Trading and Under-Diversification

Anders Anderson*

Draft 2012-03-15

Abstract

This paper documents a link between trading and diversification by using detailed trading records from a Swedish discount broker matched to individual tax records. Investors’ diversification is measured by stake size, defined as the fraction of their risky financial wealth invested in individual stocks at the broker under study. High stake investors have concentrated portfolios, trade more, and achieves lower trading performance. They share several features with those who trade excessively: they have lower income, wealth, age, and education, suggesting that they lack investment expertise. The results directly imply that trading losses in the cross-section mainly are borne by those who can least afford them.

Keywords: Investor behavior; stock trading; overconfidence; under-diversification.

JEL codes: G11, D14, C24.

*Institute for Financial Research (SIFR), Drottninggatan 89, SE-113 60 Stockholm, Sweden, Phone: +46-8-728 51 23, Fax: +46-8-728 51 30, E-mail: anders.anderson@sifr.org. Financial support from Nasdaq OMX and the Marie Curie Research Training Networks is highly appreciated. I thank an anonymous referee, Ulf Axelson, Magnus Dahlquist, Laura Frieder, Harrison Hong, Zoran Ivkovic, Lars Lochstöer, Jürgen Maurer, David Robinson, Paolo Sodini, Per Strömberg, Elu von Thadden, Martin Weber, and participants at Mannheim and Stockholm University, BI Oslo, Copenhagen Business School, SIFR, the conferences EFM Durham 2006, FMA 2006, EFA 2007, and the AFA 2007 for useful comments. I am grateful to the online broker who wishes to remain anonymous, and Finansinspektionen who provided me with the data. A previous version of this paper has been circulated under the name “Is Online Trading Gambling with Peanuts?”.
Introduction

It is widely established that individuals trade too much and hold undiversified portfolios, but so far, there is very little theoretical and empirical evidence that link these most prominent stylized facts of individual investor behavior\footnote{For evidence of excessive trading; see Barber and Odean (2000), and diversification; Goetzmann and Kumar (2008) and Polkovnichenko (2005).}. This paper aims to fill this gap. We derive a simple model which predicts that overly optimistic investors choose to trade away from the full diversification benchmark, and then show that individuals who hold concentrated portfolios also trade more, but less successfully so.

By using unique Swedish individual tax data matched with trading records, we are able to measure diversification more widely than by only using trading account data. Specifically, we find that investors with ”high stakes”, meaning that they belong to the top quintile of investors who concentrate most of their risky financial wealth into a portfolio of a few individual stocks, trade twice as much as the bottom quintile of investors who hold more diversified portfolios with a higher fraction of their wealth invested in equity mutual funds. We also show that high stakes is predicted by the same variables that have previously been considered as proxies for overconfidence, such as: gender, age, wealth, and education.

Under the standard normative finance paradigm, investors hold perfectly diversified portfolios and have little incentive to trade. Yet, many hold under-diversified portfolios and trade extensively. The literature in finance has to a large part focused in solving these puzzles separately. Trading can more generally be explained by investors holding different beliefs, as in Miller (1977), Varian (1989), and Harris and Raviv (1993). Disagreement about fundamental values instigates trading, and these effects can be amplified by investors understating the true riskiness of assets. The link between overconfidence and excessive trading has been proposed by Kyle and Wang (1997), Odean (1998b), and Hong, Scheinkman, and Xiong (2006), but these models are generally silent about to which degree they predict diversification.

On the other hand, models trying to explain high portfolio concentration generally do not predict increased trading. Incomplete information models, such as those by Merton (1987) and Arbel, Carvell, and Strebel (1983), suggest that there are some fixed costs associated
with gathering or processing information about firms, which ultimately gives a rationale for investors holding fewer stocks in their portfolio. The Merton model, as many others in this category, predicts that there is an information premium of holding such stocks in equilibrium where investors choose to “specialize” in specific stocks. The theory can explain why there is segmentation in ownership between stocks, but does not give any sharp predictions about why trading should vary systematically across stocks or investor clienteles. Behavioral finance, on the other hand, basically offers three lines of interpreting lack of diversification. First, Barberis and Huang (2008) show that investors with systematic biased beliefs about payoffs are attracted to holding assets with positively skewed payoffs, as proposed by Tversky and Kahneman’s (1992) cumulative prospect theory. Second, Huberman and Regev (2001), and Barber and Odean (2008) show that attention to events or news is an important driver for investors choosing to buy specific assets. Third, if investors have a bias towards the familiar, they may have preferences for local assets even if they do not have an information advantage, as suggested by French and Poterba (1991) and Huberman (2001). The three behavioral explanations for portfolio concentration therefore also lack a more general link between trading and diversification.

We propose a way to reconcile the two in a model in which investors are overly optimistic about their stock-picking ability. Such investors would have incentives to deviate from the benchmark portfolio, either by timing the market or to take bets in specific assets. The model shows that the key variable to identify this behavior is the fraction of risky assets allocated to this concentrated portfolio. We refer to this weight as stake size, and we investigate how high and low stake investors differ in their composition and trading behavior in the data.

The trading data comes from a Swedish online discount broker which includes the social security number of each investor. The trading data is then matched to official tax records that include detailed information of income, wealth, occupation, and education. We define stake size in the data as each individual’s stock portfolio value in the broker’s account divided by the value of all risky assets (financial wealth excluding holdings of cash and money market funds), measured at first year-end of trading. The observed stock portfolio
constitutes a relatively small part of financial wealth for many investors. Around 25% of
the investors have stocks worth 5% or less of their financial assets in the brokerage account;
for around 50% they are worth 20% or less. At the top of the distribution, 19% of the
investors have 70% or more of their financial wealth invested in stocks in these accounts,
and this constitutes all of their holdings of risky assets. These results alone shed light on
the disproportional impact measured trading can have on investors’ wealth, even if trading
returns were constant. Furthermore, if trading is mainly confined to those with low stakes,
it may rationally be explained by motives associated with learning or entertainment.

The previous evidence surrounding portfolio under-diversification, performance, and trad-
ing is somewhat fragmented. Goetzmann and Kumar (2008) find that investors with fewer
stocks in their accounts trade more on average, but also that holdings of those with fewer
stocks outperform holdings of those better diversified within the group of high-turnover in-
vestors. Using the same U.S. data, Ivkovic, Sialm, and Weisbenner (2008) also find that
households with large stock portfolios, but with few stocks, do better than less concentrated
households. These results suggest that some under-diversified households possess informa-
tion and are able to trade profitably. We can confirm some of these findings, since we also
find that wealthier investors trade more persistently and perform better than the average
investor, but we cannot establish that stake size is associated with superior performance
among the wealthiest individuals in sample. Calvet, Campbell, and Sodini (2009), who use
Swedish data from a similar time period, find that the share of funds in risky portfolio wealth
appears both to be a valid diversification proxy, and is also highly positively correlated with
Sharpe ratios in the cross-section, whereas number of stocks or funds in investors’ portfo-
lios predict much less of investors’ performance. They also show that wealthier and better
educated investors are less prone to make investment mistakes, such as diversifying poorly.
This paper is similar in that we also use a broad definition of diversification, but we instead
combine it with a detailed trading data set, in which we can calculate stock turnover and
performance very accurately.

Our first and main contribution is that we derive and document a positive correlation

\[ \text{See, for example, Dorn and Sengmueller (2009).} \]
between trading and under-diversification. Second, we show that individual characteristics
that have previously been used to proxy for overconfidence through trading, also predict
stake size. A canonical correlation analysis shows that stake size is in fact a far more
important component relative to turnover in accounting for the cross-sectional variation in
individual characteristics. Third, we find evidence that investors’ financial expertise matters.
Only wealthy and educated investors respond to past trading performance when they trade.
Unrealized losses in the portfolio is a better predictor of turnover for those who are less
sophisticated. Stake size predicts turnover for all investors, but is less pronounced for the
more sophisticated, which we argue adds further support for interpreting the results as a
consequence of a behavioral bias. Fourth, high stake investors perform worse than average
investors. The results suggest that the main channel for this underperformance is through
elevated trading, and to a lesser degree stock selection or timing. Finally, we estimate
individual profits and costs of trading in the cross-section where we show that high stake
investors also carry most of the aggregate trading losses.

The socio-economic background of high stake investors makes the documented relation
between stake size and trading particularly harmful to trading profits. We find that the top
quintile of investors with the highest stakes carry 40% of the annualized trading losses. This
loss represents 3.77% of the aggregate financial wealth within this group, which corresponds
to 1.42% of their income. These investors hardly own any equity mutual funds at all, as all of
their risky financial wealth is concentrated to a few stocks. Figure 1 summarizes the results
of the interaction between high stakes and turnover by displaying total trading costs across
groups of investors. About a quarter of investors are classified as sophisticated, meaning
that they have completed at least one year of university education and belong to the two top
wealth quintiles. These investors carry 22% of the trading losses, but together possess 61%

of the total financial wealth. At the other end, there are also about a quarter of investors
classified as unsophisticated, with less or no university education, and who belong to the
two bottom wealth quintiles. They carry 27% of the incurred losses from trading, but own
only 3% of the aggregate financial wealth. Therefore, losses are mainly borne by those who
can least afford them.

[Insert Figure 1 here]

Investors with lower education and wealth are also those with fewer investment options: the portfolios are often too small to be meaningfully diversified, and they pay disproportionately high fees. Our results shed light on the fact that these investors for all practical purposes are constrained: partly because of low wealth, but maybe more importantly, by having poor knowledge about how to save efficiently. We add to the literature on financial literacy and investment choice by showing that those who are least equipped to make complex investment decisions perform worse when managing their stock portfolios.

The paper is organized into five sections. Section 1 explains how trading and portfolio selection are related in a model where investors have, or believe that they have, good investment skills. Section 2 presents the data; section 3 derives a method of measuring trading revenue and return. Section 4 presents the results, and section 5 concludes.

See, for example, Lusardi and Mitchell (2011).
1 Self-attributed investment skill and overconfidence

Empirical work by Barber and Odean (2000) and others suggests that individual investors trade far more than can be justified by rational agents within a standard expected utility framework. Active investors earn lower returns due to fees and bid-ask spreads, but also due to unprofitable trading strategies, as found by Grinblatt and Keloharju (2000). Such irrational trading can be explained by overconfidence. The existing literature has identified several important predictors of trading related to overconfidence. For example, it is suggested that men are more overconfident than women, and therefore trade more. Investors who go on-line are young, and increase their trading after going online. Several studies find that turnover decreases with investor age and wealth.

In market settings, it usually suffices to model overconfident investors by assuming miscalibration to generate trading. More precisely, some investors understate the riskiness of assets as in, for example, Kyle and Wang (1997) and Hong, Scheinkman, and Xiong (2006). The psychology literature, however, attributes a broader meaning to overconfidence. Individuals with flawed self-assessment do not only understate risks, but also hold an overly optimistic view of their true capabilities. Some empirical evidence suggests that this distinction is important in explaining trading behavior. Using data from several UBS/Gallup Investor Surveys, Graham, Harvey, and Huang (2009) find that investors who respond that they regard themselves as competent investors both trade more frequently and diversify more widely. Glaser and Weber (2007) surveyed online investors and found a robust relation between their measure of the "better-than-average" effect and trading activity, but no relation between a traditional measure of miscalibration and trading.

To formalize the notion of why overconfident investors concentrate their portfolio holdings, and to better understand the different concepts of overconfidence, we propose a version of Treynor and Black’s (1973) model. Here, mean-variance investors allocate funds across a passive benchmark fund, and an individual stock. The decision to deviate from the passive benchmark, that is to trade, is governed by the beliefs about future expected returns of

---

5 See, for example, Dunning, Heath, and Suls (2004).
the stock. Optimism and miscalibration are modeled separately, allowing investors both to overstater the performance, as well as the riskiness, of the stock. It follows immediately that investors will tilt their risky portfolio to the stock in proportion to their optimistic view about its future performance. The decision to trade follows from adjusting the weight from the current to the optimal portfolio. Then, holding investors initial beliefs constant, we investigate the optimal portfolio strategy in a recursive setting. We show that updating beliefs about the expected stock return is inconsistent with a simple buy-and-hold strategy, which implies that trading is motivated not only in a static, but also in a dynamic setting. Finally, the analysis shows that miscalibration of investors’ beliefs is independent of the risky share of total financial assets. This motivates the choice of our measure for under-diversification as the relevant proxy for investor confidence. Below follows a more detailed description of the model and its implications.

Consider investors with standard quadratic utility functions, who subject to their risk-aversion denoted by $\gamma$, maximize utility by allocating the weight,

$$w_{P^*} = \frac{E(R_{P^*}) - R_F}{\gamma \sigma_{P^*}^2},$$  

(1)

to the optimal risky asset $P^*$, where $R_{P^*}$, $R_F$, and $\sigma_{P^*}^2$ denote the return on the risky portfolio, the risk-free asset, and the variance of $P^*$. Ex ante, the investor believes he or she is able to outperform the market by choosing to invest in a stock that generate an average excess return of $\bar{S} \geq 0$. Let there be some uncertainty about this strategy, denoted $\tilde{S}$. As in the Treynor-Black model, the weight to the stock is a function of the expected performance of the strategy, $\bar{S}$. The expectation to generate superior returns from the stock investment can be perfectly rational if the investor really is better off contesting the benchmark on average; the case of updating such beliefs if they turn out to be wrong is left to the end of this section.

The benchmark has a mean return of $\bar{M}$ with a stochastic term $\tilde{M}$ which is independent of $\bar{S}$. We allow investors to be miscalibrated such that they underweight the true risk of the benchmark portfolio with a parameter $0 < c \leq 1$. The model also allows for investors to
be miscalibrated about the variance of the stock with a parameter $0 < d \leq 1$. Assuming normality, we have

$$R_M = \bar{M} + c\tilde{M}, \ c\tilde{M} \sim N(0, c^2\sigma^2_M),$$

$$R_S = R_M + S = R_M + \bar{S} + d\tilde{S}, \ d\tilde{S} \sim N(0, d^2\sigma^2_S),$$

and the optimization problem is to find the weight $w_S$ that maximizes the utility for the portfolio

$$R_P = w_S S + R_M.$$

The solution for the optimal weight is

$$w_S^* = \frac{\bar{S}/d^2\sigma^2_S}{(\bar{M} - R_F)/c^2\sigma^2_M},$$

which is the standard solution to the model of Treynor and Black (1973), as the allocation to the stock is increasing in the appraisal ratio in the numerator. The miscalibration parameters scale variance between the portfolios, and will increase the allocation to the stock if $d < c$. Any allocation to the stock is, however, crucially dependent on investors’ average performance assessment, $\bar{S}$. Substituting (2) into (1), factorizing, and simplifying the expression gives

$$w_P^* = \frac{\frac{\bar{S}/d^2\sigma^2_S}{(\bar{M} - R_F)/c^2\sigma^2_M}}{\gamma \left( \frac{\bar{S}/d^2\sigma^2_S}{(\bar{M} - R_F)/c^2\sigma^2_M} \right)^2 d^2\sigma^2_S + c^2\sigma^2_M} = \frac{\bar{S}/d^2\sigma^2_S}{(\bar{M} - R_F)^2 (1 + \frac{c^2\sigma^2_M S^2}{d^2\sigma^2_S (\bar{M} - R_F)^2})} = \frac{(\bar{M} - R_F)^2}{\gamma c^2\sigma^2_M (\bar{M} - R_F) (1 + \frac{c^2\sigma^2_M S^2}{d^2\sigma^2_S (\bar{M} - R_F)^2})} = \frac{\bar{M} - R_F}{\gamma c^2\sigma^2_M}. \quad (3)$$

Therefore, the total weight of risky assets is only a function of the risk-premium and variance of the benchmark, risk aversion, and miscalibration of the risky benchmark return. Most

---

6 Allowing for different parameters of miscalibration, $c \neq d$, adds little to the analysis that follows, and is merely included for completeness.
notably, comparing (3) and (1), the solution is independent of \( \overline{S} \), \( \sigma^2_S \), and miscalibration of the self-assessed ability, \( d \). The intuition for this result is that the allocation to the stock portfolio scales linearly with mean and variance. The investor allocates more to the stock for a higher \( \overline{S} \), but keeps the allocation to risky assets at the whole constant. This independence may not be generalized to arbitrarily chosen utility functions, and is not the result to be stressed here. The model rather shows that it is possible to break the link between investors’ perception of risk in general, and under-diversification. Increased portfolio concentration does not necessarily imply that investors are more risk-tolerant. This becomes clear in the case when \( 0 < c < 1 \) and \( \overline{S} = 0 \). Such investors will tilt their portfolios towards risky assets due to mis-calibration, but will still stay fully diversified. The implication of this stylized model is that investors without specific views about the stock will lack any incentive to trade away from their benchmark. In contrast, those with optimistic expectations about the average return will have incentives to deviate by changing the composition of their portfolio as long as \( w_S \neq w^*_S \).

In a dynamic setting, it is difficult to see why investors, on average, can be wrong about the profitability of their trading strategies. Such behavior may seem difficult to maintain by updating and realigning expectations, but consider the simplest case of an updating rule as follows. Suppose the true average excess performance, \( S^* \), has the distribution \( N(S^*, \sigma^2_S) \). Investors have at date 0, a prior with the distribution \( N(\overline{S}_0, d^2 \sigma^2_S) \). If the individual evaluates his or her trading performance for \( t \) trading periods, and accumulates the signal of a realization of \( S \)

\[
\tilde{S}_t = \frac{1}{t} \sum_{i=1}^{t} S_i, \tag{4}
\]

it is straightforward to show that the posterior mean follows

\[
\overline{S}_{t+1} = \left[ \frac{d^2 \sigma^2_S}{d^2 \sigma^2_S + \sigma^2/S} \right] \tilde{S}_t + \left[ \frac{\sigma^2_S/t}{d^2 \sigma^2_S + \sigma^2_S/t} \right] \overline{S}_0. \tag{5}
\]

Investors who have a biased prior of the excess return will rationally adjust their expectations over time as they discover the true \( S^* \). Equation (5) shows that this process is a weighted average of the variances and the number of evaluation periods, and is therefore not
instantaneous. Except for the miscalibration term $d$, which makes convergence slower, the updating rule is rational.\footnote{7}

To show that trading will occur in a mult-period setting, we first note that trading is always motivated at each point in time, except when the portfolio weight coincides with the optimal weight, or $w_{S,t} = w^*_{S,t}$. The only recursive condition in which there would be no trading, is if investors have buy-and-hold portfolios. For a buy-and-hold portfolio to exist, we must have that the optimal weight in $t$ produces a weight in $t+1$ which itself is optimal, i.e. $w^*_{S,t} \Rightarrow w^*_{S,t+1} = w^*_{S,t+1}$. Even if this is possible, a closer look at the model reveals that this is not a general solution, since it is valid for only specific parameter values. In fact, for optimistic investors, any passive re-weighting of the portfolio, based on the realized stock return only, could never be optimal. The reason is that the updating rule is not only a function of realized stock returns, but also of the number of observations, which will only gradually change the expectation of future stock returns, and as a consequence, the optimal weight.\footnote{8}

The simple model considered in this section therefore predicts that investors with an optimistic view of the performance of an individual stock will hold concentrated portfolios in proportion to their optimistic beliefs. They will be motivated to trade because they need to reestablish the weight according to their updated expectations. The model does not describe when, or how often, investors choose to update their expectations, or for which particular stocks they may have special preferences. But given a particular shock to beliefs for a stock, we argue that it still is a useful way of describing the relation between optimism, allocation and trading. According to the model, the successful identification of optimistic and

\footnote{7}Gervais and Odean (2001) use a biased updating rule to generate endogenous overconfidence. Daniel, Hirshleifer, and Subrahmanyam (1998) model the dynamics of overconfidence by miscalibration, but are mainly concerned with pricing implications.

\footnote{8}To see this, regard all parameters except $\bar{S}$ in \cite{1} as constants. The condition of unchanged weight is fulfilled when $\bar{S} = \bar{S}_{t+1}$. This is true if the prior equals the average signal, $\bar{S}_t = S^*$, in \cite{1}. That is, if the investor’s belief about the mean is unbiased, the buy-and-hold portfolio is to always hold the same, constant weight: $w^*_{S,t} = w^*_{S,t+1}$. Then consider when a buy-and-hold portfolio could be consistent with an optimistic prior, $\bar{S}_0 > S^*$. Keeping the market return constant, a realization of the stock return $S_t$ that mechanically changes the portfolio composition to $w^*_{S,t+1}$, must then also coincide with $w^*_{S,t+1}$ in order to be optimal. But, by contradiction, $w^*_{S,t+1}$ could then never be optimal, since $w^*_{S,t+1}$ is based on a weighted average between the prior and the signal $S_t$ itself.
overconfident investors rely on determining how much of the risky wealth is concentrated to the under-diversified portfolio, or what we in the following will simply refer to as individuals’ stake size.

2 Data

The data were made available by an online broker and cover all transactions since the start in May 1999 up to and including March 2002. The data cover 324,736 transactions in common stocks distributed over 16,831 investors who enter sequentially over time. From this sample, investors are then selected if they have been active for at least 12 months and the data on wealth and income is complete. The remaining sample consists of 10,600 investors who made 224,964 common stock transactions distributed over 213,633 investor portfolio months. The average investor is therefore active for approximately 24 months.

In addition to trading data, the sample has with the help of social security numbers been matched to the Statistics Sweden database with detailed background information on the investors. These data include exact information on the market value of an investor’s total portfolio at each year-end. More importantly, all Swedish financial institutions are legally required to report the market value of all individual’s financial instruments and bank holdings directly to the Swedish Tax Authority. This means that each investor’s total portfolio is observable at yearly intervals, and can be matched to the stocks held with the particular online broker under study. Furthermore, information is available on housing wealth (taxable), total liabilities, capital insurance, and income. It is also possible to see if income originates from employment in the financial sector; this may be indicative of better financial decision making skills. Combined, this allows for a quite detailed analysis of how different investor and portfolio characteristics affect trading behavior and portfolio performance, and, in particular, for detailed controls when studying these variables in the cross-section.

See Anderson (2007) for a more comprehensive summary of this data. Total transactions include 120,734 purchases, 82,846 sales, and 21,384 deposits and redemptions of stocks. The month when investors enter the data is excluded from the analysis.
Some features of the data and matching are worth describing in more detail. Income is measured as the first observed year-end disposable income adjusted for net capital gains. Turnover is calculated as the total value of all trades each month divided by two times the portfolio value at the beginning of the month. Finance is an indicator variable if the main source of working income has been earned in the financial sector during the year. University takes the value one if the individual has completed at least one year of university studies. Wealth is broken up into three major components: financial, real estate, and debt. Financial wealth is, in turn, divided into bank holdings, money market funds, bonds, stocks, equity and mixed mutual funds, other financial instruments, capital insurance, and other financial wealth. All financial wealth is measured at market values, except for the last component that includes non-listed instruments, private equity, promissory notes, and self-reported values of possessions. The market value for real estate is not observable, but is measured as an adjusted taxable value. In general, the taxable value should correspond to 75% of the market value, but it is sometimes understated where prices are high. The reported values are conservatively adjusted by multiplying all real estate by a factor of 1.33.

Three wealth measures are used: Financial wealth, already defined, risky assets, and total wealth net of liabilities (labeled “Wealth”). Risky assets includes all financial wealth, except bank holdings and money market funds; this measure is quite wide, but has the advantage of being clearly defined. Total wealth comprises real and financial assets net of liabilities, but it is assumed that individuals can only borrow against real assets. If market values are indeed understated in spite of the correction, liabilities are assumed to be associated with real estate only, and are calculated as the maximum of real estate minus liabilities, and zero. At the heart of the analysis is the investors’ stake size. In line with the model in section 1, the proxy for portfolio concentration is investors’ observed stock portfolio value at first year-end, divided by risky assets. The trading data include all stocks traded in Sweden, where the Stockholm Stock Exchange is by far the largest exchange, but also include stocks traded at the small exchanges Nya Marknaden and Aktietorget. The tax reports contain combined market values of all stocks traded at these markets each year-end of the sample. Other equity,

10This maximum rule was imposed on 3% of investors in the sample.
such as foreign stocks, equity mutual funds, options, and warrants are excluded from the analysis with respect to trading, but is included in risky financial wealth. The full trading sample reveals that investors, however, mostly trade in Swedish equity. Trades in foreign stocks, warrants, and options represent only about 6% of the total trades. Around 44% of the investors in the sample owned stocks in other accounts during the year of matching, but on average, the sample of trading records cover 72% of the individuals’ value of stock holdings. Around 64% also held an equity (or mixed) mutual fund, and almost 65% of the investors also owned real estate, which is the most important component of total wealth. In the sample, financial wealth represents around 34% of total wealth including liabilities, which is the typical situation for Swedish citizens.

[Insert Figure 2 here]

The filled, black area of Figure 2 plots the cumulative distribution of stake size across the 10,600 investors in sample. The solid line in the same graph denotes the portfolio value normalized instead with total financial wealth, which includes riskless assets. As can be seen by reading the figure from the left, the observed stock portfolio constitutes a relatively small part of financial wealth for many investors. Around 25% of the investors have stocks worth 5% or less of their financial assets in the brokerage account; for around 50% they are worth 20% or less but constitute about 35% of risky assets. At the top of the distribution, 19% of the investors have 70% or more of their financial wealth invested in stocks in these accounts, and this constitutes all of their holdings of risky assets. 30% of the investors who collected all of their stocks at the broker do not own any other risky assets. The observed stock portfolios therefore represent very different shares of total saving across investors in sample.

Table 1 displays the data divided into quintiles with respect to stake size. Stake size is skewed: the bottom quintile of investors hold only around 2% of their risky financial assets in the broker account, whereas theses assets represent virtually all of the risky financial wealth for top stake quintile investors. Turnover across investors is also very skewed: 15% of investors never trade and the top 10% of investors turn around about 55% of their portfolio
per month (about 6.6 times per year). The average yearly turnover is 113%, which means that investors buy and sell their portfolio roughly once a year: this is almost twice the rate of the total Swedish stock market. There is a strong correlation between turnover and stake size. High turnover investors generally concentrate their portfolios more than those better diversified. Those in the lowest stake quintile, where the relevant assets should matter the least for overall performance, average turnover is around 5%, compared to the highest stake investors where turnover is almost three times higher at 14%. Turnover is decomposed both into buy and sell, where the ratio of purchases and sales are very similar across the stake quintiles. Turnover is, however, almost twice as high in the first compared to the second six month period of trading history for all investors, and this is not explained by differences in purchases and sales across time periods. This could be for different reasons. Investors enter sequentially, and more investors enter the sample during months of high market returns. One concern is therefore that there may be cohort effects in the data, where a sudden interest to trade in the stock market is followed by a period of inaction. Investors could of course also rationally learn that trading is costly, and quickly adjust their trading behavior. Finally, they may be reluctant to realize paper losses due to a disposition effect. Following a similar definition as in Odean (1998a), $PUL$ denotes Proportion Unrealized Losses, and is defined as the number of stocks with paper losses divided by the number of all stocks in the portfolio at the end of month six of their trading history. The average proportion of paper losses is 80% of the investors’ stock holdings at this point in time. The three suggested explanations for the drop in turnover are considered when paper losses and past trading returns are used among other predictors of turnover in section. Finally, high turnover can also be associated with infrequent, but large trades in small accounts. But as the data shows, persistent traders are also much more likely to have concentrated portfolios. Those who trade at least once a month six months in a row (labeled “Frequent Traders”), represent 31% of investors in the top stake size quintile, but only 15% in the smallest. Finally, the averages for the same variables for the subset of the 5,914 investors who have all of their stocks at the broker

---

11 Kaustia and Knüpfer (2011) find and analyze this effect for Finnish investors.
under study is displayed in the last column “All Stock” of Table I. These investors have slightly higher turnover than the overall sample, but also hold more concentrated portfolios on average. The average portfolio value is close to the full sample, even if they on average have lower wealth. The analysis is conducted on the full sample, because investors who collect all of their stocks to the specific broker may not be representative of its total client base. The data suggests that less affluent investors are more likely to collect all of their assets to the broker under study, presumably due to the services it provides. All the main results that follows are also valid for this subsample.

[Insert Table I here]

The portfolio values and number of stocks in the portfolio generally increases with stake size, with the exception of the highest quintile. The largest average portfolio is found in quintile four, and these investors also on average hold 17% equity funds, compared to none in the highest quintile. Investors with the highest stakes are therefore less diversified both with respect to the stock portfolio, and to the extent they invest in mutual funds. The average investor have a weight of 28% in equity mutual funds, but is higher for those in the lower stake quintiles. Similarly, when sorting on turnover rather than stake size, investors in the lowest trading quintile invest 36% of their equity in funds, which is double that of the top quintile with respect to turnover.

Investors with higher stakes have both lower income and much lower wealth. Comparing financial wealth between high and low stake investors, the latter have over ten times higher average financial wealth. In particular, equity funds constitute a large fraction of this wealth: SEK 263 Thousand for low stake investors. High stake investors are also more likely to be male, which is in line with the results of Barber and Odean (2001), who find that women trade more than men. They are also younger: low stake investors are on average 43 years old, whereas high stake investors are 7 years younger. Around 52% of the individuals in this study had completed at least one year of university studies. By this definition of the level of

\footnote{The difference between average risky assets and portfolio size in Table I is due to the matching procedure. For example, quintile 5 investors owned listed stocks at the year of matching worth on average SEK 77 Thousands. This difference is attributed to both the composition and prices of assets as they are measured at different points in time.}
education, there are a smaller number of educated investors among those with high stakes (45%), compared to low stakes (53%). A much smaller number of individuals, 3%, have employment within the financial sector. There is also a negative difference between high and low stake investors here, even if it not very big. The results from the investor characteristics of Table I therefore seem to suggest that high stake investors are less experienced managing their savings: they have lower wealth and financial wealth, and they are younger and less educated.

Finally, a comment on how the data on wealth, income, and portfolio values relate to the Swedish population in general. The average portfolio size is skewed with an average of SEK 86,183; the minimum is SEK 1,000 and the maximum exceeds SEK 19 million. The median stock portfolio in the sample is therefore much lower than the mean, at SEK 18,091. However, this is roughly the same for Swedes in general. Statistics Sweden reports that the median for the whole country is SEK 20,000 and SEK 15,000 at the end of 1999 and 2001. Similarly, diversification is also low for most Swedes. The median Swede only holds one stock. Calvet, Campbell, and Sodini (2007), who document portfolio holdings in Sweden over a similar period, show that stock market participants enjoy an approximately 20% higher income and 40% higher financial wealth than non-participants. The data under consideration match income, net wealth, and stock portfolios quite well to stock market participants among the Swedish population. For example, the average portfolio value for all Swedes is around SEK 82,145, and the median annual income is SEK 277,088 compared with a portfolio value of SEK 86,183 and income SEK 315,249 in the sample. Even if some investors self-select into trading at the discount broker under study, they do not appear to differ in any substantial way in terms of income and wealth. They do, however, have a higher education and are younger in general than Swedes who hold stocks.

3 Trading performance

It is difficult to find a proper benchmark when investors hold under-diversified portfolios. The now standard extensions of Markowitz’s (1952) single-index model proposed by Fama
and French (1992), and Carhart (1997), may fail to capture specific preferences for, for example, skewness that may be desirable to individual investors.

In the spirit of Grinblatt and Titman (1993), we let the investors instead self-select a benchmark portfolio at the beginning of each month, which is taken as given to reflect the desirable return profile. Ruling out liquidity motives, rational investors with correct expectations of future returns should on average only trade if deviations from the self-selected benchmark are profitable. Even in the presence of "mechanical" motives to trade, such as liquidity and risk management, these costs are expected to be small. If investors trade for liquidity reasons, most investors can either sell or buy a mutual fund (as there are very few funds that have loading and exit fees in Sweden), or lend, or borrow cash. If the investor is constrained, so that this is not applicable, it would still be less rational to trade too frequently at high costs, compared with trading once and depositing cash in a bank account. In any of these cases, it is difficult to understand why rational investors trade excessively at high costs.

The approach defining monthly payoffs and returns is as follows. Since the cash account is unobservable on a monthly basis, it is assumed that the investors hold unleveraged portfolios, and the capital base used when calculating returns is only increased if the transactions involve a net increase in funding. Therefore, it is important to take the exact timing of trades into account. A sale that precedes a purchase need not affect the capital base, but the reverse transaction will. Furthermore, the interest on any cash balance that is either needed for financing or for depositing cash must be properly accounted for. A detailed description of how the trading payoff and returns are calculated is given below.

3.1 Calculating trading returns

Let \( x_{n,i,t} \) be the number of shares of a stock \( n \) held by the individual \( i \) at the end of month \( t \). A transaction \( d \) during month \( t \) is denoted by \( x_{n,i,d} \), and super-indices \( B \) and \( S \) indicate whether it is a buy or a sell transaction. Similarly, associated actual purchase and sales prices including fees are denoted \( p_{n,i,d}^B \) and \( p_{n,i,d}^S \) for each of these transactions. In what
follows, we also need the closing price for stock $n$ on the last day of month $t$, which is labeled $p_{n,t}^C$. The stock position for individual $i$ at the end of month $t$ is

$$x_{n,i,t} = x_{n,i,t-1} + \sum_{d \in t} \left( x_{n,i,d}^B - x_{n,i,d}^S \right),$$

which is the position at the beginning of month $t$ plus the sum of buys and sells during the month (hereafter net purchases for short). In what follows, we will impose the restriction that $x_{n,i,t} \geq 0$, meaning that investors are not allowed to have outstanding negative positions at month-end.

### 3.1.1 Payoffs

Trading, position and total payoffs for each stock and individual are as follows. The position payoff is defined as

$$\Pi_{n,i,t}^P = x_{n,i,t-1} \cdot (p_{n,t}^C - p_{n,t-1}^C),$$

which is simply the position at the beginning of the month times the change in price. The trading payoff in stock $n$ for individual $i$ during month $t$ is given by

$$\Pi_{n,i,t}^{TR} = \sum_{d \in t} p_{n,i,d}^S \cdot x_{n,i,d}^S - \sum_{d \in t} p_{n,i,d}^B \cdot x_{n,i,d}^B + \sum_{d \in t} (x_{n,i,d}^B - x_{n,i,d}^S) \cdot p_{n,t}^C.$$

The first and second component of (8) state the net sales revenue of stock $n$ in month $t$, which is the value of sells minus buys at actual transacted prices. The third component of (8) adjusts payoffs by the value of net purchases that are already accounted for by (7). Deposits of stocks are assumed to be transacted at the beginning of the month and redemptions at the end. Therefore, $x_{n,i,t-1}$ also includes all deposits of stocks made during the month. Investors are allowed to short sell their stock with these definitions because the summation is invariant to the ordering of purchases and sales. The restriction only means that a positive holding of each stock is required at the end of the month\footnote{In the sample, this proved to be a minor problem as there were only 34 instances of open short positions at month’s end. These open positions were effectively closed by dating the corresponding future buy transaction the following month, $t + 1$, as belonging to $t$.}.
Total payoff for each investor \( i \) in stock \( n \) is the sum of trading and position payoff

\[
\Pi_{n,i,t} = \Pi_{n,i,t}^{TR} + \Pi_{n,i,t}^{P}.
\]

To find the payoff for the whole portfolio, we sum over \( n \) to obtain total portfolio payoff for individual \( i \) in month \( t \)

\[
\Pi_{i,t} = \sum_{n} \Pi_{n,i,t} = \sum_{n} \Pi_{n,i,t}^{TR} + \sum_{n} \Pi_{n,i,t}^{P}.
\]

### 3.1.2 Capital components

As with payoffs, we distinguish between position and trading capital as follows. Position capital is defined as

\[
C_{i,t}^{P} = \sum_{n} (x_{n,i,t-1} \cdot p_{n,t-1}^{C})
\]

which is simply the value of all stocks in the portfolio at the beginning of the month. The amount of capital engaged in trading is determined by listing each transaction in chronological order. The traded value of any sale or purchase on any day is

\[
TV_{i,d} = \begin{cases} p_{i,d} \cdot x_{i,d}^{J} & \text{if } J = S \\ -p_{i,d} \cdot x_{i,d}^{J} & \text{if } J = B \end{cases}
\]

such that it represents the revenue of any sales and cost of any purchase. The trade values are ordered during the month from beginning to end for each investor regardless of which stock is traded, and the cash balance is calculated at each point in time. The lowest cumulative cash balance in month \( t \) is the minimum amount needed to finance the portfolio without leverage, and is written

\[
C_{i,t}^{TR} = -\min_{d} \left[ \sum_{d \in t} TV_{i,d}, 0 \right],
\]

and expressed as a positive number since we pick out the largest negative cash balance.

Total capital is the sum of position capital and trading capital,

\[
C_{i,t} = C_{i,t}^{P} + C_{i,t}^{TR}.
\]

Therefore, the capital base only increases if trading incurs additional funding. But this is
desirable, since the investor who reallocates her investment without using additional funds will have the same capital base.

### 3.1.3 Excess returns

Excess returns are created as follows. It is assumed that the investor can borrow and deposit cash at the available 30-day T-bill rate, \( r_{t-1}^F \), in order to finance the portfolio. The interest that is attributable to the position component, \( I_{i,t}^P \), is calculated as the cost of borrowing the value of the portfolio at the beginning of the month, i.e, the first part of equation (14).

If trading occurs, we seek the net interest paid on trading capital during the month. Interest is calculated on each transaction and summed over the month creating the revenue \( I_{i,t}^{TR} \) that corresponds to the interest that is attributable to the actual timing of purchases and sales.

The excess return is therefore

\[
R_{i,t} = \frac{\Pi_{i,t} + I_{i,t}^P + I_{i,t}^{TR}}{C_{i,t} - \min \left[ I_{i,t}^{TR}, 0 \right]},
\]

where \( I_{i,t}^P \) is always 0 or negative, and \( I_{i,t}^{TR} \) is negative if there is a net cost associated with financing the monthly transactions. When trading capital is 0 but the investor is net selling, \( I_{i,t}^{TR} \) represents the interest earned on investments that are sold from the portfolio. In this way, timing of the sale is properly accounted for since positive interest is added to the return measure.

The interest on trading capital is only added to the capital base if it is negative. This is because it is assumed that interest earned is paid at the end of the month, but any costs must be covered by capital at the beginning of the month. It therefore ensures that returns are bounded at \(-1\). In the case of no trading, \( I_{i,t}^{TR} = 0 \), we obtain the familiar definition of excess returns, which is

\[
R_{i,t} = \frac{\Pi_{i,t} + I_{i,t}^P}{C_{i,t}^P} = r_{i,t} - r_{t-1}^F.
\]

\(^{15}\)Two assumptions apply: borrowing and lending rates are the same and the effect of compounding during the month is ignored.
3.1.4 Passive returns

The passive return is the return of the portfolio excluding trades:

$$R^P_{i,t} = \frac{\Pi^P_{i,t} - I^P_{i,t}}{C_{i,t}},$$

(17)

which is the position payoff divided by total capital corrected for interest. As the own benchmark return measures the return on a portfolio that excludes trading, we see that (15) and (17) coincide,

$$\Pi_{i,t} = \Pi^P_{i,t}.$$  

(18)

The passive return measure uses total capital as a base. It is therefore assumed that whatever funds used for net investments during the month are invested at the risk-free rate. The investors can only deviate from the benchmark by trading. During the month, investors can move in and out of the market as a whole or change their stock allocations. If these tactical changes in risk and reallocations are profitable, investors earn a higher excess return on the traded portfolio than their own benchmark portfolio does. Trading return is therefore defined as

$$R^{TR}_{i,t} = R_{i,t} - R^P_{i,t},$$

(19)

which is the return difference between an active and passive investment strategy. Finally, turnover cost, $TOC_i$, measures the profits per unit of traded dollar,

$$TOC_i = \frac{\sum_t (\Pi^{TR}_{i,t} - I^{TR}_{i,t})}{\sum_t \sum_d TV_{i,d,t}},$$

(20)

and is therefore an alternative measure of trading profitability, since it is normalized by turnover instead of portfolio value.

[Insert Table II here]

Table II summarizes the average returns across quintiles of stake size. The average investor has a negative trading return of around 27 basis points per month, which translates into a negative performance of around 3% per year. Investors in the highest stake size quintile lose 41 basis points per month, or about 5% per year by trading away from their benchmark portfolio. The difference between the high and low stake investors is strongly
significant, but average trading returns are quite similar for the lower quintiles. Fees are important for overall trading costs. The second row of Table II reports average trading returns excluding fees. If excluded, the remaining cost represent bid-ask spreads and intra-month market timing effects. Even though fees should be known to investors, they may not be as transparent, making such costs a much less salient feature of trading. Even after fees, the average trading return is -8 basis points (-1% per year). Roughly two-thirds of the average underperformance can therefore be attributed to fees. The returns are quite similar across quintiles, but those with the highest concentration of portfolios have considerably lower performance at -15 basis points, which is almost 9 basis points lower than for those with least concentrated portfolios. The difference in performance when excluding fees is 10 basis points lower for the high stake investors, which means that the remaining trading cost is a more important ingredient of the total cost for them. Inferior performance is therefore not explained by fees only, especially for high stake investors.

The turnover cost is also reported including and excluding fees, and is a measure of the cost per dollar traded. One can therefore interpret this measure as a normalized return, which is independent of portfolio size. On average, investors pay 2.4% for every dollar traded, which at first sight seem very high. The broker’s fee structure, as many others, involve a percentage of the value of the trade but with a minimum floor. As most portfolios are small, many transactions are priced at the minimum fee. The average fee paid per transaction is around 1.7%, which is in line with yearly management fees for most mutual funds. The turnover cost is generally decreasing in stake size, meaning that high stake investors are better at implementing their trades. But since they have a higher turnover in relation to the value of their holdings, their trading return is lower. This difference broadly survives after fees, suggesting that this result is not only attributed to fees. Overall, the data suggests that much of the cross-sectional variation in performance can be explained by portfolio characteristics, such as size, but may also be related to skills. The next section investigates

---

16 In unreported results, we find that the average return for those who had all of their stock holdings at the broker is slightly more negative with a smaller t-statistic, which is explained by a higher dispersion in performance within this group.

17 The minimum fee during the sample period was SEK 89 per trade, or approximately USD 10. The lowest percentage fee is 5 basis point per trade.
trading performance in detail while controlling for these features.

4 Results

Table I reveals a strong correlation between turnover and stake size. The analysis of this relationship proceeds as follows in four steps. First, we analyze the correlation between the variables, where we treat stake size and turnover as “choice variables” and then investigate how they are jointly related to investor characteristics, which we in turn treat as pre-determined. The main motivation for this analysis is to see to what extent the pre-determined variables explain the common source of variation for the choice variables. There is much previous research linking overconfidence to trading and investor characteristics, and the interest here is to see if also stake size is a useful proxy for this purpose. Second, we investigate to what extent stake size can predict turnover, assuming that stake size is a time-invariant proxy for biased self-attribution. In order to avoid the simultaneous decision problem, we regress future turnover on past turnover, as well as on stake size and other investor characteristics. This is a stronger test than the correlation analysis, since much of the explanatory power of investor characteristics already would be included in lagged turnover. The regression specification also allows for time dummy controls, prior losses, and past trading returns, which all may be important drivers of turnover as investors enter the sample sequentially and therefore may have different trading motives and abilities to adjust their trading based on prior outcomes. In the third step, the relation between trading returns and stake size is investigated while controlling for turnover. If stake size is a valid proxy for overconfidence, we should be able to rule out that trades originating from these investors reflect informed trading, measured by trading performance. Then, we investigate how much of the trading performance that can be attributed to turnover directly, and if there is an additional component of performance that can be attributed to stake size. Investors with concentrated portfolios could potentially be better traders on average, after controlling for their higher turnover. Finally, we aggregate trading profits and losses across the stake size quintiles in order to establish how trading costs and profits are distributed, and how they
relate to income and our measures of wealth.

4.1 Correlation analysis

The broad correlation between stake size and turnover from Table I is visualized in Figure 3. Investors are sorted on stake size on the horizontal axis, and average turnover displayed on the vertical axis. Two non-parametric methods are used to display the relation between the variables, and at the same time allow for the most flexible functional relationship: arithmetic averages, and a local regression method (LOESS). The solid double line depicts the average turnover for the stake size quintiles in Table I. The shorter segments of the line indicates that there is a concentration of investors having either small stakes (quintile 1), or large stakes (quintile 5), but also that there is considerable variation between these two extremes. A LOESS estimation of the relationship is included and plotted by the solid in the same graph, along with dotted 99% confidence bands. The regression reveal that there is some non-linearities in the regression, particularly at the end of the stake size distribution. Both estimates, however, show that the correlation is robust throughout the range of stake size across investors.\footnote{The corresponding graph for the subset of investors with all of their stock holdings at the relevant broker account is very similar, and so does not explain these cross-sectional differences.}

[Insert Figure 3 here]

The previous literature, as in Barber and Odean (2000) and others, suggests that excessive stock trading is mainly driven by overconfidence. Subsequent research have tried to link overconfidence to certain investor characteristics, such as gender, wealth, age, and education, which usually are thought of being some proxy for “investor sophistication”. Less sophisticated investors are then believed to be more easily prone to behavioral biases, such as overconfidence. The fundamental problem of identifying overconfidence through stock market behavior is that it only can be suggestive: there is little evidence of any direct links between overconfidence and trading (Glaser and Weber (2007) being an exception). The main hypothesis of this paper is that there is a positive correlation between under-diversification and trading, which is ultimately driven by a single unobservable factor: overconfidence (or
more broadly by self-attribution bias). It the hypothesis is true, we should expect that (i) investor characteristics that have previously been associated with increased turnover, loads similarly in specification for stake size; and (ii) there is indeed a single component driving most of this correlation. The analysis that follows attempts to establish statistical evidence in support for these implications.

The correlations between the choice variables stake size and turnover (labeled “dependent”) and investor characteristics are presented in Panel A of Table II. As expected, one finds that both age and education is strongly positively correlated with wealth. In line with the results of Table II, stake size is negatively correlated with wealth, age and if the investor is woman or have higher education. The corresponding correlations are significant, and of the same sign, between turnover and the characteristics. The raw correlation between stake size and turnover is also high at 24%. In Panel B of Table II, turnover and stake size is independently regressed onto the set of characteristics. The regression coefficients are significantly negative for wealth, women, and university education in both specifications, where the isolated effect of education is about one third of the gender effect. The coefficients for age are different between the specifications: it is positive for turnover, but negative for stake size despite that the raw correlation between age and turnover is negative. Unreported results suggests that age is non-linear in turnover, and peaks around 40 year-olds, to then fall with age. Including a non-linear term for age did not add much to the results, and so was dropped in the analysis for brevity. The overall fit of the regressions are low as evidenced by the R-squares. The explanatory power is, however, much larger for stake size than turnover, which may suggest that stake size better capture the common variation among the characteristics. The correlation between the residuals of the two regression are also still high at 21%, which indicates that turnover and stake size share a common component that is unaccounted for in the separate regressions.

[Insert Table II here]

The next step in the correlation analysis seeks to establish how stake size and turnover are jointly related to the common component that maximizes the correlation between the
characteristics and choice variables. The most general technique of analyzing multiple outcomes and dependent measures is Canonical Correlation Analysis (CCA), which in turn is closely related to Principal Component Analysis. The procedure groups the dependent and independent variables and finds a number of common roots, or canonical functions, that maximize the optimal linear combination between them.\footnote{For a more complete discussion of CCA, see Hair Jr., Anderson, Tatham, and Black (1998).} The system at hand have two outcomes and five independent variables (although no assumption about causality is needed for this analysis), which then limits the maximum number of common roots to two. The canonical R-square (untabulated) reveals that the overall fit between the groups of variables is 9% for the first canonical function, which explains over 91% of the common variation. The existence of a second function could not be statistically rejected, but since it does not contribute much to the total correlation, it can quite safely be disregarded in the remaining analysis. Panel C of Table \[\text{II}\] displays the canonical correlations of this function. The most important variables among the characteristics that constitute the first function is wealth, age, and university education. Stake size is strongly negatively correlated to this common factor constructed by the characteristics, measured to be -0.29. In comparison, the correlation between the common characteristic component and turnover is only -0.01. Among the dependent variables, stake size is most strongly correlated to their common component, and the same variables that load strongly onto the characteristics are found to be associated with the choice variables.

The overall conclusion of the CCA is that there is virtually only one common component driving the correlation between the choice variables and the characteristics. Stake size is the choice variable that is most strongly associated with this common root. Since turnover is also widely associated with the same variables, the results suggest that stake size and turnover is determined by the same underlying mechanism given by the characteristics. The statistical analysis can only reveal this correlation, but not identify the underlying source driving this behavior. Since overconfidence has been the most common interpretation of explaining individual investor trading, it seems natural see the results in this way. High stake investors are on average overconfident in their abilities to invest successfully, and
they trade more. They possess lower wealth, they are younger, are more likely to be men compared to women, and have lower education compared to those with less concentrated portfolios. The weaker results for turnover, is not itself indicative that high turnover is a bad proxy for overconfidence, but it is a much more noisy relationship. The results here reveals that turnover is dominated by stake size in explaining the relation with the investors’ characteristics from a pure statistical viewpoint.

4.2 Predicting turnover

The previous section documents a high contemporaneous correlation between stake size and turnover, which is consistent with the hypothesis that the outcome of these variables reflect a simultaneous decision. A potential concern, in which investors choose to enter the market endogenously, is that stake size may capture market timing rather than a cross-sectional feature of investors. A regression of stake size on turnover, however, needs some further identifying assumptions. We assume that stake size is a time-invariant proxy for investor confidence and then predict individual $i$’s average future turnover in month 7 to 12 ($TO_{i,7-12}$), while controlling for average past turnover the first six months ($TO_{i,1-6}$).\footnote{We use the transformation log(1+turnover) throughout the regression analysis of turnover and performance, which reduces the noise and influence of outliers. In a previous version of this paper, we also estimated a censored model in order to account for that turnover is bounded at 0, which did not change any of the key results. The results of running the regressions with those who actually traded in the first period only are similar to those reported for the full sample.} On average, we would expect stake size to be positively related to future turnover, even after controlling for their initial trading, if they are more confident. On the contrary, after controlling for time effects, one would not expect any significant relation between stake size and turnover if the correlation is purely mechanical, or determined by market conditions. The baseline regression predicting average turnover is

$$TO_{i,7-12} = \sum_{\tau=t}^{T} D_{\tau} + a_{1} TO_{i,1-6} + a_{2} PUL_{1-6} + a_{3} R_{i,1-6}^{TR} + \sum_{k=1}^{K} b_{k} \cdot X_{k,i} + \varepsilon_{i}, \quad i = 1, \ldots, N,$$

(21)
where $PUL_{1-6}$ denote the proportion of unrealized losses in the portfolio at the end of month six, and $RTR_{1-6}$ is the first six month trading return. There are therefore three lagged variables that are included to determine how the cross-sectional variation in individuals' past trading experience affects future turnover. The investor and portfolio characteristics from Table II are simply denoted $X_{k,i}$, which includes stake size, log portfolio size, number of stocks, log age, and a dummy for investors being woman, having financial occupation, and completed at least one year of university studies. We control for investors entering at the same point in time by the coefficient $D_{\tau}$. Each of these cohorts are therefore allowed to have different average turnover. Finally, we test interactions between some of the key variables in (21). We test if the point estimates of paper losses, trading returns, and stake size is different for subsets of investors in order to add further support for the hypothesis that stake size is related to behavioral traits.

[Insert Table IV here]

The results of regression specified in (21) are displayed in Table IV where the first column shows that both stake size and paper losses (PUL) are significant predictors of turnover, controlling for investor characteristics. Paper losses, however, does not capture much of the cross-sectional variation. In unreported results we find that stake size alone explain about 3% of investors second half-year turnover, and by including paper losses, this figure raises to 4%. The point estimates imply that those with all of their risky wealth represented by stocks at the broker have an average monthly turnover that is 7% higher than those with a negligible stake size of around 0. Similarly, those who have portfolios with all paper losses have 5% lower turnover in the second half year compared to those with no losses.

Model II includes past turnover and trading return among the independent variables. The explanatory power increases sharply from 8% to 41% as a consequence of including lagged turnover, but trading return is insignificant. Turnover is persistent, where the coefficient for lagged turnover is almost 0.5. Including lagged turnover also crowds out many of the characteristic effects, such as, portfolio size, number of stocks, wealth, and university education. The coefficients for stake size and paper losses remain significant, but there is no
evidence that past trading returns predict future turnover for the full sample. When allowing
for asymmetric effects for negative and positive returns, we find only very small differences
suggesting that past trading returns would operate differently in these domains. Rather, the
dominating mechanism is persistence, which suggests that adjusting present trading based
on past behavior is much slower than can be captured in the relative short panel under
study.\footnote{Barber, Tsung Lee, Jane Liu, and Odean (2010) do find some evidence of learning, but also high
persistence even for underperforming day traders in Taiwanese data.} Further, there is considerable noise in the data, which makes learning even more
difficult to detect. The results from the full sample may also hide differences between groups
of investors, where some respond better to the performance signal than others.

A stronger test of the hypothesis that stake size, trading return, and paper losses really
depends on investor sophistication is to interact them with other variables. The previous
literature gives some guidance of which investors can be regarded more sophisticated in that
they should be less influenced by behavioral factors: women trade less and perform better
than men, and better educated, wealthier investors tend to make less investment mistakes.\footnote{See, for example, Barber and Odean (2001) and Calvet, Campbell, and Sodini (2009).} High stake investors could presumably also be more sensitive to paper losses than to trading
returns if they follow naive investment strategies. In general, the prediction is that stake
size should interact negatively with proxies for investor sophistication, but interactions with
trading return and paper losses should be positive (adjusting turnover to past returns and
counteracting the disposition effect).

The interaction between stake size, paper losses and trading returns is reported in Model
III. The interaction between stake size and paper losses is highly significant, and economically
important. The coefficient estimate can be interpreted such that those with the highest stakes
who hold portfolios of paper losses only decrease their turnover of over 6% between the two
periods. This may not only be an effect attributed to behavioral biases. Given that people
are loss averse, we may expect paper losses to be more important or salient for high stake
investors, because their investment represents a larger fraction of their risky wealth. If this is
the case, however, it is puzzling that the coefficient for lagged trading return is insignificant,
and even negative. Rational high stake investors should put more weight on past trading

\footnote{Barber, Tsung Lee, Jane Liu, and Odean (2010) do find some evidence of learning, but also high
persistence even for underperforming day traders in Taiwanese data.}

\footnote{See, for example, Barber and Odean (2001) and Calvet, Campbell, and Sodini (2009).}
returns than paper losses.

Model IV through VII includes four other categories of investors: Wealthy, Women, University educated, and Sophisticated investors. Those classified as wealthy belong to the two top wealth quintiles, and we label them sophisticated if they are also university educated. The parameter estimates shows that wealthy investors respond positively to past trading return, meaning that they adjust their turnover based on past performance. We do not find that the interactions for paper losses and stake size are significant for this group. Women also respond positively to past trading returns, and they have on average a lower sensitivity to stake size. We find that the effect of paper losses for women is significantly positive, and the magnitude suggests that women in fact are quite insensitive to paper losses in the data. The results for university educated investors are of the same sign, but much weaker where none is statistically significant. Finally, the last regression in Model VII shows the results of the university educated, wealthy investors. We find that these investors respond significantly positively to lagged trading return, and significantly less to paper losses, although not for stake size.

To conclude, the overall results show that stake size is an important explanatory variable in the cross-section for predicting future turnover. The combined evidence supports the hypothesis that stake size is also associated with investor sophistication, which is expected under the hypothesis that stake size represents a form of overconfidence. High stake investors are more sensitive to paper losses. The marginal effect of stake size is weaker for groups of investors that are presumably better educated or less sensitive to behavioral biases, although it is only significantly different for women. We find evidence that some investors rationally adjust trading in response to past trading returns, but this effect is only supported in the results for wealthy and educated investors. Most of the cross-sectional variation in future turnover is attributed to lagged turnover and time-effects, and the observed drop in turnover in the second period of the sample is therefore not easily explained by either loss-aversion or rational learning.
4.3 Trading returns

In the next step of the analysis, we investigate how the characteristics are related to performance in the cross-section. The existing literature suggests that turnover itself is the main channel of under-performance. The model for under-diversification gives the prediction that high stake investors trade more, but does not give any sharp predictions about performance while controlling for trading. Under the hypothesis that overconfident investors trade more with higher stakes, we would therefore expect trading performance to be negatively related to stake size. By including turnover in the regression, we can efficiently separate the direct effect on performance that go through trading from the indirect effect attributed to stake size. Therefore, the theoretical justification for this analysis is to investigate if high stake investors perform below average even after controlling for turnover. This would add further support to the hypothesis that high stake investors are less sophisticated, and actually earn returns below average not only through increased trading, but due to indirect costs of trading, such as bid-ask spreads, market timing, and stock selection.

Trading performance is estimated in the panel by the following regression specification

\[ R_{i,t}^{TR} = a + \sum_{k=1}^{K} b_k \cdot X_{k,i,t} + \varepsilon_{i,t}, \quad i = 1, \ldots, N, \]
\[ t = 1, \ldots, T, \tag{22} \]

where \( R_{i,t}^{TR} \) denotes the monthly individual trading return, data \( X_{k,i,t} \) denotes the characteristic and time-series variables with associated coefficients \( b_k \). The system contains \( 1 + K \) parameters and \( i = 1, \ldots, N \) individual portfolio observations over time \( t = 1, \ldots, T \). There are 10,600 \( (N) \) investors and 35 \( (T) \) months in the data. Equation (22) implies the following \( 1 + K \) sample moment conditions,

\[ \frac{1}{T} \sum_{t=1}^{T} f_t(\theta) = \begin{bmatrix} \varepsilon_{1,t} \cdots \varepsilon_{N,t} \\ \varepsilon_{1,t}X_{1,1,t} \cdots \varepsilon_{1,t}X_{N,1,t} \\ \vdots \\ \varepsilon_{1,t}X_{1,K,t} \cdots \varepsilon_{1,t}X_{N,K,t} \end{bmatrix}' = 0 \tag{23} \]

where \( \theta \) contains the \( K + 1 \) parameters to be estimated. The point estimates can be retrieved
by OLS, but this GMM specification allows for a Newey and West (1987) robust variance-covariance matrix formed from the moment conditions\textsuperscript{23}

The result of the regression specified by (23) is presented in Table V, where all independent variables have been centralized in order for the intercept to express the average trading return in the cross-section. The first three regression models use the characteristics only, where Model IV also includes turnover among the independent variables. Therefore, the coefficient for stake size indirectly includes the higher propensity to trade among high stake investors in the first three specifications. Controlling for portfolio size, the estimate in column II suggests that stakes actually have a stronger marginal effect, since smaller portfolios pay higher fees in absolute terms. After controlling for turnover in model IV, the point estimate for stake size reveals that high stake investors in the fifth quintile have about 5 basis points lower trading performance per month compared to the low stake investors in the first quintile. Turnover itself has a large impact on performance. The coefficient estimate implies that 1% higher turnover above the mean translates into 2 basis points lower trading performance per month (24 basis points per year). This could be compared to the average underperformance in sample which is 29 basis points, or 3.5% per year, annualized. The gender effect is of about the same magnitude as the measured effect of stake size, where women outperform men by 4 basis points per month. University education also have a distinct positive effect on performance: these investors perform around 7 basis points better per month than those without university education.

The models in column V and VI repeats the regression IV, but for the highest and lowest wealth quintiles. On average, low wealth individuals have considerably lower average trading performance, -47 basis points (5.8% per year), compared to high wealth investors, which is -18 basis points (2.2% per year). Stake size is insignificant in both subsamples, suggesting that wealthy investors with higher stakes do not outperform their benchmark. Turnover itself is much more detrimental to trading performance for low wealth investors. The marginal effect of an increase in turnover is around twice as high for low wealth individuals compared

\textsuperscript{23}The panel is unbalanced, since investors enter sequentially over time. See Bansal and Dahlquist (2000) for how this is incorporated into a GMM framework.
to the whole sample. The marginal effect of portfolio size is also much greater for low wealth investors, which emphasizes the negative effect of fees in small stock portfolios. The positive effect of wealth on performance, is however mainly associated with low wealth investors. There is a small, but significantly negative, effect of holding more stocks in the portfolio. On average, investors lose 2 basis points trading return per additional stock, and this effect is only significant when controlling for investors’ wealth and turnover. Model V and VI suggest that this effect mainly can be attributed to low wealth investors and disappears when fees are excluded. The positive effect of specialization seem therefore mostly be due to lower trading costs in small portfolios, which is achieved by concentrating holdings to fewer stocks. Turnover has also a more detrimental effect on less sophisticated investors. Those labeled sophisticated underperform in sample, but only about half as much as the full sample average, since the average monthly trading return is -14 basis points. Unsophisticated investors, on the other hand, have an average trading return of -41 basis points.

The positive performance effect of portfolio size can mainly be attributed to fees (column XI). Investors with larger portfolios pay proportionally lower fees, and the positive influence of portfolio size vanishes when measuring performance excluding fees. The effect of turnover is insignificant once fees are removed from performance. The negative effect of stake size on performance is halved and insignificant when fees are excluded from the analysis. Overall wealth, however, has a positive influence on performance even after fees. The investor at the median wealth level earn about 3 basis points lower performance compared to the investor at the average wealth level. University educated individuals have 6 basis points higher monthly performance (about 0.7% per year) even after fees, and women about half of this amount. Older investors have lower performance on average even after controlling for turnover, but the effect is relatively small. The point estimate suggests that there is a 2 basis point difference in performance for a 10 year difference in age.

The absence of significant results excluding fees raises the question if investors are exposed to a saliency bias, where they trade without taking fees explicitly into account. By neglecting

---

24 Average portfolio size is SEK 22 (SEK 222) Thousands in the low (high) wealth quintiles.
fees, investors may simply understate the true cost of trading. The most active investors would be those mostly exposed to fees. Column VII shows that the top quintile of investors sorted on turnover lose 71 basis points per month by trading. Excluding fees for this group, we find that underperformance falls to 31 basis points, or about the same level as the full sample with fees included (column XII). These investors clearly lose considerable performance on average even after fees, and the coefficient for stake size is negative and strongly significant even after controlling for fees and turnover. There is no significantly negative effect of turnover itself for this group. Investors with higher stakes in the highest turnover quintile therefore perform worse due to other trading costs, which are unrelated to fees and turnover (for instance, market timing and stock selection). Saliency bias therefore do not appear to be the main motivation for excessive trading, even if this may be an important feature for explaining trading by broader investor groups.

Across the different specifications, the evidence for stake size and performance can be summarized as follows. Firstly, stake size has a clear negative impact on performance without controlling for turnover. Once controlling for turnover, the effect is still significant, but weak. This is evidence that most of the average underperformance for high stake investors go through the channel of turnover. Secondly, statistical evidence for this underperformance does not survive in treatments of subsamples, and there seem to be little differences in the impact of stake size with respect to investor sophistication. This piece of evidence suggests that one should interpret stake size as a separate channel for underperformance with some caution. Finally, when investigating high turnover investors, stake size has again a distinct effect above what is explained by turnover. One potential explanation for this result is that the effect of turnover is much lower for investors in this group, as can be seen from the estimates of turnover in column VII and XII. The performance for the average high turnover investor is therefore not as sensitive to turnover, but those with high stakes perform considerably worse.

In conclusion, there is clear evidence that the trading performance of high stake investors is below average. The main channel for this underperformance lies primarily in their higher
turnover, although the data suggests that there might be a small, additional component associated with stakes that deteriorate performance.

4.4 Aggregated trading profits

Investors with concentrated portfolios trade more and perform worse than average. Since trading profits also can be measured in absolute terms, we can analyze how gross profits are distributed across investors. This dimension is important, since stake size itself is associated with characteristics that commonly relates to investor sophistication. The gross trading profit for each investor is aggregated separately for gains and losses within each stake size quintile, and displayed in Table VI. Similarly, financial wealth, total wealth, and income are also aggregated, and total profits expressed as fractions of these measures. The results from trading returns directly carry over to this table, since wealthier investors with lower stakes both trade less and do better when they trade. The sixth row of Table VI reveals that investors in the lowest quintile of stake size are over three times as wealthy as the twenty percent in the top quintile, and possess over ten times as much financial wealth. In addition, the net trading loss is much larger for the higher quintiles. About 78% of the total loss from trading is confined to the two highest stake size quintiles, where investors in the highest alone carry 40% of this value. These investors hardly own any equity mutual funds at all, but instead concentrate their holdings to a few stocks. Therefore, the loss incurred by trading is carried by those who can least afford them. Trading losses represent 3.77% of financial wealth and 1.42% of yearly income in the highest stake size quintile. As a comparison, Swedish households with a similar income spend, on average, 14.6% of their annual disposable income on recreational activities. They spend 1.2% on sports and other hobbies, 2.4% on books and newspapers, around 2% on alcohol, and 2.6% on dining in restaurants. Trading losses are therefore far from negligible in this context.

The aggregated results do not explain how trading gains and losses are distributed on

an individual basis. In all, 57% of the investors have negative trading profits. Almost 38% (31% when fees are excluded) of the individuals have trading losses representing 1% or more of their financial wealth, and 20% over 5%. Still, about 8% of investors gain more than 5% annually from rebalancing their portfolio to their self-selected benchmark, measured as a share of total financial wealth. The effect is similar, but less dramatic, when trading profit is normalized with income: 26% lose 1% or more of their yearly income by trading, and around 8% more than 5%. Even though many investors incur relatively modest losses, there are certainly those who would feel that they matter.

Finally, investors trade more in the beginning of their careers than later. Trading costs would therefore seem to become lower going forward for many of these individuals. On the other hand, the analysis suggests that the trading frequency is higher, and performance lower, for those with higher stakes. Past trading returns are insignificant, while paper losses matter for predicting turnover among these investors. Taken together, we interpret these findings as evidence for unsophisticated investors being more prone to behavioral biases. Their behavior would in that case be more dependent on general market conditions, rather than their historical performance.

5 Conclusion

Many investors trade too frequently and most of them perform worse than their self-selected benchmark portfolio. Their behavior is puzzling, and difficult to rationalize using normative financial theory. Some explanations have been offered: investors may be overconfident and have an illusion of control, as suggested by Barber and Odean (2002), or a biased self-assessment, as evidenced by Glaser and Weber (2007), and Graham, Harvey, and Huang (2009). They may be sensation-seeking, as proposed by Grinblatt and Keloharju (2009), or simply inexperienced at financial decision making and thus unaware of the common pitfalls of stock investment.

This paper suggests that trading and under-diversification is related in that individuals who believe they are above average traders will choose to concentrate their investments into
a few stocks which they think will outperform the market. The empirical evidence shows that investors who hold concentrated portfolios (or “high stakes”), which contains little or no mutual funds and only but a few stocks, on average trade more than others. These individuals share many of the same characteristics as those who are more prone to stock trading: they have on average lower wealth and education, and are predominantly male. Those with high stakes do not perform better, as would be a rationale for concentrating assets, but actually worse when they trade. Taking this evidence together, we argue that their behavior is best explained by overconfidence.

An alternative explanation for excessive trading is that some investors enjoy trading a small part of their portfolio for their entertainment, or that they may find it rational to learn about financial markets through small investments before making larger ones. We find both these explanations unlikely for the following three reasons.

First, financial wealth is ten times greater for investors in the low, compared to high stake size quintile. If trading is for entertainment, it seems costly. In contrast, the top 100 wealthiest investors in sample own stocks worth SEK 1.3 Billion in total, of which only about 6%, or SEK 78 Million, is invested into the brokerage accounts in this study. These investors have about 2% higher turnover in their portfolios compared to average investors. This suggests that entertainment-driven stock trading may be a feature of the data, but hardly the main explanation for the amount of trading, and the cross-sectional differences in turnover, we observe.

Second, stake size predicts future trading even when controlling for the previous trading history and investor characteristics. As documented by Odean (1998a), we also find that unrealized portfolio losses is a significant determinant for turnover for those with more concentrated portfolios. On the other hand, we also find that wealthier and better educated investors are less sensitive to paper losses, but do respond to past trading performance as they trade going forward. There are therefore important differences in how investors adjust their behavior. It is tempting to draw the conclusion that unsophisticated investors are slower to learn, but this may be too hasty due to the short time-frame under study. The
results are presumably better explained by sophisticated investors starting their investment career at a higher level of knowledge.\footnote{For much more comprehensive treatments of learning by individual investors, see Seru, Shumway, and Stoffman (2010) and Nicolosi, Peng, and Zhu (2009).}

Third, aggregate losses can be substantial for high stake investors. The top quintile of investors with the most concentrated portfolios lose 3.77\% of their financial wealth on average per year in the aggregate, which represents 1.42\% of their income. The averages hides the fact that some investors lose considerably more: on an annual basis, 20\% of the investors lose more than 5\% of their financial wealth due to trading. Since the panel spans a relatively short time period of three years, these figures may be higher than the long-run average cost of trading. The analysis points to that loss-aversion, measured by paper losses, mitigates trading to some degree. We find that especially those with lower wealth and education reduce their trading when their stocks run into losses. On the other hand, the same investors appear much less sensitive to past trading returns. We interpret these findings as evidence that less experienced investors are more prone to behavioral biases, but these may not necessarily offset each other. Unsophisticated investors may continue trading at an even higher rate during more normal time periods, when the market does not fall as much as during our later sample period. Furthermore, the results may also be overstating costs, in that some investors choose to self-select into trading at the discount broker under study, and are therefore not representative of for broader group of investors. This is only possible to verify with more data. The documented cross-sectional effects are however strong, and highlights that there are still important differences how the costs from trading are being shared.

The previous literature surrounding investor behavior finds that naïve or unsophisticated investors are more prone to make investment mistakes. Our main contribution is to show that a significant proportion of investors at the discount broker under study neglect two of the most basic cornerstones of financial advice: to diversify well and to trade prudently. These mistakes, together with investors’ socio-economic background, adds to up a particularly toxic combination: the measured trading losses are carried by those who can afford them the least.
References


Barber, Brad M., Yi Tsung Lee, Yu Jane Liu, and Terrance Odean, 2010, Do day traders rationally learn about their ability?, Working paper, University of California at Berkeley.


Table I: Mean statistics sorted by stake size

The 10,600 investors are grouped into quintiles based on stake size, defined as portfolio size divided by total risky financial wealth. Average turnover is the sum of investors’ sales and purchases divided by two times the portfolio value (also reported separately for purchases and sales). Portfolio size is the value of the portfolio at the end of the first month of trading, which is also when the variables Age and Number of stocks in the portfolio are determined. All other variables are defined at first observed year-end. Stake size is the portfolio value divided by total risky financial wealth. Income is disposable income net of capital gains. Wealth includes financial wealth and adjusted housing wealth net of liabilities. Risk-free assets represent the value of bank holdings and bond funds, and Risky assets the combined value of stocks and equity funds. Matched stocks is the tax reported value of stock holdings. Women and Finance measures the frequency of women and those who earned income from a financial occupation. Frequent traders denotes the fraction of individuals who traded stocks at least six months in a row. All reported values are means, unless otherwise stated. During the sample period, $1 corresponds to about SEK 9.

<table>
<thead>
<tr>
<th>Variables</th>
<th>(Low)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Diff.</th>
<th>All</th>
<th>All Stock</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stake size, %</td>
<td></td>
<td>1.96</td>
<td>11.39</td>
<td>35.25</td>
<td>76.93</td>
<td>99.93</td>
<td>95.81</td>
<td>45.09</td>
<td>62.75</td>
</tr>
<tr>
<td>Trading measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buy Turnover, %</td>
<td></td>
<td>2.47</td>
<td>3.68</td>
<td>4.43</td>
<td>6.78</td>
<td>7.35</td>
<td>4.81</td>
<td>4.94</td>
<td>5.64</td>
</tr>
<tr>
<td>Sell Turnover, %</td>
<td></td>
<td>2.13</td>
<td>3.17</td>
<td>3.82</td>
<td>6.29</td>
<td>7.11</td>
<td>4.93</td>
<td>4.51</td>
<td>5.19</td>
</tr>
<tr>
<td>Turnover Month 1-6, %</td>
<td></td>
<td>8.43</td>
<td>11.69</td>
<td>14.12</td>
<td>21.20</td>
<td>22.35</td>
<td>13.62</td>
<td>15.56</td>
<td>17.57</td>
</tr>
<tr>
<td>Turnover Month 7-12, %</td>
<td></td>
<td>3.23</td>
<td>5.74</td>
<td>7.07</td>
<td>12.96</td>
<td>12.53</td>
<td>9.33</td>
<td>8.30</td>
<td>9.61</td>
</tr>
<tr>
<td>Losses, 6 Month PUL</td>
<td></td>
<td>0.89</td>
<td>0.82</td>
<td>0.81</td>
<td>0.78</td>
<td>0.82</td>
<td>-0.07</td>
<td>0.82</td>
<td>0.82</td>
</tr>
<tr>
<td>No. of trades per month</td>
<td></td>
<td>0.31</td>
<td>0.54</td>
<td>0.76</td>
<td>1.33</td>
<td>1.14</td>
<td>0.83</td>
<td>0.82</td>
<td>0.90</td>
</tr>
<tr>
<td>Frequent Traders, %</td>
<td></td>
<td>15.09</td>
<td>20.00</td>
<td>22.45</td>
<td>31.70</td>
<td>31.00</td>
<td>15.91</td>
<td>24.05</td>
<td>25.31</td>
</tr>
<tr>
<td>Portfolio characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Portfolio Size, SEK</td>
<td></td>
<td>32</td>
<td>46</td>
<td>94</td>
<td>158</td>
<td>100</td>
<td>88</td>
<td>86</td>
<td>89</td>
</tr>
<tr>
<td>No. of stocks</td>
<td></td>
<td>2.07</td>
<td>2.71</td>
<td>3.43</td>
<td>4.05</td>
<td>2.72</td>
<td>0.58</td>
<td>3.00</td>
<td>3.03</td>
</tr>
<tr>
<td>Diversification, eq. funds %</td>
<td></td>
<td>49.64</td>
<td>59.91</td>
<td>47.18</td>
<td>17.41</td>
<td>&lt; 0.01</td>
<td>44.29</td>
<td>28.45</td>
<td>35.32</td>
</tr>
<tr>
<td>Investor characteristics (All wealth variables in SEK thousands)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Income</td>
<td></td>
<td>378</td>
<td>346</td>
<td>348</td>
<td>327</td>
<td>313</td>
<td>-59</td>
<td>343</td>
<td>325</td>
</tr>
<tr>
<td>Wealth</td>
<td></td>
<td>2 601</td>
<td>1 121</td>
<td>1 003</td>
<td>988</td>
<td>725</td>
<td>-1 610</td>
<td>1 289</td>
<td>851</td>
</tr>
<tr>
<td>Fin. Wealth</td>
<td></td>
<td>1 507</td>
<td>376</td>
<td>263</td>
<td>269</td>
<td>118</td>
<td>-1 164</td>
<td>449</td>
<td>205</td>
</tr>
<tr>
<td>Risk-free Assets</td>
<td></td>
<td>166</td>
<td>86</td>
<td>67</td>
<td>68</td>
<td>41</td>
<td>-114</td>
<td>86</td>
<td>53</td>
</tr>
<tr>
<td>Risky Assets</td>
<td></td>
<td>1 341</td>
<td>291</td>
<td>196</td>
<td>201</td>
<td>77</td>
<td>-1 034</td>
<td>364</td>
<td>152</td>
</tr>
<tr>
<td>Equity funds</td>
<td></td>
<td>263</td>
<td>122</td>
<td>68</td>
<td>26</td>
<td>&lt; 1</td>
<td>-203</td>
<td>77</td>
<td>56</td>
</tr>
<tr>
<td>Age, years</td>
<td></td>
<td>43.18</td>
<td>38.66</td>
<td>37.07</td>
<td>36.85</td>
<td>35.49</td>
<td>-7.22</td>
<td>38.25</td>
<td>35.96</td>
</tr>
<tr>
<td>Women, %</td>
<td></td>
<td>20.52</td>
<td>19.39</td>
<td>18.44</td>
<td>15.24</td>
<td>16.56</td>
<td>-3.87</td>
<td>18.03</td>
<td>18.48</td>
</tr>
<tr>
<td>Financial occupation, %</td>
<td></td>
<td>3.40</td>
<td>3.07</td>
<td>2.97</td>
<td>3.16</td>
<td>2.78</td>
<td>-0.71</td>
<td>3.08</td>
<td>2.62</td>
</tr>
<tr>
<td>University education, %</td>
<td></td>
<td>53.21</td>
<td>53.77</td>
<td>55.38</td>
<td>50.80</td>
<td>45.42</td>
<td>-7.88</td>
<td>51.72</td>
<td>50.61</td>
</tr>
</tbody>
</table>

42
**Table II: Correlations**

Panel A display Spearman correlations between the key variables of Table 1. Panel B display the results of regressing stake size (value of stock portfolio divided by total holdings of risky assets) and portfolio turnover on the investor characteristics. Panel C presents the canonical correlation coefficients when the common component of the dependent variables stake size and turnover are projected onto the joint set of characteristics.

### Panel A: Correlations

<table>
<thead>
<tr>
<th>Variables</th>
<th>Log Wealth</th>
<th>Women</th>
<th>Log Age</th>
<th>Finance</th>
<th>University</th>
<th>Stake Size</th>
<th>Log Turnover</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log Wealth</td>
<td>1.00</td>
<td>0.02**</td>
<td>0.47***</td>
<td>0.05***</td>
<td>0.16***</td>
<td>-0.28***</td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>1.00</td>
<td>1.00</td>
<td>0.08***</td>
<td>-0.02**</td>
<td>0.04***</td>
<td>-0.04***</td>
<td>1.00</td>
</tr>
<tr>
<td>Log Age</td>
<td>0.08***</td>
<td>1.00</td>
<td>-0.02**</td>
<td>0.03***</td>
<td>0.02***</td>
<td>-0.04***</td>
<td>1.00</td>
</tr>
<tr>
<td>Finance</td>
<td>0.05***</td>
<td>-0.02**</td>
<td>1.00</td>
<td>0.03***</td>
<td>0.02***</td>
<td>-0.04***</td>
<td>1.00</td>
</tr>
<tr>
<td>University</td>
<td>0.16***</td>
<td>0.04***</td>
<td>0.02***</td>
<td>0.03***</td>
<td>1.00</td>
<td>0.24***</td>
<td></td>
</tr>
<tr>
<td><strong>Dependent</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stake Size</td>
<td>-0.28***</td>
<td>-0.04***</td>
<td>-0.21***</td>
<td>-0.01</td>
<td>-0.06***</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Log Turnover</td>
<td>-0.04***</td>
<td>-0.11***</td>
<td>-0.02**</td>
<td>0.01</td>
<td>-0.04***</td>
<td>0.24***</td>
<td>1.00</td>
</tr>
</tbody>
</table>

### Panel B: Regressions

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Log Wealth</th>
<th>Women</th>
<th>Log Age</th>
<th>Finance</th>
<th>University</th>
<th>R-sq</th>
<th>Residual corr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Turnover</td>
<td>-0.19**</td>
<td>-2.84***</td>
<td>2.14***</td>
<td>0.09</td>
<td>-0.94***</td>
<td>0.01</td>
<td>0.21***</td>
</tr>
<tr>
<td></td>
<td>(-2.34)</td>
<td>(-9.92)</td>
<td>(4.29)</td>
<td>(1.34)</td>
<td>(-3.43)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stake Size</td>
<td>-5.32***</td>
<td>-3.11***</td>
<td>-7.37***</td>
<td>0.11</td>
<td>-1.36*</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-25.66)</td>
<td>(-3.35)</td>
<td>(-5.53)</td>
<td>(0.71)</td>
<td>(-1.86)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Panel C: Canonical correlations (1st common root)

<table>
<thead>
<tr>
<th>Canonical variables</th>
<th>Log Wealth</th>
<th>Women</th>
<th>Log Age</th>
<th>Finance</th>
<th>University</th>
<th>Stake Size</th>
<th>Log Turnover</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characteristics</td>
<td>0.98</td>
<td>0.10</td>
<td>0.64</td>
<td>0.05</td>
<td>0.18</td>
<td>-0.29</td>
<td>-0.01</td>
</tr>
<tr>
<td>Dependent</td>
<td>0.29</td>
<td>0.03</td>
<td>0.19</td>
<td>0.01</td>
<td>0.05</td>
<td>-0.98</td>
<td>-0.03</td>
</tr>
</tbody>
</table>

Significance of a t-test (test statistics within parenthesis) are marked by ***, **, and *, indicating rejection at the 1%, 5%, and 10% level. Standard errors in the regression of Panel B are clustered on postal zip codes.
Table III: Average trading returns and turnover costs

Investors are grouped into quintiles of stake size, which is the stock portfolio value divided by the value of total risky assets. The rows display the mean trading returns, with and without fees, and the turnover costs defined in the main text.

<table>
<thead>
<tr>
<th>Mean monthly returns</th>
<th>Quintiles by Stake size</th>
<th>Difference Q5-Q1</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 (Low)</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>(-11.12)</td>
<td>(-11.85)</td>
<td>(-10.55)</td>
</tr>
<tr>
<td>Trading return (w/o fees), BP</td>
<td>-6.64***</td>
<td>-7.06***</td>
<td>-4.69***</td>
</tr>
<tr>
<td></td>
<td>(-3.56)</td>
<td>(-3.90)</td>
<td>(-2.73)</td>
</tr>
<tr>
<td>Turnover cost, %</td>
<td>-3.12***</td>
<td>-2.81***</td>
<td>-2.23***</td>
</tr>
<tr>
<td></td>
<td>(-10.22)</td>
<td>(-10.44)</td>
<td>(-10.58)</td>
</tr>
<tr>
<td>Turnover cost (w/o fees), %</td>
<td>-1.29***</td>
<td>-1.02***</td>
<td>-0.77***</td>
</tr>
<tr>
<td></td>
<td>(-4.46)</td>
<td>(-4.10)</td>
<td>(-3.92)</td>
</tr>
</tbody>
</table>

Significance of a t-test (test statistics within parenthesis) are marked by ***, **, and *, indicating rejection at the 1%, 5%, and 10% level.
Table IV: Predictive regressions of turnover

This table reports the regression results where future turnover is predicted by past turnover and a set of portfolio and investor characteristics. The dependent variable is the average log transformation of turnover, log(1+turnover), for each investor measured during trading month 7 to 12 in sample. Lagged Turnover and Trading Return denote the average turnover and trading return in trading month 1 to 6 as defined in the main text. PUL denotes the investors proportion of unrealized losses in the portfolio. Stake size is the stock portfolio value divided by the total value of risky assets. Portfolio Size and Number of Stocks denote the log of portfolio value and number of stocks in the portfolio. Log Age and Wealth denote the natural logarithm of age and total net wealth of investors. Women, Finance, and University are indicator variables that takes the value 1 if investors are female, are employed within the financial sector, or have at least one year of completed university studies: zero otherwise. Model III through VII include interactions with PUL and investor groups: “Wealthy” are those who belong to the top two wealth quintiles; “Sophisticated” also completed at least one year of university studies. “Cohort controls” refers to indicator variables for the month investors enter the sample. There are 10,600 observations.

<table>
<thead>
<tr>
<th>Turnover, month 7 to 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model number</td>
</tr>
<tr>
<td>Interaction (IA)</td>
</tr>
<tr>
<td>Lagged Turnover</td>
</tr>
<tr>
<td>Lagged Trading Return</td>
</tr>
<tr>
<td>PUL</td>
</tr>
<tr>
<td>Stake Size</td>
</tr>
<tr>
<td>Log Portfolio Size</td>
</tr>
<tr>
<td>No. of stocks</td>
</tr>
<tr>
<td>Log Wealth</td>
</tr>
<tr>
<td>Women</td>
</tr>
<tr>
<td>Log Age</td>
</tr>
<tr>
<td>Finance</td>
</tr>
<tr>
<td>University</td>
</tr>
<tr>
<td>IA / Stake Size</td>
</tr>
<tr>
<td>IA / Lagged Trading Return</td>
</tr>
<tr>
<td>IA / PUL</td>
</tr>
</tbody>
</table>

Cohort controls | No | Yes | Yes | Yes | Yes | Yes | Yes |
R-squared | 0.08 | 0.41 | 0.41 | 0.41 | 0.41 | 0.41 | 0.41 |

Reported t-statistics in parentheses are based on standard errors clustered on 5,350 investor zip codes. Asterisks ***, **, and * denote significance at the 1%, 5%, and 10% level.
Table V: Trading returns in the cross-section

This table reports the coefficients from a panel regression model with the investor’s monthly trading returns as the dependent variable. Turnover is the monthly gross value of purchases divided by portfolio value at the beginning of month. Stake size is the stock portfolio value divided by the total value of risky assets. Portfolio Size and Number of Stocks denote the log of portfolio value and number of stocks in the portfolio. Log Age and Wealth denote the natural logarithm of age and total net wealth of investors. Women, Finance, and University are indicator variables that takes the value 1 if investors are female, are employed within the financial sector, or have at least one year of completed university studies: zero otherwise. The first four columns labeled Model I to IV use the full sample of 10,600 investors, and Model V through XII report the results from investors sorted across subsamples of wealth, turnover, gender, education, and sophistication. Those labelled “Sophisticated” are university educated who belong to the top two wealth quintiles. “Unsophisticated” are investors with less education belonging to the bottom two quintiles of wealth. The last two columns, denoted Model XI and XII, use the trading return excluding fees as the dependent variable in the regression. All independent variables have been centralized. There are 213,633 observations in the full panel sample where there are 10,600 and 35 months.

<table>
<thead>
<tr>
<th>Model number</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>VI</th>
<th>VII</th>
<th>VIII</th>
<th>IX</th>
<th>X</th>
<th>XI</th>
<th>XII</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>All</td>
<td>All</td>
<td>All</td>
<td>All</td>
<td>Wealth 1</td>
<td>Wealth 5</td>
<td>Turnover 5</td>
<td>Women</td>
<td>Sophisticated</td>
<td>Unsophisticated</td>
<td>No Fees</td>
<td>No Fees</td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.25***</td>
<td>-0.25***</td>
<td>-0.31***</td>
<td>-0.29***</td>
<td>-0.47***</td>
<td>-0.18***</td>
<td>-0.71***</td>
<td>-0.22***</td>
<td>-0.14***</td>
<td>-0.41***</td>
<td>-0.11**</td>
<td>-0.31**</td>
</tr>
<tr>
<td></td>
<td>(-5.03)</td>
<td>(-4.94)</td>
<td>(-4.74)</td>
<td>(-5.85)</td>
<td>(-10.87)</td>
<td>(-3.59)</td>
<td>(-4.31)</td>
<td>(-4.28)</td>
<td>(-4.05)</td>
<td>(-6.76)</td>
<td>(-2.37)</td>
<td>(-2.01)</td>
</tr>
<tr>
<td>Stake Size</td>
<td>-0.13***</td>
<td>-0.21***</td>
<td>-0.14***</td>
<td>-0.06**</td>
<td>0.02</td>
<td>-0.06</td>
<td>-0.17**</td>
<td>-0.03</td>
<td>-0.06</td>
<td>-0.06</td>
<td>-0.03</td>
<td>-0.22***</td>
</tr>
<tr>
<td></td>
<td>(-4.26)</td>
<td>(-6.39)</td>
<td>(-6.41)</td>
<td>(-3.30)</td>
<td>(0.74)</td>
<td>(-1.23)</td>
<td>(-2.11)</td>
<td>(-0.58)</td>
<td>(-1.53)</td>
<td>(-0.87)</td>
<td>(-1.15)</td>
<td>(-2.77)</td>
</tr>
<tr>
<td>Log Portfolio Size</td>
<td>0.08***</td>
<td>0.06***</td>
<td>0.10***</td>
<td>0.15***</td>
<td>0.05***</td>
<td>0.18***</td>
<td>0.10***</td>
<td>0.06***</td>
<td>0.13***</td>
<td>0.02</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.08)</td>
<td>(2.86)</td>
<td>(4.24)</td>
<td>(3.62)</td>
<td>(2.62)</td>
<td>(4.63)</td>
<td>(4.13)</td>
<td>(3.34)</td>
<td>(3.14)</td>
<td>(1.00)</td>
<td>(0.50)</td>
<td></td>
</tr>
<tr>
<td>No. of stocks</td>
<td></td>
<td>-0.01</td>
<td>-0.01*</td>
<td>-0.05**</td>
<td>-0.01</td>
<td>0.01</td>
<td>-0.02***</td>
<td>-0.01*</td>
<td>-0.03*</td>
<td>-0.01</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-0.75)</td>
<td>(-1.71)</td>
<td>(-2.49)</td>
<td>(-0.58)</td>
<td>(0.25)</td>
<td>(-3.00)</td>
<td>(-1.74)</td>
<td>(-1.91)</td>
<td>(-0.21)</td>
<td>(1.09)</td>
<td></td>
</tr>
<tr>
<td>Log Wealth</td>
<td></td>
<td>0.04***</td>
<td>0.04***</td>
<td>0.11***</td>
<td>-0.05</td>
<td>0.10***</td>
<td>0.02</td>
<td>-0.02</td>
<td>0.09***</td>
<td>0.02**</td>
<td>0.04*</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(5.32)</td>
<td>(4.76)</td>
<td>(6.46)</td>
<td>(4.45)</td>
<td>(2.12)</td>
<td>(-0.46)</td>
<td>(5.73)</td>
<td>(2.42)</td>
<td>(1.95)</td>
<td>(2.06)</td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td></td>
<td>0.08***</td>
<td>0.04***</td>
<td>0.10**</td>
<td>-0.02</td>
<td>0.17*</td>
<td></td>
<td>0.04</td>
<td>0.11**</td>
<td>0.03**</td>
<td>0.19**</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(5.33)</td>
<td>(3.32)</td>
<td>(2.24)</td>
<td>(-0.50)</td>
<td>(1.91)</td>
<td></td>
<td>(0.60)</td>
<td>(2.27)</td>
<td>(2.44)</td>
<td>(2.06)</td>
<td></td>
</tr>
<tr>
<td>Log Age</td>
<td>-0.14***</td>
<td>-0.09***</td>
<td>-0.20**</td>
<td>0.02</td>
<td>-0.44***</td>
<td>-0.10*</td>
<td>0.03</td>
<td>-0.20</td>
<td>-0.09***</td>
<td>-0.38***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-5.14)</td>
<td>(-3.17)</td>
<td>(-2.83)</td>
<td>(0.24)</td>
<td>(-3.79)</td>
<td>(-1.80)</td>
<td>(0.85)</td>
<td>(-2.61)</td>
<td>(-3.07)</td>
<td>(-3.47)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finance</td>
<td>-0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>-0.07</td>
<td>0.04</td>
<td>-0.08</td>
<td>-0.04</td>
<td>-0.12</td>
<td>0.02</td>
<td>0.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.06)</td>
<td>(0.19)</td>
<td>(0.10)</td>
<td>(-0.68)</td>
<td>(0.23)</td>
<td>(-0.71)</td>
<td>(-0.43)</td>
<td>(-0.88)</td>
<td>(0.34)</td>
<td>(0.38)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University</td>
<td>0.09***</td>
<td>0.07***</td>
<td>0.09***</td>
<td>0.08**</td>
<td>0.25***</td>
<td>0.07</td>
<td></td>
<td>0.06**</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.80)</td>
<td>(2.62)</td>
<td>(2.08)</td>
<td>(2.13)</td>
<td>(2.02)</td>
<td>(1.39)</td>
<td></td>
<td>(2.09)</td>
<td>(1.55)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log Turnover</td>
<td>-0.28***</td>
<td>-0.43***</td>
<td>-0.95***</td>
<td>0.05***</td>
<td>-2.03***</td>
<td>-1.08***</td>
<td>-3.16***</td>
<td>-0.54</td>
<td>-0.21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-4.47)</td>
<td>(-7.45)</td>
<td>(-2.33)</td>
<td>(-2.59)</td>
<td>(-4.00)</td>
<td>(-2.31)</td>
<td>(-6.48)</td>
<td>(-1.29)</td>
<td>(-0.21)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. Obs. (T x N)</td>
<td>213633</td>
<td>213633</td>
<td>213633</td>
<td>213633</td>
<td>40938</td>
<td>44213</td>
<td>42953</td>
<td>38507</td>
<td>53539</td>
<td>45322</td>
<td>213633</td>
<td>42953</td>
</tr>
</tbody>
</table>

Reported t-statistics in parentheses are based on standard errors robust to autocorrelation and heteroscedasticity as described in Newey and West (1987). Asterisks ***, **, and * denote significance at the 1%, 5%, and 10% level.
Table VI: Aggregate trading profits

The 10,600 investors in the sample are grouped into quintiles of stake size, which is the portfolio value observed at first year-end divided by total financial wealth. The net trading payoff for each investor is aggregated into total losses and gains for each of the two groups. The four bottom rows display and normalize net gains with total financial wealth, total wealth, and income net of capital gains for each investor group. Values are in thousands of Swedish kronor (SEK), where $1 corresponds to about SEK 9.

<table>
<thead>
<tr>
<th>Quintiles by Stake Size (Account value/Risky Assets)</th>
<th>1 (Low)</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5 (High)</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total losses</td>
<td>-4 763</td>
<td>-6 486</td>
<td>-8 685</td>
<td>-24 632</td>
<td>-18 147</td>
<td>-62 713</td>
</tr>
<tr>
<td>Total gains</td>
<td>3 281</td>
<td>4 006</td>
<td>7 623</td>
<td>15 575</td>
<td>8 739</td>
<td>39 224</td>
</tr>
<tr>
<td>Net gains</td>
<td>-1 482</td>
<td>-2 480</td>
<td>-1 062</td>
<td>-9 057</td>
<td>-9 408</td>
<td>-23 489</td>
</tr>
<tr>
<td>Financial wealth</td>
<td>3 194 982</td>
<td>797 738</td>
<td>557 960</td>
<td>569 483</td>
<td>249 819</td>
<td>5 369 982</td>
</tr>
<tr>
<td>Net gains / fin. wealth, %</td>
<td>-0.05</td>
<td>-0.31</td>
<td>-0.19</td>
<td>-1.59</td>
<td>-3.77</td>
<td>-0.44</td>
</tr>
<tr>
<td>Wealth</td>
<td>5 513 064</td>
<td>2 376 052</td>
<td>2 126 930</td>
<td>2 094 166</td>
<td>1 537 530</td>
<td>13 647 742</td>
</tr>
<tr>
<td>Net gains / wealth, %</td>
<td>-0.03</td>
<td>-0.10</td>
<td>-0.05</td>
<td>-0.43</td>
<td>-0.61</td>
<td>-0.17</td>
</tr>
<tr>
<td>Income</td>
<td>802 106</td>
<td>734 175</td>
<td>738 657</td>
<td>692 596</td>
<td>663 299</td>
<td>3 630 833</td>
</tr>
<tr>
<td>Net gains / income, %</td>
<td>-0.18</td>
<td>-0.34</td>
<td>-0.14</td>
<td>-1.31</td>
<td>-1.42</td>
<td>-0.65</td>
</tr>
</tbody>
</table>
Figure 1: Distribution of Financial Wealth and Trading Costs
This figure plots the distribution of financial wealth and trading costs across three categories of the 10,600 investors in the sample. There are 24% classified as sophisticated investors, who are university educated and belong to the 40% wealthiest investors. There are 22% labeled unsophisticated investors, who belong to bottom 40% of the wealth distribution lacking university education. The middle bars represent the remaining 54% investors in sample.

Figure 2: Stake size in the cross-section of investors
The cumulative frequency distribution of stake size is plotted for the 10,600 investors in the sample. Portfolio size is defined as portfolio value at first observable year-end, and is divided by either total market value of risky assets (defined as stake size, dotted line) or total financial wealth (dashed line). The solid 45 degree line plots an imposed uniform distribution.
Figure 3: Correlation between Stake size and Turnover

This figure shows the 10,600 investors in sample sorted on stake size on the horizontal axis, and their estimated turnover on the vertical axis. Stake size is the portfolio value divided by total risky financial wealth as defined in the main text. The solid bold line plots the relationship using non-parametric LOESS estimate along with confidence bands (dotted lines). The solid double line plots average turnover for each quintile of investors sorted on stake size.