we only look at the decision to hold these financial products[35]. Thereby, we do not refer to financial wealth but only focus on active accounts. We are able to distinguish two risky financial products, such as equity-indexed annuities[36] ("Annuities") and retirement plans[37] ("Retirement"). Retirement plans are not comparable between France and the US at all since in France there is a pay-as-you-go system

[32]Due to little information about assets hold in retirement accounts, Hong et al. (2004) do not include them into wealth measure.

[33]For preserving a significant sample size, the net monthly income reported in this study is extracted from the MiFID questionnaire answers. Indeed, the bank enumerates different income brackets that are reported in Table 3. For computing the average income of the whole sample, we use "monetary" codes which correspond to the median values of income brackets, except for the first income which is equal to zero. For the last income bracket, we use the lower bound, i.e. 10,000 euros. This treatment allows us considering the variable "Income" as being a continuous variable. In this paper, we use these codes for better interpreting our results. Indeed, the average amount is consistent with that corresponding to the whole French population. Furthermore, we also apply numerical modalities to income brackets which are coded from 0 to 5. We find anew that retail clients earn, on average, between 1,500-3,000 euros (coded 2) since the mean computed by using modalities is about 1.90.

[34]For the variable "Credit", we apply the same treatment than "Income", i.e. we use "monetary" codes which correspond to the median values of brackets. Lower bound is used coded for the last bracket, i.e. 100,000 euros. Moreover, credit brackets are attributed to numerical modalities which are coded from 0 to 3. Among indebted retail clients, we computed the mean by using modalities and find 2.08 which correspond to the credit bracket 10,000-100,000 euros.

[35]In banking records, we also know whether retail clients hold bond(s) or not. However, we do not include this financial product into our analysis due to a weak proportion of bondholders (which is lower than 1%).

[36]This kind of contract may generate losses since it depends on financial markets' performance. Hence, individuals opting for this financial product accept to take a risk of loss in capital.

[37]The variable "Retirement plans" gathers "Popular Retirement Savings Plan", i.e. *Plan Epargne Retraite Populaire* (*PERP*) in French and "Retirement Savings Plan", i.e. *Plan Epargne Retraite* (*PER*) in French. Both products represent contracts in which the customer indirectly invest amounts on financial supports such as stocks, mutual funds, etc. In our database, they are associated to an ISIN code. Therefore, it implies financial market participation.

whereas in the US there are defined benefit plan and defined contribution plan[38]. Therefore, our study distinguishes from US studies. In our sample, 16.83% of retail clients invest in equity-indexed annuities ("Annuities") like in the US (Bricker et al., 2014). Due to retirement plan differences, we also notice a remarkable disparity between France and the US since about half of US households hold retirement accounts[39] whereas there are only 1.37% of retail clients who hold them in our sample

[38] In France, pay-as-you-go system is a retirement scheme where individuals are free to decide how much they would like to contribute at this retirement account. This system is similar to the US system, defined contribution plan. However, in defined contribution plan, the benefit depends on the amount contributed in this scheme but also on their investment performance as individuals are free to invest a fraction of employee's salary or a specific amount until their retirement. Those funds are generally invested in mutual funds or annuities that are included inside the retirement plan. We can cite for example 401(k), which is the most commonly defined contribution pension used in the US. In defined benefit plan, benefits are computed by using a fixed formula depending on earnings history, tenure of service and age. Paid by an employer, this plan guarantees a monthly payment at retirement.

[39] According to Bricker et al. (2014) who analyze the periods 2010 and 2013.

Table 3 – Descriptive statistics

		N	$\overline{X}$/ %	std	min	max
Retail clients		77,365	100%	-	-	-
Dependent variable						
Stocks		77,365	11.05%	-	-	-
Independent variables						
Panel A : MiFID indicators						
Risk tolerance		71,461	0.32	0.50	0	2
Accepting			69.35%$^{(0)}$	-	-	-
SeekBetter			28.90%$^{(1)}$	-	-	-
SeekHigh			1.75%$^{(2)}$	-	-	-
Losses		71,745	2.71	0.78	1	4
SellingAll			14.29%$^{(1)}$	-	-	-
SellingPart			6.24%$^{(2)}$	-	-	-
Waiting			73.93%$^{(3)}$	-	-	-
Investing			5.54%$^{(4)}$	-	-	-
Panel B : Socio-demographics indicators						
Gender		77,365	51.24%	-	-	-
Age		77,365	47.97	17.55	18	105
Native		77,365	84.59%	-	-	-
Paris		77,365	12.26%	-	-	-
Matrimonial		77,365	10.30%	-	-	-
Occupations						
Self-employed		77,365	12.61%	-	-	-
Salaried		77,365	55.36%	-	-	-
Retired		77,365	15.59%	-	-	-
No occupation		77,365	16.44%	-	-	-
Panel C : Wealth and patrimony indicators						
Income		77,365	2,418.07	2,192.97	0	10,000
			1.90	1.11	0	5
Income brackets :	Codes :					
0	0		7.28%$^{(0)}$	-	-	-
<1,500	750		31.62%$^{(1)}$	-	-	-
1,500-3,000	2,250		36.67%$^{(2)}$	-	-	-
3,000-5,000	4,000		15.32%$^{(3)}$	-	-	-
5,000-10,000	7,500		6.72%$^{(4)}$	-	-	-
>10,000	10,000		2.39%$^{(5)}$	-	-	-
Credit		77,365	28,668.91	38,960.65	0	100,000
			1.04	1.18	0	3
Credit brackets :	Codes :					
0	0		50.08%$^{(0)}$	-	-	-
<10,000	5,000		13.51%$^{(1)}$	-	-	-
10,000-100,000	55,000		18.70%$^{(2)}$	-	-	-
>100,000	100,000		17.71%$^{(3)}$	-	-	-
Annuities		77,365	16.83%	-	-	-
Retirement		77,365	1.37%	-	-	-

Table 3 presents descriptive statistics of the variables in Panels A, B and C. The first column reports all variables. For each variable, the second column reports the number of retail clients (N) for which the data is available. The third column reports the percentage (%) of retail clients for which the corresponding variable is equal to one for binary variables and the mean ($\overline{X}$) for continuous variables. For MiFID dummy variables, "Income" and "Credit", we provide in addition the percentage corresponding to each of the modalities indicated in parentheses and superscript. Description of MiFID modalities is detailed in Table 2. The fourth column reports the standard deviation (std). The fifth and sixth columns indicate minimum (min) and maximum (max) values.

After having described our datasets, we focus on our empirical results (Section 4) for identifying stock market participation determinants.

4 Empirical results

In Section 4, we perform a multivariate analysis by using Binary Logistic Regressions (BLR) for assessing the impact of independent variables on our interest variable "Stocks" since this latter refers to 2 attributes: holding stock(s) (1) or not (0). We choose to provide a deepened analysis of stock market participation around 3 models. Our preliminary data analysis (Appendix A) has shown that the Spearman rank correlation between attitudes towards risk and attitudes towards losses is highly significant (r_{sp}=0.25). Therefore, we decide to analyze them separately. The 3 models differ by the number of independent variables included into the analysis and their sample size. They present the following characteristics:

- Model 1 focuses only on Panels B and C. This model can be considered like being a "basic model" since we do not include MiFID questionnaire answers (Panel A). Thereby, results would be easily compared to prior studies. This model contains the highest sample size, i.e. 77,365 retail clients.

- Model 2 focuses on Panels A (only "Risk tolerance"), B and C. Complementing Model 1, Model 2 incorporates retail clients' own preferences towards risk for assessing stock market participation. The sample size differs from Model 1 since there are 5,904 retail clients who did not answer the question dealing with attitudes towards risk. Therefore, Model 2 focuses on 71,461 retail clients.

- Model 3 focuses on Panels A (only "Losses"), B and C. The sample size is lower than that of Model 1 since 5,620 retail clients did not answer the question dealing with attitudes towards losses. The sample size is thus reduced to 71,745 retail clients.

Before analysing the three models, we check the presence of possible multicollinearity problem by using "BKW indicators" introduced by Belsley et al. (1980), i.e. "condition index" and "variance-decomposition proportions". The first indicator allows to measure multicollinearity level existing between independent variables whereas the second one allows detecting on which coefficients we face such problem[40]. As condition index of our models does not exceed 18, we conclude that we do not

[40]De Bourmont (2012) demonstrates the strength of BKW indicators for testing the presence of multicollinearity problem. Another method consists in detecting such a problem is illustrated by the Variance Inflation Factor (VIF). Specifically, the VIF of any independent variable quantifies how much the variances of the estimated regression coefficients and the standard errors of other independent variable. Chatterjee et al. (2000) denote that a VIF larger than 10 and/or a mean VIF greater than or equal to 2 indicate the presence of multicollinearity problem. Results obtained from VIF are reported in Table 10 in Appendix B. Since results respect both conditions, we conclude that we do not face multicollinearity problem.

face such problem since it is below the critical threshold of 30[41].

Table 4 presents our results. We report Average Marginal Effects (denoted AMEs henceforth) instead of coefficients[42]. A positive (negative) AME of x% means that an increase of one unit of the independent variable value leads to an increase (decrease) of the probability corresponding to the dependent variable by x%. Furthermore, AME is an appropriate measure for both discrete and continuous variables[43] (Longhi and Nandi, 2015).

We first notice that almost all independent variables are statistically significant at all reasonable significance levels whatever the model we analyze. We also point out that including MiFID indicators (Models 2 and 3) into a basic model (Model 1) allows increasing the goodness-of-fit measures[44].

[41] According to Erkel-Rousse (1995), multicollinearity problem may occur as condition index is greater than 20. Respecting this critical threshold, we confirm that we do not face multicollinearity problem. Furthermore, even by replacing "Credit" and "Credit" by their numerical codes (indicated in parentheses and superscript in Table 3), the maximum value of condition index is about 15.67 leading us to confirm that we do not face multicollinearity problem.

[42] In a logit model, coefficients value only indicates the direction of the average effect and not the magnitude effect which gives more accurate information about the relationship existing between dependent and independent variables.

[43] The marginal effects at means (MEMs) is another kind of measure used by Hoffmann et al. (2015). However, this measure is only appropriate in the case of continuous variables, i.e. the variables for which there is an infinitesimal change (Longhi and Nandi, 2015).

[44] We apply different measures for ensuring the quality of goodness-of-fit of our models and their degrees of prediction. These results are available in Appendix C.

Table 4 – BLR results

	Model 1		Model 2		Model 3	
	AMEs	**std**	**AMEs**	**std**	**AMEs**	**std**
Dependent variable						
Stocks						
Independent variables						
Panel A: MiFID indicators						
SeekBetter			0.1000***	0.0022		
SeekHigh			0.1821***	0.0053		
Accepting			(omitted)			
SellingAll					-0.0817***	0.0049
SellingPart					-0.0215***	0.0048
Investing					0.0633***	0.0037
Waiting					(omitted)	
Panel B: Socio-demographic indicators						
Gender	0,0146***	0.0021	0.0086***	0.0022	0.0127***	0.0023
Age	0.0037***	0.0001	0.0036***	0.0001	0.0038***	0.0001
Native	0.0454***	0.0033	0.0398***	0.0034	0.0444***	0.0035
Paris	0.0385***	0.0029	0.0368***	0.0030	0.0352***	0.0031
Matrimonial	0.0295***	0.0029	0.0224***	0.0030	0.0281***	0.0031
Self-employed	0.0091***	0.0031	0.0086***	0.0032	0.0096***	0.0033
Retired	-0.0215***	0.0033	-0.0189***	0.0034	-0.0216***	0.0035
No occupation	0.0118***	0.0039	0.0074*	0.0041	0.0119***	0.0042
Salaried	(omitted)		(omitted)		(omitted)	
Panel C: Wealth and patrimony indicators						
ln(Income)	0.0150***	0.0010	0.0087***	0.0010	0.0133***	0.0011
ln(Credit)	-0.0006***	0.0002	-0.0010***	0.0002	-0.0009***	0.0002
Annuities	0.1320***	0.0020	0.0985***	0.0022	0.1280***	0.0021
Retirement	0.0858***	0.0058	0.0737***	0.0059	0.0839***	0.0061
N	77,365		71,461		71,745	
LR Chi2	10,906.17		12,675.22		10,919.84	
Proba>chi2	0.0000		0.0000		0.0000	
Pseudo-R2	0.2028		0.2446		0.2102	
Log likelihood	-21,440.37		-19,576.17		-20,511.58	

Table 4 reports results of BLR that aim at determining stock market participation during the sample period covering 2007 to 2015. Three models (1, 2 and 3) are presented. The dependent variable "Stocks" indicates whether a retail client directly or indirectly held at least one stock (1) or not (0) between 2007 and 2015. All variables are described in Table 1. We report the average marginal effects (AMEs) of independent variables. For interpretating "Income" and "Credit" in monetary unit, we take the logarithm of monetary codes reported in Table 3. The category variables "Salaried", "Accepting" and "Waiting" are reference category variables since they concentrate the highest number of retail clients. Statistical significance levels are fixed at 1%, 5% and 10% that are denoted by ***, ** and * respectively.

Second, in Panel A, we find that MiFID indicators exhibit greater magnitude effects than those of variables commonly studied in the literature such as gender, age and income. Moreover, AMEs increase with the ordinal ranking of the categories of the two MiFID indicators. Retail clients who prefer bearing a capital risk ("SeekBetter") or a significant part of capital risk ("SeekHigh") are more likely to invest in stocks than those who prefer no taking risk on the invested capital ("Accepting"). This indicates that the probability of investing in stock market increases with retail clients' risk tolerance. This result is consistent with the findings of Hoffmann et al. (2015) as they show that retail investors with higher levels of and upward revisions in risk tolerance have a high tendency to trade. Looking at attitudes towards losses, we find that retail clients preferring to sell the entire portfolio ("SellingAll") or a part of the portfolio ("SellingPart") are less inclined to hold stocks during the downturn compared to those preferring to wait until their investment values increase ("Waiting"). Indeed, the probability of stockholding decreases by 8.17% (2.15% respectively) if retail clients prefer selling their entire (a part of their, respectively) portfolio. Retail clients preferring to accumulate further financial securities ("Investing") during the downturn have a high tendency to hold stocks than those preferring to wait. Therefore, we show that the higher the willingness to hold financial securities during the downturn is, the higher the probability of participating on stock market is. Specifically, the probability of participating in stock market increases by 6.33% for retail clients who are attracted by purchasing additional financial securities during the downturn. In other words, retail clients who are less sensitive towards losses during the downturn have a high tendency to own stocks. Therefore, introducing MiFID indicators provides an accurate explanation of stock investment decision.

In Panel B, all classical determinants of stock market participation exhibit consistent and significant AMEs in the three models. Male retail clients are more likely to participate in stock market than women like in van Rooij et al. (2011) and Almenberg and Dreber (2015) because women have a more conservative approach than men Jacobsen et al. (2014). Looking at AMEs, we find that being a man increases by 1.46% the percentage of owning stocks in Model 1. By including MiFID indicators (Models 2 and 3), we find that, on average, being a man increases by 1.07% the percentage of stockownership. Age also affects retail clients' behaviour on stock market since the probability of owning stocks is, on average, 0.37% higher among elderly retail clients than younger ones.

We then examine the new variables of our study: "Native", "Paris" and "Matrimonial". The geographical variables "Native" and "Paris" are significant at all reasonable significance levels whatever the model we focus on. Indeed, native-born French retail clients are more prone to invest in stock market than for immigrants. By analysing AMEs, we find that the probability of owning stocks is, on average, 4.32% greater among natives than immigrants. Our result is in line with that of Osili and Paulson (2004) who underline a low financial asset holding among immigrants. As for "Paris",

we show that retail clients living in the most developed region of the whole country are more likely to hold stocks than those living in province. Living in Parisian region increases, on average, by 3.68% the probability of stockholding. Indeed capital region displays the highest economic performance of the whole country thus concentrating big firms. Our result can be explained by geographic proximity. Grinblatt and Keloharju (2001) indicate that investors have a high tendency to hold stocks of firms that are located close to their place of living. Matrimonial regime also affects retail clients' stockholding. Indeed, retail clients opting for the separation of property regime are more inclined to participate in stock market. Looking at AMEs, the probability of participating in stock market is, on average, 2.67% higher among retail clients opting for the separation regime. Since there is no joint-ownership between spouses, each spouse is free to manage his/her own goods. Any gain and loss generated by stockholding does not impact the wealth of the other spouse. That financial independence promotes stock market participation.

By looking at retail client's professional occupations, we find that self-employed retail clients are more likely to own stocks than salaried ones. The probability of owning stocks is, on average, 0.91% higher among self-employed compared to salaried ones. Indeed, self-employed are financially independent since they perceive their income directly from their own professional activity whereas salaried perceive straight salary work from their employer. Therefore, salaried bear a financial constraint that might limit stock market participation. Being retired decreases, on average, by 2.07% the probability of owning stocks compared to salaried individuals. Being professionally active on labor market increases the tendency to participate in stock market as professional activity promotes social interactions. According to Hong et al. (2004), social households are more likely to invest in stock market than non-social households. They argue that individuals are more attracted by stock market participation when more of their peers do. As for "No occupation", we find that the probability of investing in stock market is, on average, 1.04% higher among those exercising no professional activity than salaried. For these non-occupied individuals, we can not make inferences on their social interactions and the only remaining explanation is the time free that they can allocate to stock investment.

In Panel C, we notice that the probability of investing in stock market increases with the net monthly income. Specifically, we find that the probability of investing in stocks increases, on average, by 0.12% given a 10% increase in the net monthly income. Our result is consistent with that of Barber and Odean (2001). Indeed, retail clients perceiving a higher income level can afford to own stocks since their disposable income is greater. Besides, they are able to recover losses incurred by that stockholding leading them to be more risk tolerant than those perceiving a lower income level. Illustrating the indebtedness situation, we find that the credit amount remaining to reimburse also impacts stock market participation. The probability of participating in stock market decreases, on

average, by 0.01% given a 10% increase in the credit amount remaining to reimburse. This negative relationship can be explained in the following manner. Credit amount remaining to be reimbursed limit financial investment opportunities since retail clients are credit-constrained. Therefore, stock market participation is lower among indebted retail clients. Concerning the different financial products that reflect movable patrimony, we find that holding equity-indexed annuities ("Annuities") and/or retirement accounts ("Retirement") influence positively retail clients' stock market participation. Specifically, the magnitude effects (AMEs) of "Annuities" and "Retirement" are the highest in Panel C whatever the model we focus on. By holding these products, the probability of participating in stock market is, on average, 12% and 8% greater respectively. Then they have a greater impact on stock market participation. Holding at least one of these financial products means that retail clients are willing to accept a risk of loss in capital since both products depend on financial markets' performance. Besides, it also refers to retail clients' portfolio composition. Indeed, by holding one of both financial products, retail clients hold more diversified portfolios leading them to be more financially sophisticated (Boolell-Gunesh et al., 2009).

In order to complement our empirical results, we perform robustness checks for assessing the consistency of our results (Section 5).

5 Robustness checks

In Section 5, we carry out two robustness checks (noted RC henceforth) in order to test the consistency of our findings. Results are reported in Table 5.

Robustness check 1 (RC1) introduces a new variable "Account tenure" in Panel C. This variable measures the length of time (in years) during which an individual has been a retail client of the European commercial bank. It is computed by using the date of arrival in the bank and the date of extraction of banking records which corresponds to the 07/31/2015. This variable may be considered like being a proxy of retail clients' experience in terms of investment[45]. In our sample, individuals are clients within the bank, on average, since 14 years. This period is much longer than those reported by Dorn and Sengmueller (2009) and Hoffmann et al. (2015), i.e. 3 and 4 years respectively. Thereby, our retail clients are considered like being more experienced. In RC1, we do not use "Age" as "Age" and "Account tenure" are highly correlated (Pearson correlation coefficient is 0.59).

Robustness check 2 (RC2) focuses on retail clients having answered thrice the MiFID questionnaire (N=11,839). This subsample of retail clients is then more familiar with the MiFID questionnaire content and with the stock market (about 30% of these retail clients invested in stocks between 2007 and 2015). Given that the proportion of stockholding is three times higher than that of the

[45]By using retail investors' answers, Glaser and Weber (2007) and Merkle (2015) introduce the variable "Experience" for referring the length of time during which retail investors have been directly investing in stock market.

whole retail clients (see Table 3), we check whether our empirical results are still valid on this specific subsample.

In RC1, we first notice that AMEs of "Account tenure" are similar to those of "Age" (see Table 4). Significant at all reasonable significance levels, both variables indicate that the probability of stockholding increases with experience. As for the other independent variables, we notice that results are quietly similar to our main results (Table 4), except for sign change of "Retired" which only represents a professional category among the three others. Account tenure has already been taken into account by preceding studies such as Dorn and Sengmueller (2009), Nicolosi et al. (2009) and Hoffmann et al. (2015). The probability of trading derivatives is shown to be higher among older and more experienced investors (Bauer et al., 2009 and Hoffmann et al., 2015). Their results are in line with that previously demonstrated by Corter and Chen (2005). Besides, the trade quality, i.e. average raw and excess buy-minus-sell returns, becomes higher with experience (Nicolosi et al., 2009). In our empirical analysis, we decide to include "Age" instead of "Account tenure" since it is commonly used and identified as a key driver of investment decision like gender and income in behavioural finance works.

In RC2, AMEs of quite all independent variables are greater than those corresponding to the whole sample (see Table 4). The findings are consistent with the results corresponding to the whole sample, except for "SellingPart", "No occupation" and "Gender" (Models 2 and 3) which are now insignificant. In all, we notice that results corresponding to this subsample are more accurate and greater than those corresponding to the whole sample leading us to reinforce our findings. Besides, administering several times the MiFID questionnaire is useful for twofold reasons. First, in the third time questionnaire, retail clients are more familiar and confident with the MiFID questionnaire. Moreover, this may explain the decrease of the number of unreported answers we observe between two successive questionnaires. Second, investment service providers are more likely to offer suited advices and financial products to their clients since they are able to collect further information on them. Therefore, this "pedagogical" aspect fits MiFID requirements.

Table 5 – RC results

	RC1						RC2					
	Model 1		Model 2		Model 3		Model 1		Model 2		Model 3	
	AMEs	std	AMEs	std	AMEs	std	AMEs	std	AMEs	std	AMEs	std
Panel A: MiFID indicators												
SeekBetter			0,0988***	0,0022					0.2033***	.0077		
SeekHigh			0,1817***	0,0052					0.3376***	.0173		
Accepting			(omitted)						(omitted)			
SellingAll					-0,0816***	0,0048					-0,1832***	0,0202
SellingPart					-0,0217***	0,0048					-0,0180	0,0172
Investing					0,0603***	0,0037					0,1090***	0,0126
Waiting					(omitted)						(omitted)	
Panel B: Socio-demographic indicators												
Gender	0,0107***	0,0021	0,0050**	0,0022	0,0088***	0,0022	0,0176**	0,0080	0,0051	0,0077	0,0127	0,0080
Age							0,0065***	0,0003	0,0058***	0,0003	0,0064***	0,0003
Native	0,0290***	0,0033	0,0241***	0,0034	0,0276***	0,0035	0,0844***	0,0123	0,0727***	0,0118	0,0773***	0,0122
Paris	0,0345***	0,0029	0,0334***	0,0030	0,0315***	0,0031	0,0669***	0,0119	0,0687***	0,0115	0,0632***	0,0118
Matrimonial	0,0374***	0,0029	0,0300***	0,0030	0,0363***	0,0031	0,0596***	0,0113	0,0435***	0,0108	0,0543***	0,0112
Self-employed	0,0278***	0,0031	0,0266***	0,0032	0,0288***	0,0032	0,0286**	0,0120	0,0194*	0,0115	0,0244**	0,0119
Retired	0,0211***	0,0027	0,0227***	0,0028	0,0232***	0,0029	-0,0315***	0,0121	-0,0283**	0,0117	-0,0345***	0,0120
No occupation	0,0107***	0,0038	0,0070*	0,0040	0,0109***	0,0040	0,0165	0,0154	0,0041	0,0149	0,0145	0,0153
Salaried	(omitted)		(omitted)		(omitted)		(omitted)		(omitted)		(omitted)	
Panel C: Wealth and patrimony indicators												
ln(Income)	0,0198***	0,0010	0,0133***	0,0010	0,0183***	0,0011	0,0360***	0,0042	0,0231***	0,0037	0,0320***	0,0040
ln(Credit)	-0,0006***	0,0002	-0,0009***	0,0002	-0,0008***	0,0002	-0,0014*	0,0008	-0,0018**	0,0008	-0,0017**	0,0008
Account tenure	0,0039***	0,0001	0,0039***	0,0001	0,0040***	0,0001						
Annuities	0,1186***	0,0020	0,0861***	0,0022	0,1150***	0,0021	0,2039***	0,0069	0,1360***	0,0073	0,1901***	0,0070
Retirement	0,0777***	0,0057	0,0655***	0,0059	0,0757***	0,0060	0,1356***	0,0174	0,1089***	0,0167	0,1260***	0,0173
N	77,365		71,461		71,745		11,839		11,818		11,824	
LR Chi2	11,762.76		13,481.78		11,718.33		2,198.01		2,960.72		2,377.72	
Prob>Chi2	0.0000		0.0000		0.0000		0.0000		0.0000		0.0000	
Pseudi-R2	0.2187		0.2601		0.2256		0.1526		0.2058		0.1652	
Log likelihood	-21,012.08		-19,172.89		-20,112.34		-6,103.38		-5,712.89		-6,007.37	

Table 5 reports robustness checks results. Robustness check 1 (RC1) takes into account "Account tenure" instead of "Age" and focuses on the whole sample of retail clients. Robustness check 2 (RC2) refers to a subsample of retail clients having answered thrice the MiFID questionnaire and takes into account the same independent variables included in our empirical analysis. Statistical significance levels are fixed at 1% and 10% that are denoted by *** and * respectively.

6 Conclusion

The Markets in Financial Instrument Directive (MiFID) requires investment service providers to get a deep knowledge of their clients by addressing a questionnaire, called *MiFID questionnaire*, in order to offer advices and financial products perfectly suited to their situation. We simultaneously analyze two datasets, MiFID questionnaires answers and banking records, corresponding to more than 70,000 retail clients of a big European commercial bank.

In the MiFID questionnaire, we particularly focus on two indicators which are self-declared attitudes towards risk and attitudes towards losses for understanding their impact on stock investment decision while controlling variables commonly studied in the literature such as gender, age and income. Specifically, we find that the probability of investing in stocks increases with risk tolerance and the willingness to hold further financial securities during the downturn. Besides, MiFID indicators display greater magnitude effects than those of classical variables. Consistent with prior studies on the usual determinants of stock market participation, our results are complemented by the use of specific variables such as geographical variables, matrimonial regime and holding of other risky financial products such as equity-indexed annuities and retirement accounts.

Looking at socio-demographic indicators, we find that the probability of stockholding is higher among males, elderly, natives and those living in capital region. Besides, this probability is greater among retail clients opting for the separation of property regime. Unlike retired retail clients, self-employed and those exercising no professional activity are more prone to invest in stock market than salaried ones. Focusing on wealth and patrimony indicators, we find that the probability of stockholding increases with income and decreases as the credit amount remaining to reimburse increases. Furthermore, holding other risky financial products, such as equity-indexed annuities and/or retirement accounts, increases the probability of investing in stock market.

Our results contribute to the actual debate between professionals, regulators and academics about the usefulness of MiFID indicators. de Palma and Picard (2010) have realized a first diagnosis of 14 MiFID questionnaires provided by 10 financial intermediaries in France. They suggested that a quantitative measurement of risk-taking preferences is necessary for ensuring that investment service providers offer suited advices and financial products to their clients. This claim is now reached as preferences assessed in the MiFID questionnaire explain stock market participation.

Since there is no regulatory constraint on MiFID questionnaire content, investment service providers are free to establish their questionnaire. Therefore, questionnaire length may be different from a bank to another one. When they exist, questions dealing with individual characteristics (e.g. gender, age and marital status) are, in general, placed at the beginning of the questionnaire. However, these pieces of information are already recorded in the bank database like the other financial information (e.g. wealth including net monthly income and holding of financial assets). Even if

each question brings additional information about retail clients, fatigue effect or mistrust should be taken into account since the longer the questionnaire, the less retail clients pay attention to the last questions of the MiFID questionnaire. Therefore, the probability of getting correct answers may decrease with the questionnaire length due to a misinterpretation of questions or proposals. Besides, questionnaire length may also influence the number of unreported answers which may restrict the ability of investment service providers to offer suited advices and financial products to their clients. Therefore, we recommend that MiFID questionnaire mainly focus on retail clients' own preferences which have been demonstrated like being key drivers of stock investment decision in this study.

References

Ackert, L., Church, B. K. and Englis, B. (2002). The asset allocation decision and investor heterogeneity: a puzzle? *Journal of Economic Behavior & Organization* 47: 423–433.

Agnew, J., Balduzzi, P. and Sunden, A. (2003). Portfolio choice and trading in a large 401 (k) plan. *The American Economic Review* 93: 193–215.

Almenberg, J. and Dreber, A. (2015). Gender, stock market participation and financial literacy. *Economics Letters* 137: 140 – 142.

Antonides, G. and Van Der Sar, N. L. (1990). Individual expectations, risk perception and preferences in relation to investment decision making. *Journal of Economic Psychology* 11: 227–245.

Arrondel, L., Debbich, M. and Savignac, F. (2015). Stockholding in France: the role of financial literacy and information. *Applied Economics Letters* 22: 1315–1319.

Attanasio, O., Banks, J. and Tanner, S. (2002). Asset Holding and Consumption Volatility. *Journal of Political Economy* 110: 771–792.

Bajtelsmit, V. L., Bernasek, A. and Jianakoplos, N. A. (1999). Gender differences in defined contribution pension decisions. *Financial Services Review* 8: 1–10.

Bajtelsmit, V. L. and VanDerhei, J. L. (1997). Risk aversion and pension investment choices. *Positioning pensions for the twenty-first century* 45: 66.

Baker, M. and Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance* 61: 1645–1680.

Baker, M. and Wurgler, J. (2007). Investor sentiment in the stock market. *The Journal of Economic Perspectives* 21: 129–151.

Bakshi, G. S. and Chen, Z. (1994). Baby boom, population aging, and capital markets. *Journal of Business* 67: 163–202.

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

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

Barber, B. M. and Odean, T. (2002). Online investors: do the slow die first? *Review of Financial Studies* 15: 455–488.

Bauer, R., Cosemans, M. and Eichholtz, P. (2009). Option trading and individual investor performance. *Journal of Banking & Finance* 33: 731–746.

Bellofatto, A., De Winne, R. and D'Hondt, C. (2014). Beyond the Disposition Effect: Evidence from the 1999-2012 period. In *The 2014 EFM Merton H Miller Doctoral Seminar of the European Financial Management Association (EFMA)*.

Belsley, D. A., Kuh, E. and Welsch, R. E. (1980). Regression diagnostics: Identifying influential observations and sources of collinearity.

Bernasek, A. and Shwiff, S. (2001). Gender, Risk, and Retirement. *Journal of Economic Issues* 35: 345–356.

Bernheim, B. D., Skinner, J. and Weinberg, S. (2001). What Accounts for the Variation in Retirement Wealth among U.S. Households? *American Economic Review* 91: 832–857.

Bertocchi, G., Brunetti, M. and Torricelli, C. (2011). Marriage and other risky assets: A portfolio approach. *Journal of Banking & Finance* 35: 2902–2915.

Bodie, Z. and Crane, D. B. (1997). Personal Investing: Advice, Theory, and Evidence from a Survey of TIAA-CREF Participants. SSRN Scholarly Paper ID 36158, Social Science Research Network, Rochester, NY.

Booij, A. S. and Kuilen, G. Van de (2009). A parameter-free analysis of the utility of money for the general population under prospect theory. *Journal of Economic psychology* 30: 651–666.

Boolell-Gunesh, S., Broihanne, M.-H. and Merli, M. (2009). Disposition effect, investor sophistication and taxes: Some French Specificities. *Finance* 30: 51–78.

Boolell-Gunesh, S., Broihanne, M.-H. and Merli, M. (2012). Sophistication of Individual Investors and Disposition Effect Dynamics. *Finance* 33: 9–37.

Booth, A. and Nolen, P. (2012). Salience, risky choices and gender. *Economics Letters* 117: 517–520.

Bricker, J., Dettling, L. J., Henriques, A., Hsu, J. W., Moore, K. B., Sabelhaus, J., Thompson, J. and Windle, R. A. (2014). Changes in us family finances from 2010 to 2013: Evidence from the survey of consumer finances. *Federal Reserve Bulletin* 100: 1–41.

Brooks, P. and Zank, H. (2005). Loss Averse Behavior. *Journal of Risk and Uncertainty* 31: 301–325.

Brown, G. W. and Cliff, M. T. (2005). Investor sentiment and asset valuation. *The Journal of Business* 78: 405–440.

Cahuzac, E. and Bontemps, C. (2008). *Stata par la pratique : statistiques, graphiques et elements de programmation*. No. splp in Stata Press books. StataCorp LP.

Cardak, B. A. and Wilkins, R. (2009). The determinants of household risky asset holdings: Australian evidence on background risk and other factors. *Journal of Banking & Finance* 33: 850 – 860.

Charness, G. and Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization* 83: 50 – 58, gender Differences in Risk Aversion and Competition.

Chatterjee, S. (2009). Individual Stockownership in the United States: Native-Immigrant Gap and the Role of Risk Tolerance. *International Research Journal of Finance and Economics* 28: 160–168.

Chatterjee, S., Hadi, A. S. and Price, B. (2000). *Regression Analysis by Example*. New York : Chapman & Hall.

Chatterjee, S. and Zahirovic-Herbert, V. (2011). Homeownership and Housing Equity: An Examination of Native-Immigrant Differences in Housing Wealth. *International Advances in Economic Research* 17: 211–223.

Chatterjee, S. and Zahirovic-Herbert, V. (2014). A road to assimilation: immigrants and financial markets. *Journal of Economics and Finance* 38: 345–358.

Chen, G., Kim, K. A., Nofsinger, J. R. and Rui, O. M. (2007). Trading performance, disposition effect, overconfidence, representativeness bias, and experience of emerging market investors. *Journal of Behavioral Decision Making* 20: 425–451.

Chen, G.-M., Kim, K. A., Nofsinger, J. R. and Rui, O. M. (2004). Behavior and performance of emerging market investors: Evidence from China. *unpublished Washington State University Working paper (January)* .

Cho, I. (2014). Homeownership and investment in risky assets in europe. *Review of European Studies* 6.

Cohn, R. A., Lewellen, W. G., Lease, R. C. and Schlarbaum, G. G. (1975). Individual investor risk aversion and investment portfolio composition. *The Journal of Finance* 30: 605–620.

Constantinides, G., Donaldson, J. B. and Mehra, R. (2002). Junior can't borrow: A new perspective on the equity premium puzzle. *The Quarterly Journal of Economics* 117: 269–296.

Corter, J. E. and Chen, Y.-J. (2005). Do investment risk tolerance attitudes predict portfolio risk? *Journal of Business and Psychology* 20: 369.

Cummings, R. G., Brookshire, D. S., Bishop, R. C. and Arrow, K. J. (1986). Valuing environmental goods: an assessment of the contingent valuation method.

De Bourmont, M. (2012). La résolution d'un problème de multicolinéarité au sein des études portant sur les déterminants d'une publication volontaire d'informations : proposition d'un algorithme de décision simplifié basé sur les indicateurs de Belsley, Kuh et Welsch (1980). In *Comptabilités et innovation*. Grenoble, France, cd–rom.

Dhar, R. and Zhu, N. (2006). Up Close and Personal: Investor Sophistication and the Disposition Effect. *Management Science* 52: 726–740.

Dimmock, S. G. and Kouwenberg, R. (2010). Loss-aversion and household portfolio choice. *Journal of Empirical Finance* 17: 441–459.

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

Dorn, D. and Sengmueller, P. (2009). Trading as entertainment? *Management Science* 55: 591–603.

Dwyer, P. D., Gilkeson, J. H. and List, J. (2002). Gender differences in revealed risk taking: evidence from mutual fund investors. *Economics Letters* 76: 151–158.

Erkel-Rousse, H. (1995). DÃ©tection de la multicolinÃ©aritÃ© dans un modÃšle linÃ©aire ordinaire: quelques Ã©lÃ©ments pour un usage averti des indicateurs de belsley, kuh et welsch. *Revue de Statistique AppliquÃ©e* 43: 19–42.

Faff, R., Mulino, D. and Chai, D. (2008). On the linkage between financial risk tolerance and risk aversion. *Journal of Financial Research* 31: 1–23.

Fan, J. X. and Xiao, J. J. (2006). Cross-cultural differences in risk tolerance: A comparison between Chinese and Americans. *Journal of Personal Finance* 5: 54–75.

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

Feng, L. and Seasholes, M. S. (2008). Individual investors and gender similarities in an emerging stock market. *Pacific-Basin Finance Journal* 16: 44–60.

Foucault, T., Sraer, D. and Thesmar, D. (2011). Individual investors and volatility. *Journal of Finance* 66: 1369–1406.

Fratantoni, M. C. (1998). Homeownership and investment in risky assets. *Journal of Urban Economics* 44: 27–42.

Fuertes, A.-M., Muradoglu, G. and Ozturkkal, B. (2014). A behavioral analysis of investor diversification. *The European Journal of Finance* 20: 499–523.

Glaser, M. and Weber, M. (2007). Overconfidence and trading volume. *The Geneva Risk and Insurance Review* 32: 1–36.

Goetzmann, W. and Kumar, A. (2008). Equity Portfolio Diversification. *Review of Finance* 12: 433–463.

Grable, J. E. (2000). Financial risk tolerance and additional factors that affect risk taking in everyday money matters. *Journal of Business and Psychology* 14: 625–630.

Grable, J. E. and Joo, S.-h. (1999). Factors related to risk tolerance: A further examination. *Consumer Interests Annual* 45: 53–58.

Grable, J. E. and Joo, S.-H. (2004). Environmental and biophysical factors associated with financial risk tolerance. *Journal of Financial Counseling and Planning* 15.

Grable, J. E. and Lytton, R. H. (1998). Investor risk tolerance: Testing the efficacy of demographics as differentiating and classifying factors. *Financial Counseling and Planning* 9: 61–74.

Grinblatt, M. and Keloharju, M. (2001). What Makes Investors Trade? *The Journal of Finance* 56: 589–616.

Grinblatt, M. and Keloharju, M. (2009). Sensation Seeking, Overconfidence, and Trading Activity. *Journal of Finance* 64: 549–578.

Grinblatt, M., Keloharju, M. and Linnainmaa, J. (2011). Iq and stock market participation. *The Journal of Finance* 66: 2121–2164.

Guiso, L., Jappelli, T. and Terlizzese, D. (1996). Income Risk, Borrowing Constraints, and Portfolio Choice. *American Economic Review* 86: 158–172.

Guiso, L., Sapienza, P. and Zingales, L. (2008). Trusting the stock market. *The Journal of Finance* 63: 2557–2600.

Guiso, L. and Sodini, P. (2013). *Household Finance: An Emerging Field*. Elsevier, Handbook of the Economics of Finance *2*, chap. 0. 1397–1532.

Haliassos, M. and Bertaut, C. C. (1995). Why do so few hold stocks? *Economic Journal* 105: 1110–29.

Hallahan, T., Faff, R. and McKenzie, M. (2003). An exploratory investigation of the relation between risk tolerance scores and demographic characteristics. *Journal of Multinational Financial Management* 13: 483–502.

Hallahan, T. A., Faff, R. W. and McKenzie, M. D. (2004). An empirical investigation of personal financial risk tolerance. *Financial Services Review* : 57–78.

Hanna, S. D., Gutter, M. S. and Fan, J. X. (2001). A measure of risk tolerance based on economic theory. *Journal of Financial Counseling and Planning* 12: 53.

Hariharan, G., Chapman, K. S. and Domian, D. L. (2000). Risk tolerance and asset allocation for investors nearing retirement. *Financial Services Review* 9: 159 – 170.

Hinz, R. P., McCarthy, D. D. and Turner, J. A. (1997). Are Women Conservative Investors? Gender Differences in Participant Directed Pension Investments. Pension Research Council Working Papers 96-17, Wharton School Pension Research Council, University of Pennsylvania.

Hirshleifer, D. (2001). Investor psychology and asset pricing. *The Journal of Finance* 56: 1533–1597.

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

Hoffmann, A. O., Post, T. and Pennings, J. M. (2015). How investor perceptions drive actual trading and risk-taking behavior. *Journal of Behavioral Finance* 16: 94–103.

Holt, C. and Laury, S. (2002). Risk aversion and incentive effects. *American Economic Review* 92: 1644–1655.

Hong, H., Kubik, J. D. and Stein, J. C. (2004). Social Interaction and Stock-Market Participation. *The Journal of Finance* 59: 137–163.

Jacobsen, B., Lee, J. B., Marquering, W. and Zhang, C. Y. (2014). Gender differences in optimism and asset allocation. *Journal of Economic Behavior & Organization* 107: 630–651.

Jianakoplos, N. and Bernasek, A. (1998). Are Women More Risk Averse? *Economic Inquiry* 36: 620–30.

Korniotis, G. M. and Kumar, A. (2011). Do Older Investors Make Better Investment Decisions? *The Review of Economics and Statistics* 93: 244–265.

Kremelberg, D. (2011). *Practical Statistics: A Quick and Easy Guide to IBMÂ® SPSSÂ® Statistics, STATA, and Other Statistical Software*. SAGE Publications.

Kumar, A. (2007). Do the diversification choices of individual investors influence stock returns? *Journal of Financial Markets* 10: 362–390.

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

Lease, R. C., Lewellen, W. G. and Schlarbaum, G. G. (1974). The Individual Investor: Attributes and Attitudes. *The Journal of Finance* 29: 413–433.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics* 47: 13–37.

Long, J. S. and Freese, J. (2006). *Regression Models for Categorical Dependent Variables using Stata, 2nd Edition*. StataCorp LP, 2nd ed.

Longhi, S. and Nandi, A. (2015). *A Practical Guide to Using Panel Data*. SAGE Publications.

Love, D. and Schmidt, L. (2015). Comprehensive Wealth of Immigrants and Natives. Working Papers wp328, University of Michigan, Michigan Retirement Research Center.

Maccrimmon, K. and Wehrung, D. A. (1986). *The management of uncertainty: Taking risks.*. New york ed.

Magron, C. (2012). Performance of individual investors and personal investment objectives. Working Papers of LaRGE Research Center 2012-07, Laboratoire de Recherche en Gestion et Economie (LaRGE), UniversitÃ© de Strasbourg.

Magron, C. (2014). The Trading Performance of Individual Investors. *Bankers, Markets & Investors* : 43–52.

Mankiw, N. G. and Zeldes, S. P. (1991). The consumption of stockholders and nonstockholders. *Journal of Financial Economics* 29: 97–112.

Markowitz, H. (1959). *Portfolio Selection: Efficient Diversification of Investments*. New York : Wiley.

McInish, T., Ramaswami, S. N. and Srivastava, R. K. (1993). Do More Risk-Averse Investors Have Lower Net Worth and Income? *The Financial Review* 28: 91–106.

Merkle, C. (2015). Financial loss aversion illusion. *Browser Download This Paper* .

Mitchell, R. C. and Carson, R. T. (1989). Using surveys to value public goods: the contingent valuation method.

Mitton, T. and Vorkink, K. (2007). Equilibrium underdiversification and the preference for skewness. *Review of Financial studies* 20: 1255–1288.

Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica* 34: 768–783.

Neumann, J. von and Morgenstern, O. (1947). *Theory of games and economic behavior*. Princeton University Press.

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

Odean, T. (1999). Do Investors Trade Too Much? *American Economic Review* 89: 1279–1298.

Osili, U. O. and Paulson, A. (2004). Immigrant-native differences in financial market participation. *Federal Reserve Bank of Chicago, WP* 18.

Palma, A. de and Picard, N. (2010). Evaluation of MiFID qquestionnaire in France. Tech. rep., AMF, study for the Autorite des Marches Financiers.

Poterba, J. M. and Samwick, A. A. (1995). Stock Ownership Patterns, Stock Market Fluctuations, and Consumption. *Brookings Papers on Economic Activity* 26: 295–372.

Powell, M. and Ansic, D. (1997). Gender differences in risk behaviour in financial decision-making: An experimental analysis. *Journal of Economic Psychology* 18: 605–628.

Rau, H. A. (2014). The disposition effect and loss aversion: Do gender differences matter? *Economics Letters* 123: 33 – 36.

Richards, D., Rutterford, J. and Fenton-O Creevy, M. (2011). Do stop losses work? the disposition effect, stop losses and investor demographics. *The Open University Business School Working Paper* .

Riley, W. B. and Chow, K. V. (1992). Asset allocation and individual risk aversion. *Financial Analysts Journal* 48: 32–37.

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

Roszkowski, M. J., Snelbecker, G. E. and Leimberg, S. (1993). Risk tolerance and risk aversion. *The tools and techniques of financial planning* 4: 213–225.

Schmidt, U. and Traub, S. (2002). An Experimental Test of Loss Aversion. *Journal of Risk and Uncertainty* 25: 233–49.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance* 19: 425–442.

Shum, P. and Faig, M. (2006). What explains household stock holdings? *Journal of Banking & Finance* 30: 2579 – 2597.

Sung, J. and Hanna, S. (1996). Factors related to risk tolerance. *Journal of Financial Counseling and Planning* 7: 11.

Tekce, B. and Yilmaz, N. (2015). Are individual stock investors overconfident? Evidence from an emerging market. *Journal of Behavioral and Experimental Finance* 5: 35–45.

Thomas, B. and Reza, S. (2010). Outstanding debt and the household portfolio. *Review of Financial Studies* 23: 2900–2934.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5: 297–323.

Wachter, J. A. and Yogo, M. (2010). Why do household portfolio shares rise in wealth? *Review of Financial Studies* 23: 3929–3965.

Wang, H. and Hanna, S. D. (1997). Does risk tolerance decrease with age? *Financial Counseling and Planning* 8: 27–31.

Weber, M., Weber, E. U. and Nosic, A. (2013). Who takes risks when and why: Determinants of changes in investor risk taking*. *Review of Finance* 17: 847–883.

Weber, M. and Welfens, F. (2007). An Individual Level Analysis of the Disposition Effect: Empirical and Experimental Evidence. Sonderforschungsbereich 504 Publications 07-45, Sonderforschungsbereich 504, Universitat Mannheim.

Appendix

Appendix A: Measures of association between variables

We focus on the study of relationships existing between variables as defined in Table 1. This preliminary step is necessary in order to build our estimation models carefully. We use 4 appropriate measures to analyze associations between variables: Phi coefficient (Φ), Cramer's V (V)[46], point biserial correlation coefficient (r_{pb}) and Spearman's rank correlation coefficient (r_{sp}).

We use Phi coefficient for studying the association between two binary variables, Cramer's V for studying two nominal variables (at least one of which has more than two modalities), point biserial correlation coefficient for binary/continuous pairwise variables and Spearman's rank correlation coefficient for binary/ordinal and continuous/ordinal pairwise variables. Contrary to the other association measures, Spearman's rank correlation is a non-parametric measure that does not assume normal distribution and linear relationship between variables. Spearman's rank correlation is used for variables with ordinal measurement levels.

These association measures are presented separately in order to provide a specific explanation for each measure. In all, we analyze 71,188 retail clients for which all independent variables are available. We particularly look at binary associations having significant coefficients exceeding 0.20 in absolute value and we carefully study the potential for multicollinearity issues between independent variables if coefficients exceed 0.50 (De Bourmont, 2012) in Section 4.

- *Phi coefficient*

Following Kremelberg (2011), we report the adjusted Phi value instead of Phi coefficient value due to unequal marginal distributions of binary variables[47]. Table 6 displays all binary pairwise variables.

[46] Unlike the three binary association measures, Cramer's V is comprised between 0 and 1.

[47] Phi coefficient value is equivalent to Pearson correlation coefficient for binary variables. It represents the square root of Chi-squared statistic divided by the sample size. However, it is assumed that each category approximately contains 50% of individuals. By violating this condition, Phi coefficient maximum value may be lower than 1 leading thus to misinterpretation of the strength of the binary association. For getting Phi adjusted value, we divided Phi value by its maximum possible value. This maximum value is computed by using the following formula : $\frac{\sqrt{p_j - p_j p_i}}{\sqrt{p_i - p_i p_j}}$ where $p_j$is the row or column containing the lowest proportion and p_i is the row or column containing the second lowest proportion. Thereby, this measure takes into account marginal distribution of each binary variable. Like Pearson correlation coefficient, the upper bound of this adjusted value is fixed to 1.

Table 6 – Phi coefficients

	Gender	Native	Paris	Matrimonial	Stocks	Annuities	Retirement
Gender	1						
Native	0.05***	1					
Paris	0.03***	-0.21***	1				
Matrimonial	0.06***	0.01***	0.05***	1			
Stocks	0.05***	0.06***	0.05***	0.09***	1		
Annuities	0.02***	0.04***	0.03***	0.07***	0.41***	1	
Retirement	0.08**	0.03***	0.06***	0.06***	0.33***	0.50***	1

Table 6 presents Phi coefficients corresponding to each binary variables association. Chi-squared test is used for assessing coefficients significance. Statistical significance levels are fixed at 1%, 5% and 10% that are denoted by ***, ** and * respectively.

Table 6 shows that the two geographical variables are negatively associated (Φ=-0.21). We also notice that there is a strong association between risky financial products. We can hypothesize that the probability of stockholding increases with the probability of holding equity-indexed annuities and/or retirement plans. Indeed, our main variable "Stocks" is positively associated to "Annuities" (Φ=0.41) and to "Retirement" (Φ=0.33). In the same manner, "Annuities" and "Retirement" display a high Phi coefficient value (Φ=0.50). These relationships are consistent since all these financial products are risky.

- *Cramer's V*

Cramer's V allows to assess relationships between professional occupations ("Occupations") and binary variables. Results are displayed in Table 7.

Table 7 – Cramer's V coefficients

	Occupations
Gender	0.18***
Native	0.03***
Paris	0.04***
Matrimonial	0.17***
Stocks	0.17***
Annuities	0.14***
Retirement	0.05***

Table 7 presents Cramer's V coefficients corresponding to each binary/nominal qualitative pairwise variables. Chi-squared test is used for assessing coefficients significance. Statistical significance levels are fixed at 1%, 5% and 10% that are denoted by ***, ** and * respectively.

In Table 7, we notice that there is no strong association between professional occupations and binary variables. Indeed, the highest values are 0.18 and 0.17 indicating that professional occupations are weakly associated to "Gender", "Matrimonial" and "Stocks".

- *Point biserial correlation coefficient*

This coefficient is computed for assessing relationships between age and binary variables. Results are reported in Table 8.

Table 8 – Point biserial correlation coefficient

	Age
Gender	-0.05***
Native	-0.03***
Paris	0.00
Matrimonial	0.10***
Stocks	0.26***
Annuities	0.22***
Retirement	0.03***

Table 8 displays point biserial correlation coefficients corresponding to each binary/continuous pairwise variables. t-test is used for assessing coefficients significance. Statistical significance levels are fixed at 1%, 5% and 10% that are denoted by ***, ** and * respectively.

Table 8 indicates that age is positively correlated to "Stocks" (r_{pb}=0.26) and "Annuities" (r_{pb} =0.22) leading us to hypothesize that stockholding increases with age.

- *Spearman rank correlation coefficient*

By applying this association measure, we look at the relationship between ordinal qualitative variables (i.e. MiFID indicators, "Income" and "Credit") and banking records, including our main variable "Stocks". We note that "Credit" and "Income" are ordinal qualitative variables. For these two variables, we use their numerical modalities (indicated in parentheses and superscript in Table 3) instead of their monetary codes (see "CODES" in Table 3).

Table 9 – Spearman rank correlation coefficients

	Risk tolerance	**Losses**	**Income**	**Credit**
Risk tolerance	1			
Losses	0.25***	1		
Gender	0.07***	0.05***	0.16***	
Age	0.16***	0.06***	0.27***	
Native	0.06***	0.03***	0.06***	
Paris	0.02***	0.06***	0.08***	
Matrimonial	0.10***	0.06***	0.22***	
Income	0.24***	0.14***	1	
Credit	0.10***	0.08***	0.41***	1
Stocks	0.30***	0.13***	0.18***	-0.03***
Annuities	0.32***	0.12***	0.12***	-0.02***
Retirement	0.09***	0.05***	0.07***	0.04***

Table 9 displays Spearman rank correlation coefficients corresponding to each binary/ordinal and ordinal/ordinal pairwise variables. T-test is used for assessing coefficients significance. Statistical significance levels are fixed at 1%, 5% and 10% that are denoted by ***, ** and * respectively.

Table 9 shows that "Income" and "Credit" display a strong and positive association (r_{sp}=0.41) since banks only lend to clients whose income allows them to repay their debt. "Risk tolerance" is positively associated to "Stocks" (r_{sp}=0.30) and to "Annuities" (r_{sp}=0.32), both being two risky financial products. "Age" and "Income" are also positively associated reflecting wealth accumulation across time (r_{sp}=0.27). Furthermore, MiFID indicators are positively correlated with each other

(r_{sp}=0.25). "Risk tolerance" and "Income" also display a positive and non-surprising association (r_{sp}=0.24). For "Losses", no high significant association is found with banking records variables.

Appendix B: Variance Inflation Factor

Table 10 – Variance Inflation Factors (VIF)

	VIF		
	Model 1	**Model 2**	**Model 3**
	N=77,365	**N=71,461**	**N=71,745**
Panel A : MiFID indicators			
SeekBetter		1.20	
SeekHigh		1.03	
Accepting		(omitted)	
SellingAll			1.05
SellingPart			1.02
Investing			1.03
Waiting			(omitted)
Panel B : Socio-demographic indicators			
Gender	1.03	1.04	1.04
Age	2.94	2.88	2.88
Native	1.06	1.06	1.06
Paris	1.04	1.05	1.05
Matrimonial	1.07	1.07	1.07
Self-employed	1.15	1.14	1.14
Retired	1.86	1.87	1.87
No occupation	2.30	2.22	2.23
Salaried	(omitted)	(omitted)	(omitted)
Panel C : Wealth and patrimony indicators			
ln(Income)	2.60	2.54	2.53
ln(Credit)	1.37	1.36	1.36
Annuities	1.08	1.18	1.09
Retirement	1.03	1.03	1.03
Mean VIF	**1.54**	**1.48**	**1.43**

Table 10 reports VIF corresponding to independent variables. We do not face multicollinearity problem since all VIF are below the critical threshold of 10 (Chatterjee et al., 2000).

Appendix C: Goodness-of-fit measures

We apply different measures for ensuring the quality of goodness-of-fit of our BLR and their degrees of prediction. We begin by interpreting statistical measures reported in Table 4 before conducting a deepened analysis.

In Table 4, we report the likelihood ratio chi-square (noted LR Chi2). This numerical measure of fit allows comparing the goodness-of-fit of each model to the intercept-only model, also called empty model. Then, we find that our models predict better than the intercept-only model. We also report the p-value corresponding to each model. Whatever the model we focus on, p-value is always equal to 0.00 leading us to conclude that our models have a whole fit better than an empty model at all reasonable significance levels. Besides, pseudo-R^2 increases by adding MiFID variables into Model 1 (0.2028) leading thus to an increase of the quality of goodness-of-fit in models 2 (0.2446) and 3 (0.2102). Indeed, the closer the pseudo-R^2 is to 1, the stronger is the predictive power of a model. Furthermore, by adding MiFID variables into Model 1, the log likelihood increases and then converges to 0 in Models 2 and 3 meaning that we converge to a good model (Cahuzac and Bontemps, 2008).

Table 11 reports two other statistical measures that allows deepening our analysis.

Table 11 – Goodness-of-fit measures

Goodness-of-fit Measures	**Model 1 N=77,365**	**Model 2 N=71,461**	**Model 3 N=71,745**
Correct classification rate	89.23%	88.92%	88.66%
AUC	0.82	0.84	0.82

Table 11 presents results obtained from goodness-of-fit measures corresponding to each model performed in Table 4. We report correct classification rate and Area Under the Curve (AUC).

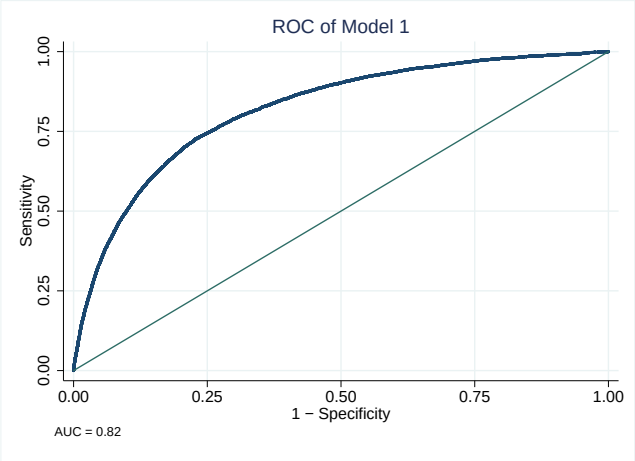
By comparing concordances and discrepancies between estimated values and observed values, we determine the explanatory power of all models, i.e. the correct classification rate. Specifically, we analyze our models' sensitivity and specificity. The sensitivity (specificity respectively) refers to the proportion of individuals who are positively (negatively respectively) and correctly classified. In our case, the sensitivity refers to the proportion of retail clients declared positive (i.e. stock holding) by the model and who are in reality. The specificity refers to the probability of retail clients declared negative (non-stock holding) by the model and who are in reality (i.e. they do not hold stock). Thereby, for detecting the correct classification rate, we have to sum the number of retail clients (positively and negatively) correctly classified and divide it by the sample size. In our models, the

correct classification rate is about 89% meaning that estimated values and observed values tie in 89%.

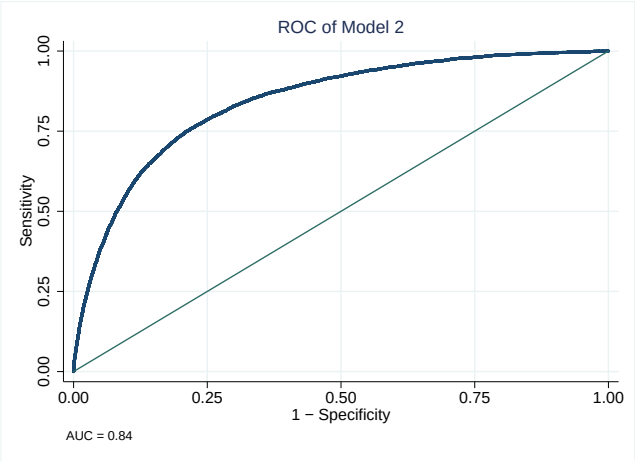
A graphical representation can also be realized for assessing the quality of a regression model. Indeed, we can use ROC (Receiver Operating Characteristics) curve for viewing the performance of a model. In other words, ROC represents graphically the discriminatory quality of a model. In our study, we aim at distinguishing retail clients holding stock from those who do not hold stock. For performing a ROC curve, we need to compute the sensitivity and the specificity. Graphically, "1 - specificity" is reported on the x-axis and "sensitivity" is reported on the y-axis. ROC curve provides a synthetic index, called Area Under the Curve (AUC). In our case, if AUC is equal to 1, then we can show that our model discriminates retail clients holding stock(s) from those who do not hold stock(s) in 100% of cases. Therefore, there is a strong discrimination. However, AUC equal to 0.5 means that the probability of discriminating both retail clients is 50%. Therefore, the model is not informative since it is equivalent to a random selection. Graphically, an AUC equal to 0.50 represents a bisector. As its name suggests, we interest in the area comprised between ROC curve and the bisector. Then, the further away we locate from the bisector, the greater is the discriminatory quality of a regression. Therefore, we should obtain an AUC close to 1 for ensuring that the model discriminates stock holding from non-stock holding. In Table 11, we notice that the smallest AUC value is equal to 0.82 and the highest one is equal to 0.84. Graphical representation is reported in figure 1. According to Long and Freese (2006), an AUC comprised between 0.80 and 0.90 means that there is a good discrimination. Therefore, we can conclude that all models reported in this study allow correctly discriminating retail clients holding stock(s) from those who do not hold any stock during the sample period.

Figure 1 – ROC and AUC

(a) Model 1



(b) Model 2



(c) Model 3

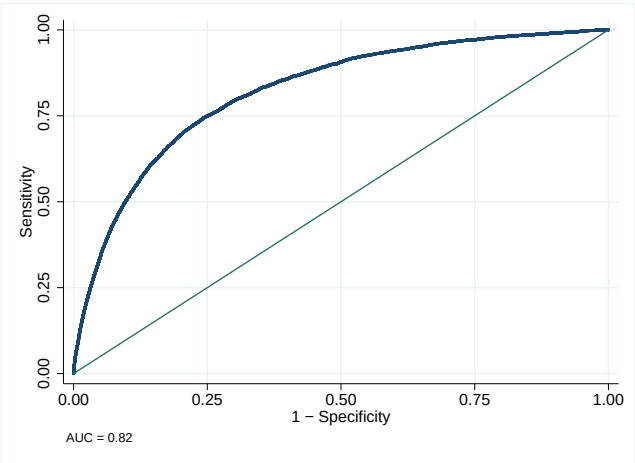


Figure 1 reports ROC and AUC corresponding to our models.