Individual Investors' Behavior: Are Small Investors Different?

First Draft: 16 April 2009 – This Draft: 15 January 2010

Enrico Maria Cervellati[1], Pino Fattori[2], Pierpaolo Pattitoni[3]

Abstract

Individual investors' stock trading has increased over time, thus attracting the attention of academics. The greater availability of information and the development of e-trading platforms are among the main determinants of this phenomenon. The behavioral finance literature has pointed out the dangers related to individual trading, mainly regarding overconfidence. This leads to excessive trading and consequent under-performance due to the effect of commissions on net returns. Since it has be shown that individual characteristics are important in determining the degree of overconfidence of a person, in this paper, we investigate the relationship between age, gender, job type, and income with the number of stock trades made by the clients of a small Italian cooperative bank. The aim of the analysis is to verify if being the client of such a bank affects somehow the trading behavior. In particular, given the emphasis that these cooperative banks put on trust and transparency with their client we could expect different trading attitudes. Alternatively, it could be that the trading behavior is mainly affected by individual characteristics irrespectively of being client of a small or big bank. Using a unique dataset for the three-year period 2005-2007, we propose bivariate and multivariate analysis using the Negative Binomial model. Our results show that the number of trades depends on individual characteristics, in line with previous research focusing on big banks. It seems therefore that neither the bank size, nor the cooperative nature of the bank is relevant when it comes to the determinants of trading behavior.

JEL Classification: D03, G11, C25

Keywords: behavioral finance, overconfidence, investors' behavior, individual characteristics, banks.

We would like to thank the "Banca di Credito Cooperativo di Gatteo" that allowed us to use its clients' data as well as for the help in the database construction. We thank Massimiliano Castellani, Riccardo Ferretti and Ugo Rigoni for their precious comments, but also the Finance Group members at the Department of Management of the University of Bologna for their suggestions on an earlier draft of this paper. We are responsible for errors and omissions.

[1] University of Bologna, Department of Management. Email: enrico.cervellati@unibo.it, Ph.: +390512098103, Fax: +39051246411. **Presenting author**.

[2] Banca di Credito Cooperativo di Gatteo, Credit and Finance Consultant. Email: pino.fattori@gatteo.bcc.it, Ph. +390541933053, Fax: +390541933583.

[3] University of Bologna, Department of Management. Email: pierpaolo.pattitoni@unibo.it, Ph.: +390512098404, Fax: +39051246411.

1. Introduction

Individual investors' trading has grown over time and has attracted the attention of academics. The growing behavioral finance literature has investigated individual investors' behaviors focusing on the effect of overconfidence on trading. Odean (1999) proposes a theoretical framework in which overconfident traders tend to trade too much, thus lowering their net performances. Barber and Odean (2000) empirically verify this idea analyzing the monthly reports of a discount brokerage house[4] on 78,000 households and 158,034 accounts in the period 1991-1996. Dividing the clients in quintiles of number of trades, the authors show that individuals with higher portfolio turnover recorded lower performances due to the high transaction costs associated with excessive trading.

Online trading, thus, seems to induce people to overtrade. Barber and Odean (2002) analyze a sample of 1,607 online investors in the period 1991-1996, showing that those individual investors that switched from traditional (phone or desk trading) to online trading, on average, recorded lower performances compared to other investors with similar characteristics that instead didn't switch to online trading. Not only the former investors tend to increase their portfolio turnover once they go online, but also the trades appear to be more speculative.

The behavioral literature has shown that overconfidence tend to increase when two psychological biases play a role in the investors' behavior: self-attribution bias and the illusion of control.[5]

Apart from psychological traits, however, other individual characteristics may influence the trading behavior of individuals. Several experimental studies, for example, show how men tend to be more overconfident than women, especially in tasks generally perceived as masculine [Deaux, Farris (1977)] such as trading. Barber and Odean (2001) compare the performances of males and females finding that since men tend to trade more than women, they end up with lower net returns.

A recent Borsa Italiana S.p.A.[6] report [Coraggio, Filippa, Franzosi (2008)] highlights that also in Italy individual investors' trading has grown over time. Between 2003 and 2008, about three million households directly invested in stocks, with an average number of trades per investors of 16 (that was only 12 in 2003). The main reasons of this growth in individual trading are considered to be the higher availability of financial information in the news and the development of e-trading platforms: e.g. 12% of Italian individual investors use the internet to buy stocks.

[4] A discount brokerage house merely transmit its clients' order to the market, usually without consulting and at reduced costs. This is why the authors claim to analyze the true clients' intentions, not influenced by the broker.

[5] Self-attribution bias means that people tend to attribute favorable outcomes to themselves while blaming other for failures. The illusion of control, instead, is the evidence that people believe that they can control things that instead are random, like the behavior of stock markets.

[6] Borsa Italiana S.p.A. is the Italian Stock Exchange management company.

Through a telephone survey, Alemanni and Franzosi (2006) analyze a sample of 203 online traders, comparing the individual characteristics of these investors with the ones of traders only using traditional trading channels. Their results highlight that, from a socio-demographic perspective, online traders differ from the average investor. The typical online traders are men (97% of the sample, *vs* 49% of the Italian population), they live in north-west regions[7] (36% of the sample, *vs* 25% of Italians) and are between 35 and 44 years old (35% in the sample, *vs* 22% in Italy). Furthermore, online traders are usually self-employed, professionals or entrepreneurs (38% of the sample, *vs* only 5% of the population), have a high level of education (49% of the sample has an undergraduate degree, *vs* a low 7% of the population), and have a rather good income. The results of this survey are that the probability of being an heavy trader[8] depends on being: retiree,[9] self-employed and be a resident of a region in the south of Italy.[10]

However, a well-known problem of surveys is the so-called "selection bias", i.e. those willing to do the survey could have certain individual characteristics not shared by the overall set of people under investigation. For example, the investors that are more available to respond to a survey could be the ones that recorded higher performances and are happy to show them. Even if a good choice of the interviewees could partly reduce the selection problem, the self-attribution bias could render the results of the surveys less reliable.

Likely, in a subsequent paper, Alemanni and Franzosi (2007) perform a similar analysis using data obtained from five big Italian online brokers. The brokers made available 500 accounts (100 per each broker) considered as representative of their clients. The results are in line with the 2006 survey, but one important point of the analysis is that is based on actual and not on survey data. Even though the period analyzed is quite limited (1st July – 31st December 2005), the number and details of information are really informative. A possible drawback, nevertheless, is that the analysis is concentrated on 2005, and thus the results could somehow be biased and only represent that particular year. It seems, however, that the results are in line with others found in the international literature and, therefore, they are quite robust.[11]

This paper analyzes the stock trading behavior of the clients of a small cooperative bank located in the north of Italy between 2005 and 2007. We perform both bivariate and multivariate analysis of the relationships between individual characteristics and the number of trades.

[7] It should be emphasized here that the north-west regions are among the wealthier in Italy.
[8] Being a heavy trader means trading stocks daily.
[9] Or, at least, it seems that this probability increases with age. A potential explanation could be that retiree, or older people have more free time to spend in online trading as well as more accumulated wealth.
[10] Therefore, while online traders in general are more concentrated in north-west regions, the ones in southern regions seem to be more active.
[11] Alemanni and Franzosi (2006), comparing other researches on investors in other European countries, highlight that online traders share very similar individual characteristics, irrespectively of their nationality.

In particular, in line with previous studies in the literature with regard to individual characteristics, we focus on gender, age, level of income and job position as direct explanatory variable of the number of trades, but also as possible proxies for overconfidence. We also analyze the effect of being an online trader to verify the presence of excessive trading driven by overconfidence between clients of such a small cooperative bank.

Our analysis contributes to the literature in several respects. First, in spite of the relevance of the issues related to individual investors' trading, due to limitations imposed by access to real data, the available empirical studies that can be found in the literature are still few, and they are even fewer in countries that are not the United States,[12] or with a focus on small banks' clients. Second, our dataset is built directly from the informative system of the bank, and it includes all the data related to the trades in stocks of its clients as well as information on individual characteristics. Thus, we don't need to infer any data or information, or to formulate hypotheses. Lastly, we adopt an alternative econometric methodology that allows to properly treat counting variables such as the number of trades. Since our dependent variable is always non-negative and discrete and that our preliminary analysis shows that the number of trades is usually low, it is not appropriate to use the Ordinary Least Squares (OLS) method.[13]

The paper is structured as follows: section 2 introduces the methodology; section 3 describes the database; section 4 presents an explorative analysis of the relationship between the number of trades and the individual characteristics; section 5 shows the results; section 6, finally, concludes.

2. Methodology

In this paper, we look at individual trading activity focusing on the number of trades in stocks. The dependent variable, therefore, can only assume non negative values. Besides, in our data we record a not negligible number of zeros, i.e. several clients of the bank do not trade in stocks. In addition, the large majority of the clients that are active usually trade few stock. In other words, the mode of our distribution of trades is quite low.

We indicate with y_i the number of trades made by client *i* and with $\mathbf{x}_i$ the vector of explicative variables, and we are interested at explaining the expected value of y_i given $\mathbf{x}_i$, i.e. $\mathrm{E}(y_i \mid \mathbf{x}_i)$.

The peculiarities of our dependent variable (non negative values, a large number of zeros, low average value) impose us to deviate from Ordinary Least Squares (OLS) methodology.

[12] In particular in Italy. Furthermore, we should highlight that the studies by Barber and Odean mostly come from the same database. Other studies on other data will definitely deepen our understand of individual investors' behavior.

[13] Even if the OLS method is based on the assumption of a normal distribution, if the mean of the dependent variable was high enough, the use of this traditional methodology could be appropriated. Unfortunately, in our case, since the average number of trades is very low, we could not apply the OLS method.

More formally, since we must be sure that $\mathrm{E}(y_i \mid \mathbf{x}_i) \geq 0$ for every $\mathbf{x}_i$ the OLS estimator of $\boldsymbol{\beta}$ in the simple linear model $\mathrm{E}(y_i \mid \mathbf{x}_i) = \mathbf{x}_i^T \boldsymbol{\beta}$ is not a suitable candidate, since it allows negative values.

At the same time, since in our dataset y_i assumes the value zero with a not negligible frequency, it is not possible to use a log-linear model.

For counting data (like the ones we have), it is preferable to choose a functional form that insure that $\mathrm{E}(y_i \mid \mathbf{x}_i) \geq 0$. One of most frequent choices in the literature [Wooldridge (2002)] is the exponential function: $\mathrm{E}(y_i \mid \mathbf{x}_i) = \exp(\mathbf{x}_i^T \boldsymbol{\beta})$.

Furthermore, the regression model often assumes that the counting variable y_i follows a Poisson distribution conditionally on $\mathbf{x}_i$ with expected value $\mathrm{E}(y_i \mid \mathbf{x}_i) = \exp(\mathbf{x}_i^T \boldsymbol{\beta})$. The main limit of this model, however, is that it assumes that the conditional variance is equal to the conditional expected value, i.e. $\mathrm{Var}(y_i \mid \mathbf{x}_i) = \mathrm{E}(y_i \mid \mathbf{x}_i) = \exp(\mathbf{x}_i^T \boldsymbol{\beta})$. In case of violation of this condition, the standard errors of the coefficients are underestimated, as well as the associated p-values. Unfortunately, economic data are often characterized, as in our dataset, by over-dispersion [Verbeek (2008)], i.e. $\mathrm{Var}(y_i \mid \mathbf{x}_i) > \exp(\mathbf{x}_i^T \boldsymbol{\beta})$.

An alternative and more general model to treat counting data that solves the over-dispersion issue, is the Negative Binomial [NegBin II in Cameron, Trivedi (1986)], that can be derived from a not observed heterogeneity model in a Poisson model. This heterogeneity could result from errors in the specification of the model, due to the omission of relevant exogenous variables or even, more simply, to be due to the intrinsic unpredictability of the analyzed phenomenon.

If we indicate with h_i the not observed heterogeneity, with $h_i > 0$, and we also assume that $y_i \mid \mathbf{x}_i \sim \mathrm{Poisson}[h_i \exp(\mathbf{x}_i^T \boldsymbol{\beta})]$ where h_i is independent from $\mathbf{x}_i$ and follows a Gamma distribution with average equal to one, and variance $1/\theta$, then it is possible to demonstrate that the distribution of y_i given $\mathbf{x}_i$ is a negative binomial with conditional expected value $\mathrm{E}(y_i \mid \mathbf{x}_i) = \exp(\mathbf{x}_i^T \boldsymbol{\beta})$ and conditional expected variance $\mathrm{Var}(y_i \mid \mathbf{x}_i) = \exp(\mathbf{x}_i^T \boldsymbol{\beta}) + 1/\theta \exp(\mathbf{x}_i^T \boldsymbol{\beta})^2$. This means that the degree of over-dispersion increases with the conditional expected value.

Parameters estimation in the negative binomial model is generally obtained through the maximum likelihood method since it yields robust estimates even in case of errors of specification in the distribution. Thus, if the conditional average is well specified, then the maximum likelihood estimator of $\boldsymbol{\beta}$ in the negative binomial model is consistent.

3. Dataset description

Our database includes both the data on stock trades[14] in the three-year period 2005-2007 and the socio-demographic information on the banks' clients. With regard to the latter, we have available the following information: gender, age,[15] income,[16] and job description. Furthermore, we add a dummy variable that takes the value of one if the client use the internet, and zero otherwise.

Table 1 presents some descriptive statistics related to the number of trades in the considered period. The average (median) number of trades per year for each client is seven (two), even if the standard deviation of trades is three times the average, highlighting a high degree of dispersion in clients' behavior. As a matter of fact, few clients do a lot of trading. The maximum number of trades made by a client is around 200: 166 in 2005, 173 in 2006 and 253 in 2007 (on average, a trade per open markets day). With respect to an average total number of trades per year of about 2,000 (1,989 in 2005, 2,181 in 2006 and 2,318 in 2007), the most active client accounted for about 8% of the total in 2005 and 2006, and about 11% in 2007. Also the overall number of clients slightly increased in the three-year period: 278 in 2005, 299 in 2006 and 312 in 2007.

Table 1. Descriptive statistics for number of trades

Number of trades	**2005**	**2006**	**2007**
Average	7.15	7.29	7.43
Standard Deviation	19.89	20.86	22.46
Minimum	0	0	0
Median	2	2	2
Maximum	166	173	253
Total	1,989	2,181	2,318
Number of clients	278	299	312
Concentration (Gini) [17]	0.7533	0.7529	0.7516

4. Explorative Analysis

In this section, we propose an explorative analysis of the bivariate relationships between the number of trades and the variables in the dataset.

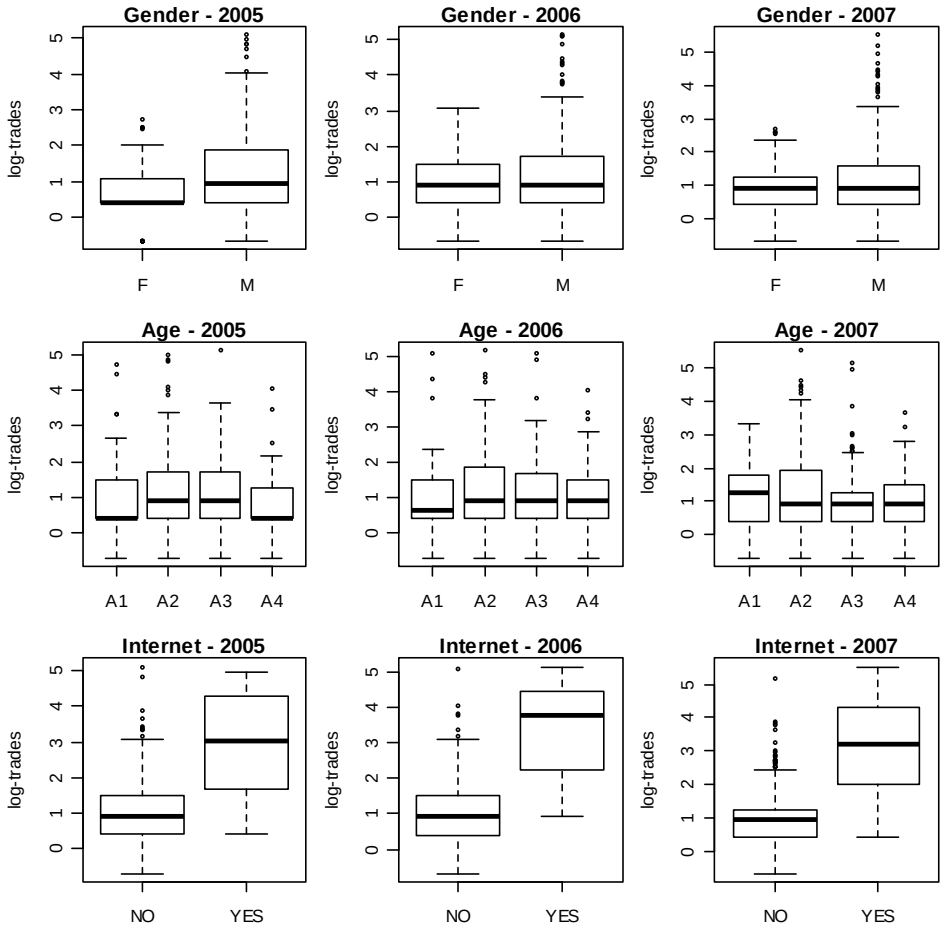
[14] By trade we mean every buy or sell transactions done by the clients via internet, phone, or directly at the bank desk.
[15] Age is calculated, for every year, as the difference between the year of birth and the considered year.
[16] Information on income are certain for employees and retirees, while for self-employed persons are extrapolated from the movements in their checking and saving accounts.
[17] We recall that when the Gini coefficient equals one, there is maximum concentration, i.e. one client makes all the trades, while if it is equal to zero there is equi-distribution, i.e. all clients do the same number of trades.

In figures 1 and 2 we report, for each year, the box-and-whisker graphs in which the number of trades is related to the socio-demographic variables. This graphical representation allows to highlight the degree of dispersion, its asymmetry, and the existence of extreme events. Furthermore, it is possible to compare several distributions, even conditionally to different groups of interest (e.g. gender, age etc.).

In the graphs, the boxes have as extremes, respectively, the first and third quartiles (Q_1 and Q_3) and, therefore, the vertical height of the box represents the inter-quartile difference ($\delta_Q = Q_3 - Q_1$), an index of variability. The horizontal segments inside the boxes represent the medians, while the two external lines (whiskers) correspond to, respectively, $Q_1 - 1.5\delta_Q$ and $Q_3 - 1.5\delta_Q$. Observations beyond these lines indicate the presence of extreme values.

In all graphs (except for the last row in Figure 2 in which we show the frequency of trades), we represent in the vertical axis the logarithmic transformation of the number of trades, with correction of continuity, $\ddot{y} = \log(\text{number of trades} + 0.5)$, to account for zeros [Zeileis et al. (2008)].

Figure 1. Number of trades by gender (F: Female, M: Male), Age (A1: <35, A2: 36-50, A3: 51-65, A4: >65) and existence of an online account. In the vertical axis: logarithm of the number of trades, log(number of trades + 0.5)



In the top row graphs of figure 1, we show the relationship between the number of trades and the gender of clients. From a theoretical point of view, in line with what found in the experimental literature on overconfidence, we would expect a higher number of trades made by men compared to the one of women.[18] The evidence, however, is not clear cut: while in 2005 the median number of trades made by men was definitely greater than the one made by women, in 2006 and 2007 the median is almost the same. It is true that the dispersion is always higher for man than it is for women, but we note that in the men class there are several outliers. The presence of outliers increase the average number of trades for men (8 per year) that is almost three-fold the one found for women (only 3 per year). Thus, while on average men seem more overconfident than women, this evidence could be due to a rather small number of particularly active male traders.

In the middle row of Figure 1, we show the relationship between our dependent variable and the clients' age. We divide clients into four groups depending on their age: below 35 years-old (A1); between 36 and 50 years-old (A2); between 51 and 65 years-old (A3); above di 65 years-old (A4). The two most numerous classes are the central ones, each counting for a third of the total clients, while the extreme classes are about one sixth of the total.[19]

The relationship between age and number of trades in not obvious. On one hand, with age, also the degree of overconfidence could increase due, for example, to greater experience (not only on stock trading, but in general), or to a higher wealth. Also, the amount of time that can be devoted to trading once retired could be higher if compared to the working period in one's life. These two effects would increase the number of trades. On the other hand, however, since we are only focusing on stock trading, we can predict the percentage of stocks in portfolio will decrease as age increases. These two competing forces could yield a peculiar pattern in the relationship between age and stock trading. We forecast that the number of trades in stocks will be relatively limited in the younger age, mostly due to low income and low wealth available as well as the limited time to devote to trading, having to work. Thereafter, the income and wealth conditions for middle-aged persons generally increase, allowing more resources to devote to investments in stocks. Finally, once retired, or approaching retirement, people will tend to decrease the amount invested in stocks. Besides, we believe it is reasonable to think that overconfidence usually reaches its maximum level when a person is at the peak of his/her career.

[18] We note that, in our dataset, men are 73% of the total (664 clients).

[19] In 2005 the four classes A1, A2, A3 and A4 accounted for, respectively 15%, 38%, 31%, and 16% of the total. In 2006: 13%, 37%, 33%, and 17%. In 2007: 10%, 37%, 33%, and 20%.

The evidence seems to support our intuition,[20] at least for the years 2005 and 2006. However, in 2007 the pattern shown in the graph is rather different: it seems, in fact, that the most active traders are the youngest in our dataset.

It should be highlighted that we are focusing on the number of trades, not on monetary amounts, thus it could be that in 2007 the younger group was doing the greater number of trades, but not in terms of amounts traded.[21]

Our results seem in contrast to what found by Alemanni and Franzosi (2006, 2007). However, we should highlight that they show that the probability of being an heavy trader, not an online trader in general, increases with age. In other words, their focus is different from ours. Furthermore, the authors look at several kinds of financial investments, considering not only domestic stocks (as it is our case), but also international ones and derivatives. In fact, as the 2007 study allows to do, focusing on domestic stock trading only it is possible to see that the most active traders are the middle-aged persons, in line with what we found.[22]

Finally, in the last row of Figure 1, we show how the number of trades differ if a client use the internet or not.

We present our considerations on the potential endogeneity of this variable in the section that follows, given the importance of the consequences that this could have in terms of inverse causality for the estimates. Here, we just highlights that, in all the years analyzed, the clients that use the internet for stock trading (about the 5% of the total in 2005 and 2006, and about the 7% in 2007) do trade more. This is in line with previous studies in the literature [see, for example, Barber and Odean (2002)]. It is immediate to see the difference in the number of trades of online *vs* traditional investors in our sample.

Of course, all these relationship are useless if we do not insert in our analysis a measure the clients' income since it is so influential. While the multivariate regression analysis will better take into account the importance of the various explanatory variables, we present here the simple relationship between classes of income and number of trades.

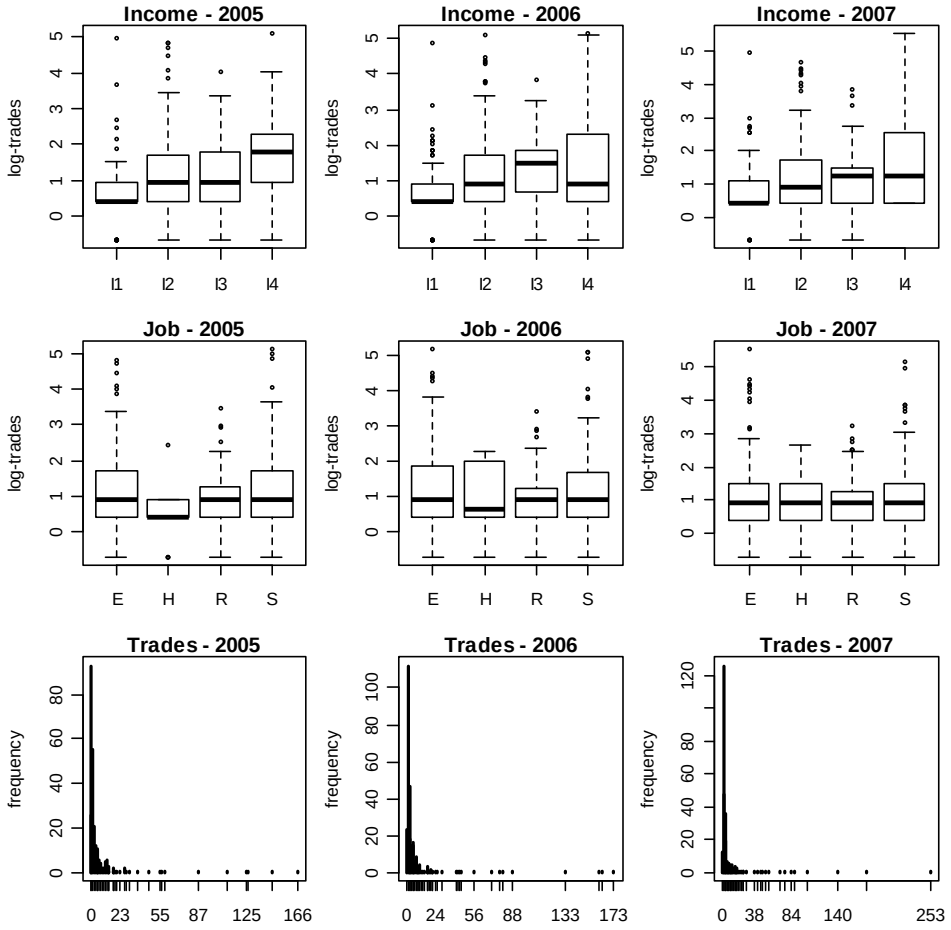
We divide the clients if four income classes: less than 15,000 Euros (I1); between 15,001 and 30,000 Euros (I2); between 30,001 and 55,000 Euros (I3); and greater than 55,000 Euros (I4). In Figure 2, we show that these classes represent, respectively: 33%, 51%, 11% and 6% of the total.

[20] We are well aware, however, that intuition and confirmation bias are very risky, therefore we are presenting more objective support in the regression analysis.

[21] Our data, so far, only allow us to have access to the number of trades, not also to the amounts traded. However, the bank, agreed to give us further data, in terms of kind of information (prices, commissions etc.) and for more years.

[22] The two authors focus on monetary amounts, not only on number of trades. They show that the median level for people in the classes 35-44, 45-55 and more than 55 years-old are, respectively: 12,532; 15,885 and 9,360 Euros.

Figure 2. The graphs in the first two rows report the number of trades for each income class (I1: <15,000; I2: 15,001-30,000, I3: 30,001-55,000, I4: >55,000) and type of job (S: Self-Employed, E: Employee, H: Housewife, R: Retiree). On the vertical axis, in the first two row we report the logarithm of the number of trades or, more formally: log(number of trades + 0.5). Instead, in the last row, in the vertical axis we show the frequency distribution of the number of trades.



We expect that as the income increases, *ceteris paribus*, also the number of stock trades will.[23] People with low annual income levels would have a lower amount to devote to stock trading, even if they would be willing to do so.

In Figure 2, we show how the median number of trades in each year is growing as the income increases.[24]

[23] We do not know the overall clients' wealth. However, we claim that it is quite hard to found data on this variable. It would be possible to prepare a survey to distribute to the clients, but it is very difficult that the clients would honestly answers to the questions related to their total wealth. Furthermore, we could also take into account not only the financial, but also the human capital as suggested by Samuelson (1994). In this respect, we could assume that human capital is higher for young people compared to older ones, but it is really difficult to measure. These two reasons led us to concentrate on annual income for which, besides, we have data.

[24] The only exception is class I4 in 2006 for which the median number of trades is actually lower than the one for I3. However, we note that the dispersion in I4 is higher than the other classes. As a matter of fact, the average number in class I4 is definitely higher (about twenty) than the one of the other classes (from four to eight).

Another important variable of interest in determining the trading behavior is the client's job. Of course, it is related to his/her annual income, but it also influences the amount of free time to dedicate to investing. Furthermore, it could somehow represent a proxy for the degree of risk-aversion. For example, an entrepreneur or a self-employed person usually have a quite risk-seeking attitude.

We present the relation between number of trades and job type in the second row of Figure 2.

We consider four main types of jobs depending on the client being: Self-employed (S),[25] Housewife (H); Employee (E); Retiree (R).[26] The four job types account for 32%, 6%, 41% and 21% of the total number of clients, respectively. In contrast to our expectations, at least from this basic descriptive analysis,[27] we are unable to observe a clear relationship between job type and number of trades.

Finally, as mentioned before, in the bottom row of Figure 2, we show the frequency distribution of trades. As it is clear, we have a not negligible number of zeros (clients not active in stock trading) and the majority of the clients tend to trade little.

5. Results

In this section, we apply the regression model introduced in the methodology part above, with the number of trades, y_i, expressed as $\mathrm{E}(y_i \mid \mathbf{x}_i) = \exp(\mathbf{x}_i^T \boldsymbol{\beta})$. The vector of explanatory variables $\mathbf{x}_i$ contains a constant [*constant*], a dummy variable that equals one if the client is a man [*man*], two variables to verify the influence of age, a linear one [*age*], and a quadratic one [age^2], a dummy variable that equals one if the client uses the internet [*internet*], the natural logarithm of income [*ln(income)*] and other three dummy variables that equals one, respectively, when the client is a housewife [*housewife*], an employee [*employee*] or a retiree [*retiree*]. Since our job category of reference is composed by the self-employed clients we expect to find negative coefficients for the other categories, showing a low propensity to stock trading.

With regard to the other explanatory variables, and based on our preliminary analysis presented in the above sections, we expect to obtain positive coefficients for the variables *man*, *internet*, *ln(income)*, and *age*, and a negative one for age^2.

[25] In this class we include both entrepreneurs and professionals.

[26] We started dividing self-employed people from employees, but then we realized that the number of housewives and retirees trading in stocks was not negligible, so we decided to include these two additional classes. We also tested several others, more detailed, distinctions, but the results were similar to the ones we report here, and there was no relevant additional information apart from adding the housewives and retirees categories.

[27] The regression analysis will provide additional insight on this issue.

Including two variables to measure the relationship between age and number of trades allows us to account for the fact that we expect the number of trades in stocks to initially increase and then to follow as retirement approaches.[28] In our analysis, the variables of interest are *man*, *age*, age^2 and *internet*, while *ln(income)*, *housewife*, *employed* and *retiree* are control variables. Furthermore, to account for the clients turnover, we report the regression results separately for each year, instead of using a panel approach [see Table 2].

Table 2. The table reports the negative binomial regression models in 2005, 2006 and 2007. The dependent variable is the annual number of trades. *Man* is a dummy variable that equals one if the client is a man; *age* and age^2 measure, respectively, the number of years from the clients' birth year and its squared value; *internet* is a dummy variable that equals one if the client uses the internet for stock trading; *ln(income)* is the natural logarithm of the income; *housewife*, *employee* and *retiree* are three dummy variables that equal one when, respectively, the client is a housewife, an employee or a retiree. At the bottom of the table we report two Likelihood Ratio tests (LR). The first one tests the hypothesis that all the coefficients, excluding the intercept, are simultaneously equal to zero. The second one tests the hypothesis that the dispersion parameter $\alpha = 1/\theta$ is zero.

	No. Trades 2005			No. Trades 2006			No. Trades 2007		
	Coeff.	Std. Error	Sign.	Coeff.	Std. Error	Sign.	Coeff.	Std. Error	Sign.
Constant	-5.617	1.673	***	-6.072	1.502	***	-5.348	1.484	***
Man	0.615	0.200	***	0.374	0.179	**	0.443	0.179	**
Age	0.089	0.042	**	0.016	0.037		0.024	0.036	
$Age^2 (\times 10^4)$	-8.925	3.950	**	-0.011	0.335		-1.193	3.221	
Internet	2.239	0.314	***	2.759	0.273	***	2.693	0.243	***
Ln(income)	0.480	0.130	***	0.643	0.118	***	0.568	0.112	***
Housewife	-0.210	0.429		0.677	0.372	*	0.470	0.369	
Employee	-0.317	0.184	*	-0.055	0.166		-0.493	0.162	***
Retired	-0.197	0.261		-0.189	0.235		-0.294	0.217	
θ	0.7609			0.9499			1.0059		
N	278			299			312		
AIC	1,508.30			1,575.50			1,646.70		
Test LR $\left(\chi_8^2\right)$	122.02		***	189.33		***	211.54		***
Test LR $\left(\chi_1^2\right)$	2,676.80		***	2,073.10		***	2,336.80		***

***, ** and * indicate, respectively, significance levels at 99, 95 and 90 percent

The choice of adopting the negative binomial model, rather that the Poisson, is due to the over-dispersion problem mentioned above. A first naïve test to verify the presence of over-dispersion is to compare the mean and the variance of the dependent variable.

[28] Following the so-called principle of time diversification when people are young, i.e. the greater is the investment horizon, the higher should be the portfolio fraction invested in stocks. See Rigoni (2006).

When the variance-mean ratio is greater than two it means that there is a high degree of over-dispersion that the inclusion of explanatory variables could not eliminate. In our case, this ratio is equal to 55.33, 59.69 and 67.89 in 2005, 2006 and 2007, respectively. A more formal test of over-dispersion [Cameron and Trivedi (1990)] consists in testing the equi-dispersion hypothesis against the alternative holding when $\mathrm{Var}(y_i \mid \mathbf{x}_i) = \mathrm{E}(y_i \mid \mathbf{x}_i) + \alpha \cdot m(\mathrm{E}(y_i \mid \mathbf{x}_i))$, where m is a positive function of $\mathrm{E}(y_i \mid \mathbf{x}_i)$, and $\alpha > 0$ if there is over-dispersion. In our case we set $m(\mathrm{E}(y_i \mid \mathbf{x}_i)) = \mathrm{E}(y_i \mid \mathbf{x}_i)^2$, in line with the negative binomial model described above. Following Cameron and Trivedi (1990) we estimated the coefficient α, for the three years, through an OLS auxiliary regression and we used the correspondent statistics *t* that, under the null hypothesis of equi-dispersion, is asymptotically normal. Considering a test size of 0.05, the values of the test statistics in the three years (1.9573, 1.7983 and 2.9282) allow us to reject the null hypothesis.

At the bottom of table 2, we report a further *Likelihood-Ratio* (LR) test of over-dispersion. Indicating with $\ell(\text{Poisson})$ and $\ell(\text{negative binomial})$ respectively the log-likelihood calculated for a Poisson model,[29] and for the negative binomial model, we show that the test statistics $\mathrm{LR} = -2[\ell(\text{Poisson}) - \ell(\text{negative binomial})]$ converges in distribution to a χ_1^2 under the null hypothesis of equi-dispersion.

Since the Poisson model can be obtained imposing the parametric restriction $\alpha = 1/\theta = 0$ in the negative binomial model, the test consists in testing the hypothesis that the dispersion parameter α is zero ($\theta \to \infty$). The values of the statistics in the three years (respectively: 2,676.80; 2,073.10 and 2,336.80) once more suggest to reject to Poisson model in favor of the negative binomial one.

The coefficients of the variables *housewife*, *employee* and *retiree* have, as expected, a negative sign, even if they are not always significantly different from zero. Since the residual job type is *self-employed*, our results are in line with what found by Alemanni and Franzosi (2006).[30]

At the bottom of table 2, we report a LR test statistics to test the hypothesis that all the coefficients in our model, apart from the intercept, are simultaneously equal to zero. The values of the statistics in the three considered years are, respectively, 122.02, 189.33, 211.54), thus we reject the null hypothesis that the conditional mean is constant and independent from the explanatory variables.

The results largely confirm our expectations. The positive coefficient associated with the *man* variable suggests that men actually do more trades than women, in line with what found in previous studies in the literature [Barber e Odean (2001)].

[29] We do not report the results for the Poisson model for brevity.

[30] The only exception is the housewife coefficient in 2006 that is positive and significantly different from zero. It seems therefore that in that year the housewives trade more in stocks than the other clients of the bank. This could be explained by the fact that housewives could have more free time to devote to stock trading. However, it does not explain why this is the case only in 2006.

The coefficients related to the variables *age* and *age* ^{2} also have the expected signs (positive in the first case and negative in the second one) in all the years considered.[31] Our intuition of a non-linear effect of age on the number of stock trades is therefore confirmed.

The coefficient of the *internet* variable is also in line with the literature [Barber e Odean (2002)] suggesting that online trading pushes the number of trades.

Finally, also the coefficient of the variable labeled [*ln(income)*] is positive, suggesting that the higher is the income, the higher the money that can be devoted to stock trading.

To verify the existence of multicollinearity between the explanatory variables, we calculated the Variance Inflation Factor (VIF) for every regressor in our models. Even though the VIF values for each variable do not exceed the critical value of ten, the variables *age* and age^2 record very high values (around seven). This could be why the coefficients for these two variables are not significantly different from zero in 2006 and 2007.

Lastly, we verified the existence of a causality relationship between the number of trades and the *internet* variable. The endogeneity problem is particularly relevant in our analysis, i.e. it could be that the choice of using internet is determined by an increase in the number of trades and not vice versa. In other word, it could be that only the most active clients decide to put their order via internet for example to save in commission, that are usually lower using the internet.

We therefore performed an endogeneity test following Wooldridge (2002). The procedure requires two steps. In the first one, we regressed the variable *internet* on all the other explanatory variables (*constant*, *man*, *age*, age^2, *ln(income)*, *housewife*, *employee*, and *retiree*) and a new dummy variable that equals one when a client has a job for which is more probable that he or she will use the internet. To construct this variable we analyzed in detail the information on the clients' job descriptions. Then, we have added the residuals of this regression as an additional explanatory variable in our original model. If the *internet* variable is endogenous, the coefficient associated to the residuals would significantly be different from zero. In our case the p-values associated with this test are, respectively, 0.937, 0.4082, and 0.4915 in the three considered years, suggesting that the inverse causality problem does not characterize our models.

6. Conclusions

Given the increasing trading of individual investors, it is of primary importance to understand how their personal characteristics influence their behaviors.

[31] Even if they are significantly different from zero only in 2005.

In this paper, using a unique dataset obtained from a small cooperative bank, we propose an analysis of the relationship between number of stock trades and individual characteristics like age, gender, job type, income, but also the possibility to use the internet.

The behavioral finance literature, as well as several psychological experiments, have demonstrated how these characteristics may influence the degree of overconfidence of an individual and, thus, his or her trading behavior.

Since the dependent variable in our analyses (the number of trades) can only assume non-negative values, and given the evidence that usually the clients in our dataset make, on average, a small number of trades in stocks, we have chosen the Negative Binomial model for our regression analysis since neither the OLS, nor the Poisson model are suitable candidates in this case.

Our results are in line with previous study in the literature, showing the importance of our explanatory variables. In particular, we show that the number of trades is higher for men, for the clients with higher income and that use the internet for trading. Furthermore, we analyze both a linear and a quadratic relationship between our dependent variable and the clients' age. Our results, in line with our intuition, show that while for young people the number of trades is usually quite low, it increases for middle-aged clients, and thereafter it decreases for retirees or for individuals approaching retirement. Finally, also the job type is important in our framework. This is the case not only because it is correlated with income, but also since it could be taken as a proxy for either free time to devote to stock trading, and for risk-seeking behavior. As a matter of fact, self-employed individuals tend to me the class of clients making more transactions.

An important aspect of our analysis is that it shows that the clients of a small bank are not different from the ones of biggest financial institutions. To the best of our knowledge, previous studies in the literature mostly rely on the cases of the clients of big financial institutions.

Furthermore, the bank in our case is a cooperative one, but this does not alter the clients' behaviors. We think that our results can be generalized to other small banks, that constitute an interesting research field, not only for Italy.

Our results show that investors' trading behaviors seem to be dictated by individual characteristics rather than from being the clients of a big or small bank (at least when the bank avoid any consulting activity). However, this conclusion holds in a period that can be considered quite "normal". It would be interesting to see the effect of the recent financial crisis on the bank's clients. In this case, we could truly analyze the effect of trust on trading behavior.

Of course, we are analyzing only one particular case of a small cooperative bank, and it would be interesting to enlarge the analysis the a greater number of similar institutions.

References

Alemanni B., Franzosi A. (2006) "Portfolio and psychology of high frequency online traders. Second report on the Italian market", *Borsa Italiana Bitnotes*, n. 16.

Alemanni B., Franzosi A. (2007) "Traders' online behaviour. Come operano i trader online italiani sui mercati. Terzo rapporto sul mercato italiano", Borsa Italiana, April, 11, 2007.

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

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

Barber B.M., Odean T. (2002) "Online Investors: Do the Slow Die First?", Review *of Financial Studies*, 15: 455-489.

Cameron A.C., Trivedi P.K. (1986) "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests", *Journal of Applied Econometrics*, 1: 29-53.

Cameron A.C., Trivedi PK (1990) "Regression-Based Tests for Overdispersion in the Poisson Model", *Journal of Econometrics*, 46: 347-364.

Coraggio V., Filippa L., Franzosi A. (2008) "L'investitore *retail* e la Borsa. Quale possibile relazione con le società quotate. Terzo rapporto di Borsa Italiana sullo *shareholding* in Italia".

Deaux K., Farris E. (1977) "Attributing causes for one's performance: The effects of sex, norms, and outcomes", *Journal of Research Personality*, 11: 59-72.

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

Rigoni U. (2006) *Finanza comportamentale e gestione del risparmio*, Giappichelli.

Samuelson P.A. (1994) "The long-term case for equities", *Journal of Portfolio Management*, Fall, 15-24.

Verbeek M. (2008) *A Guide to Modern Econometrics*. Wiley.

Wooldridge J.M. (2002) *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA.

Zeileis A., Kleiber C., Jackman S. (2008) "Regression Models for Count Data in R", *Journal of Statistical Software*, 27 (8): 1-25.