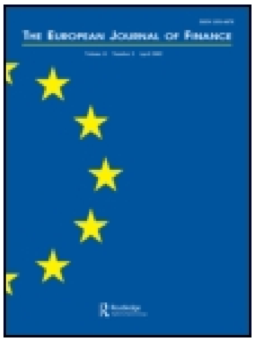

The European Journal of Finance

ISSN: 1351-847X (Print) 1466-4364 (Online) Journal homepage: https://www.tandfonline.com/loi/rejf20

A behavioral analysis of investor diversification

Ana-Maria Fuertes, Gulnur Muradoglu & Belma Ozturkkal

To cite this article: Ana-Maria Fuertes, Gulnur Muradoglu & Belma Ozturkkal (2014) A behavioral analysis of investor diversification, The European Journal of Finance, 20:6, 499-523, DOI: 10.1080/1351847X.2012.719829

To link to this article: https://doi.org/10.1080/1351847X.2012.719829

Published online: 28 Nov 2012.

The European Journal of Finance, 2014
Vol. 20, No. 6, 499–523, http://dx.doi.org/10.1080/1351847X.2012.719829

A behavioral analysis of investor diversification

Ana-Maria Fuertes[a], Gulnur Muradoglu[b]* and Belma Ozturkkal[c]

[a]*Faculty of Finance, Cass Business School, City University, London ECIY 8TZ, England;* [b]*Behavioral Finance Working Group (BFWG), Faculty of Finance, Cass Business School, City University, London ECIY 8TZ, England;*
[c]*School of Economics and Administrative Sciences, Department of International Finance, Kadir Has University, Istanbul, Turkey*

(*Received 2 July 2012; final version received 3 August 2012*)

This paper studies the link between individual investors' portfolio diversification levels and various personal traits that proxy informational advantages and overconfidence. The analysis is based on objective data from the largest Turkish brokerage house tracking 59,951 individual investors' accounts with a total of 3,248,654 million transactions over the period 2008–2010. Wealthier, highly educated, older investors working in the finance sector and those trading relatively often show higher diversification levels possibly because they are better equipped to obtain and process information. Finance professionals, married investors, and those placing high-volume orders through investment centers show poorer diversification possibly as a reflection of overconfidence. Our analysis reveals important nonlinear effects, implying that the marginal impact of overconfidence on diversification is not uniform across investors but varies according to the investor's information gathering and processing abilities.

Keywords: individual investor; behavioral finance; diversification; portfolio risk; emerging market

JEL Classification: G01; G11; G24

1. Introduction

Investors can benefit from portfolio diversification by mitigating large return correlations between assets (Markowitz 1952). For a well-diversified portfolio, investors should hold a number of assets somewhere between 10 (Evans and Archer 1968) and 30 or 40 (Statman 1987). Empirical evidence has shown for more than three decades now that the typical US individual investor's portfolio contains a much smaller fraction of the optimal portfolio size (Blume and Friend 1975). Recent studies for other countries are also consistent with much smaller portfolio sizes in terms of diversification than theory predicts. The average size of individual investors' portfolios is about two in Finland (Grinblatt and Keloharju 2009), between four and five in Germany (Dorn and Huberman 2005; Dorn and Sengmueller 2009), and about seven in the Netherlands (Hoffman and Sheffrin 2011).

The above studies investigate trading performance and its relation to trading choices and personal traits. There is a paucity of empirical research regarding the nexus between *portfolio diversification* of individual investors and their demographic and trading characteristics. The goal of this paper is to contribute toward filling this gap. Diversification is the most naïve, and almost costless, method of risk reduction. It is important to understand who diversifies better. This paper is one of the first to directly examine the determinants of the apparent failure to hold

*Corresponding author. Email: g.muradoglu@city.ac.uk

a well-diversified portfolio. The existing empirical literature uses a number of objective personal traits such as age, education, employment, income, and gender, in explaining aspects of investment behavior such as frequency of trading (Barber and Odean 2001), trading activity (Grinblatt and Keloharju 2009), portfolio turnover (Dorn and Sengmueller 2009), stock market participation (Grinblatt, Keloharju, and Linnainmaa 2011), objective strategies and performance (Hoffmann and Shefrin 2011), and self-reported risk aversion (Dorn and Huberman 2005). None of them directly seeks to map personal traits into diversification.

We depart from the previous literature in formulating and assessing the empirical validity of new hypotheses around the issue of how individual attributes effect portfolio diversification. First, we build on information theory to hypothesize that better informed investors diversify more (this is referred to as the *informational advantage* hypothesis, H_{0A}). We rely on traditional theory of investor behavior to hypothesize that overconfident investors are characterized by poor portfolio diversification (this is termed the *overconfidence bias* hypothesis, H_{0B}). Second, we further conjecture that there are nonlinear effects, namely, interactions between investors' information-processing capacity and their degree of overconfidence can be influential for their portfolio diversification decisions (this is referred to as the *information-overconfidence interaction* hypothesis, H_{0C}).

Traditional finance theory recommends that individuals hold a well-diversified portfolio of stocks, but there is ample empirical evidence suggesting that the typical retail investor fails to do so. Explanations vary as to why. The first hypothesis in this paper, H_{0A}, hinges on the information-processing ability of individual investors; it states that individuals who are more able to obtain and process economic and financial information are more likely to invest in stocks and hold better diversified portfolios. Our empirical findings show that this is indeed the case. The second hypothesis, H_{0B}, draws from traditional theory of investor behavior. Risk-aversion levels and overconfidence are the most frequently cited psychological attributes in trading behavior (Dorn and Huberman 2005; Glaser and Weber 2009). Overconfidence and risk-taking behavior are intertwined in the sense that too confident investors are prone to take higher risks. We hypothesize that poor portfolio diversification is related to overconfidence. Our inferences largely bear this out. The third and final hypothesis in the paper, H_{0C}, states that the impact of investors' information-processing capacity on portfolio diversification is not constant across investors but instead it is influenced by their overconfidence. Overconfidence is shown plausibly to hinder the portfolio diversification levels of investors that are well able to obtain and process information.

Finally, we corroborate that more diversified investors earn better returns on their investments. Poor portfolio diversification is hazardous to individual investors' wealth. Benartzi and Thaler (2007) discuss that when investors diversify, they tend to use naïve diversification strategies such as the $1/n$ rule. When investors are faced with a limited number of 'n' options, they simply divide the assets evenly across the options. DeMiguel, Garlappi, and Uppal (2009) show that the $1/n$ rule can indeed be difficult to beat as a portfolio allocation strategy. Our work differs from this line of research in that we investigate whether the information-processing ability and overconfidence of individual investors are significant factors in explaining the cross-section and time-series variation in their portfolio diversification levels, and assess the marginal impact of diversification on trading profitability.

Goetzman and Kumar (2008) use age and income as two key variables to proxy investor sophistication. They report that younger and lower income individuals hold less-diversified portfolios. We would expect younger investors to have low information-processing ability due to lack of experience. Low-income investors are unlikely to pay for financial advice and information, whereas, on the other hand, financially wealthy individuals have the means and willingness to do so. Existing evidence suggests that wealthier investors hold better diversified portfolios (Dhar and Zhu 2006; Vissing-Jorgensen 2004).

In this paper, we dig deeper into information-processing theory of investor behavior and conjecture that education, wealth, and job sector also matter in processing information. Less-educated investors are not as well equipped to gather and process financial information as highly educated individuals. Wealthy investors are better positioned to allocate resources to gathering and processing of financial information. The type of investor's job has a less clear-cut effect on portfolio diversification. Individuals working in the finance industry are better placed to obtain and evaluate information for their investment decisions.[1] If the enhanced information-processing capacity of investors leads to better portfolio decisions then individuals who are working in the finance sector would have better diversified portfolios. However, if an investor has a job that is financial in nature she could be relatively overconfident which may hinder diversification. Thus, the link between portfolio diversification and an investor's information-processing ability hinges on her overconfidence, and the relationship between portfolio diversification and investor's overconfidence hinges on her information-processing ability.

Both Benos (1998) and Odean (1998) argue that overconfidence induces excessive trading. Overconfident traders engage in more frequent transactions because they overestimate the precision of their own signals compared with the precision of other traders' signals. To put it more generally, investors with 'excess' of confidence tend to overestimate their trading skills. We would expect overconfident individuals to be less diversified in their portfolio behavior. Barber and Odean (2002) analyze brokerage clients that switch from phone-based to online trading. They argue that online traders become more overconfident because of self-attribution bias, illusion of knowledge, and control. However, it is possible too that by placing trading orders through investment centers which provide financial advice, investors feel more satisfied, and re-assured and thus become overconfident. Thus, *a priori* the effect of type of order (through investment center, phone, or online) is not clear cut.

In various domains of life, anecdotal evidence suggests that women and married people exhibit less confidence and higher risk-aversion levels than men and single people, respectively; for instance, it has often been reported in the press that women are underpaid compared with men on the same jobs for the same level of experience and education level. Barber and Odean (2001) find that married US investors have lower turnover, lower return volatility, and lower market risk choices, married women trade less than married men, and women have higher stock market return compared with men. The conjecture that overconfidence leads to poor portfolio diversification leads naturally to expect that men are less diversified than women, and married people are better diversified than singles. However, recent evidence for the Finnish stock market conveys a slightly different message: Grinblatt and Keloharju (2009) report that married Finnish investors have higher number of trades and higher portfolio turnover indicating overconfidence. Likewise, Grinblatt, Keloharju, and Linnainmaa (2011) document that married Finnish investors tend to be less diversified *ceteris paribus*.

In traditional theory of investor behavior, overconfidence is proxied by measures of trading activity. Odean (1998) and Barber and Odean (2000, 2001) argue that frequent US traders do considerably worse than less-active (and passive) traders when transaction costs are taken into account. Conventional motives of trading such as savings and risk sharing cannot be used to explain this, whereas overconfidence provides a logical explanation. Trading activity has two dimensions: frequency of trading and volume of trading. Overconfidence has been often directly linked to the former. However, the frequency of trading could be higher due to better information-processing capacity. Investors who can efficiently process more signals are likely to trade more frequently. There may be also a learning curve or 'learning by trading' process by which investors who trade very frequently learn to process financial and economic information more efficiently (less costly). Thus, it is possible to argue that investors with higher number of trades end up diversifying better. On the other hand, volume of trading can plausibly signal the overconfidence an investor places in her bets; accordingly, one should expect individual investors with higher volume of trades to diversify less well.

The previous literature does not take into account the interaction between an investor's information-processing capacity and her overconfidence. An investor with a post-graduate degree is better equipped to process economic and financial information better than another individual with only school-level education. However, having a finance-related job could boost her confidence as an investor. Heightened information-processing capability can plausibly improve diversification, whereas overconfidence will have the opposite effect. The aforementioned interaction would work so as to make the positive effect of information gathering (post-graduate education) on portfolio diversification decrease with the level of overconfidence (finance-related job). Our empirical analysis suggests that overconfidence proxies interacted with personal traits that proxy the investor's information-gathering-and processing ability are significant nonlinear determinants of portfolio diversification.

A final strength of our paper comes from relying on objective investor traits. Recent papers employ surveys to elicit investor attributes (Dorn and Huberman 2005; Glaser and Weber 2007; Graham, Campbell, and Huang 2009). The use of surveys raises several issues such as inaccurate responses (Campbell 2003), misunderstood questions (Bertrand and Mullainathan 2001), and non-response biases. Individual responses could be domain specific and poorly correlated among the proxies (Weber, Blais, and Betz 2002). We use objective investor attributes as proxies for psychological traits (Barber and Odean 2001). All the variables that we use to characterize an investor's profile are of this objective nature and are obtained from brokerage house records. The remainder of the paper unfolds as follows. The next section describes the dataset. Section 3 presents the methodology. Section 4 discusses the findings from our empirical analysis. Section 5 concludes and provides directions for further research.

2. Data description

Our analysis exploits two databases consisting, respectively, of individual trading information and end-of-quarter portfolio positions for 59,951 individual investors with accounts at a major Turkish brokerage house (*Garanti Yatirim*) from 31 March 2008 to 26 February 2010. This period spans a total of $T = 697$ trading days distributed over eight quarters. Such a disaggregated and comprehensive database from a European market has not been exploited before in a behavioral analysis of individual investor diversification. The dataset contains demographic information such as age, education, occupation, financial wealth, city of residence, gender, and marital status. It includes also trading information for all stocks that are bought and sold during the research period and end-of-quarter portfolio compositions for each of the individual (retail) investors. For each share in the individual investors' portfolios, we obtain prices, returns, and market capitalization from *Datastream International*.

Although the cross-section in our sample represents only about 6% of all retail investors in Turkey, it provides a fairly good representation of the one million total individual investors in the Turkish stock market for various reasons. The data are provided by one of the largest brokerage houses in the country.[2] Including their portfolio management branch, the company has 16% market share in assets under management and 6% market share on Istanbul Stock Exchange (ISE) trading volume. Clients can trade either through the investment center, Internet, or via their call centers, where they can place their orders over the phone. The cross-section in our sample represents about 6% of all retail investors in Turkey.

Our sample can be cast as a broad representative of the overall Turkish retail stock market for other reasons. A comparison of broad descriptive statistics for our data and those reported by the Association of Capital Market Intermediary Institutions of Turkey (TSPAKB) bears this out. About 70% of Turkish investors are in the 30–54 age group. Regarding gender, 73% of domestic investors are male and 27% are male.[3] In our sample, 66% of the investors fall in the 30–54 group with an average age of 40; the total cross-section contains 83% male and 17% female investors. TSPAKB states that most individual investors (54%) and those with the largest stock holdings (84%) are from the three largest cities in the country, namely, Istanbul, Ankara, and Izmir. In our sample, a total of 65% of retail investors are from the three major cities. According to TSPAKB reports, about 33% of Turkish investors' total portfolio is allocated to four stocks comprising three major banks and one telecom stock (Garanti, İş, Ak, and Tcell). Our cross-section of investors has 8% of their portfolios allocated in those four shares. Appendix 1 provides a brief 'anatomy' of the Turkish Stock Market. Appendix 2 gives the cross-section of investors (out of the maximum 59,951 investors sampled) with stock portfolio holdings at the eight end-of-quarter snapshots in our 2-year observation period.

2.1 *Main variables and preliminary statistics*

Our analysis of the relationship between portfolio diversification (dependent variable) and individual investors' information-processing and overconfidence characteristics (independent variables) is organized around the first two hypotheses presented in Section 1 that hinge on *information theory* (H_{0A}) and *behavioral theory* (H_{0B}). These hypotheses motivate various covariates, which can be broadly grouped as proxies for the information gathering and processing ability of investors, on the one hand, and proxies for the level of overconfidence, on the other. A third set that complements the above two includes realized profit/loss measures. Appendix 3 provides a full list of the variables with brief definitions.

The main focus of the analysis is to map the level of diversification, an important aspect of portfolio composition, into an investor's profile. In order to increase the robustness of our conclusions, we consider four measures of diversification.[4] Two of them, referred to as HHI_$Q(t)$ (HHI, Herfindahl–Hirschmann Index) and DIVERSIFY_$Q(t)$ have both a cross-section (client) and time (end-of-quarter) dimension, whereas the other two referred to as DIVERSIFY_Q and DIVERSIFY have only a cross-section dimension. Hence, an observation is defined as a client-quarter data point in the context of the former two measures of diversification, whereas an observation is a client data point in the context of the latter two measures. More specifically, for the first two diversification measures, each data point represents a client whose portfolio holding is observed at a specific quarter-end snapshot. Since some clients may temporarily liquidate their portfolio, they may not hold any stock at such specific snapshots so the sample is unbalanced. Furthermore, some client-quarter (or client) observations may not be available for some of the independent variables, and this further reduces the effective sample size available for the estimation of the regression model parameters. Each diversification measure is formally presented next.

Our first diversification measure is the HHI defined as the sum of the squared normalized portfolio allocation weights following Dorn and Sengmueller (2009), Dorn, Huberman, and Sengmueller (2008), Dorn and Huberman (2005), Hoffmann and Shefrin (2011), and Hoffmann, Shefrin, and Pennings (2010) *inter alia*. Formally, this index is computed as follows:

$$\text{HHI\_}Q(t) \sum_{i=1}^{n(t)} \left(\frac{N_{it} * P_{it}}{\sum_{i=1}^{n} N_{it} * P_{it}} \right)^2, \tag{1}$$

where $n(t)$ is the number of different stocks in the investor's portfolio at quarter-end t, N_{it} is the number of shares in stock i at quarter-end t and P_{it} is the price of each of those shares. Higher values indicate better portfolio diversification; the lower bound of the index is 0 as $n(t) \to \infty$, and the maximum value is 1 when there is maximum concentration (no diversification) and the entire portfolio is allocated to one stock, *i.e.* $0 < \text{HHI\_}Q(t) \leq 1$. By assuming equal-weight allocation to different shares, it is possible to map the average HHI into the average number of different stocks in the investor's portfolio over time as $n \approx 1/T^{-1} \sum_{t=1}^{T} \text{HHI\_}Q(t)$ where $t = 1, \ldots, T$ are quarters with $n(t) > 0$.

Several investors hold no portfolios at some of the end-of-quarter snapshots, i.e. $n(t) = 0$. As detailed in Appendix 2, the number of investors holding a non-zero portfolio at any quarter-end point is 29,649. Moreover, some share prices to calculate the HHI_$Q(t)$ are unavailable because of discontinued trading. Thus, in effect a total of 63,682 client-quarter points are available for the HHI_$Q(t)$ variable to use in our subsequent modeling exercise.

Our second diversification variable, DIVERSIFY_$Q(t)$, is an end-of-quarter snapshot of the portfolio held by each of the investors in the sample. Formally, we denote it by

$$\text{DIVERSIFY\_}Q(t) \equiv n(t), \tag{2}$$

where $n(t)$ is the number of different stocks held in the investor's portfolio at the end-of-quarter t. A total of 74,824 client-quarter observations are available for this variable.

Our third measure of diversification, DIVERSIFY_Q, is an *average* of the number of different shares held in end-of quarter portfolios for each investor. We conceptualize this cross-sectional diversification measure as the cumulative number of different stocks held in the investor's portfolio each end-of-quarter t divided over the number of quarter-end points when the investor is holding a non-zero portfolio, $n(t) > 0$, as defined below

$$\text{DIVERSIFY\_}Q \equiv \frac{\sum_{t=1}^{T^*} n(t)}{T^*}, \tag{3}$$

where $T^* \leq T$ and T denotes the eight quarter-end points available in our 2-year sample. This diversification measure has observations available for a total of 23,345 clients.

Our final diversification measure, DIVERSIFY, is a *time-weighted average* of the number of different shares in the investor's portfolio constructed on the basis of continuous (daily) information. It is computed as the cumulative number of different stocks held by the investor weighted by the holding duration (length in days) divided by the total inventory duration or total days of the sample period when the client is holding some shares. Formally, we have

$$\text{DIVERSIFY} \equiv \frac{n_0 \times \text{days}_0^1 + n_1 \times \text{days}_1^2 + \cdots + n_{N-1} \times \text{days}_{N-1}^N)}{\text{days}_1^0 + \text{days}_1^2 + \cdots + \times \text{days}_{N-1}^N}, \tag{4}$$

where $n_0 \geq 1$ is the number of different stocks held on day 0 (initial day in the sample period), days_0^1 is the numbers of days between day 0 and the first trading day (called day 1), $n_1 \geq 1$ is the updated number of different stocks in the portfolio after the day 1 trades have been accounted for, days_1^2 is the number of days between day 1 and the second trading day (called day 2), and so forth. The denominator amounts to the total inventory duration, dur_inv $\leq T_d$, where $T_d = 697$ is the number of days in our 2-year sample period. The number of observations available for DIVERSIFY are for a total of 56,263 clients.

Summary statistics for the four diversification measures over the entire 2-year sample can be seen in Table 1. End-of-quarter summary statistics for HHI_$Q(t)$ and DIVERSIFY_$Q(t)$ are presented in Table 2. The two panel measures, HHI_$Q(t)$ and DIVERSIFY_$Q(t)$, contain overlapping information albeit not fully and are negatively associated (Pearson correlation −0.4209) low portfolio concentration HHI_$Q(t)$ is tantamount to high diversification. The two cross-section measures, DIVERSIFY_Q and DIVERSIFY, not only contain overlapping information,but is imperfectly too (Pearson correlation 0.6751).

The mean of HHI_$Q(t)$ is 0.86 which corresponds to 1.2 shares on average for individual investors. The other three measures, DIVERSIFY_$Q(t)$, DIVERSIFY_Q, and DIVERSIFY, suggest an average diversification of 2.06, 1.73, and 2.26 shares, respectively. Overall the four diversification proxies indicate that the number of stocks in Turkish investors' portfolios is

Table 1. Preliminary data analysis. Summary statistics from 31 March 2008 to 28 February 2010.

	Diversification variables				Profit and loss variables	
	HHI_$Q(t)$	DIVERSIFY_$Q(t)$	DIVERSIFY_Q	DIVERSIFY	PROFIT-LOSS	PROFIT-LOSS dummy
Mean	0.86	2.06	1.73	2.26	0.02	0.55
Median	1.00	1.00	1.00	1.00	0.01	1.00
Max.	1.00	342.00	299.88	242.58	0.99	1.00
Min.	0.02	1.00	1.00	1.00	−0.18	0.00
St. Dev	0.24	4.73	3.13	3.09	0.27	0.50
Obs.	63,682	74,824	23,345	56,263	132,742	59,951

	Informational advantage proxies							
	AGE	EDU_High School	EDU_College	EDU_Postgrad	SECTOR	WEALTH(t)	TRADES(t)	GNP_CITY
Mean	40.05	0.28	0.55	0.09	0.04	87,259	63.93	3430
Median	38.00	0.00	1.00	0.00	0.00	8,818	13.00	3711
Max.	95.00	1.00	1.00	1.00	1.00	102,000,000	11,278	7468
Min.	18.00	0.00	0.00	0.00	0.00	0.00	1.00	688.33
St. Dev	10.94	0.45	0.50	0.29	0.20	762,769	186.13	1069
Obs.	52,395	52,395	52,395	52,395	52,395	374,564	161,965	52,650

	Overconfidence Proxies				
	PROFESSION	MARRIED	GENDER	VOLUME	ORDER_TYPE
Mean	0.06	0.70	0.83	1.37E+06	0.23
Median	0.00	1.00	1.00	37259.68	0.00
Max.	1.00	1.00	1.00	2.36E+09	1.00
Min.	0.00	0.00	0.00	0.30	0.00
St. Dev	0.24	0.46	0.38	1.58E+07	0.42
Obs.	52,395	52,395	52,395	59,951	59,951

Notes: Variables with time t in parenthesis have client and quarter (or end-of-quarter) dimensions for the panel regressions. Appendix 3 defines each variable.

Table 2. Preliminary data analysis. End-of-quarter statistics for panel diversification measures.

	31.03.2008	30.06.2008	30.09.2008	31.12.2008	31.03.2009	30.06.2009	30.09.2009	31.12.2009
Porfolio concentration HHI$_Q(t)$								
Mean	0.864	0.867	0.869	0.861	0.853	0.860	0.860	0.858
Median	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Max.	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Min.	0.029	0.026	0.030	0.021	0.022	0.023	0.023	0.026
St. Dev	0.241	0.237	0.237	0.243	0.249	0.246	0.247	0.249
Obs. (clients)	7013	7362	7378	8331	8939	8241	8371	8047
Number of different shares DIVERSIFY$_Q(t)$								
Mean	2.022	2.037	1.990	2.080	2.128	2.092	2.048	2.046
Median	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Max.	164.000	281.000	309.000	312.000	317.000	340.000	342.000	334.000
Min.	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
St. Dev	3.558	4.378	4.445	4.772	4.793	5.702	4.923	4.767
Obs. (clients)	8398	8694	8712	9681	10242	9688	9827	9582

Notes: The table reports summary statistics for two panel diversification measures based on end-of-quarter portfolios. Variable definitions are in Appendix 3.

about 2 on average. The cumulative distribution shown in Table 3 indicates that the percentage of our original cross-section of 59,951 individual investors that hold more than 10 stocks on average during the 2-year sample period is 0.4% according to the HHI_$Q(t)$ measure using the approximation of equal-weight allocation, 1.4% according to DIVERSIFY_$Q(t)$, 0.9% using DIVERSIFY_Q, and 1.6% using DIVERSIFY. These descriptive statistics are in line with the average number of stocks held in individual Finnish investors' portfolios, reported at about 2 by Grinblatt and Keloharju (2009). The average number of stocks reported for other countries is slightly higher, at about 4 in Germany (Dorn and Huberman 2005; Dorn and Sengmueller 2009), and about 7 in the Netherlands (Hoffman and Sheffrin 2011). The portfolio of a US household has been estimated to comprise 4 stocks (Barber and Odean 2001) or 4.7 stocks (Goetzmann and Kumar 2008). The latter study documents that over 45% of individual US investors have less than three different stocks in their portfolio and that 10% of individual investors hold more than 10 stocks on average. Individual investor under-diversification appears to be a universal problem.

As potential determinants of diversification, we consider a combination of quantitative (continuous) and qualitative (discrete) variables seeking to provide a complete characterization of an investor's profile. Our previous theoretical discussion motivates three categories. A first category of variables is linked to the *informational advantage* hypothesis $H_{0\mathrm{A}}$. AGE is a quantitative variable giving the age of the investor at the end of the sample period. Education is measured discretely on a four-point scale to represent the highest education level of the investor at the end of the sample period: elementary school, high school, university, or post-graduate degree. Hence, three dummy variables are included in the analysis: EDU_HighSchool equal to 1 if the investor has up to high-school education, EDU_College equal to 1 if the investor has up to BSc degree, and EDU_Postgrad if the investor has up to PhD or MSc degree. The brokerage house provided us with 23 different codes for the sectors their clients are working on and 435 different codes for their clients' profession at the end of the sample period. On this basis, we created two qualitative variables: SECTOR and PROFESSION. The former takes value 1 if the investor is working in the finance sector and 0 otherwise. PROFESSION is explained below since it belongs to our second category of variables. WEALTH is the size in Turkish Lira of the investor's financial asset portfolio measured as end-of-quarter values. This information is provided by the brokerage house

Table 3. Preliminary data analysis. Cumulative distribution of diversification.

	HHI_$Q(t)$	DIVERSIFY_$Q(t)$	DIVERSIFY_Q	DIVERSIFY
Number of shares	Cumulative frequency %			
1	58.91	64.54	59.09	27.01
2	89.43	82.31	84.55	67.58
3	95.12	89.85	92.44	82.85
4	97.19	93.35	95.46	89.64
5	98.12	95.22	96.95	93.26
10	99.64	98.49	99.13	98.37
20	99.90	99.50	99.78	99.67
30	100	99.7	99.9	99.9

Notes: The table reports the cumulative distribution of diversification. The percentage of individual investors holding at least three shares in their portfolios on average during the 2-year sample period is 95.1% according to HHI_$Q(t)$, 90.0% according to DIVERSIFY_$Q(t)$, 92.4% according to DIVERSIFY_Q, and 82.9% according to DIVERSIFY. For the conversion of HHI_$Q(t)$ values to number of shares, an assumption of equal-weighted portfolio is made for simplicity. Variables definitions are in Appendix 3.

and includes investments not only in shares but also bonds and other financial assets including savings and checking accounts. TRADES reflects the investor's trading 'activity', a proxy for her propensity to speculate, and is computed as total transactions per quarter. GNP_CITY is the wealth, measured as gross national product (GNP) per capita of the city where the investor's stock trading account was opened.

Our second group of explanatory variables is linked to the *overconfidence* hypothesis H_{0B}. PROFESSION is a discrete binary variable taking value 1 if the investor has a finance-related job, and 0 otherwise, at the end of the sample period. GENDER equals 1 for males and 0 for females. MARRIED takes value 1 if the investor is married at the end of the sample period, and 0 otherwise. ORDER TYPE is another binary variable that represents the distribution channel: equal to 1 if the investor predominantly trades stocks through an investment center (branch), thus, having access to an expert's financial advice at the time of placing the order, and 0 if he predominantly opts for call center or Internet trading which precludes personal investment advice. VOLUME is a continuous variable defined as the total volume of shares traded in Turkish Lira (bought or sold) by the investor each quarter.

For completeness of our empirical analysis, in order to illustrate the importance of diversification on *performance*, we include two realized *profit or loss* variables. One is the investor's profit or loss (PROFIT–LOSS) associated with each sell transaction aggregated over quarters (or panel regressions) or over the entire sample period (for cross-section regressions). For this purpose, we utilize as purchase price the average inventory price at the time of each sell transaction, and as sell price the actual market price at the point the transaction was made. More specifically, the profit or loss per share sold in the jth transaction conducted at time t (for simplicity, the symbol t is omitted from the righthand side of the following equation and made synonymous with j) of the investor is computed as $\text{Sshares}(j) \times [\text{Sp}(j) - \text{INV}p(j)]$ where $\text{Sshares}(j)$ is the total number of shares sold in the jth transaction at market price per share $\text{Sp}(j)$ in Turkish Lira, and $\text{INVp}(j)$ is the average inventory price in Turkish Lira. Thus, the variable used is

$$\text{PROFIT-LOSS}(t) \equiv \frac{\sum_{j=1}^{J} \text{Sshares}(j) \times [\text{Sp}(j) - \text{INVp}(j)]}{\sum_{j=1}^{J} \text{Svolume}(j)}, \tag{5}$$

where $\text{Svolume}(j)$ denotes the total volume of the jth sell transaction in Turkish Lira and J is the total number of sell transactions carried out by the investor each quarter (for the panel regressions) or over the entire 2-year sample period (for the cross-section regressions).

Over time, the inventory price level is updated as follows. Let j now denote a point in time when shares for a given stock are bought, the average inventory price is updated then as

$$\text{INVp}(j) = \frac{\text{INVp}(j^*) \times \text{INVshares}(j^*) + \text{Bp}(j)\text{Bshares}(j)}{\text{INVshares}(j^*) + \text{Bshares}(j)}, \tag{6}$$

where $\text{INV}p(j^*)$ represents the average inventory price updated at the *time* of the previous purchase denoted j^*, that is, the inventory price before the current purchase of shares at time j; and $\text{INVshares}(j^*)$ is the previous inventory level. $\text{Bshares}(j)$ is the number of shares bought at time j and $\text{Bp}(j)$ is the market price of each share bought. The inventory price on day 0 is dictated by closing market prices on 31 March 2008 as purchase prices.

Our second profit or loss variable is binary (PROFIT-LOSS Dummy) and takes a value of 1, where on average over each quarter (panel regressions) or the entire sample (cross-section regressions) the investor incurred a net profit by trading shares and 0 otherwise. Transaction costs are taken into account in the PROFIT-LOSS(t) and PROFIT-LOSS Dummy (t) variables by applying a commission charge per transaction which varies with the distribution channel: 0.18% for Internet or call-center trade orders and 0.1% for investment center trade orders.[5]

Summary statistics on all the above explanatory variables are set out in Table 1. The mean (median) investor's age is about 40 (38) years. Most investors have college education (55%) with a clear minority (8%) having education level lower than high school. Only a very small proportion of investors in our sample have jobs in the finance sector (4%) or finance-related professions (6%). The mean (median) financial wealth of the investor is 87,259 (8818) Turkish Liras. The mean (median) volume of trade of the investors during the 2-year duration of the research period is 1,370,000 (37,259) Turkish Liras. The majority of investors are male (83%) or married (70%). The mean (median) number of trades over the 2-year sample period is 64 (13) and most investors trade through Internet or call centers (77%). In order to rule out collinearity issues, we examined the degree of linear dependence between the explanatory variables. Although most of the pairwise correlations, shown in Table 4, are statistically significant they are rather small economically. For instance, for the quantitative variables, the largest correlation is between total volume (VOLUME) and total number of trades (TRADES) at 0.21 which is very small.

3. Research methodology

We use both cross-section and panel regression analyses to uncover significant links between individual Turkish investor's portfolio diversification levels and their demographic/trading characteristics. Ultimately, our research goal is to test the *informational advantage* hypothesis (H_{0A}), behavioral *overconfidence bias* hypothesis (H_{0B}), and *information–overconfidence interaction* hypothesis (H_{0C}) motivated earlier in Section 1.

We consider a panel framework to model HHI_$Q(t)$ and DIVERSIFY_$Q(t)$, for which each observation is an investor-quarter pair. We consider a cross-sectional approach to model DIVERSIFY_Q and DIVERSIFY, for which each observation pertains to a different investor. The investor-quarter observations for the covariates VOLUME and TRADES are summed over quarters and WEALTH is time-averaged for the DIVERSIFY_Q and DIVERSIFY regressions. All explanatory variables (other than the dummies) are in logarithms.[6] Parameter estimation is by ordinary least squares. Inferences are based on White heteroskedasticity-consistent standard errors in the cross-section regressions, and on White-period standard errors in the panel regressions. The latter are robust to heteroskedasticity, and within-cluster (cross-section) and serial correlation (Wooldridge 2002, 148–153).

Our analysis extends previous studies by accommodating interaction effects among two types of variables, those that proxy *informational advantage* and those that proxy *overconfidence* which amounts to allowing for nonlinear effects. This methodology enables a test of the informational–overconfidence interaction hypothesis (H_{0C}). As an illustration, an interaction variable such as EDU_postgrad $\times$ PROFESSION can capture the following nonlinearity: the impact of postgraduate education (acting as proxy for information-processing ability) on diversification is no longer constant across investors but depends instead on the investor's profession. Likewise, the impact of a finance-related profession (acting as proxy for overconfidence) on diversification is moderated by the individual's education level.

Through ordinary and logit regressions we map diversification levels into performance. The dependent variable in these reduced-form empirical models is either PROFIT-LOSS(t) as defined in Equation (5), or its binary version PROFIT-LOSS Dummy. The independent variables are either HHI_$Q(t)$, DIVERSIFY_$Q(t)$, DIVERSIFY_Q, or DIVERSIFY so that the ordinary/logit

Table 4. Preliminary data analysis. Pearson correlations.

	AGE	MARRIED	EDU_High School	EDU College	EDU Postgrad	GENDER	SECTOR	PROFESSION	VOLUME	TRADES	ORDER TYPE	WEALTH	GNP_CITY
AGE	1												
MARRIED	0.3264	1											
	(0.000)												
EDU_High School	0.0222	0.0389	1										
	(0.000)	(0.000)											
EDU_College	−0.1317	−0.0784	−0.6760	1									
	(0.000)	(0.000)	(0.000)										
EDU_Postgrad	0.0316	−0.0030	−0.1961	−0.3489	1								
	(0.000)	(0.498)	(0.000)	(0.000)									
GENDER	−0.0668	0.0570	−0.0143	0.0107	0.0217	1							
	(0.000)	(0.000)	(0.001)	(0.015)	(0.000)								
SECTOR	0.0175	0.0329	−0.0493	0.0450	0.0474	−0.0707	1						
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)							
PROFESSION	−0.0927	−0.0529	−0.1009	0.1091	0.0328	−0.1078	0.3820	1					
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)						
VOLUME	0.0361	0.0179	0.0034	−0.0075	0.0022	0.0176	−0.0037	−0.0127	1				
	(0.000)	(0.000)	(0.445)	(0.087)	(0.613)	(0.000)	(0.398)	(0.004)					
TRADES	−0.0192	0.0201	0.0084	0.0080	−0.0177	0.0624	−0.0087	−0.0254	0.2059	1			
	(0.000)	(0.000)	(0.055)	(0.070)	(0.000)	(0.000)	(0.048)	(0.000)	(0.000)				
ORDER_TYPE	0.4454	0.1232	0.0618	−0.1282	−0.0040	−0.0705	−0.0449	−0.0818	0.1117	−0.0473	1		
	(0.000)	(0.000)	(0.000)	(0.000)	(0.357)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)			
WEALTH	0.1103	0.0264	−0.0124	0.0001	0.0255	0.0023	−0.0020	−0.0115	0.1697	0.0228	0.1496	1	
	(0.000)	(0.000)	(0.005)	(0.978)	(0.000)	(0.596)	(0.657)	(0.009)	(0.000)	(0.000)	(0.000)		
GNP_CITY	0.0006	−0.0457	−0.0005	−0.0062	0.0289	−0.0440	0.0298	0.0258	−0.0130	−0.0196	−0.0533	0.0102	1
	(0.896)	(0.000)	(0.907)	(0.156)	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)	(0.000)	(0.000)	(0.020)	

Notes: This table presents pairwise correlations for the explanatory variables in our study as defined in Appendix 3. Significance *p*-values in parenthesis.

regressions using the former two variables are panel type and exploit investor-quarter observations, whereas the ordinary/logit regressions for the latter two are cross-section type.

4. Empirical findings

The estimation and inference results for the linear panel and cross-section models are set out in Table 5, whereas those for the nonlinear counterpart models are reported in Table 6. For the most part, the sign and statistical significance of the coefficients reveal identical associations between investor's demographics/trading attributes and diversification levels, irrespective of the specification (and dependent variable) chosen for the analysis.

Beginning the discussion with Table 5, the first panel regression for HHI_$Q(t)$ shows that demographic and trading variables are able to explain about 12% of the overall variation in diversification levels, and the model is overall significant according to a standard F-test.

The coefficients of all the covariates acting as proxies for the information gathering and processing ability of individual investors are strongly significant at the 1% level, and have signs consistent with the *informational advantage* hypothesis (H_{0A}). Our findings are in line with those reported in Dorn and Sengmueller (2009) suggesting that the diversification level of Dutch investors increases significantly with age and education levels. Campbell (2006) documents that a minority of US retail investors appear to be uneducated and poorer and make significant investment mistakes. We also observe that investors who are employed in the finance sector have better diversified portfolios, and this may be because they have better access to financial information. Our results suggest also that wealthier investors hold better diversified portfolios in line with Goetzmann and Kumar (2008). Investors who live in wealthier cities with higher GNPs diversify better; the higher level of wealth in these cities could facilitate better economic/financial information gathering and processing opportunities. Investors with higher numbers of trades also diversify better, which contrasts with the previous findings of Odean (1998) and Barber and Odean (2001). Our evidence from the Turkish emerging market endorses the view that more active (i.e. frequent) traders have more information to act upon. Learning-by-trading mechanisms could make frequent traders more experienced and lead them to process information more efficiently and, in turn, to diversify better. Dorn and Huberman (2005) document that less-experienced German investors tend to churn poorly diversified portfolios. In our regression analysis, we differentiate between the two components of trading activity, number of trades and volume of trades. As discussed below, the sign of the coefficient on trade volume is consistent with the notion that on average the size of trading orders reveals the degree of overconfidence.

The coefficient estimates for all of our overconfidence proxies are also clearly significant at the 1% level and generally provide support for the behavioral *overconfidence bias* hypothesis (H_{0B}). Investors who are employed in finance-related jobs have poor portfolio diversification. Working in the finance industry can facilitate access to economic and financial information which improves portfolio diversification, but having a finance-related job increases overconfidence which reduces portfolio diversification *ceteris paribus*. Additionally, married investors seem to display poorer portfolio diversification which stands in contrast to Barber and Odean (2001) for US investors, but is in line with Grinblatt, Keloharju, and Linnainmaa (2011) for Finnish investors. Married Turkish investors exhibit higher overconfidence behavior. This could be due to the context of modernization in Turkey where autonomy and relatedness in the family context are compatible (Kagitcibasi 2005) and although family structures are nuclear, important members of the extended family typically reside nearby and maintain functional relationships (Georgas et al. 2001). Thus,

Table 5. Linear models of portfolio diversification.

	Dependent variable							
	Panel regressions				Cross-section regressions			
	HHI_$Q(t)$		DIVERSIFY_$Q(t)$		DIVERSIFY_Q		DIVERSIFY	
Regressor	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error
C	1.5790***	0.0465	−9.6135***	1.0768	−0.7247	0.5914	−4.2100***	0.3911
AGE	−0.0808***	0.0063	1.9072***	0.1934	0.7082***	0.1358	1.3276***	0.0836
MARRIED	0.0106***	0.0031	−0.3168***	0.0501	−0.1478***	0.0423	−0.1630***	0.0344
EDU_High School	−0.0148***	0.0050	0.2640***	0.0921	0.0509	0.0734	0.0292	0.0519
EDU_College	−0.0338***	0.0047	0.2288***	0.0613	0.0509	0.0587	0.0482	0.0467
EDU_Postgrad	−0.0611***	0.0062	0.6242***	0.1132	0.1853**	0.0904	0.2270***	0.0704
GENDER	0.0020	0.0036	0.1447**	0.0729	0.0419	0.0564	−0.0522	0.0340
SECTOR	−0.0232***	0.0069	1.1039***	0.3511	0.2438	0.2363	0.2477**	0.1098
PROFESSION	0.0242***	0.0058	0.3709	0.2698	0.0985	0.1864	−0.0307	0.0808
VOLUME	0.0421***	0.0009	−0.5497***	0.0436	−0.4713***	0.0556	−0.4897***	0.0222
TRADES	−0.0835***	0.0014	1.2487***	0.0846	0.8687***	0.0927	1.0127***	0.0400
ORDER_TYPE	0.0230***	0.0043	−0.1119**	0.0542	0.0646	0.0437	0.1709***	0.0333
WEALTH	−0.0381***	0.0031	0.4619***	0.0471	0.2056***	0.0266	0.3941***	0.0138
GNP_CITY	−0.0229***	0.0039	0.2081***	0.0521	0.0178	0.0411	0.0632**	0.0307
Adjusted R^2 (%)	11.86		5.22		6.15		12.56	
F statistic	381.07		176.80		110.46		553.18	
p-value (F stat)	0.00		0.00		0.00		0.00	
Obs.	36,713		41,516		21,725		49,998	

Notes: This table reports the estimates, standard errors, and diagnostics for panel and cross-section regressions of diversification on informational advantage and overconfidence proxies. The underlying cross-section observations are 59,951 clients with equity portfolio holdings who traded at least once during the sample period from 31 March 2008 to 28 February 2010. The panels for the first two regressions are unbalanced since there are missing observations on the (in)dependent variables in some quarters. Variable definitions are as in Appendix 3. White-period standard errors which are robust to heteroskedasticity and within-cluster and serial correlation are reported for the panel regressions. White heteroskedasticity-consistent standard errors are reported for the cross-section regressions. All explanatory variables (other than the dummies) are in logarithms.

***Significant at the 1% level.

**Significant at the 5% level.

*Significant at the 10% level.

Table 6. Nonlinear models of portfolio diversification.

	Dependent variable							
	Panel regressions				Cross-section regressions			
	HHI_$Q(t)$		DIVERSIFY_$Q(t)$		DIVERSIFY_Q		DIVERSIFY	
Regressor	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error
C	2.1797***	0.0554	−17.5543***	1.4370	−6.1005***	1.0463	−10.4025***	0.5787
AGE	−0.0785***	0.0248	0.7704**	0.3579	0.6867**	0.2922	1.0470***	0.1857
EDU_Postgrad	−0.0116	0.0216	0.3881	0.7505	0.3201	0.5377	0.0658	0.3130
SECTOR	−0.0159	0.0297	2.4326**	1.1238	0.1734	0.6680	0.2298	0.2926
WEALTH	−0.0586***	0.0120	1.2679***	0.1816	0.1026**	0.0443	0.2024***	0.0277
GNP_CITY	−0.0601***	0.0139	0.4618**	0.1840	0.6587***	0.1923	1.0316***	0.1156
TRADES	−0.1532***	0.0046	2.0411***	0.1837	1.4561***	0.1614	1.7195***	0.0975
AGE*PROFESSION	−0.0074	0.0260	5.8336**	2.2928	1.2310	1.3249	0.4261	0.5030
EDU_Postgrad*PROFESSION	0.0262*	0.0151	−1.6509***	0.4285	−0.4840*	0.2700	−0.2567	0.1648
SECTOR*PROFESSION	0.0079	0.0147	2.1265**	0.8646	0.6864	0.6407	0.2394	0.2615
WEALTH*PROFESSION	0.0056	0.0103	−1.8712***	0.7188	−0.0628	0.0695	−0.0625*	0.0340
GNP_CITY*PROFESSION	−0.0024	0.0126	−0.2002	0.2504	−0.5795	0.6479	−0.1476	0.2434
TRADES*PROFESSION	−0.0020	0.0042	1.2691**	0.5346	0.2895	0.3783	0.0674	0.1605
AGE*GENDER	−0.0187	0.0148	1.1607***	0.2592	0.2791	0.2305	0.1839	0.1354
EDU_Postgrad*GENDER	−0.0239*	0.0130	0.0771	0.2521	0.0562	0.2004	0.0246	0.1425
SECTOR*GENDER	−0.0145	0.0144	2.2090***	0.7015	0.6078	0.5306	0.0739	0.2076
WEALTH*GENDER	0.0024	0.0063	−0.4765***	0.1093	−0.0067	0.0241	−0.0164	0.0160
GNP_CITY*GENDER	0.0060	0.0081	0.1088	0.0987	−0.1412	0.1073	−0.0852	0.0651
TRADES*GENDER	−0.0022	0.0025	0.1685**	0.0701	0.0553	0.0527	0.0302	0.0328
AGE*MARRIED	−0.0271**	0.0130	0.6089**	0.2380	0.1056	0.1954	0.3932***	0.1270
EDU_Postgrad*MARRIED	−0.0033	0.0097	−0.4227	0.2585	−0.2315	0.2120	0.0125	0.1450
SECTOR*MARRIED	0.0018	0.0154	0.3789	0.4719	0.2821	0.3267	−0.0371	0.2041
WEALTH*MARRIED	−0.0123**	0.0051	−0.0955	0.0863	0.0330	0.0304	0.0250	0.0159

(*Continued*)

Table 6. Continued

	Dependent variable							
	Panel regressions				Cross-section regressions			
	HHI_$Q(t)$		DIVERSIFY_$Q(t)$		DIVERSIFY_Q		DIVERSIFY	
Regressor	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error
GNP_CITY*MARRIED	0.0301***	0.0062	−0.1843**	0.0885	−0.0833***	0.0102	−0.2095***	0.0610
TRADES*MARRIED	0.0028	0.0020	0.0493	0.0719	−0.0370	0.0473	−0.0328	0.0325
AGE*VOLUME	0.0050*	0.0022	−0.0816*	0.0470	−0.0311	0.0239	0.0010	0.0172
EDU_Postgrad*VOLUME	−0.0004	0.0019	0.0413	0.0713	−0.0010	0.0438	0.0142	0.0244
SECTOR*VOLUME	−0.0004	0.0027	−0.4580**	0.1993	−0.0782	0.1159	0.0343	0.0417
WEALTH*VOLUME	−0.0001	0.0009	0.0169	0.0184	0.0115***	0.0035	0.0216***	0.0024
GNP_CITY*VOLUME	0.0019	0.0012	−0.0349*	0.0183	−0.0426***	0.0106	−0.0733***	0.0086
TRADES*VOLUME	0.0067***	0.0004	−0.0978***	0.0178	−0.0506***	0.0104	−0.0577***	0.0064
AGE*ORDER_TYPE	−0.0258*	0.0154	0.1449	0.2010	−0.1559	0.1314	−0.4086***	0.1166
EDU_Postgrad*ORDER_TYPE	0.0486***	0.0128	−0.6380***	0.1881	−0.1897	0.1186	−0.2534**	0.1236
SECTOR*ORDER_TYPE	0.0382**	0.0192	−0.7139***	0.2766	−0.2726	0.1832	0.1790	0.1804
WEALTH*ORDER_TYPE	0.0353***	0.0056	−0.5004***	0.0876	−0.0117	0.0185	−0.0065	0.0171
GNP_CITY*ORDER_TYPE	−0.0391***	0.0083	0.7270***	0.1140	0.1563*	0.0595	0.2381***	0.0568
TRADES*ORDER_TYPE	0.0045*	0.0026	−0.2144***	0.0363	−0.1324***	0.0228	−0.0329	0.0236
Adjusted R^2 (%)	13.11		7.03		7.05		13.34	
F statistic	154.91		88.25		46.74		214.96	
p-value (F stat)	0.00		0.00		0.00		0.00	
Obs.	36,713		41,516		21,725		49,998	

Notes: The table reports estimates, standard errors, and diagnostics of panel and cross-section regressions to explain diversification on the basis of informational advantage proxies and their interactions with overconfidence proxies. See note to Table 5.

married Turkish investors have additional support from their extended family should they need it to cover investment losses.

Investors who place larger volume orders on average have poorer portfolio diversification endorsing the view that trading volume acts as proxy for overconfidence (Benos 1998). Thus, the size of the investor's transaction orders, rather than the frequency of trading, reveals the degree of overconfidence. Hoffmann, Shefrin, and Pennings (2010) note that investors with speculative instincts have higher turnover. Previous research for well-developed stock markets has shown that Internet-trading individual investors are more overconfident (Barber and Odean 2002; Glaser and Weber 2007; Graham, Campbell, and Huang 2009; Grinblatt and Keloharju 2009), and trade excessively compared with others. After controlling for both trading frequency and volume, we find that those Turkish investors who use investment centers as distribution channel (ORDER_TYPE) tend to be less diversified. A rationale for this finding is that, by receiving personal financial advice at the time of placing the trade order, such investors feel reassured and more satisfied with their actions which may lead to overconfidence and worse diversified portfolios. On the other hand, call center or Internet investors may perceive themselves as knowledgeable enough so as not to necessitate such financial advice. Hoffmann and Shefrin (2011), and Dorn and Huberman (2005) show that investors who *think* of themselves knowledgeable about financial securities hold better diversified portfolios.

Similar findings are obtained when the dependent variable in the panel regression model is DIVERSIFY_$Q(t)$. All the information variables – investor's age, college and postgraduate education level, job sector, number of trades, wealth, and GNP of the city – are strongly significant at the 1% level or better. Among the overconfidence variables, marital status and volume remain strongly significant at the 1% level. However, gender is now revealed as statistically significant in this model at the 5% level. This result may indicate that Turkish female investors tend to be less diversified which may reflect overconfidence relative to their male peers; this result will be corroborated later in the interaction terms of nonlinear models. In this respect, our evidence from the Turkish market also contrasts with previous studies where men are reported to be more overconfident than women (Barber and Odean 2001). However, Dorn and Huberman (2005) report less-significant findings on risk attitude for gender and age than for other investors' demographic characteristics. These findings on gender could be intrinsic to the emerging market under study. In fact, the republican culture in Turkey largely promotes women rights as a symbol of westernization and modernity, which leads to gender roles redefinition (Toktas and Cindoglu 2006) and the use of empowerment and resistance strategies by women (Cindoglu and Toktas 2002).

We now turn to the cross-section regressions. If the dependent variable is DIVERSIFY_*Q*, three information variables lose significance: job sector, profession, and distribution channel. Older, better educated, wealthier individuals, and investors who trade more frequently hold on average a larger number of different stocks suggesting a positive link between investors' informational advantage and their portfolio diversification level. Married investors and those who place large orders have poorer portfolio diversification, possibly because those variables act as proxies for Turkish investors' overconfidence as discussed above. The model DIVERSIFY confirms most of our previous findings. In general, diversification is positively linked with proxies for information-processing ability and negatively linked with overconfidence proxies. Among the latter, only gender and profession have a negligible effect on diversification. Among the education proxies for informational advantage, EDU_Postgrad is the most significant, both economically and statistically, revealing that Turkish investors with postgraduate education tend to hold better diversified equity portfolios *ceteris paribus*. The influence of distribution channel (ORDER_TYPE) is significant only in one of the two cross-section models, for DIVERSIFY, but the sign is at odds with that shown earlier by the panel regressions that exploit both the cross-section and time-series variation of the data.

The linear models thus far considered impose constant coefficients and hence implicitly assume that the effect of the different covariates is identical across investors. To illustrate, they do not accommodate the possibility that the strength of the association between trading volume (overconfidence proxy) on diversification varies among investors depending, for instance, on their education level. To make the modeling framework more general and for a better understanding of the link between *information* and *overconfidence* variables, we allow for non-constant effects in a relatively parsimonious model. Table 6 reports estimates and diagnostics for the nonlinear panel/cross-section regressions. The regressor set contains the informational advantage proxies and their interactions with the overconfidence proxies.

A comparison between Tables 5 and 6 suggests that the explanatory power of the model for HHI_$Q(t)$ increases from 11.86% (linear) to 13.11% (nonlinear). Various *informational advantage* proxies such as AGE, GNP_CITY, and TRADES remain strongly significant and exert a favorable influence on diversification but EDU_Postgrad and SECTOR are no longer significant. However, the latter two variables still influence investor's diversification through their moderating effects on overconfidence as revealed through various interaction terms.[7]

A clear-cut finding is that the strength of the link between diversification and frequency of trading appears notably different from investor to investor according to trading volume: high information-processing ability (stemming from the learning experience acquired by frequent trading) enhances diversification but less so for overconfident investors (as signaled by high-trading volume). Another consistent result across models is the significantly positive coefficient of the interaction between marital status and GNP suggesting that the favorable city wealth effect on diversification (informational advantage) is lessened for married people who may be more overconfident. The sign of the interactions with the distribution channel (ORDER_TYPE) suggests, essentially, that the positive effect of informational advantage proxies on diversification is reduced for investors who place their orders through investment centers. In particular, if investors trade through investment centers that offer financial advice (which possibly boosts confidence in investor's actions), there is a lessening of the positive effect on diversification derived from having post-graduate education, a finance sector job, being wealthy and trading frequently. One clear exception to this finding arises from the interaction between order type and city GNP: the positive impact of living in a wealthy city on diversification becomes stronger when investors choose investment centers to execute trading orders. Although frequent traders (information-processing ability) are better diversified, this positive link weakens for high-volume (overconfident) investors. Finally, there is evidence from the panel model for DIVERSIFY_$Q(t)$ that the beneficial influence of informational advantage (AGE, SECTOR, and TRADES) on diversification is, if anything, lower for female investors possibly because they are less-risk averse than their male peers.

To sum up, the linear (constant elasticity) models presented earlier in Table 5 show that better information-processing ability improves portfolio diversification *ceteris paribus*, and overconfidence reduces portfolio diversification *ceteris paribus*. The subsequent nonlinear (interaction) models presented in Table 6 reveal that the negative effect of overconfidence on portfolio diversification plausibly varies from investor to investor according to their capacity to gather and process economic/financial information. For instance, as investors get older, the adverse impact of overconfidence (signaled, for instance, by volume of trades) on portfolio diversification is mitigated. Similarly, the favorable effect of higher education (information-processing ability) on diversification is lessened by the overconfidence bias associated, for instance, with being finance professional or with trading through investment centers.

Finally, for completeness, we map the degree of portfolio diversification into trading profitability. For this purpose, we estimate ordinary regression models where the dependent variable is PROFIT-LOSS(t), measured as in Equation (5), and logit regression models for PROFIT-LOSS Dummy (t); see variable definitions in Appendix 3. Diversification is our conditioning variable throughout.[8] All the variables are in level form. Table 7 reports the results of both types of models, ordinary and logit, estimated by ordinary least squares and maximum likelihood, respectively. In the ordinary regressions, the marginal effect of diversification on trading profits (losses) is provided by the slope, whereas in the logit model it is a nonlinear function of diversification. The marginal effect reported for the logit regressions is evaluated at the mean of each of the explanatory (diversification) variables. For the panel regressions, the observations for the covariate PROFIT-LOSS(t) are quarterly aggregates for each investor; for the cross-section regressions, they are total time aggregates. As suggested by the summary statistics reported in Table 1, about 55% of the 59.951 clients in our sample experience a net trading profit over the 2-year sample period.

Irrespective of the diversification measure considered, the results of the top exhibit (ordinary panel/cross-section regressions) confirm that better diversified investors tend to earn higher profits on average. Considering the first ordinary (panel) regression for HHI_$Q(t)$, the significantly negative coefficient of HHI_$Q(t)$ suggests that a decrease in portfolio concentration, HHI_$Q(t)$, by 1 unit entails an increase in profits of 2.7 percentage points. Take the average investor with an average HHI_$Q(t)$ of 0.86, which amounts roughly, under the assumption of identical weight allocation to different stocks in the portfolio, to holding 1.2 shares. If the average portfolio diversification increases by one more share (to 2.2 shares) corresponding to an average HHI_$Q(t)$ of 0.45, then her average profits will increase by 1.14%. Similarly, the significantly positive slope coefficients of DIVERSIFY_Q and DIVERSIFY indicate that better diversification increases profits. The average investor holds 1.73 shares in her equity portfolio according to the DIVERSIFY_Q measure as shown in Table 1. Our analysis suggests that a decision to hold one more share (i.e. 2.73 shares on average) will increase profitability by 0.12 percentage points. For the average investor that earns 2% on average (Table 1), this corresponds to a non-negligible 6% increase in trading profits. Our DIVERSIFY measure indicates that investors hold 2.3 shares on average and, according to the estimates in Table 7, by adding one more share to their portfolios (i.e. 3.3 shares on average) can boost profits by 0.37 percentage points representing a 18% increase.

Similar conclusions can be gleaned from the logit regressions. If portfolio concentration, as measured by HHI_$Q(t)$, increases by one unit (i.e. less diversification) then the probability of earning a profit falls by about 11.5%. If the investor's average level of diversification as suggested by the HHI_$Q(t)$ measure increases from 1.2 to 2.2 shares (or roughly under an equal weights assumption, HHI_$Q(t)$ decreases from 0.86 to 0.45), the probability of earning a profit increases by 4.7% (i.e. 0.41*0.1153), which is a significant change both statistically and economically. According to the DIVERSIFY_$Q(t)$ measure (see Table 1), the average Turkish investor holds 2.06 shares and can increase the probability of making a profit by 0.5% by holding one more share. For the DIVERSIFY_Q and DIVERSIFY measures, increasing the average portfolio size by one more share increases the probability of making trading profits by 0.8% and 2.2%, respectively. These are all economically significant increases in the probability of realizing profits by improving diversification levels.

Table 7. Stock trading profitability and portfolio diversification.

	Explanatory variable							
	Panel regressions				Cross-section regressions			
	HHI_$Q(t)$		DIVERSIFY_$Q(t)$		DIVERSIFY_Q		DIVERSIFY	
	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error
I. Ordinary regression (dependent variable: PROFIT-LOSS)								
Intercept	0.0256***	0.0068	0.0033	0.0055	0.0296***	0.0017	0.0107***	0.0017
Slope	−0.02770**	0.0124	−0.00061	0.0012	0.00122***	0.0003	0.0037***	0.0006
Adjusted R^2 (%)	0.00		0.00		0.03		0.18	
F statistic	2.17		0.54		5.90		95.75	
p-value (F stat)	0.14		0.46		0.02		0.00	
Obs.	29,004		33,102		21,796		52,075	
II. Binary logit regression (dependent variable: PROFIT-LOSS dummy)								
Intercept	−0.4023***	0.0305	−0.9711**	0.0095	0.4263	0.0182	0.1589***	0.0129
Slope	−0.5567***	0.0345	0.02350**	0.0024	0.03280***	0.0074	0.0899***	0.0046
Marginal effect	−0.1153		0.0048		0.0077		0.0218	
McFadden R^2 (%)	0.33		0.14		0.08		0.01	
LR statistic	255.65		125.97		24.78		487.41	
p-value (LR stat)	0.00		0.00		0.00		0.00	
Obs.	63,682		74,824		23,345		56,263	

Notes: The table reports estimates and standard errors of ordinary and logit regressions estimated by ordinary least squares and maximum likelihood, respectively. The variables are defined in Appendix 3. The marginal effect in the logit regression is calculated at the mean of the diversification variable. White heteroskedasticity-consistent standard errors are reported. All variables are in levels. The underlying cross-section observations are 59,951 clients with equity portfolio holdings who traded at least once during the sample period from 31 March 2008 to 28 February 2010.

***Significant at the 1% level.

**Significant at the 5% level.

*Significant at the 10% level.

5. Conclusions

This study sheds light on how individual investor's informational advantage and overconfidence attributes affect portfolio diversification levels. For this purpose, we rely on objective investor traits by using a unique dataset from one of the major brokerage houses in Turkey. Our measures of diversification suggest that individual Turkish investors hold about two shares in their equity portfolios on average. Although this is far below the 10–30 shares suggested in the literature for a well-diversified portfolio, it is relatively close to the averages documented for other European markets such as Finland and Germany.

Portfolio diversification is a simple and costless strategy of risk reduction. To the best of our knowledge, this is the first paper that seeks to map a large number of individual investor's *informational advantage* proxies and *overconfidence* bias proxies on diversification levels. Our second contribution is that this is carried out in a modeling framework that allows for their impact to be non-identical (or non-linear) across investors. Moreover, our entire analysis relies on objective investor traits instead of self-reported attributes from survey data.

Our findings support both the *informational* advantage hypothesis and the *overconfidence bias* hypothesis by suggesting that better informed investors diversify better, whereas overconfident investors diversify less. Better educated, older and higher income individuals who have jobs in the financial sector, and those who live in richer cities may possess better information gathering and processing ability and, at the same time, are better diversified. Overconfidence that can stem from being finance professional and trading through investment centers (branches) that facilitate personal financial advice reduces diversification. Our analysis further controls for two aspects of trading activity. Individuals who trade very frequently, which is likely due to being able to process relevant information efficiently, show better diversified portfolios *ceteris paribus*. In contrast, those who engage in high-volume trades, possibly revealing overconfidence, tend to be less diversified *ceteris paribus*. The evidence stemming from our dataset challenges previous studies suggesting that women and married investors are less confident than men and single investors and, in turn, more risk averse and better diversified. Our results do not reveal a strong association between diversification and gender. Moreover, both panel and cross-section analyses reveal that married investors tend to be less diversified in line with recent evidence from Finnish data.

Our modeling framework allows us also to demonstrate empirically that the strength of the link between investor's diversification and demographics/trading attributes is not constant. Instead, the positive association between informational advantage and diversification is moderated by the influence of overconfidence. Our findings suggest, for instance, that more frequent traders tend to diversify better *ceteris paribus* but less so if they are also high-volume traders. Finally, we show that poor diversification is costly. Adding one more share to an individual investor's portfolio can materialize into profitability increases of up to 18% on average over a 2-year trading period and the likelihood of realizing net profits rises by up to 2.2%. Low diversification levels thus materialize into foregone profit opportunities.

In further research, it would be interesting to see whether the evidence presented here is intrinsic to the Turkish market or extends also to other world markets. Although there is considerable work on the effects of biases such as overconfidence on trading activity, worldwide evidence on portfolio diversification is still scant. An obvious drawback of our analysis is that it does not rely on directly observed psychological traits or emotional states. Although we define overconfidence using several proxies, it is difficult to assess the relative importance of other psychological traits with the data at hand.

Acknowledgements

The authors are grateful to Garanti Bank and Ismail Bayarslan for providing the data and to Cass Pump-Priming and Finans Yatırım Menkul Değerler AŞ for financially supporting this research. We also acknowledge comments from Warren Bailey, Jerry Coakley, Richard Fairchild, and participants at the 2012 *Behavioural Finance Working Group* Conference at Cass Business School, the 2011 International Conference on *Computational and Financial Econometrics* (CFE'11) at Birkbeck College, University of London, the 2011 Annual Meeting of the *Academy of Behavioral Finance and Economics* at UCLA, and Bilkent University Finance Seminar in May 2011.

Notes

1. For instance, an investor working as human resources director in a bank is regarded as having a job in the finance industry but as having a non finance-related job. An investor working as an accountant in a bank is categorized as having both a job in the finance industry and a finance-related job.
2. The clients can only have one account with one broker according to requirements by the Capital Markets Board of Turkey. For further details, see legal information at http://www.cmb.gov.tr
3. See TSPAKB December 2009 report at http://www.tspakb.org.tr
4. We do not account for equity investment through mutual funds because of data unavailability. Nevertheless, the bias that this could introduce should be small given that Turkish investors' mutual portfolio holdings represent only 3% of total financial investable assets and only 4% of these funds are allocated to equities. See http://www.tspakb.org.tr and http://www.spk.gov.tr
5. Cost reasons induce banks to channel low trading-volume (small portfolio) clients to call centers and Internet. For the larger portfolio clients who are serviced through branches, the market is very competitive in Turkey and that may be the reason why banks seek to avoid charging high commission to clients at the investment centers.
6. The diversification (dependent) variables are entered in levels in the subsequent panel and cross-section regressions and hence, strictly speaking, the slope coefficients do not have the interpretation of elasticities.
7. We do not report the estimation results for more elaborate (less parsimonious) models that incorporate *all* the information and overconfidence variables and their interactions. The findings are fairly similar. In general, the effect of overconfidence is most clearly revealed through interactions and hence, most of the overconfidence proxies appear significant when interacted with information variables but less so individually.
8. Other factors over and above diversification can influence investor's P&L. Hence, the slope coefficients may suffer from omitted variable bias. Notwithstanding this caveat, we proceed to interpret the results.
9. http://www.tspakb.org.tr/tr/DesktopDefault.aspx?tabid=151 (*Monthly Bulletin*, November 2010).
10. http://www.ise.org/Data/Consolidated.aspx. There are seven companies which are temporarily delisted by ISE Board of Directors.
11. A total of 45,214 companies are listed and the total market capitalization is 50,200 billion USD on the World Exchanges. There is 47,127 billion USD trading in World markets. Major exchanges in the US, NASDAQ, and NYSE in total had 15 trillion USD, UK 2.8 trillion USD, Germany 1.3 trillion USD, and Greece 113 billion USD market capitalization as of 2009. See January–September 2010, World Federation of Stock Exchanges. http://www.world-exchanges.org/statistics/key-market-figures and http://www.world-exchanges.org/statistics/annual/2009/equity-markets/domestic-market-capitalization
12. http://www.world-exchanges.org/statistics/annual/2009/equity-markets/turnover-velocity-domestic-shares
13. http://www.turkstat.gov.tr/PreTablo.do?tb_id=39&ust_id=11
14. http://www.tspakb.org.tr/tr/DesktopDefault.aspx?tabid=151 (*Monthly Bulletin*, November 2010).

References

Barber, B.M., and T. Odean. 2000. Trading is Hazardous to your wealth: The common stock investment performance of individual investors. *Journal of Finance* 40, no. 2: 773–806.

Barber, B.M., and T. Odean. 2001. Boys will be boys: Gender, overconfidence, and common stock investment. *Quarterly Journal of Economics* 116, no. 1: 261–92.

Barber, B.M., and T. Odean. 2002. Online investors: Do the slow die first? *Review of Financial Studies* 15, no. 2: 455–87.

Benartzi, S., and R. Thaler. 2007. Heuristics and biases in retirement savings behavior. *Journal of Economic Perspectives* 21, no. 3: 81–104.

Benos, A.V. 1998. Aggressiveness and survival of overconfident traders. *Journal of Financial Markets*, 1, nos. 3–4: 353–83.
Bertrand, M., and S. Mullainathan. 2001. Do people mean what they say? Implications for subjective survey data. *American Economic Review* 91, no. 2: 67–72.
Blume, M.E., and I. Friend. 1975. The asset structure of individual portfolios and some implications for utility functions. *Journal of Finance* 30, no. 2: 585–603.
Campbell, J.Y. 2006. Household finance. *Journal of Finance* 61, no. 4: 1553–604.
Cindoglu, D., and S. Toktas. 2002. Empowerment and resistance strategies of working women in Turkey. *European Journal of Women's Studies* 9, no. 1: 31–48.
DeMiguel, V., L. Garlappi, and R. Uppal. 2009. Optimal versus Naive diversification: How inefficient is the 1/*N* portfolio strategy? *Review of Financial Studies* 22, no. 5: 1915–53.
Dhar, R., and N. Zhu. 2006. Up close and personal: Investor sophistication and the disposition effect. *Management Science* 52, no. 5: 726–40.
Dorn, D., and G. Huberman. 2005. Talk and action: What individual investors say and what they do. *Review of Finance* 9, no. 4: 437–81.
Dorn, D., and P. Sengmueller. 2009. Trading as entertainment? *Management Science* 55, no. 4: 591–603.
Dorn, D., G. Huberman, and P. Sengmueller. 2008. Correlated trading and returns. *Journal of Finance* 63, no. 2: 885–920.
Evans, J.L., and S.H. Archer. 1968. Diversification and the reduction of dispersion: An empirical analysis. *Journal of Finance* 23, no. 5: 761–7.
Georgas, J., K. Mylonas, S. Bafiti, Y.H. Poortinga, S. Christakopoulou, C. Kagitcibasi, K. Kwak et al. 2001. Functional relationships in the nuclear and extended family: A 16-culture study. *International Journal of Psychology* 36, no. 5: 289–300.
Glaser, M., and M. Weber. 2007. Overconfidence and trading volume. *Geneva Risk and Insurance Review* 32, no. 1: 1–36.
Glaser, M., and M. Weber. 2009. Which past returns affect trading volume? *Journal of Financial Markets* 12, no. 1: 1–31.
Goetzmann, W.N., and A. Kumar. 2008. Equity portfolio diversification. *Review of Finance* 12, no. 3: 433–63.
Graham, J.R., R.H. Campbell, and H. Huang. 2009. Investor competence, trading frequency, and home bias. *Management Science* 55, no. 7: 1094–106.
Grinblatt, M., and M. Keloharju. 2009. Sensation seeking, overconfidence, and trading activity. *Journal of Finance* 64, no. 2: 549–78.
Grinblatt, M., M. Keloharju, and J.T. Linnainmaa. 2011. IQ and stock market participation. *Journal of Finance* 66, no. 6: 2121–64.
Hoffmann, A.O.I., and H.M. Shefrin. 2011. *Online investors: What they want, what they do and how their portfolios perform*. Mimeo: Maastricht University School of Business and Economics.
Hoffmann, A.O.I., H.M. Shefrin, and J.M.E. Pennings. 2010. Behavioral portfolio analysis of individual investors. SSRN working paper no. 1629786 (unpublished).
Kagitcibasi, C. 2005. Autonomy and relatedness in cultural context implications for self and family. *Journal of Cross-Cultural Psychology* 36, no. 4: 403–22.
Markowitz, H.M. 1952. Portfolio selection. *Journal of Finance* 7, no. 1: 77–91.
Odean, T. 1998. Volume, volatility, price, and profit when all traders are above average. *Journal of Finance* 53, no. 6: 1887–934.
Statman, M. 1987. How many stocks make a diversified portfolio? *Journal of Financial and Quantitative Analysis* 22, no. 3: 353–63.
Toktas, S., and D. Cindoglu. 2006. Modernisation and gender: A history of girls in technical education in Turkey since 1927. *Women's History Review* 15, no. 5: 737–49.
Vissing-Jorgensen, A. 2004. Perspectives on behavioral finance: Does irrationality disappear with wealth? Evidence from expectations and actions. In *NBER Macroeconomics Annual*, ed. M. Gertler and K. Rogoff. vol 18, 140–208. The MIT Press. http://www.nber.org/books/gert04-1.
Weber, E.U., A.-R. Blais, and N.E. Betz. 2002. A domain-specific risk-attitude scale: Measuring risk perceptions and risk behavior. *Journal of Behavioral Decision Making* 15, no. 4: 263–90.
Wooldridge, J.M. 2002. *Econometric analysis of cross section and panel data*. Cambridge, MA: The MIT Press.

Appendix 1. A brief anatomy of the Turkish stock market

The ISE was established in 1986 and has expanded over the years reaching today a total of 343 listed shares with a free float rate of 32%.[9] Transactions take place in an electronic trading environment. Daily average trading volume is 1.6 billion USD and yearly total volume of trade is 341 billion USD. Its total market capitalization is 336.5 billion USD as of October 2010.[10] This makes the Turkish stock market the seventh largest market in Europe in terms of market capitalization.[11]

Turkey is peculiar with respect to its trading volume. There are not many markets in the world which can beat Turkey (165.6% in 2009) in terms of turnover velocity besides China, and Taiwan.[12] These statistics are astonishing if the free float rate is taken into consideration.

The population of Turkey is 72.5[13] million so only about 2% of the population has accounts in the 144 brokerage houses that operate in Turkey. In 2010, there are a total of 1,027,732 investors representing a 2.7% increase from last year.[14] Out of these investors, 1,104,071 (99%) are individual investors and 11,243 are institutional investors (0.3%). As of October 2010, foreign investors own 67.3% of market capitalization and the remaining 32.7% is owned by domestic investors. In terms of trading activity, 67.3% of the transactions are made by domestic individuals, 18.4% by domestic institutions and 14.3% by the foreign investors. Most of the individual investors are small investors where top 10,000 stock market investors represent 67.5% of the local investors with respect to total portfolio size.

Appendix 2. Investors with portfolio holdings at quarter-end snapshots

The total number of investors in the sample is 59,951. However, not all of them hold a portfolio of stocks during the entire observation period from 31 March 2008 to 26 February 2010. The table below reports the number of clients in the database that are holding a portfolio of stocks at specific quarter-end snapshots. The last row reports the number of clients holding a portfolio at the end of at least one of the eight quarter-end snapshots.

Date	Investors
31.03.2008	14,605
30.06.2008	15,228
30.09.2008	15,389
31.12.2008	16,415
31.03.2009	17,370
30.06.2009	15,154
30.09.2009	15,303
31.12.2009	15,093
Any quarter-end	29,649

Appendix 3. Variable names and definitions

Date	Investors
31.03.2008	14,605
30.06.2008	15,228
30.09.2008	15,389
31.12.2008	16,415
31.03.2009	17,370
30.06.2009	15,154
30.09.2009	15,303
31.12.2009	15,093
Any quarter-end	29,649

A.1 *Dependent variable*

A.1.1 Diversification measures

HHI_$Q(t)$	Portfolio concentration measured as the normalized sum of squared portfolio allocation weights of the different stocks held at the end of each quarter t
DIVERSIFY_$Q(t)$	Number of different shares in the investor's portfolio at the end of each quarter t
DIVERSIFY_Q	Diversification measured as number of different stocks held at the end of each quarter averaged over the quarters when a portfolio is held
DIVERSIFY	Diversification measured as number of different stocks in the investor's portfolio weighted by the number of days the portfolio is held

A.2 *Independent variables*

A.2.1 Information gathering and processing ability proxies

AGE	Age of investor at the end of the sample period
EDU_High School	Dummy. High School Education of investor: 1 if highest level of education at the end of sample period is High School, 0 else
EDU_College	Dummy. College education of investor: 1 if up to university degree or 2-year further education after High School, 0 else
EDU_Postgrad	Dummy. Post-graduate education of investor: 1 if up to graduate, masters or PhD degree, 0 else
SECTOR	Dummy. Investor's job sector at the end of the sample period: 1 if finance and 0 non-finance 0 (total: 51 sectors in the sample)
WEALTH(t)	Investors's financial wealth: asset values in Turkish Lira reported by the brokerage bank as end-of-quarter values for panel regressions or total average over the 2-year sample period for cross-section regressions
TRADES(t)	Trade frequency: total number of transactions (buys or sells) per quarter for panel regressions or over the entire 2-year sample period for cross-section regressions
GNP_CITY	City wealth: GNP per capita in Turkish Lira of city where account is opened

A.2.2 Overconfidence proxies

PROFESSION	Dummy. Investor's profession at the end of the sample period: 1 if finance and 0 if non-finance (total: 435 professions in the sample)
GENDER	Dummy. Gender: 1 if male and 0 if female
MARRIED	Dummy. Marital status of investor at the end of sample period: 1 married, 0 else
ORDER_TYPE	Dummy. Distribution channel: 1 if investment center, 0 if call center or Internet
VOLUME(t)	Total volume of shares traded (bought/sold) in Turkish Lira per quarter for panel regressions or over the 2-year sample period for cross-section regressions

A.3 *Profit and loss variables*

PROFIT_LOSS(t)	Profit or loss from each sale over total sale volume, on aggregate over each quarter for panel regressions or over the 2-year sample period for cross-section regressions
PROFIT_LOSS Dummy(t)	Binary variable equal to 1 if the client made a net profit from Sale transactions in a given quarter (or over the two-year sample period), and 0 else