P. JOAKIM WESTERHOLM and MARTIN OLLILA

Impact of Gender, Age and Language on Investment Strategy[1]

ABSTRACT

It has been proposed that the portfolio returns of individual private investors suffer due to excessive trading activity[2] Behavioral finance theory explains this as a result of investor overconfidence. According to traditional portfolio theory a well diversified portfolio is mainly subjected to market risk, when stock specific risk is diversified away. Based on the assumption that sophisticated investors make efficient investment decisions, we propose that portfolio diversification is a measure of investor sophistication. To investigate if investors of different gender, age and language background are characterized by systematic differences in trading activity, diversification and portfolio value, we propose a model of investor sophistication in which we measure trading activity and portfolio diversification for a large sample of individual investors. The data set used for the study includes the share holding records for 11795 individual private investors that have been randomly selected from the Finnish Central Securities Depository (FCSD). The data contains daily records of the trades of all market participants over the period January 1, 1995 through May 31, 2000 and includes a set of demographic vari-

1 Acknowledgements: We are indebted to the Finnish Central Share Depository, for providing us with the data. Financial support from OKOBANK Group Research Foundation is gratefully acknowledged.
2 Odean (1998).

P. JOAKIM WESTERHOLM, Dr, Senior Lecturer
School of Business H69, The University of Sydney • e-mail: j.westerholm@econ.usyd.edu.au
MARTIN OLLILA, M.Sc. (econ)
• e-mail: martin.ollila@alandsbanken.fi

ables for the investors. Male investors trade more frequently and are more diversified than female investors. Investors in the age group (26–45) trade the most. Investors in the age group (46–65) and investors with large portfolios are the most diversified. Finnish speaking investors trade more frequently than Swedish speaking investors. We suggest that the poor diversification observed for most investors is due to mental accounting[3]. Earlier studies have reported poor performance for more active investors and explain this as a result of excess activity. The difference in performance between more and less active investors might also in the light of our study be caused by differences in diversification. Investors that trade less are also less diversified and are thus compensated for their higher risk by higher return. Investors that trade more are better diversified and are not performing as well in rising markets since their risk is also lower. We find little evidence of excessive trading by Finnish investors.

Keywords: *Diversification, Gender, Investment Strategy, Overconfidence*
JEL classification: G11

1. INTRODUCTION

Portfolio diversification is a cornerstone in the portfolio theory developed by Markowitz (1952). In a stock portfolio, idiosyncratic risk can be minimized due to that returns for different stocks are rarely perfectly correlated. Optimally, a well diversified portfolio is only subject to market risk once idiosyncratic risk is diversified away. In a study by Karhunen and Keloharju (2001), it is observed that Finnish individual investors have poor portfolio diversification. Barber and Odean (2001) report similar results in their study using data from a US securities firm. In their study households held in average 4 stocks in their portfolio. Another study by Benartzi (2001) shows that US pension contributions are also very poorly diversified and tend to be allocated to the employer's stock. Behavioral models can provide guidance to explaining the poor diversification among individual investors. According to the Prospect Theory developed by Kahnemann and Tversky (1979), a person has a varying attitude toward risk. Idiosyncratic returns, similar to lotteries, offer the investor a small possibility to win and in this situation the investor is ready to accept more risk than he or she would accept making other investment decisions. If we assume that sophisticated investors make efficient investment decisions in the mean-variance frame, both acknowledging the return and variance of their portfolios, then it is justified to propose that portfolio diversification is a measure of investor sophistication.

Overconfidence has lately been subject to a large quantity of research in behavioral finance, see e.g. Odean (1998 and 1999). Overconfidence has been explained as a characteris-

3 The tendency to categorize funds or items of value even though there is no logical basis for the categorization. For example, individuals often segregate their savings into separate accounts to meet different goals even though funds from any of the accounts can be applied to any of the goals (Thaler 1985).

tic of human information processing. A common human feature is that we constantly learn about our abilities by observing our actions. It is, in short, a "trial-and-error" process. But when processing information, a heuristic bias causes us to take too much credit for our successes and blame failures on unmanageable external forces. Overconfidence is e.g. a Darwinian mechanism that helps us to survive in a competitive environment. Another interesting finding is that overconfidence is greater in areas that are demanding and that lack direct and clear feedback. Investing in the stock market can be seen as such an area that lacks clear and direct feedback. Barber and Odean (2000 and 2001), conclude that excessive trading is a direct symptom of overconfidence. They find that overconfident investors overestimate the precision of their knowledge about the value of a security. This is why overconfident investors engage in frequent trading because they believe that they achieve a superior return. Barber and Odean (2001) investigate the hypothesis of overconfidence by dividing investors by gender. Their study show that male investors trade more frequently and by doing so they lower their returns due to transaction costs and the choice of securities.

In earlier studies of Finnish investors Karhunen and Keloharju (2001) report low diversification and Grinblatt and Keloharju (2000) conclude that private investors follow less sophisticated investment strategies. Grinblatt and Keloharju (2001) investigate why investors trade as much as they do. According to portfolio theory and behavioral finance both diversification and trading activity are relevant for the performance of investment portfolios. Whether Finnish individual investors trade excessively and how well they diversify their investment portfolios has not been directly investigated. This is why we in this study include both these factors in a model of investor sophistication. We study investor sophistication by measuring three features of investment strategy: trading activity, portfolio diversification and portfolio value for a large sample of individual investors. Since portfolio value may be a related to investment strategy or to the initial wealth of an investor we focus on trading activity and diversification. We investigate how differences in trading activity and portfolio diversification are related to gender, age, language and portfolio value. Our aim is to investigate if high trading activity is characteristic for investors of a certain, gender, age, language, wealth and portfolio diversification. Our aim is also to investigate if low diversification is characteristic for investors of a certain, gender, age, language, wealth and trading activity. This study will give insight into if the tendencies of certain type of investors to trade excessively and diversify poorly observed in earlier studies are persistent when we analyze a representative sample of a whole market.

Analyzing investor behavior in a small securities market, market specific conditions such as how concentrated the market is and if the trading is organized differently to earlier investigated markets must be considered. In Finland, which is dominated by one high-capitalization company, Nokia, special caution must be taken when analyzing trading activity. A high trad-

ing volume in such a securities market can partially be due to portfolio rebalancing with the aim to improve portfolio diversification. In this respect, portfolios with securities that have increased their weight so that diversification effects are lost must be rebalanced. During the time period of the current study the stock markets in Finland experienced an exceptional bull market. This is largely as a result of the success of the Nokia stock, which has a weight of over 50% in the Finnish securities market. During the study period the stock price increased by almost 2,800%. For a Nokia investor, in order to achieve an optimal diversification, a frequent rebalancing is needed which would lead to a larger trading activity. Categorizing these investors as overconfident without considering other factors than trading activity would be an oversimplification.

The data set used for the study consists of 11795 individual private investors that have been randomly selected from the Finnish Central Securities Depository (FCSD) over the period January 1, 1995 through May 31, 2000. We find that male investors trade significantly more and are more diversified than female investors. Investors in the age group (26–45) trade the most. Investors in the age group (46–65) and investors with large portfolios are the most diversified.

The remainder of the paper is organized as follows: In section 2, the data, a diversification measure and the empirical models are presented. In section 3 the findings are presented and the analysis of trading activity, portfolio diversification and portfolio value are discussed in separate subsections. Section 4 summarizes the findings and presents the conclusions.

2. DATA AND METHODOLOGY

2.1 Data

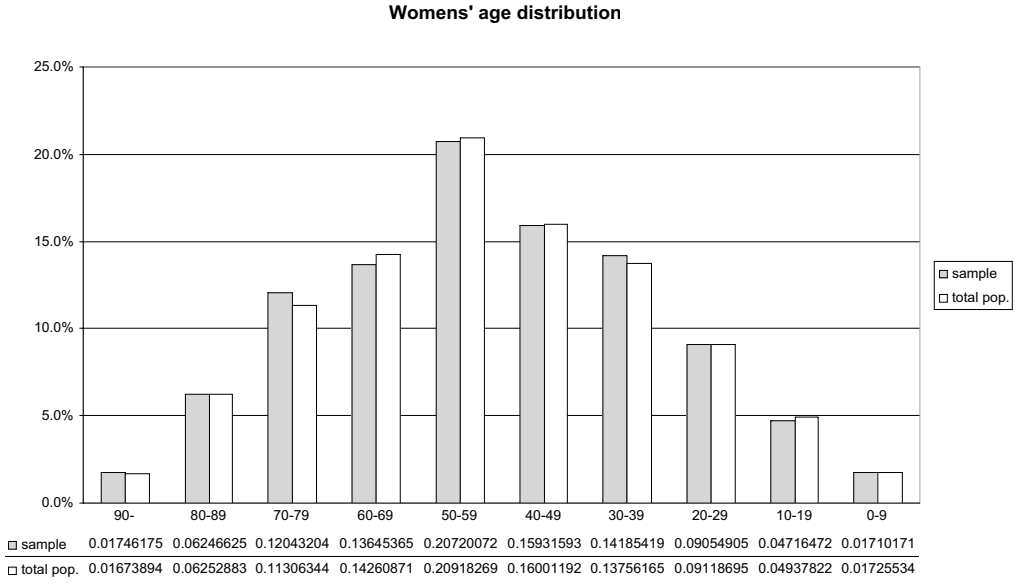
The data set used for the study is a sample consisting of 11795 individual private investors that have been randomly selected from the Finnish Central Securities Depository (FCSD) using a random number generator. The FCSD records the trades of all market participants on a daily basis and includes a set of demographic variables of the investor. The time period for the data set covers January 1, 1995 through May 31, 2000. Compared to survey data and data from a single securities firm the prime advantage is that the data does not suffer from potential problems with how representative it is. Also, since the shareholdings are recorded at a daily basis, it is much more exact and extensive than brokerage accounts, which at best provide data at a quarterly level[4]. The data set provides records of the investors' demographic characteristics.

4 A more extensive description of the Finnish book-entry system can be found on the Helsinki Stock Exchange's homepage at: www.hexgroup.fi.

For instance, the age, gender, mother-tongue, and the area of residence of each investor is included in the data set, thereby providing an excellent research base for investment behavior related studies. The FCSD data is also used e.g. in Grinblatt and Keloharju (2000 and 2001) and in Karhunen and Keloharju (2001).

The sample's distribution of male and female investors as well as age categories is illustrated in Figures 1 and 2. The division is rather even between the genders, there are 5494 women and 6301 men in the sample. The ratio of female to male investors in the sample is thus 0.47 per investor. This is close to the gender distribution of 45,9% female investors and 54,1% male investors in the whole population of individual investors in the FCSD. The age distribution of the sample corresponds well to the age distribution in the whole population as well. We conclude that the investigated sample is representative of the investor population in Finland.

The sample selection method in the current study limits a bias in the results by also including investors who have opened a book-entry account during the study period as opposed to selecting only investors who had opened a book-entry account before the study period. By including later registered investors who may have different investor behavior one limits the risk of a representative bias in the sample. We calculate variables to measure investment be-



	90-	80-89	70-79	60-69	50-59	40-49	30-39	20-29	10-19	0-9
sample	0.01746175	0.06246625	0.12043204	0.13645365	0.20720072	0.15931593	0.14185419	0.09054905	0.04716472	0.01710171
total pop.	0.01673894	0.06252883	0.11306344	0.14260871	0.20918269	0.16001192	0.13756165	0.09118695	0.04937822	0.01725534

FIGURE 1. Age distribution of sample and whole population for female investors.

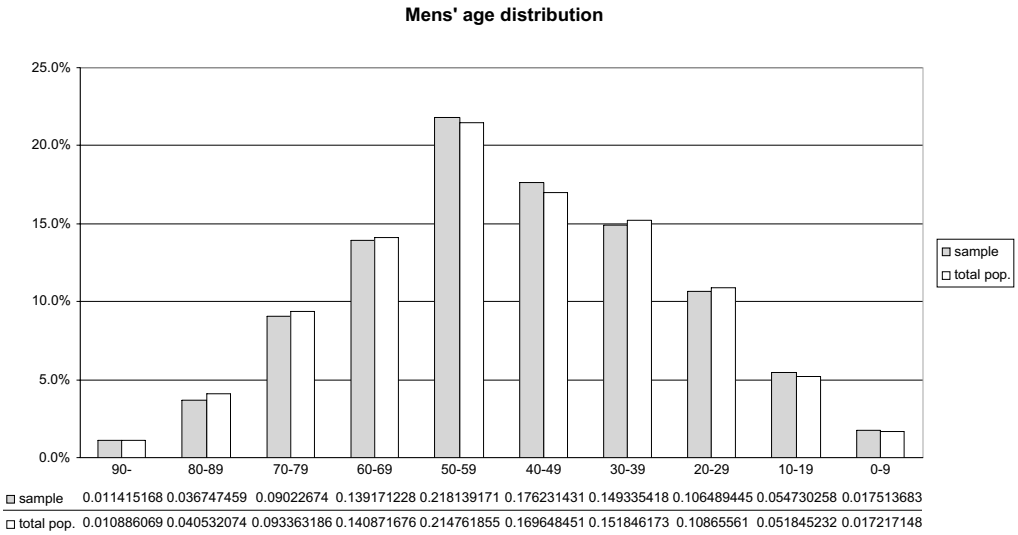


FIGURE 2. Age distribution of sample and whole population for male investors.

havior: the number of trades, a measure of portfolio diversification and the average portfolio value of the investor during the investigated period.

2.2 A measure of portfolio diversification

Portfolio diversification could be measured simply by the number of different stocks in a portfolio. This would however not take into account the weights of the holdings in these stocks. We use a value-weighted diversification measure, Herfindahl-Hirshmann index [H], to calculate the investor's portfolio diversification on a daily basis. Anti-trust authorities use the index as a measure of market concentration, but it can equally well be applied to securities in a portfolio. The lower the H index, the more diversified the portfolio is. The value of the H index is the sum of the squares of the weights of all securities in a portfolio:

$$H_i = \sum_{i=1}^{N} (w_i)^2 \qquad (1)$$

where:

H is the Herfindahl-Hirschmann index for an investor, measuring the value-weighted diversification of the portfolio and abbreviated H index in the text.

w is the weight of a security in the portfolio

i is the security

N is the number of securities in the portfolio

The H index does not measure and account for the variance of the portfolio. It is merely a value-weighted measure of the number of stocks in a portfolio. As such this measure has the weakness that if a portfolio holds several stocks that are highly correlated it would exaggerate the benefits of diversification. However to the extent that we can expect individual stocks to not be perfectly correlated the H index will be inversely related to diversification. The advantage of the index is that it can be easily and accurately calculated for a large sample of portfolios.

2.3 Empirical model of investor sophistication

Our aim is to investigate if there are systematic differences in number of trades, diversification and portfolio value between investor groups with different demographic characteristics. Such systematic differences would increase the understanding of why investors follow certain investments strategies. Some seemingly irrational strategies (e.g. very active trading) may be possible to explain in the light of what type of investors (e.g. well diversified) are following these strategies. We propose an empirical model of investor sophistication that includes the demographic determinants that may influence trading activity, portfolio diversification and portfolio value. As demographic variables we include age, gender and language. We create a set of age dummies in order to better capture the influence of age on trading activity and portfolio diversification. We aim to investigate if different age groups of investors follow different investments strategies.

Firstly we analyze the impact of possible determinants of trading activity and estimate Equations (2), (3) and (4). We measure trading activity as the number of trades executed by each investor during the study period. This measure is comparable between investors with different size portfolios since small trades have the same weight as large trades. We analyze the relationship between number of trades and gender and the relationship between number of trades and language. As control variables in these equations we include portfolio diversification and portfolio value to allow for portfolio sophistication and size related to other factors than gender and language.

$$T_i = \alpha_1 D_{g,i} + \delta H_i + \chi P_i + \varepsilon_i \quad (2)$$

$$T_i = \alpha_1 D_{L,i} + \delta H_i + \chi P_i + \varepsilon_i \quad (3)$$

T_i is the number of trades during study period. $D_{G,i}$ is the gender dummy with 1 for male investor, 0 otherwise. $D_{L,i}$ is the language dummy variable taking the value of 1 if the investor is Finnish speaking, 0 otherwise. H_i is the portfolio diversification index with a maximum of 1 and minimum of 0. P_i is the average portfolio value of the investor during the investigated period. ε_i is the error term.

The model describes the relationship between number of trades and portfolio diversification and the relationship between number of trades and portfolio value. We expect these relationships to be positive since more diversified portfolios need to trade more to rebalance and larger portfolios need to trade more due to the larger size. In addition we include a dummy variable for gender in Equation (2) to determine the relationship between gender of the investor and trading activity. Based on Barber and Odean (2001) we expect men to trade more than women and $D_{G,i}$ to take a positive value. Men also own larger portfolios than women as reported in Karhunen and Keloharju (2001) which would also lead to higher number of trades.

We include a dummy variable for language in Equation (3) to determine the relationship between language background of the investor and trading activity. Based on Karhunen and Keloharju (2001), that show that Swedish speaking investors hold larger portfolios, we might expect Finnish speaking investors to trade more than Swedish speaking investors when diversification and size of the portfolio is controlled for and $D_{L,i}$ to take a positive value. This would be a sign of a tendency of less experienced investors to trade more than what is required to rebalance the portfolio.

Then we include a set of age dummy variables in Equation (4) and analyze the relationship between trading activity and these age variables in combination with the portfolio diversification and portfolio value control variables to allow for portfolio sophistication and size related to other factors than age.

$$T_i = \alpha_1 + \alpha_2 D_{1,i} + \beta_1 D_{2,i} + \beta_2 D_{3,i} + \beta_3 D_{4,i} + \beta_4 D_{5,i} + \delta H_i + \chi P_i + \varepsilon_i \quad (4)$$

T_i is the number of trades during study period. $D_{1,i}$ denotes the age dummy that takes the value 1 for investors below 25 years of age and 0 otherwise.$D_{2,i}$ denotes an age dummy for investors between 26 and 45 years of age. $D_{3,i}$ denotes the age dummy for investors between 46 and 65 years of age. $D_{4,i}$ denotes the age dummy for investors between 66 and 85 years of age. $D_{5,i}$ denotes the age dummy for investors 86 years and older. H_i is the portfolio diversification index with a maximum of 1 and minimum of 0. P_i denotes the average portfolio value of the investor. ε_i is the error term.

We estimate Equation (4) excluding and including the age dummies and perform a joint F-test to assess if the age variables improve the model. We also compare the F-values and the R^2 estimates excluding and including age dummies. We then analyze the intercept terms and the coefficients for the age dummy variables obtained in estimations when one age dummy variable at a time is dropped. There is some controversy as to how far age dummy variables can be interpreted beyond the effect they have on the overall explanatory power of the equation. In this context we isolate the effect of the age dummy variables by analyzing their impact separately from other variables. We do consider the sign and significance of the age

dummy variables to be informative, particularly when they are compared to the dropped variable.

Secondly we analyze the impact of possible determinants of portfolio diversification and estimate Equations (5), (6) and (7). Portfolio diversification is measured as the H index defined in Equation (1).

$$H_i = \alpha_1 D_{G,i} + \delta T_i + \chi P_i + \varepsilon_i \quad (5)$$

$$H_i = \alpha_1 D_{L,i} + \delta T_i + \chi P_i + \varepsilon_i \quad (6)$$

H_i is the H index with a maximum of 1 and minimum of 0, a lower value for H_i implies a higher level of diversification. $D_{G,i}$ is the gender dummy with 1 for male investor, 0 otherwise. $D_{L,i}$ is the language dummy variable taking the value of 1 if the investor is Finnish speaking, 0 otherwise. T_i is the number of trades during study period. P_i is the average portfolio value of the investor during the investigated period. ε_i is the error term.

Then we include a set of age dummy variables and analyze the relationship between portfolio diversification and these age variables in combination with the number of trades and portfolio value control variables to allow for portfolio sophistication and size related to other factors than age.

$$H_i = \alpha_1 + \alpha_2 D_{1,i} + \beta_1 D_{2,i} + \beta_2 D_{3,i} + \beta_3 D_{4,i} + \beta_4 D_{5,i} + \delta T_i + \chi P_i + \varepsilon_i \quad (7)$$

H_i is the H index measuring diversification, T_i is the number of trades, P_i is the average portfolio value and the age dummy variables are the same as in Equation (4). We estimate Equation (7) excluding and including the age dummies and perform a joint F-test to assess the significance of the age variables. We also compare the F-values and the R^2 estimates excluding and including age dummies. We then analyze the F-values, the intercept terms and the coefficients for other dummy variables obtained for estimations when one age dummy variable at a time is dropped.

Thirdly we analyze the impact of possible determinants of portfolio value and estimate Equations (8), (9) and (10)

$$P_i = \alpha_1 D_{G,i} + \delta H_i + \chi T_i + \varepsilon_i \quad (8)$$

$$P_i = \alpha_1 D_{L,i} + \delta H_i + \chi T_i + \varepsilon_i \quad (9)$$

P_i is the average portfolio value of the investor during the investigated period. The average is obtained by first calculating the daily portfolio value for each investor in the sample for each day during the period and then computing the arithmetic average of these values. H_i is

the portfolio diversification index with a maximum of 1 and minimum of 0. T_i is the number of trades during study period. $D_{G,i}$ is the gender dummy with 1 for male investor, 0 otherwise. $D_{L,i}$ is the language dummy variable taking the value of 1 if the investor is Finnish speaking, 0 otherwise. ε_i is the error term.

Then we include a set of age dummy variables and analyze the relationship between portfolio value and these age variables in combination with the portfolio diversification and number of trades control variables to allow for portfolio sophistication and size related to other factors than age.

$$P_i = \alpha_1 + \alpha_2 D_{1,i} + \beta_1 D_{2,i} + \beta_2 D_{3,i} + \beta_3 D_{4,i} + \beta_4 D_{5,i} + \delta H_i + \chi T_i + \varepsilon_i \quad (10)$$

P_i is the average portfolio value for each investor during the investigated period. H_i is the portfolio diversification index, T_i is the number of trades and the age dummy variables are the same as in Equation (4). We estimate Equation (10) excluding and including the age dummies and perform a joint F-test to assess if the age variables improve the model. We also compare the F-values and the R^2 estimates excluding and including age dummies. We then analyze the F-values, the intercept terms and the coefficients for other dummy variables obtained for estimations when one age dummy variable at a time is dropped.

3. FINDINGS

3.1 Analysis of impact of gender and language

The results of the estimations of Equations (2), (3), (5), (6), (8) and (9) are reported in Table (2). The R^2 of the estimations are 1.5%, undefined and 0.9% when the dependent variables are number of trades, H index and portfolio value respectively. These estimations are done separately for each variable without intercept to isolate the effects of each variable and to avoid problems with dummy variables interfering with each other. Due to the exclusion of the intercept term the R^2 estimations are less reliable and no F-value is calculated. The gender variable (male investor) is significantly and positively related to number of trades and portfolio value. The language variable (Finnish speaking investor) is positively related to number of trades. The estimations of Equations (5) and (6) with the H index as the dependent variable has no explanatory power.

3.2 Analysis of trading activity

Significant differences in the number of trades executed by female investors compared to male investors are visible already in the descriptive Table 1, Panel A. Female investors in the sample trade on average 3.1 times during the period. On an annual level female investors trade

TABLE 1. Number of trades and diversification.

Panel A

	Women Number of trades	Men Number of trades	Women H index	Men H index
Mean	3.06	7.78	0.8592	0.8092
Median	1	2	1.0	1.0
Standard Deviation	8.28	31.92	0.203	0.224
Minimum	0	0	0.15	0.09
Maximum	202	1327	1.0	1.0
Number of Investors	5494	6301	5494	6301

Panel B

Year registered in FCSD	Average number of trades		Difference	
	Men	Women	t-statistic	p-value
1995 or before	7.29	3.53	7.24***	0.000
1996	10.36	3.74	2.11**	0.018
1997	8.94	4.30	3.88***	0.000
1998	10.27	3.65	3.46***	0.000
1999	8.17	1.68	6.12***	0.000
Jan 2000–May 31 2000	10.99	1.97	2.29**	0.011

Number of trades is the average of the number of transactions each investor makes during the investigated period. H index is a measure inversely related to portfolio diversification. *,**,* denote significance at the 1, 5 and 10% level, respectively. Tests are based on the two-sampled T-test assuming unequal variance.**

0.56 times. Male investors are clearly more active investors and trade on average 7.8 times, 1.4 trades annually. The median number of trades for male and female investors is 2 and 1, respectively for the whole period. Buy and sell trades as well as subscriptions to initial public offerings (IPOs) are considered as trades. When investors who have been registered in the FCSD a similar length of time are compared in Table 1, Panel B, each group of male investors trade significantly more than female investors. The proportion of female investors in the sample is 0.47 on average during the investigated period and in the end of the period they hold 37% of the combined invested wealth by individuals, as reported in Karhunen and Keloharju (2001). The difference in share ownership is not large enough to explain whey men trade two and a half times more frequently than women.

Equation (4), where number of trades is the dependent variable is estimated with and without age dummy variables. In the joint F-test the statistic is 11.7 which is significant on 1% level. This confirms that the age dummy variables improve the model. The calculation of the joint F-value is not reported. Table 3 reports the estimations of Equation (4). The F-value is significant on 1% level for both equations with and without the age dummy variables. The

TABLE 2. Relationship between number of trades, diversification, value and gender, language.

Regressions Variables	Dependent Number trades	Dependent Number trades.	Dependent H index	Dependent H index	Dependent Portfolio value	Dependent Portfolio value
Equation	(2)	(3)	(5)	(6)	(8)	(9)
Gender:	7.94***		0.8104***		43213**	
Male investor=1 Female investor=0	(15.2)		(288.1)		(2.04)	
Language: Finnish speaking investor=1		10.738***		0.7987***		–8228.0
Swedish speaking investor=0		(8.88)		(55.1)		(–0.193)
Activity: Number of trades			–0.347E-03 (–3.04)***	–0.151E-02 (–3.72)***	4756.8*** (3.01)	4946.1*** (2.99)
H index (inversely related to diversification)	–0.5106*** (–2.66)	–7.2164*** (–5.99)			40034*** (3.88)	73158 (1.49)
Portfolio value	0.228E-05 (2.24)**	0.236E-05 (2.36)**	–0.130E-07 (1.98)**	0.733E-08 (2.45)**		
Adjusted R^2	0.0128	0.0145	n/a	n/a	0.0094	0.0090
n	11795	11795	11795	11795	11795	11795

The table reports estimations of Equations (2), (3), (5), (6), (8) and (9) The Gender variable is a dummy variable that takes the value 1 for male and 0 for female investors. The Language variable is a variable that takes the value 1 for Finnish speaking investors and 0 for Swedish speaking investors. *,**,* denotes significance at the 1, 5 and 10% level, respectively.**

adjusted R^2 is 8.4% before and 8.8% after he age dummy variables are included. The age variable takes the highest value and significance for the age groups (–25) and (26–45), intercept 29.3 and t-value 22.3 in the second column and intercept 32.5 and t-value 21.4 in the third column. When the dummy variable for age group (26–45) is dropped all other age variables take a significantly negative coefficient. The age group 26 to 45 thus emerges as the most active investors. Portfolio value is positively related to activity and the H index is negatively related to activity. The significantly negative sign of the H index coefficient confirms that investors that trade more are better diversified than less active investors. (A lower diversification index indicates more diversification as explained in Section 2.2.).

3.3 Analysis of portfolio diversification

In Table 1, Panel A the descriptive statistics for the Herfindahl-Hirshmann measure of diversification for our sample is reported separately for female and male investors. The average H

TABLE 3. Relationship between number of trades, portfolio characteristics and age.

Regressions / Variables	No Age dummy variables	Age Group (–25) dropped	Age Group (26–45) dropped	Age Group (46–65) dropped	Age Group (66–85) dropped	Age Group (86–) dropped
Intercept	30.787	29.265	32.48	30.71	28.82	27.47
(t-value)	(21.7)	(22.3)	(21.4)	(20.8)	(22.5)	(21.3)
H index (inversely related	–30.484	–30.237	–30.237	–30.237	–30.237	–30.237
to diversification)	(–20.8)	(–21.0)	(–21.0)	(–21.0)	(–21.0)	(–21.0)
Portfolio value	0.184E-05	1.85E-06	1.85E-06	1.85E-06	1.85E-06	1.85E-06
	(2.24)	(2.21)	(2.21)	(2.21)	(2.21)	(2.21)
Age of the investor (years):						
–25		dropped	–3.2149	–1.4449	0.44652	1.7947
			(–6.23)	(–3.06)	(1.31)	(3.90)
26–45		3.2149		1.77	3.6615	5.0097
		(6.23)	dropped	(2.81)	(7.31)	(8.52)
46–65		1.4449	–1.77		1.8914	3.2396
		(3.06)	(–2.81)	dropped	(4.20)	(5.96)
66–85		–0.44652	–3.6615	–1.8914		1.3482
		(–1.31)	(–7.31)	(–4.20)	dropped	(3.22)
86–		–1.7947	–5.0097	–3.2396	–1.3482	
		(–3.90)	(–8.52)	(–5.96)	(–3.22)	dropped
Adjusted R^2	0.0844	0.0877	0.0877	0.0877	0.0877	0.0877
F-value	544.33	189.9	189.9	189.9	189.9	189.9
p-value	0.00	0.00	0.00	0.00	0.00	0.00
n	11795	11795	11795	11795	11795	11795

Equation (4) with the dependent variable number of trades is first estimated without age dummies and then by dropping each of the age dummy variables one at a time. A White (1980) heteroskedasticity-consistent covariance matrix is used in the regression. T-values are reported in parentheses beneath the coefficients. For this sample a t-value of approximately 2.0 denotes significance on 5% and a higher t-value indicates significance on 1% level. P-values for the F-tests reported under the F-statistic.

index, is 0.86 for women and 0.81 for men. The most striking result looking at the distribution of the H index (full H distribution not reported), is that 6523 investors have an H of 1, which is about 55% of the 11795 investors in the sample. Finnish stock portfolios are thus on average poorly diversified.

Equation (7), where portfolio diversification is the dependent variable is estimated with and without age dummy variables. In the joint F-test the statistic is 10.7 which is significant on 1% level. This confirms that the age dummy variables improve the model. The calculation of the joint F-value is not reported. Table 4 reports the estimations of Equation (7). The F-value is significant on 1% level for both equations with and without the age dummy variables. The adjusted R^2 is 7.9% before and 8.2% after he age dummy variables are included. The highly

TABLE 4. Relationship between H index (diversification), portfolio characteristics and age.

Regressions / Variables	No Age Dummies	Age Group (–25) dropped	Age Group (26–45) dropped	Age Group (46–65) dropped	Age Group (66–85) dropped	Age Group (86–) dropped
Intercept	0.84689	0.87177	0.84939	0.83327	0.85305	0.84723
(t-value)	(230.6)	(161.5)	(147.8)	(187.5)	(175.2)	(73.2)
Number of trades	–0.246E-02	–0.244E-02	–0.244E-02	–0.244E-02	–0.244E-02	–0.244E-02
	(–3.79)	(–3.76)	(–3.76)	(–3.76)	(–3.76)	(–3.76)
Portfolio value	–0.704E-08	–0.685E-08	–0.685E-08	–0.685E-08	–0.685E-08	–0.685E-08
	(–1.617)	(–1.58)	(–1.58)	(–1.58)	(–1.58)	(–1.58)
Age of investor (years):						
–25		dropped	2.24E-02	3.85E-02	1.87E-02	2.45E-02
			(3.39)	(6.41)	(2.82)	(1.95)
26–45		–2.24E-02		1.61E-02	–3.67E-03	2.16E-03
		(–3.39)	dropped	(3.26)	(–0.60)	(0.17)
46–65		–3.85E-02	–1.61E-02		–1.98E-02	–1.40E-02
		(–6.41)	(–3.26)	dropped	(–3.60)	(–1.16)
66–85		–1.87E-02	3.67E-03	1.98E-02		5.82E-03
		(–2.82)	(0.60)	(3.60)	dropped	
86–		–2.45E-02	–2.16E-03	1.40E-02	–5.82E-03	
		(–1.95)	(–0.173)	(1.16)	(–0.47)	dropped
Adjusted R^2	0.0786	0.0816	0.0816	0.0816	0.0816	0.0816
F-value	504.0	175.7	175.7	175.7	175.7	175.7
p-value	0.00	0.00	0.00	0.00	0.00	0.00
n	11795	11795	11795	11795	11795	11795

Equation (7) with the dependent variable H index (diversification) is first estimated without age dummies and then by dropping each of the age dummy variables one at a time. A White (1980) heteroskedasticity-consistent covariance matrix is used in the regression. T-values are reported in parentheses beneath the coefficients. For this sample a t-value of approximately 2.0 denotes significance on 5% and a higher t-value indicates significance on 1% level. P-values for the F-tests reported under the F-statistic.

significant intercept term in these estimations indicate that the model is not complete and any observations made here should be interpreted with caution and compared to the estimations of Equations (4) and (10).

Investors in the age group (46–65) are the most diversified. In the fourth column in Table 4 where the dummy variable for age group (46–65) is dropped all other age variables take a positive coefficient, all significant except for the age group over 85. The coefficient for number of trades is significantly negative across all regressions in Table 4 implying that investors making frequent trades are more diversified. The coefficient for portfolio value is negative across all regressions but the significance is low. A larger portfolio value leads naturally to a larger incentive to diversify when costs of rebalancing are lower than the benefits of diversification.

TABLE 5. Relationship between average portfolio value, portfolio characteristics and age.

Regressions / Variables	No Age Dummies	Age Group (–25) dropped	Age Group (26–45) dropped	Age Group (46–65) dropped	Age Group (66–85) dropped	Age Group (86–) dropped
Intercept	0.235E+06	0.204E+06	0.199E+06	0.252E+06	0.248E+06	0.258E+06
(t-value)	(4.28)	(–3.44)	(2.84)	(5.27)	(4.26)	(4.14)
H index	–0.196E+06	–0.191E+06	–0.191E+06	–0.191E+06	–0.191E+06	–0.191E+06
(diversification)	(–3.30)	(–3.13)	(–3.13)	(–3.13)	(–3.13)	(–3.13)
Number of trades	4144.8	4179.9	4179.9	4179.9	4179.9	4179.9
	(2.66)	(2.66)	(2.66)	(2.66)	(2.66)	(2.66)
Age of investor (years):						
–25		dropped	5126.7	–48301	–43651	–54248
			(0.189)	(–1.88)	(–2.18)	(–1.78)
26–45		–5126.7		–53428	–48777	–59374
		(–0.189)	dropped	(–1.76)	(–1.93)	(–1.72)
46–65		48301	53428		4650.7	–5946.2
		(1.88)	(1.76)	dropped	(0.202)	(–0.185)
66–85		43651	48777	–4650.7		–10597
		(2.18)	(1.93)	(–0.202)	dropped	(–0.366)
86–		54248	59374	5946.2	10597	
		(1.78)	(1.72)	(0.185)	(0.366)	dropped
Adjusted R^2	0.0115	0.0117	0.0117	0.0117	0.0117	0.0117
F-value	69.7	24.30.00	24.30.00	24.30.00	24.30.00	24.30.00
p-value	0.00	0.00	0.00	0.00	0.00	0.00
n	11795	11795	11795	11795	11795	11795

Equation (10) with the dependent variable average portfolio value (in FIM) is first estimated without age dummies and then by dropping each of the age dummy variables one at a time. A White (1980) heteroskedasticity-consistent covariance matrix is used in the regression. T-values are reported in parentheses beneath the coefficients. For this sample a t-value of approximately 2.0 denotes significance on 5% and a higher t-value indicates significance on 1% level. P-values for the F-tests reported under the F-statistic. 5.94573 FIM corresponds to one EURO.

This effect may be picked up by the number of trades variable that is positively correlated with portfolio value (correlation coefficient 0.10).

3.4 Analysis of portfolio value

Equation (10), where portfolio diversification is the dependent variable, is estimated with and without age dummy variables. In the joint F-test the statistic is 1.46 which is not significant. The age dummy variables do not improve the model in a significant way. Table 5 reports the estimations of Equation (10). The F-value is significant on 1% level for both equations with and without the age dummy variables. The adjusted R^2 is 1.15% before and 1.17% after the age dummy variables are included.

The H index is negatively related to portfolio value with the t-value –3.13, significant on 1% level. (A lower H index indicates higher diversification.) This confirms that investors with larger portfolios are better diversified as would be expected, since they have better opportunities to cost efficiently include a larger number of shares in their portfolios. The number of trades is positively related to portfolio value with the coefficient 4179.9 and the t-value 2.66, significant on 1% level. (One more trade is made per each FIM4179.9 of average portfolio value.) It appears that investors with larger portfolios perform adjustments to their portfolios more often to keep their portfolios better diversified. There is no reason to expect that larger portfolio size would automatically lead to significantly larger number of trades since these larger portfolios could simply make larger trades. We interpret the larger activity as a result of more active portfolio rebalancing. The age groups (46–65) and (86–) have the largest portfolios, see the fourth and the sixth columns in Table 5.

The relatively low R^2 of 1.17% implies that the model describes actual portfolio value quite poorly. This is expected since some relevant explanatory variables are excluded from the model. E.g. information on the salary of the investor would probably give more explanatory power to the model. Since investor identities are not disclosed in the data, this is not possible. (5.94573 FIM corresponds to one EURO)

4. CONCLUSIONS

It has been proposed that the portfolio returns of individual private investors suffer due to excessive trading activity. We also know that a well diversified portfolio is mainly subjected to market risk and impacted less by stock specific risk. To investigate if investors of different gender, age and language background are characterized by systematic differences in trading activity, diversification and portfolio value, we propose a model of investor sophistication in which we measure trading activity and portfolio diversification for a large sample of individual investors. The data set used for the study consists of 11795 individual private investors that have been randomly selected from the full Finnish Central Securities Depository (FCSD) over the period January 1, 1995 through May 31, 2000. The data set includes all shareholdings on the Helsinki Exchanges for these investors.

In our sample, male investor trade 7.8 times on average during the period while female investors trade 3.1 times on average. The difference is significant on the 1% confidence level. In estimations of our model the variable male investor is positively related to number of trades with an adjusted R^2 of 1.3%. The most active traders are in the age group 26 through 45 years of age with an adjusted R^2 of 8.8%. The variable Finnish language background is positively related to number of trades with an adjusted R^2 of 1.5 %.

We measure portfolio diversification using the Herfindahl-Hirschmann index (H). The average H index is 0.81 for men and 0.86 for women, indicating that men are more diversified than women. The majority of the sample fails to diversify their portfolio at all, since they have all their direct shareholder wealth invested into one security. When we estimate the model of the relationship between diversification and age, diversification and language, diversification and number of trades; investors between 46 and 65 years of age are the most diversified, Finnish speaking investors are less diversified than Swedish speaking investors and investors that trade more have more diversified portfolios. The estimations on language and number of trades have a poor explanatory power due to that there are very few variables in the equation, while the estimation on age has an adjusted R^2 of 8.2 %.

Explaining the observed poor diversification is a challenging task. Behavioral theories on mental accounting may illuminate the source of the phenomenon, see e.g. Thaler (1985). Instead of viewing the return and variance of the total financial assets, the investor puts different investments into different mental accounts. These mental accounts are viewed differently by the investor and serve different purposes for the investor. One mental account may include low risk bearing assets such as bank deposits and real estate, which are common investments in Finland. An investor may view this part of the financial assets as a part of a long term savings strategy. The other mental account of the investor may contain high risk bearing assets which are undiversified, representing a small part of the investor's total financial assets. This part is viewed as a type of "lottery ticket" or an option with a high expected return. The investor may find that optimal diversification decreases the expected return of the investment and thus decides to leave the portfolio undiversified. The more experienced and sophisticated investors are, the less likely they are to be affected by mental accounting. We would expect to find less evidence of mental accounting for larger portfolios, since larger portfolios may have been actively managed for a longer time and more resources can be spent on gaining experience. Our findings support this line of thinking, since portfolio size is significantly related to higher diversification. Based on the conservative way that Finnish households allocate their financial assets, keeping most of their savings in bank accounts, it is not unexpected to find that individual private investors view their shareholdings in a naive way that leads to poor diversification.

In earlier studies it has been suggested that poor performance by investors in comparison to the market index is explained by excess activity, see Barber and Odean (2000). Since we find that investors that trade more are also better diversified, the outperformance by less active investors might at least partly be a result of poor diversification and thus higher risk taking. We suggest that since investors that trade less also have a lower degree of diversification they appear to do better, not because they save on transaction costs, but because they assume more

stock specific risk. Overall Finnish investors are not very frequent traders (in our sample they trade between 3 and 8 times on average over 5 and a half years) and also since the high concentration of the market justifies a certain level of rebalancing activity, Finnish investors do not appear to trade excessively. The more frequently trading investors in our investigated sample appear to be more sophisticated investors since they have more diversified portfolios and their trading is not too frequent to qualify as portfolio rebalancing trades. ■

REFERENCES

BARBER B. and T. ODEAN, 2000, "Trading is hazardous to your wealth: The common stock investment performance of individual investors", Journal of Finance 55, 773–806.

BARBER, B. and T. ODEAN, 2001, "Boys will be boys: Gender, overconfidence and common stock investment", Quarterly Journal of Economics, 261–292.

BENARTZI, S., 2001, "Excessive extrapolation and the allocation of 401(k) accounts to company stock", Journal of Finance 56, 1747–1764.

GRINBLATT M. and M. KELOHARJU, 2000, "The investment behavior and performance of various investor types: A study of Finland's unique data set", Journal of Financial Economics 55, 43–67.

GRINBLATT M. and M. KELOHARJU, 2001, "What makes investors trade?", Journal of Finance 56, 589–616.

KAHNEMANN, D. and A. TVERSKY, 1979, "Prospect theory: An analysis of decision under risk", Econometrica 47, 263–291.

KARHUNEN J. and M. KELOHARJU, 2001, "Share ownership in Finland 2000", Finnish Journal of Business Economics 50, 188–226.

MARKOWITZ, H., 1952, "Portfolio selection", Journal of Finance 7, 77–91.

ODEAN T., 1998, "Volume, volatility, price and profit when all traders are above average", Journal of Finance 53, 1887–1934.

ODEAN T., 1999, "Do investors trade too much?", American Economic Review 89, 1279–1298.

THALER, R. 1985, "Mental accounting and consumer choice", Market Science 4, 199–214.

WHITE, H., 1980, "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity", Econometrica 48, 817–838.