Behavioral Biases and Investment

Massimo Massa*

INSEAD

Andrei Simonov

Stockholm School of Economics

October 5, 2002

Abstract

We use a new and unique dataset to investigate the way investors react to prior gains/losses and the so called "familiarity" bias. We distinguish between different behavioral theories (loss aversion, house-money effect, mental accounting) and between behavioral and rational hypotheses (pure familiarity and information-based familiarity). We show that, on an yearly horizon, investors react to previous gains/losses according to the house-money effect. No evidence is found of narrow accounting as investors consider wealth in its entirety and risk taking in the financial market is affected by gains/losses in overall wealth, as well as financial and real estate wealth. In terms of individual stock picking, we provide evidence in favor of the information-based theory and show that familiarity can be considered as a proxy for the availability of information as opposed to a behavioral heuristics.

JEL classification: G11, G14.

Keywords: *Behavioral finance, portfolio investment, loss aversion, familiarity bias, information.*

*Corresponding author: M. Massa, Finance Department, INSEAD, Boulevard de Constance, 77305 Fontainebleau Cedex, France. Tel: (33)1 60 72 44 81 Fax: (33)1 60 72 40 45. Email: massimo.massa@insead.edu. We thank for helpful comments O.Bondarenko, F.DeJong, B.Dumas, H.Hau, P.Hillion, M.Lettau, P.Maenhout, J.Peress, P.Sodini, M.Suominen, L.Tepla, P.Veronesi, M.Weber and the participants of the Summer Financial Markets Symposium at Gerzensee. We are grateful to Sven-Ivan Sundqvist for numerous helpful discussions and for providing us with the data. Andrei Simonov acknowledges financial support from the Stockholm Institute for Financial Research and Jan Wallander och Tom Hedelius Stiftelse. All the remaining errors are ours.

1 Introduction

Behavioral motivations have been advocated as a main driving force in investment portfolio choice. In particular, two behavioral phenomena have emerged as relevant: the way investors react to prior gains and losses and the so called "familiarity" bias. The combined effect of these two phenomena - potentially inconsistent with standard "rational" investment theories - redefines the way we think of investor behavior. The "behavioral investor" decides how much to invest in risky assets mainly on the basis of prior gains and losses and selects the individual risky securities on the basis of his familiarity with them. Hedging does not play any role.

However, this behavioral approach, while well grounded in experiments, still does not provide a consistent unified view. It seems clear that investor behavior is affected by prior outcomes and by the changes in wealth as opposed to the mere level of it. But, the "direction" of the reaction to prior gains/losses is not well defined as different psychological theories advocate different reactions. Prior losses increase risk taking in the case of loss aversion and a decrease it in the case of house-money effect. Moreover, in the case where the direction of the impact is clear (i.e., familiarity bias), the behavioral stylized evidence that supports it can also be explained in terms of a standard rational theory.

Efforts to empirically address these issues have been hindered by the limitation of the data that has made it impossible to test the different behavioral theories one against the other and to compare them against their alternative "rational" counterparts.

We bridge this gap by using a new and unique dataset. Given the richness of the data - that contains basically any information on the holdings, wealth, broken down into *all* of its components, income and demographic characteristics of a very representative sample of the Swedish population - we are able to properly control for most components of the investor's decision. This allows us to study how investors react to prior gains/losses and how familiarity affects their portfolio choices.

In particular, in the case of the reaction to prior gains/losses, we test loss aversion, and house-money effect by directly inspecting investor reactions to different definitions of gains and losses (i.e., overall wealth, financial gains and losses and real estate gains and losses). We also investigate the issue of narrow framing and mental accounting by considering how gains and losses in a category of wealth (e.g., real estate) affects changes in holdings in other categories (e.g., financial assets). We provide evidence in favor of the house-money effect and against narrow framing. That is, investors change their holdings of risky assets as a function of both financial and real estate gains. Prior gains increase risk taking, while prior losses reduce it.

In the case of familiarity, we group investors into more and less informed ones and we trace their exposure to the familiarity bias. We provide evidence that familiarity is more based on information constraints than on behavioral heuristics. This is, to our knowledge, the first time that such an analysis has been attempted.

The remainder of the paper is articulated as follows. In Section 2 we describe the problem and refers to the literature. In Section 3 we describe out approach. In Section 4 we describe the datasets we use. In Section 5 we report the way we construct the variables and our measures of familiarity. In Section 6 we discuss our identification restrictions, the econometric issues and the methodology we employ. In Section 7 we report the results of the tests of hedging versus familiarity and provide evidence for the familiarity hypothesis. A brief conclusion follows.

2 The problem

We focus on two salient moments of investor's behavior - risk-taking and stock-picking - and analyze the impact of behavioral biases on them. In particular, we study the relationship between risk taking and prior gains and losses and the relationship between stock-picking and familiarity. We consider the two moments separately for mere expositional purposes. In fact, they are very much intertwined.

Risk-taking

Behavioral theory argues that prior gains induce a different behavior from prior losses. Loss aversion hypothesizes that prior losses increase risk taking, while prior gains reduce it. Investors have the "tendency to seek risk when faced with possible losses, and to avoid risk when a certain gain is possible". Loss aversion is based on the psychological grounding that a decline in utility arising out of the realization of losses relative to gains induces investors not to sell losing stocks relative to winning ones (Kahneman and Tversky, 1979, DeBondt and Thaler, 1995). Empirical evidence of it has been found by Odean (1998) and Barber and Odean (1999) who have shown that investors tend to "hold on to the losers and sell the winners".

The alternative house-money effect suggests that prior gains, by providing investors with a "cushion" that makes future losses less painful, in fact increase risk taking. Thaler and Johnson (1990) show that "while a prior gain can increase subjects willingness to accept gambles, ...prior losses sensitize people to subsequent losses of similar magnitude." Barberis, Huang and Santos (1999) use this behavioral finding in order to explain the size of the equity premium and patterns in stock volatility. They argue that "previous capital gains reduce investors' sensitivity to risk...while previous losses, by making any new loss more painful, increase risk aversion".[1]

Both loss aversion and house-money effect theories are vague about the definition of "gains" and "losses" - i.e., whether they apply to the overall investor wealth or to a limited subset of it (i.e., financial wealth, real estate, single stocks). In fact, investors may react to the gains and losses in different categories of wealth differently, depending on their categories. This theory can be defined as "mental accounting" or "narrow framing". For example, Barberis and Huang (2000) have suggested that investors apply mental accounting to stock holdings and react separately to gains and losses for different stocks.

While experimental literature has provided ample evidence of the conditions under which loss aversion, house-money effect and mental accounting occur (Shefrin and Statman, 1985, Tversky and Kanheman, 1981, Glaser and Weber, 2002, Weber and Camerer, 1998, Weber and Zuchel, 2001), the three theories have never been simultaneously tested using field data. Given that the implications of these theories are often very highly correlated, separate partial tests may fail to provide a proper identification that avoids spurious correlation.

In fact, these three theories deliver a set of easily testable cross-restrictions. For example, loss aversion postulates a *negative* correlation between investment in stocks and prior *overall* changes in wealth. The house-money effect postulates a *positive* correlation between investment in stocks and prior *overall* changes in wealth. Mental accounting assumes that *only changes in stock market wealth* affects the investment in stocks. That is, if investors categorize gains and losses on the basis of narrow categories, there should be a positive correlation between investment in a particular category of wealth and previous capital gains/losses in such a category. Other gains and losses are irrelevant. Joint tests can exploit these cross-

[1] This approach, while opposite of the one based on loss aversion, would also be consistent with standard utility theory in the case of a utility function characterized by decreasing risk aversion. Indeed, previous losses, by reducing wealth, increase risk aversion, while previous gains, by increasing wealth, reduce risk aversion. Recent evidence finding that risk aversion is a decreasing function of endowment (Guiso and Paiella, 1999) is along this line.

restrictions to effectively distinguish between theories.

The main obstacle to the implementation of these tests has been, up to now, the lack of good-quality data *on the overall wealth of the investors for a representative sample of investors in the market*. In general, the analysis has been focused on a subset of investor's wealth and limited to a particular subset of investors . For example, Barber and Odean (2000, 2001, 2002) and Odean (1998, 1999) rely on a dataset that contains accurate information on all the trades and holdings of individual investors with a big discount broker. Coval and Moskowitz (1999 and 2001) and Frieder and Subrahmanian (2001) focus on the stock holdings of institutional investors. Only Grinblatt and Kelloarji (2000, 2001a, 2001b) use a dataset that contains, for the first time, the entire stock portfolio of the investors. However, they focus on issues such us geographical preferences and momentum trading, without considering the overall dimension of the portfolio problem (wealth, real estate, income).

Recently Jackson (2002), using Australian data, shows that loss-aversion is a short-term phenomenon, suggesting that loss aversion is a better predictor of short term behavior, while house-money effect mostly applies to long run behavior. However, also in this case, the analysis is based just on stock holdings and does not account for the wealth of the investor in its entirety. Detailed information on the overall wealth of the investor is required not only for the variables needed to carry out the tests (i.e., changes in wealth, capital gains/losses, risk taking), but especially to construct proper control variables for all the other factors affecting investor's behavior. Indeed, the main issue that empirical studies *based on field data* face is the "coeteris paribus" condition or spurious correlation.

The use of information limited to a subset of the entire stock-portfolio with no control for the other sources of wealth or income of the investor (i.e., labor income, entrepreneurial income) may be problematic. For example, if the change in wealth of the investor is due to income shocks or real estate capital gains, a test of investor behavior *based only on changes in portfolio holdings and stock market capital gains/losses* would not be able to distinguish loss aversion from standard wealth effects.

Maybe, the biggest omission is represented by real estate. Changes in wealth due to a real estate capital gain/loss may swing investor behavior regardless of the financial capital gains/losses and therefore make any behavioral study focused only on financial holdings problematic. Genesove and Mayer (2001) shed some light on this topic by analyzing loss aversion and seller behavior in the housing market.[2]

Stock-picking

The second moment of investor behavior is the choice of the assets in which to invest. There is increasing evidence that investors, far from choosing their portfolio according to standard theory, tend to invest in firm that are "close by". Huberman (2001) argues that there is a "general tendency of people to have concentrated portfolios, ...to hold their own company' s stock in their retirement accounts...invest in stocks of their home country. Together, these phenomena provide compelling evidence that people invest in the familiar while often ignoring the principles of portfolio theory".

There is now a lot of evidence to support this "familiarity bias". Huberman (2001) finds that workers in a Regional Bell Company tend to buy stock of the firm where they work but not of similar firms present in other regions. Frieder and Subrahmanyam (2002) report that individual investors tend to hold disproportionate amounts of stocks with high brand recognition.

Also, investors may choose the stocks of the company for which they work because familiarity induces them to optimistically extrapolate past returns (Benartzi and Thaler, 1995, Benartzi, 2001). Alternatively, investors may display a home bias and invest in stocks of companies headquartered close to where they live (Coval and Moskowitz, 1999, 2001, Hau, 2001, Huberman, 2001) or of the country they come from (Bhattacharya, 2001).

Behavioral theories relate familiarity bias to the findings in psychology that show that human beings use heuristic simplifications in their decision making process. One of those heuristics is the saliency or availability bias. This is the tendency to focus heavily on information that is salient or is often mentioned, rather that information that is blended in the background. We will define this hypothesis, entirely grounded on behavioral heuristics, as "pure familiarity".

The alternative approach is the "information-based familiarity".[3] This states that "investors buy and hold only those securities about which they have enough information." The revealed portfolio formation under information-based familiarity is observationally equivalent to that under exogenous portfolio constraints (Merton, 1987, Shapiro, 2002) as information about a stock affects investment decision by altering the perceived expected pay-off in a

[2] However, while they are the first ones to bring behavioral theories to the data by using real estate data, they do not consider the other aspects of the investor's wealth problem.

[3] Alternatively defined as the "investor recognition hypothesis".

rational portfolio decision.[4]

While there exists sufficient empirical evidence for the existence of familiarity, there has been no direct test aimed at distinguishing between the rational and the behavioral explanation of it. Unlike the case of reaction to past gains/losses, here the problem is compounded by the apparent "observational equivalency" of the rational and behavioral theories. The standard testing approach relies on indirect inference based on the observation of the financial anomalies. However, it can be shown that, in the case of indirect inference, theories based on information ("rational structural uncertainty") are observationally equivalent to the ones based on behavioral biases ("behavioral theories"). Although the two sets of theories "relax opposite assumptions of the rational expectations ideal, their mathematical and predictive similarities make them difficult to distinguish." (Brav and Heaton, 2002).

Moreover, as it was the case of reaction to past gains/losses, the analysis is complicated by the confounding effect of income and wealth shocks as well as by specific individual characteristics. For instance, let's consider the standard test of the impact of familiarity on investment. If the investor is subject to the shocks of the geographical area where he lives, he is likely to have more funds available to be invested in stocks at the very time when the local stocks are performing well. If the stocks are selected on the basis of performance, there is a spurious correlation between portfolio allocation and geographical allocation that may be properly explained in terms of income shocks as opposed to behavioral heuristics.[5]

3 Our approach

We try to assess the relative merit of the different theories by focusing on stock holdings and investor characteristics. Unlike the seminal papers of Barber and Odean (2000, 2001, 2002) and Odean (1998, 1999), which focus on transaction data and short-term behavior, we focus on holdings and long-term behavior. To be as close as possible to theory and existing experimental evidence (Benartzi and Thaler, 1995, Barberis, Huang and Santos, 2000 and Barberis and Huang, 2001), we focus on yearly horizons. This allows us to operate at a frequency where we can properly account for all the other sources of income and changes in wealth of the investor. Our study therefore complements the seminal ones done at higher

[4] Also, investor decision may be motivated by "rational structural uncertainty" about the payoff of the assets (Brav and Heaton, 2002).

[5] This would not be the case for institutional investors such as mutual funds (Moskowitz 1999) or dealers on a stock market (Hau 2001). In such cases we can safely assume that the income/wealth shocks of such investors are more equally distributed across the country as a whole.

frequency by Barber and Odean (2000, 2001, 2002) and Odean (1998, 1999).

The availability of a unique dataset with detailed information on investor wealth, income and asset holdings allows us to bring the alternative theories to the data. We are able to run a direct horse race, *after controlling for investor's other income and wealth shocks.* In particular, in the case of the reaction to prior gains/losses, we test one against the other: loss aversion, house-money effect and mental accounting. In the case of familiarity, we take a direct approach. Instead of relying on indirect inference based on the observation of the financial anomalies, we directly test how differentially informed investors are affected by familiarity. In particular, we use information on the wealth and the degree of liquidity in the investor portfolio in order to identify classes of informed investors. We then see whether the exposure to the familiarity bias changes with the degree of informativeness of the investor.

We will proceed as follows. First, we consider the impact of changes in wealth on risk taking. Then, we consider the stock-picking decision and compare rational and behavioral theories.

3.1 Risk-taking: Loss aversion, house-money effect and mental accounting

The standard formulation of prospect theory assumes that risk taking is a function of prior gains/losses. Let's consider a simplified reduced form that relates the fraction of the *ith* investor's wealth invested in risky *financial* assets (h_i) to prior positive and negative changes in his wealth (respectively $\Delta^+ W_i$ and $\Delta^- W_i$):

$$\Delta_t h_i = \beta_f \Delta^+_{t-1} W^f_i + \gamma_f \Delta^-_{t-1} W^f_i + \beta_{re} \Delta^+_{t-1} W^{nf}_i + \gamma_{re} \Delta^-_{t-1} W^{nf}_i + \boldsymbol{\delta} \mathbf{C}_{it} \tag{1}$$

where, for each *ith* investor, W^f_i and W^{nf}_i are, respectively, capital gains/losses in financial assets and non-financial assets (e.g., real estate). The operator Δ_{t-1} represents the change in the interval $[t-1, t-2]$. $\mathbf{C}_{it}$ is a vector of control variables. It contains all the alternative factors that affect the portfolio choice of the *ith* investor (e.g., labor income risk, level of wealth, ...). We follow Benartzi and Thaler (1995), Barberis, Huang and Santos (2000) and Barberis and Huang (2001) and consider the unit of measure equal to one year.

Equation 1 says that the change in holdings of risky assets is related to the capital gains/losses in the previous year (i.e., positive and negative changes in wealth). It nests three theories: loss aversion, house-money effect and mental accounting (or narrow framing). Let's see this more in detail.

H1a *Loss aversion with no mental accounting would require that:*

$$\beta_f < 0, \beta_{nf} < 0, \ \gamma_f > 0, \gamma_{nf} > 0. \tag{2}$$

The house-money effect with no mental accounting would require that:

$$\beta_f > 0, \beta_{nf} > 0, \ \gamma_f < 0, \gamma_{nf} < 0. \tag{3}$$

That is, we expect that investors react to a negative (positive) change in wealth by taking more (less) risk in the case of loss aversion and by taking less (more) risk in the case of the house-money effect. In both cases, investors do not distinguish between the source of income that has generated the loss/gain and the investment decision is a function of both.

H1b Mental accounting would require that:

$$\beta_f \neq \beta_{nf}, \ \gamma_f \neq \gamma_{nf}, \tag{4}$$

that is, investors react differently prior gains/losses, depending on the source of income that has generated them. In its more extreme form, mental accounting hypothesizes that investors categorize different sources of wealth in different accounts and that gains/losses in a particular account only affect the risk taking in the specific account. That is,

$$\beta_f \neq 0, \beta_{nf} = 0, \ \gamma_f \neq 0 \text{ and } \gamma_{nf} = 0. \tag{5}$$

Equations 2, 3, 4 and 5 provide the set of restrictions we will be testing.

3.2 Stock-picking: Information-based familiarity and pure familiarity.

Let's now focus on the specific stocks and consider the following simplified reduced form:

$$h_{ijt} = \boldsymbol{\beta}\mathbf{F}_{ijt} + \boldsymbol{\delta}\mathbf{C}_{ijt} \tag{6}$$

where h_{ijt} is fraction of the portfolio of risky financial assets of the *ith* investor invested in the *jth* asset and $\mathbf{F}_{ijt}$ proxies for the familiarity bias. This is a set of motives affecting the choice of a particular stock that are related to geographical and professional proximity as well as to other forms of proximity. They may be either information-based or mere behavioral heuristics. $\mathbf{C}_{ijt}$ is a vector of control variables that contains all of the alternative factors that affect the portfolio choice of the *ith* investor in the *jth* asset (e.g., labor income risk, level of wealth, ...).

Specification 6 allows us to perform two types of tests. First, we directly assess the impact of the familiarity bias in the portfolio choice once we control for other factors.

H2a If familiarity plays a role, we expect that:

$$H_0 : \beta = 0,\ H_A : \beta \neq 0. \tag{7}$$

A second set of tests aims at separating information-based familiarity hypothesis from the pure familiarity one. In this case, we identify the informed investors and assess how the impact of familiarity varies between informed and uninformed investors, under the null of no change in the case familiarity is a behavioral heuristics.

H2b If familiarity (i.e., F_{it}) proxies for information, the sensitivity of the investors to it should change with their degree of informativeness. Therefore, the null of no change of pure familiarity is tested against the alternative of information-based familiarity. That is,

$$\begin{cases} H_0 : \left|\beta_{\text{high info}}\right| = \left|\beta_{\text{low info}}\right| ;\ H_a : \left|\beta_{\text{high info}}\right| \neq \left|\beta_{\text{low info}}\right| \\ H_0 : \left|\beta_{\text{high info}}\right| = \left|\beta_{\text{low info}}\right| ;\ H_a : \left|\beta_{\text{high info}}\right| \neq \left|\beta_{\text{low info}}\right| \end{cases} \tag{8}$$

Finally, we can also venture on defining the direction of the reaction of the investors in terms of their degree of informativeness. We consider two types of familiarity variables. The ones that proxy for publicly available information (e.g., age of the company, time of IPO, geographical and professional proximity) and the ones that proxy for privately generated information (e.g., holding period of the stock). If the information theory holds, we expect informed investors to be less affected by the former variables and more affected by the latter ones. In particular, uninformed investors should place higher weight on public information (i.e., geographical and professional proximity as well as age of the company), while informed investors will weigh more private information (i.e., the information that is related to stake-holding).

This implies that less informed investors's holdings should be positively related to proxies of publicly available information (e.g., age of the company, time of IPO, geographical and professional proximity) and negatively related to the proxies of privately generated information (e.g., holding period of the stock). In equilibrium, this also implies that informed investors's holdings should be negatively related to proxies of publicly available information (e.g., age of the company, time of IPO, geographical and professional proximity) and positively related to the proxies of privately generated information (e.g., holding period of the stock).

4 Data description

We use data from different sources. The data contain information on both sides of the market: the investors as well as the companies. For each investor we have detailed information of his individual holdings of stocks (broken down at the stock level), mutual funds, bank accounts, real estate and other types of wealth.

We also have available information on the income profile of the investor provided by the fiscal authorities, as well as his demographic and family characteristics. This information has been matched *at the individual level,* so that, for each investor we have available all the investment and income information. For each stock we have detailed information on the company as well as stock market data (price, volume and volatility). We also use aggregated data on Swedish macro-economic conditions and on the indexes of the real estate market. Let's see the sources more in detail.

4.1 Individual stockholding

We use the data on individual shareholders collected by Värdepapperscentralen (VPC), the Security Register Center. The data contain both stockholding held directly and on the street name, including holdings of US-listed ADRs. In addition, SIS Ägarservice AB collects information on ultimate owners of shares held via trusts, foreign holding companies and alike (for details see Sundin and Sundquist, 2002).

Our data cover the period 1995-2001. Overall, the records provide information about the owners of 98% of the market capitalization of publicly traded Swedish companies. For the median company, we have information about 97.9% of the equity, and in the worst case we have information on 81.6% of market capitalization of the company. The data provided by SIS Ägarservice AB were linked by Statistics Sweden with the LINDA dataset described below.

4.2 LINDA

LINDA (Longitudinal INdividual DAtaset for Sweden) is a register-based longitudinal data set and is a joint endeavour between the Department of Economics at Uppsala University, The National Social Insurance Board (RFV), Statistics Sweden, and the Ministries of Finance and Labor. It consists of a large panel of individuals, and their household members, which is representative for the population during 1960 to 1999. For each year, information on all family members of the sampled individuals are added to the data set. Apart from being a panel which is representative of the population in general, the sampling procedure ensures that the data are representative for each year. Moreover, *the same family* is traced over time. This provides a real time series dimension in general missing in surveys based on different cohorts polled over time.

The variables available include individual background variables (sex, age, marital status, country of birth, citizenship, year of immigration, place of residence detailed at the parish level, education, profession, employment status), housing information (type and size of housing, owner, rental and occupation status, one-family or several-family dwelling, year of construction, housing taxation value) and tax and wealth information. In particular, the income and wealth tax registers include information on labor income, capital gains and losses, business income and losses, pension contributions, taxes paid and taxable wealth. A detailed description of the dataset is provided by Edin and Fredriksson, (2000) and is available on the web site http://linda.nek.uu.se/.

The tax part deserves more detailed discussion. In Sweden, in addition to usual income taxation, there exists an additional wealth tax which is paid by every investor with net worth in excess of 900,000 SEK (about US$90,000). The taxable wealth includes tax-accessed value of real estate, 75% of market value of publicly listed securities [6], balance of bank accounts and fair value of valuable possessions (including jewelry, cars, antiques, etc.). For the purposes of this paper, we compute the current market value of housing using the tax-accessed value provided by LINDA. We evaluate it at current prices by using the average ratio of market value to tax-accessed value that is provided for each year and county by the Swedish Office of Statistics.[7]

The combined LINDA/Shareholding dataset covers the period 1995-1999. In addition, we also use 1994 data from LINDA. The overall sample we use contains 1,487,602 observations. In Table 1 we report some descriptive statistics. In particular, Panel A contains the general demographic characteristics (number of households for each year, members in household, adults in household, age of the oldest member of household, percent of the sample with

[6] With the exception of the companies listed on o-list (OTC-list) of Stockholm Stock Exchange. These companies are mostly small companies with very short operating history. The disclosure rules for these companies are lax and most of the stocks are quite illiquid. A detailed description is provided on the web site of the Stockholm Stock Exchange, http://www.stockholmsborsen.se. There is no estimate of market value of privately held companies.

[7] It may lack precision for summer houses if they are located in a county different from the one in which the household is residing as no information about the location of summer houses is provided.

secondary and higher education, percent of immigrants). Panels B and C report, respectively, the regional distribution of the sample and its age and gender distribution. Panel D reports the percentage of the households holding risky securities, having real estate holdings, having debt, and running entrepreneurial activities.

4.3 Firm-level information and other data

In order to derive information on individual security returns (including dividends) and to track the overall market index (SIX market index), we use the SIX Trust Database. For information on the various firm-level characteristics, we use the Market Manager Partners Databases. These two databases are the equivalent of, respectively, CRSP and COMPUSTAT for the US. In addition, Market Manager Partners Databases contain information at the plant level, including location of the plant (detailed at the level of municipality).

We use the set of Swedish residential real estate indices provided by P. Englund. The indices were computed at the county level and are based on resale value of the properties.[8] The consumer confidence index is provided by Statistics Sweden. Geographical coordinates are supplied by Swedish Postal Service and contain latitude and longitude of Swedish Postal Offices (on 3-digit level).

5 Construction of variables

5.1 Measures of familiarity

We consider several measures of the degree of familiarity. The first is related to "*professional proximity*". It is a dummy taking the value 1 if the investor's profession is in the same area of activity as the stock under consideration and zero otherwise. We use the one digit SNI92 codes (similar to SIC codes) to identify the areas of activities. For example, in the case an investor working in the mining sector holds a stock of a mining company, the dummy would be equal to 1.

The second measure is related to "*geographical proximity*". It is defined as the inverse of the distance between the residence of the investor and the place where the company is located. In particular, we use two different measures: the first one is the distance between the ZIP code of the investor and the ZIP code of the closest branch/subsidiary of the company whose stock we consider. As an alternative measure, we use the inverse of the distance between the ZIP

[8] The methodology of construction of the indices is described in Englund, Quigley and Redfearn (1998).

code of the investor and the ZIP code of the company headquarters. Given that the results do not differ and the variables are highly collinear, we report only the first specification. These measures are analogous to the one brought forward by Coval and Moskowitz (1999, 2001) in the study of geographical preferences in mutual fund investment. The greater the value of the variable, the closer the investor is located to the stock.

Variables of familiarity can also be based on the degree of notoriety of the company. Given that companies that have been around for a long time are more likely to be familiar to the investors, we construct a variable based on the age of the company from the date of incorporation (*time since incorporation*). Also, given that IPO is a strong informational event for the market, we construct a variable based on the time distance from an IPO (*time since IPO*). Let's see this point in detail. In a standard learning framework, uncertainty about an asset is negatively related to the flow of information on the asset that has accrued over time. Barry and Brown (1984, 1985) and Pastor and Veronesi (2002) use this rationale to explain, respectively, the small firm effect and the book-to-market pricing anomalies. The intuition is that, the more information is generated about a particular stock, the lower the uncertainty about its average profitability and therefore the higher the demand. A proxy for the public information is the time the firm has been around. Old firms are better known and their business is easier to evaluate as they tend to operate in areas with tested technologies.[9]

These variables - *geographical and professional proximity, time since incorporation and time since IPO* - proxy for public, "cheap" information. We argue that a proxy for private information is related to the time the investor has held the stock. That is, investors are more likely to be informed about stocks they already own than about stocks that are not yet part of their portfolio. Indeed, once the stock is in the portfolio, investors follow it more closely, reading the reports, paying attention to the earning announcements and so on. The longer they hold the stock, the more attention they tend to pay to the reports and announcements of the company. That is, stockholding may proxy for "selective attention". This selective attention would be consistent with the information-based familiarity hypothesis but not with the pure familiarity hypothesis.

Alternatively, it is reasonable to assume that the amount of information purchased is directly related to the stockholding.[10] The active purchase of information by the investors

[9] Therefore, the fact that older firms tend to be bigger and operating in more mature areas - i.e., size and book-to-market ratio increase with time - would explain why big and value firms should require lower returns.

[10] Indeed, if we assume a standard information technology (Peress, 2002) the wealthy investor would be willing to spend more to purchase information on a particular stock than a less wealthy investor, because the

about the stocks they hold reduces their sensitivity to risk (Bawa and Brown, 1984, 1985, Pastor and Veronesi, 2002) and therefore increases their propensity to invest in such an asset. We therefore construct a variable based on the time since when a stock entered the investor's portfolio (*holding period*).

5.2 Control Variables

We consider four types of control variables: measures of income and wealth, demographic variables, company specific characteristics, and regional and macroeconomic variables.

The *measures of income* include the variance of labor and entrepreneurial income of the investor and their correlations with their financial as well as real estate income. To construct proxies for permanent non-financial income, its volatility and its correlation to financial and real estate income, we use the approach of Carrol and Samwick (1997) and Vissing-Jørgensen (2001). We consider as non-financial income: labor income and entrepreneurial income. In the Appendix we provide a detailed description of the procedure.

As an additional robustness check, we replicate our results by using the actual levels of income, their volatility and the correlation of incomes and stock returns. This replaces the measures of permanent income, volatility of income and their correlations with portfolio returns that had been constructed according to the Carrol and Samwick (1997) methodology we described earlier. Given that the results are consistent, we will report only the ones based on the Carrol and Samwick methodology.[11]

We also include among the measures of income the mean and variance of the investor's portfolio return in the previous 12 months, financial and real estate capital gains and losses. These are the ones reported for tax purposes on stocks, bonds, mutual funds and real estate.

The *measures of wealth* include the overall level of wealth of the investor and its breakdown into its components. Overall wealth is defined as the sum of financial and real estate wealth. We also consider two types of borrowing constraints. The first one is the ratio of investor debt to total income and the second one is the ratio of investor debt to total wealth. Both of them are constructed at the investor level at time t. Also, we include a variable that accounts for the number of positions (type of assets) held by the investors.

The *demographic variables* include: the profession of the investor, his level of education, broken down into high-school and university level. We also include the age of the oldest

stake invested is relatively bigger than the one of the less wealthy investor.

[11] The results are available upon request.

member of the family of the investor and its value squared. This latter variable is consistent with standard results (Guiso and Jappelli, 2002, Vissing-Jørgensen, 2002) which find a non-linear relationship between age and the degree of stock market participation.

We also include a dummy that takes the value of 1 if the investor lives in the capital and 0 otherwise and the investor' immigration status. The latter takes the value 0 if all the members of the household are native Swedes, and 1 if at least one member of household immigrated from abroad.[12] Furthermore, we construct a variable to proxy for the ability of the investor in his occupation. This is based on the difference between his income and the average income of his profession. The assumption is that the higher the income of the investor relative to the average income of the other investors in the same area, the higher his ability should be.

The *company-specific characteristics* include the book-to-market value, the stock market capitalization of the company and its dividend yield. We also consider the affiliation of the stock with a glamour industry, under the assumption that it may breed familiarity (*glamor industry*). We focus on the high-tech industry and use a dummy that takes the value of 1 if the stock belongs to a high-tech company and zero otherwise. We also include dummies to control for the fact that the company has its headquarter in Stockholm or is O-listed. Location in Stockholm confers the company visibility, while O-listing (i.e., OTC) attributes the company some tax advantages. We want to control for both of them.

Finally, the *regional and macroeconomic variables* include an Index of Consumer Confidence and a set of dummies that account for the regional location of the investor as well as the industry in which he works. We consider 8 geographical areas and 10 industries.[13]

6 Identification and econometric issues

6.1 Identification of informed investors

In order to identify the informed investors, we use wealth and portfolio liquidity. These are two variables that are strongly related to the degree of informativeness of the investor and, presumably, independent of his behavioral heuristics. Let's start by considering how

[12] We also tried two alternative specifications. In the first one, we used the sum of the immigration statuses of the members of the household. That is, if two members of the household are immigrants, the variable takes value 2. In a second specification, we used the inverse of the number of years since the oldest immigrant in the household arrived in Sweden. These two alternative specifications deliver results that are qualitatively analogous to the ones reported. These results are available upon request from the authors.

[13] Geographical area definitions are based on the *NUTS2* classification for Sweden. An additional dummy for public sector workers is added to the industrial classification of households.

differences in wealth may affect the familiarity effect. Rational theories have a role for wealth. Higher wealth may relax informational constraints and, by making it easier to purchase more information, may reduce investor sensitivity to each unitary piece of information. Indeed, a wealthier investor would have the resources to consider a wider menu of assets. This would make him less dependent on "cheap" and publicly available information.[14] In general, an informed investor is less influenced by public sources of information (i.e., familiarity) as he can rely on his "private" one. Therefore, if familiarity is a proxy for (cheap) information, the demand of stocks of wealthier investors should be less sensitive to it.

On the contrary, behavioral theories are mute about the role of wealth. That is, investors are in general assumed to suffer from biases (e.g. familiarity), regardless of their level of wealth. Indeed, it makes sense to assume that if we are really dealing with human biases, saliency and behavioral heuristics should affect informed money managers (Coval and Moskowitz, 1999, 2001), as well as traders (Hau, 2001), managers (Gervais, Heaton and Odean, 2000) and individual investors (Benartzi 2001, Batthacharya, 2001).

For example, in the case familiarity rests on geographical proximity, a limited information story posits that investors are more likely to invest in stocks located near them simply because scarce resources restrict the processing of information. Geographical proximity provides a cheap way of acquiring information. A behavioral story, instead, postulates that geographical proximity is relevant as it enhances "confidence" in a particular stock. A change in wealth would not necessarily affect this.

Therefore, wealth provides a good starting point to distinguish these theories. Let's now consider the degree of liquidity of the investor's portfolio, that is the fraction invested in liquid assets. Liquidity impacts the investor's decision to acquire information and therefore, indirectly, the portfolio choice. The two forms of wealth that are characterized by opposite degrees of liquidity are financial assets (highest liquidity) and real estate (highest illiquidity)

Empirical findings show that access to professional financial investment advice is positively related to his net financial wealth and negatively related to the share of real estate in the investor overall portfolio. In particular, the "illiquidity of housing has a strong negative effect on the equity-value ratio and the relative share of housing equity in total wealth. Access to professional investment advice has negative effects on the housing share, and positive ones on that of net financial wealth" (Ioannides, 1989). That is, it is more likely that the investors

[14] Also, if investors hedge against learning uncertainty (Brennan, 1997, Xia, 2001), a change in wealth affects the desire to hedge and therefore the sensitivity of investment to information.

with the highest ratio of liquid to illiquid assets are also the more informed ones. The intuition is that, for an investor who has a bigger proportion of his wealth invested in financial assets (i.e., "liquid investor"), more information may reduce the uncertainty about a bigger fraction of his overall wealth.[15]

The positive mapping between information and the degree of liquidity of the investor portfolio suggests that we can use the ratio of liquid over illiquid assets as a proxy for his informativeness. Therefore, if the familiarity bias is just related to publicly available information (e.g., geographical or professional proximity to the stock), liquid investors, being more informed, would be less affected by it. In the case of heuristics, on the contrary, the impact of familiarity on stock holding should not change with the degree of liquidity of the overall portfolio.

In order to operationalize our approach, we consider four different samples: the overall sample and two subsamples constructed on the basis of either the wealth or the liquidity of the investor's overall portfolio. In particular, we define as high-wealth investors all the investors who, *in the previous year*, paid the wealth tax. We define low-wealth investors all the other ones. The high-wealth investors represent approximately 10% of the overall sample.

Then, we split the high-wealth investors into illiquid and liquid ones. In order to do this, we rank all the high-wealth investors in terms of the ratio of illiquid assets (i.e., real estate) over total wealth. Illiquid investors are the ones who, *in the previous year*, belonged to the top quintile, while liquid investors are the ones who, *in the previous year*, belonged to the bottom quintile.

In order to assess whether our identification is correct, we consider the profits of the different classes of investors. We expect informed investors - i.e., high-wealth and liquid investors - to make more profits than the uninformed ones - i.e., low-wealth and illiquid investors. We construct three types of profits for the different groups of investors: the change in overall wealth ($\Pi_{Wt} = Wealth_t/Wealth_{t-1} - 1$), the financial realized capital gains and losses ($\Pi_{Ft} = (\textit{Financial Capital Gains-Financial Capital Losses}_t)/Wealth_{t-1})$) and the real estate realized capital gains and losses ($\Pi_{Rt} = (\textit{Real Estate Capital Gains-Real Estate Capital Losses}_t)/Wealth_{t-1})$). In Table 2, Panel A, we report the mean and median values for the different classes, as well as their standard deviation and inter-quartile range (I.Q.R.). We consider raw profits and risk-adjusted profits - i.e., net of risk. Risk is constructed as the

[15] Indeed, lower information uncertainty (i.e., "estimation risk") increases the investment in the risky asset (Brennan, 1998).

standard deviation of the returns of the assets held, weighted according to their holdings. In Table 2, Panels B and C, we report tests of differences between classes. We compare high-wealth investors to the low-wealth ones and the liquid investors to the illiquid ones. Given the characteristics of the distribution of the investors, we consider three different tests: *t-tests*, Wilcoxon test and the Kolmogorov-Smirnov test.

In the case classes are defined in terms of wealth (Panel B), all the tests consistently show that high-wealth investors make more profits than the low-wealth ones. Also, in the case the classification is based on the degree of liquidity of the overall portfolio (Panel C), liquid investors seem to make more profits than the illiquid ones[16] This confirms our identification, as there is a direct mapping between "informed investors" and profits generated on financial assets. That is, the investors we consider as more informed (i.e., liquid and high-wealth investors) are also the ones who make more profits. This further supports our identification.

6.2 Econometric issues

We now consider the econometric issues we face in testing the two sets of restrictions and we lay-out the specifications we estimate. The biggest problem is generated by the selection bias. As we can see from Table 1, Panel D, there is a strong selection bias as only a minority of the households enters the stock market. The fact that we do not observe the investment decision of investors who do not participate in the financial market and the fact that such a participation is endogenous makes the standard estimates biased (Maddala, 1983, Nijman and Verbeek, 1996).

To address this issue we use Heckman (1979) two-stage procedure and separately estimate what induces the investor to enter the stock market and what influences his choices of assets. The decision to enter the market can be represented as:

$$P_{it} = \alpha_1 + \boldsymbol{\beta}_1 \mathbf{X}_{it} + \varepsilon_{1,it}, \tag{9}$$

where P_{it} is a dummy that takes the value of 1 if the investor participates in the financial market and zero otherwise and $\mathbf{X}_{it}$ is a vector of variables that affect the probability of entry. We include among them all the control variables defined before. The probability that the investor enters the financial market (i.e., $Prob(P_{it}^* = 1)$) is modeled as a normal c.d.f. F(·).

[16] If we use the tests based on the median (i.e., Wilcoxon and Kolmogorov-Smirnov), the more liquid investors always make more profits than the less liquid ones. If we consider the *t-test*, however, it appears that there is no significant difference for the gains defined in terms of overall wealth, while less liquid investors make more real-estate capital gains. Given the skewness of the distribution of the investors, we consider the median-based tests to be more reliable.

In order to estimate this probability we need to consider a bigger dataset based on the whole sample universe: i.e., both households who hold financial assets and households who do not. The expanded dataset includes the totality of the households tracked over time over each of the sample years 1995 through 1999, regardless of whether they invested in the stock market. It includes a total of 1,487,602 households tracked over time.

From the estimation of 9 we can construct a variables (λ_{it}, or "Heckman's lambda") that we use to control for the problem of omission of variables due to self-selection.[17] Given that equation 9 is just an auxiliary regression only needed for the proper estimation of the second stage, but out of the scope of this paper, we will not discuss the results.

Restriction H1

The portfolio, *conditional on market participation,* can be represented as:

$$\Delta_t h_i = \beta_f \Delta_{t-1}^{+} W_i^{f} + \gamma_f \Delta_{t-1}^{-} W_i^{f} + \beta_{re} \Delta_{t-1}^{+} W_i^{nf} + \gamma_{re} \Delta_{t-1}^{-} W_i^{nf} + \theta \lambda_{it} + \mu h_{it-1} + \boldsymbol{\delta} \mathbf{C}_{it} + \varepsilon_{2,it}, \tag{10}$$

where λ_{it} is the vector that control for selection bias and $\Delta_t h_{it}$ is the percentage change in the holdings of risky assets in the period ($[t, t-1]$). Risky assets have been standardized by the overall assets. $\mathbf{C}_{ijt}$ is the vector of control variables and $\Delta_{t-1}^{+} W_i$ and $\Delta_{t-1}^{-} W_i$ are, respectively, the positive and negative changes in wealth in the prior period ($[t-1, t-2]$).

We consider three measures of changes in wealth: change in overall wealth, security capital gains/losses and real estate capital gains/losses. Overall wealth contains both financial and real estate gains/losses. This specification allows us to directly identify the variables that affects investor's choice. It corresponds to equation 1 and restrictions 2 and 3. It is worth pointing out that a change in wealth is a less precise measure of gains/losses, as it also contains the saving/dissaving decisions of the investor. In order to account for this we include among the explanatory variables the main moments (mean, variance and correlation with financial and real estate income) of the other sources of income. Provided that the saving/income ratio of the investor does not change wildly over time, this should control for it.

To test mental accounting and restriction 4, we separately consider financial gains/losses and real estate gains/losses. We consider the three sets of control variables we described before: investor' income and wealth variables, demographic variables and macro-variables.

[17] It is derived from the first stage and is constructed as $\lambda_{it} = \frac{\phi(\beta_2 X_{2it})}{\Phi(\beta_2 X_{2it})}$, where X_{2it} is the vector that stacks all the explanatory variables.

Restriction H2

The stock-picking decision can be represented as:

$$h_{ijt} = \boldsymbol{\beta}\mathbf{F}_{ijt} + \boldsymbol{\delta}\mathbf{C}_{ijt} + \theta\lambda_{it} + \mu h_{it-1} + \varepsilon_{2,it}, \tag{11}$$

where λ_{it} is the vector that control for selection bias, h_{ijt} is fraction of the portfolio of risky financial assets of the *ith* investor invested in the *jth* asset and $\mathbf{F}_{ijt}$ is a vector that contains the "*familiarity variables*", that is the ones that proxy for the information the *ith* investor has about the *jth* stock. The control variables ($\mathbf{C}_{ijt}$) include the same control variables defined before as well as the mean and variance of the *jth* stock held by the *ith* investor. We also include the company-specific dummies.

We consider two specifications: in the first one we consider all the risky financial assets, in the second one we eliminate the mutual funds. This is done because mutual funds and stocks have some important differences that can affect our analysis. First, investment in a mutual fund is closer to investment in an index than to investment in individual stocks. Also, in the case of mutual funds, is more difficult to identify the familiarity variables, as the distance from the company is not necessarily a good proxy for the familiarity with the type and quality of investment of the company. Moreover, the period the fund has been held is not a goody proxy for the information about the quality of the fund, given the fact that "winners do not repeat themselves". It is important to note that, in the case in which we focus only on stocks, market participation (i.e., the first stage in the Heckman procedure) is also modified accordingly. That is, market participation is defined as investing in stocks only.

For both equations 10 and 11, we consider a dynamic specification in order to account for possible feedback effects [18] from past values of the dependent variable. Equations 10 and 11 are the empirical analogues of, respectively, equations 1 and 6.

Given that most of the hedging variables, as well as the parameter λ_{it}, have been generated on the basis of a previous estimation, we need to properly account for the bias induced by the "generated regressors". This problem is exacerbated by the existence of the lagged dependent variable. One way of dealing with this problem is to use an instrumental variable estimation. However, the endogeneity issues further complicates the task of finding proper instrumental variables, as only strictly exogenous variables or predetermined ones can be used in the

[18] At the aggregate level, market participation as well as portfolio choice are a function of asset returns. Asset returns are themselves a function of market demand. Given that we are effectively considering the demand of all the investors in the economy, we expect it to affect stock returns and therefore future demand.

case in which the variables are predetermined. We therefore follow the standard literature (Arellano, 1989, Arellano and Bond, 1991 and Kiviet, 1995) and the previous applications to finance (Vissing-Jørgensen, 2002) and use as instruments a combination of strictly exogenous variables (i.e., demographic variables) and the lagged values of the potential endogenous variables. Moreover, in order to control for heteroskedasticity, we correct the standard errors in the second stage regression.

Specifications 10 and 11 are estimated by using a two stage least squares with consistent variance-covariance matrix. We use data disaggregated at the individual investor level as well as disaggregated at the household level. The results do not differ from the ones based on households, so we will report only the latter.

The significance of the estimate of θ provides a test the null of no sample selection bias. In all the specifications we find a high degree of significance of the coefficient of the Heckman's Lambda suggests that self-selection is indeed important in the sample.

7 Empirical findings

We articulate the analysis as follows. First, we consider the aggregate risk taking decision and test loss aversion against the house-money effect and mental accounting. Then, we investigate stock-picking and test information-based familiarity versus pure familiarity.

7.1 Loss aversion, house-money effect and mental accounting

We start by comparing the alternative behavioral theories in terms of reaction to prior gains/losses. The results of the estimation of equation 10 are reported in Table 3. Panel A reports the result for full-fledged specification, while Panels B and C report robustness checks based on alternative specifications (without contemporaneous profits, without demographic variables). We report the results for estimates performed on the sample of participants with Heckman correction. We will discuss the results of the fully-fledged specification (Panel A) and leave the others as robustness checks.

The results support the house-money effect hypothesis, as investors react to prior positive changes in wealth by increasing risk taking. This holds for different classes of investors and for changes in overall wealth, as well as for financial and real estate capital gains.

In particular, for the overall sample, 1% increase in wealth in the prior year ($\Delta\, Wealth^{+}$) increases risk taking in the stock market by 0.26%, while in the case of a reduction of wealth

($\Delta Wealth^{-}$), 1% decrease in wealth reduces risk taking by a mere 0.05%. This strong asymmetry holds across different categories. It is interesting to note that the investors who react more strongly are the high-wealth ones and the liquid investors, that is the ones we define as "informed".

In the case of capital gains and losses, instead, the reaction varies across classes of investors. On average, a realized capital gain in the prior year equal to 1% of the investor's (overall) wealth increases risk taking in the stock market by 0.17%, while a realized capital loss in the prior year equal to 1% of the investor's (overall) wealth reduces risk taking in the stock market by 0.79%. However, these figure hide a strong heterogeneity across classes. Indeed, while low-wealth investors display a stronger reaction to prior gains (2.18%) than to prior losses (0.33%), high-wealth investors are only slightly more sensitive to prior losses (5.06%) than to prior gains (4.17%).

Interestingly, the illiquid investors do not seem to be affected by either gains or losses. They have most of their wealth invested in real estate and this may explain this lower reaction reaction. In the case of the liquid investors, instead, the reaction to gains and losses is stronger and approximately identical (i.e., 2.81% in the case of gains and 2.99% in the case of losses).

The fact that for the low-wealth investors the reaction is disproportionately stronger for capital gains as opposed to capital losses suggests that there may be borrowing constraints. In fact, in the case of low-wealth investors, both measures of borrowing constraints (i.e., debt-to-wealth ratio and debt-to-total income ratio*)* display a significant negative correlation with the investment into risky assets. In the case of the high-wealth investors, however, borrowing constraints seem to affect only the illiquid investors, presumably the ones more exposed to mortgages given their higher investment in real estate. The liquid investors are not affected by borrowing constraints.

If we now consider the test of mental accounting, we see that both financial gain/losses and real estate gain/losses affect overall risk taking in the stock market. On average, a realized real estate gain in the prior year equal to 1% of the investor's (overall) wealth increases risk taking in the stock market by 0.31%, while a realized real estate loss in the prior year equal to 1% of the investor's (overall) wealth reduces risk taking in the stock market by 0.18%.

At a disaggregated level, this holds for both high-wealth and low-wealth investors. The reaction to gains is less strong than the one to losses, for all classes with the exception of the illiquid ones. In particular, the reaction to prior real estate gains ranges between 1.7%

to 11.5%, while the reaction to prior real estate losses ranges between 0.49% to 0.90%. This feedback from real estate to financial market are consistent with a wealth effect due to the real estate (Case, Quigley and Shiller, 2001).

It is worth noting that these results employ data at a lower frequency that the previous findings based on transaction data (Odean, 1998 and Barber and Odean, 1999). These results are consistent with recent findings (Jackson, 2002) who show that, for Australian data, the disposition effect fades away after approximately 200 days. This suggests that maybe loss aversion is a better predictor of short term behavior, while house-money effect mostly applies to long run behavior.

7.2 Information-based vs. pure familiarity.

Let's now consider the stock-picking decision. The results of the estimation of equation 11 are reported in Table 4. In Panel A we report the results for the full sample and the high-wealth and low-wealth investors, while in Panel B we report the results for the liquid and illiquid investors.

The results provide strong evidence in favor of the familiarity effect. In general, investors tend to buy stocks that are either geographically or professionally close to them, or that have been around longer. This holds across the different specifications (with and without mutual funds) and for different sets of control variables. This provides evidence in terms of restriction 7. There is also a strong heterogeneity across investors. Indeed, the impact of the proximity variables for low-wealth and illiquid investors is different from the impact for high-wealth and liquid investors. As we will see later, these differences are statistically significant. This provides evidence in terms of restriction 8. We will see a formal test of it later on.

Let's now analyze more in detail the heterogeneity across investors. The first thing to note is that low-wealth investors tend to invest in stocks that are professionally and geographically close to them and that have been along and listed for a longer period. This provides evidence of familiarity.

The high-wealth investors, do quite the opposite, holding stocks not closely related to them. Moreover, the size of the coefficient for the high-wealth investors is always much smaller than the size of the coefficient for the low-wealth investors (on average ten times smaller). This suggests that familiarity does not affect the high-wealth investors on a similar scale. Wealthier investors rely less on the information based on professional proximity because they can afford better quality information. In fact, it seems that in aggregate, the high-wealth investors act as a counterpart to the low-wealth ones and allowing them to hold the stocks with which they are more familiar. The "fee" for this service should be part of the higher profits that they enjoy with respect to the low wealth investors and that we described in Table 3.

If then we break the sample of the high wealth investors into liquid and illiquid ones, we find a strong heterogeneity. The liquid investors behave as the high-wealth investors, while the illiquid investors behave as the low wealth ones. Indeed, illiquid investors are positively affected familiarity, while liquid investors tend to be less affected and in general in a negative way. These results are consistent with our identification of high-wealth and liquid investors with informed ones and illiquid as uninformed.

Very different is the behavior, if we consider the variable that proxy for the length of the period the stock has been held in the portfolio of the investor. We assumed that this variable proxies for the amount of private information purchased by the investor, as well as for his selective attention. As expected, high-wealth and liquid investors tend to invest more in stocks they have held for a long period. On the contrary, low-wealth and illiquid investors are less attracted by the stocks they have held longer. This provides further support for the information story.

The results of the specification where stock-picking includes both stocks and mutual funds are consistent with the ones that includes only stocks. The only noticeable difference is the even stronger impact of the familiarity variables in the case only stocks are considered. This is intuitive as the variables that proxy for familiarity are better defined in case of the investment in stocks than in the case of investment in mutual funds (for which their definition is more problematic).

Finally it is interesting to note that investors do not choose stocks on the basis of the correlation of their returns with either labor income or entrepreneurial income. Indeed, the correlation coefficients are never significant regardless of the level of wealth and the degree of liquidity of their overall portfolios. Stock-picking does not seem to be influenced by hedging motivations.[19]

Let's now proceed to test restriction 8 by testing the difference between the sensitivity

[19] These results on these variables pertain to the "control variables" and are not reported for space reasons. They are, however, available upon request from the authors.

of different classes of investors to familiarity. That is, we assess whether the familiarity effect is due to "information constraints" (i.e., $|\gamma_{high\ wealth}| = |\gamma_{low\ wealth}|$ or $|\gamma_{high\ liquidy}| = |\gamma_{low\ liquidy}|$) or to behavioral biases (i.e., $|\gamma_{high\ wealth}| \neq |\gamma_{low\ wealth}|$ or $|\gamma_{high\ liquidy}| \neq |\gamma_{low\ liquidy}|$).

The results are reported in Table 6. In the first column we report the results of the tests of high-wealth versus low-wealth investors, while in the second column we report the results of the tests of liquid versus illiquid investors. As before, we consider the case of investment in stocks only and the case of investment in stocks and mutual funds.

These tests are performed by re-estimating equation 11 on the entire sample and adding an interactive dummy that accounts for the fact that the investor is high-wealth or has an illiquid portfolio. The statistics on these dummies are relevant measures of our test.[20]

The results confirm our previous findings and show that there is a strong and statistically significant difference of the impact familiarity, depending on the level of wealth and on the degree of liquidity of the overall portfolio. The results go in the direction previously discussed. They show that the impact of familiarity depends on the degree of informativeness of the investor. More informed investors are less affected by the familiarity with the stock.

These results hold for all the specifications (high-wealth vs. low-wealth, liquid vs. illiquid investors), both when we limit the analysis to investment in stocks and to the case when we also consider mutual funds.

These results suggest that familiarity is a substitute for better information. Its importance decreases when the investor has access to more information. Alternatively, this may be seen as evidence that the role of behavioral biases might diminish with investor sophistication and/or information.

This has a big practical importance, given that the information-based familiarity hypothesis and the pure familiarity one have different normative and operational implications. Behavioral biases are related to human characteristics and are equally likely to be present in different countries and across markets. Informational constraints and market frictions are, instead, more likely to be affected by institutional as well as endowment differences. If familiarity is information-based, we may expect it to lose importance as the degree of sophistication of the investors or their relative wealth increase. This provides an additional dimension to understand, for example, processes such as globalization and financial integration and their

[20] In the case of the "overall" test of liquid versus illiquid investors, the sample is restricted to the sum of the liquid and illiquid. That is, the top and bottom 20% investors.

impact on market prices.

8 Conclusions

We focused on the determinants of portfolio choice, by directly comparing and testing different behavioral theories (loss aversion, house-money effect, mental accounting) and between behavioral and rational explanations of investor's reaction to familiarity. We provided evidence that shows that investors react to prior gains/losses according to what postulated by the house-money effect. That is, previous gains increase investor risk taking, while previous losses reduced reduce it. No evidence is found of mental accounting. It seems that investors consider wealth in its entirety and risk taking in the financial market is affected by gains/losses in overall wealth, financial wealth and real estate wealth.

In terms of individual stock picking we provided evidence in favor of the information-based familiarity hypothesis and showed that investor stock choice is mostly driven by the availability of information. Familiarity can be considered more as a proxy for the availability of information than a behavioral heuristics.

These results provide evidence that investors approach portfolio choice in a more complicated fashion we used to think of. Indeed, risk taking is affected by behavioral biases and, in particular, by the house-money effect. Individual stock picking, instead, seems to be mostly affected by the amount of information available. Reliance on familiarity with a stock is a cheap way of acquiring information about it that fades away when better and more expensive information is provided.

9 Appendix: Construction of income-related variables

Here, we briefly describe the methodology we follow to construct proxies for permanent non-financial income, its volatility and its correlation to financial and real estate income. We follow the approach of Carrol and Samwick (1997) and Vissing-Jørgensen (2001). We consider as non-financial income: labor income and entrepreneurial income. In particular, we define the relevant moments of long term investor's non-financial income:

$$E(\omega_{it}|\omega_{it-1}, X_{it-1}),\ Var(\omega_{it}|\omega_{it-1}, X_{it-1}) \text{ and } \rho_{it}, \tag{12}$$

where ω_{it} is the non-financial income of investor i at time t, X_{it-1} are the variables that can be used to predict income next period and ρ_{it} is the conditional correlation between shocks to log non-financial income and the log stock return. We assume that non-financial income follows:

$$\ln \omega_{it} = p_{it} + \varepsilon_{it}, \tag{13}$$

$$\begin{aligned} \text{where, } p_{it} &= g_{it} + p_{it-1} + \eta_{it},\ \varepsilon_{it} \sim N(0, \sigma^2_{\varepsilon i}),\ \eta_{it} \sim N(0, \sigma^2_{\eta i}), \\ \text{and } \operatorname{cov}(\varepsilon_{it},\ \varepsilon_{is}) &= 0,\ \operatorname{cov}(\eta_{it},\ \eta_{is}) = 0,\ \operatorname{cov}(\varepsilon_{it},\ \eta_{is}) = 0 \text{ for each } t, s. \end{aligned}$$

The variable p_{it} represents the permanent income component of non-financial income. It has a drift term (g_{it}) that is known and based on the information available at *t-1.* This allows us to write:

$$\begin{aligned} \ln \omega_{it} - \ln \omega_{it-1} &= p_{it} - p_{it-1} + \varepsilon_{it} - \varepsilon_{it-1} = g_{it} + \varepsilon_{it} - \varepsilon_{it-1} + \eta_{it} && (14) \\ \text{or } \ln \omega_{it} &= \ln \omega_{it-1} + g_{it} + \eta_{it} + \varepsilon_{it} - \varepsilon_{it-1}. && (15) \end{aligned}$$

This implies:

$$\begin{cases} E(\omega_{it}|\omega_{i,t-1}, X_{i,t-1}) = \ln \omega_{it-1} + g_{it} = \omega_{i,t-1} G_{it} \exp\{0.5 J_{it}\} \\ V(\omega_{it}|\omega_{i,t-1}, X_{i,t-1}) = J_{it} = (\omega_{i,t-1} G_{it})^2 \exp(J_{it})\{\exp(J_{it}) - 1\}, \end{cases} \tag{16}$$

where $G_{it} = \exp(g_{it})$, $J_{it} = \sigma^2_{\eta i} + 2\sigma^2_{\varepsilon i}$ and $X_{i,t-1}$ is the set of variables usable to predict g_{it}.

In order to estimate $E(\omega_{it}|\omega_{i,t-1}, X_{i,t-1})$ and $V(\omega_{it}|\omega_{i,t-1}, X_{i,t-1})$, we use income data for the period 1994-2000. Following Carrol and Samwick (1997) and Vissing-Jørgensen (2001) methodology, we regress $\ln \omega_{it} - \ln \omega_{it-1}$ on the set of explanatory variables $X_{i,t-1}$ and use the predicted values of such a regression as an estimate of g_{it} and the residuals as an estimate of $\eta_{it} + \varepsilon_{it} - \varepsilon_{it-1}$.[21] We then use the sample variance to construct $\sigma^2_{\eta i} + 2\sigma^2_{\varepsilon i}$. In order to control for measurement and estimation errors we use instrumental variables.

The correlation between financial and non-financial income (ρ_{it}) is constructed as the correlation between $\eta_{it} + \varepsilon_{it} - \varepsilon_{it-1}$ and the logarithm of gross stock return.[22]

References

[1] Arellano, M., 1990, Testing for Autocorrelation in Dynamic Random Effects Models. Review of Economic Studies, 57 127-134.

[21] The set of variables contained in $X_{i,t-1}$ are: demographic variables (secondary education, higher education, age, age squared, marriage status, size of the household, number of adults belonging to the household), changes in the demographic variables, industry dummies for the company the investor is working for (e.g., oil industry), dummies for the type of profession of the investor (e.g., doctor), emigration status.

[22] Following Vissing-Jorgenson, given the limited sample size, we use the entire sample period to estimate it.

[2] Arellano, M. and Bond, S.R., 1991, Some Tests of Specification for Panel Data. Monte Carlo Evidence and an Application to Employment Equations, Review of Economic Studies, 58, 277-297.

[3] Barber, B. M.and Odean, T., 2000, Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors, Journal of Finance, 55, 773-806.

[4] Barber, B. M. and Odean, T., 2001, Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment, Quarterly Journal of Economics, 116, 261-292.

[5] Barber, B. M. and Odean, T. 2002, Online Investors: Do the Slow Die First?, Review of Financial Studies, 15, 455-487.

[6] Barry, C.B. and Brown, S.J., 1984, Anomalies in Security Returns and the Specification of the Market Model, Journal of Finance, 39, 807-815.

[7] Barry, C.B. and Brown, S.J., 1985, Differential Information and Security Market Equilibrium, Journal of Financial and Quantitative Analysis, 20, 407-422.

[8] Brav, A. and Heaton, J.B., 2002, Competing Theories of Financial Anomalies, The Review of Financial Studies, 15, 575-606.

[9] Benartzi, S. and Thaler, R., 1995, Myopic Loss Aversion and the Equity Premium Puzzle, Quarterly Journal of Economics, 110, 73-92.

[10] Benartzi, S., 2001, Excessive Extrapolation and the Allocation of 401(k) Accounts to Company Stock, Journal of Finance, 56, 1747-1764.

[11] Benartzi, S. and Thaler, R., 2001, Naive Diversification Strategies in Defined Contribution Saving Plans, American Economic Review, 91, 79-98.

[12] Bhattacharya, U. and Groznik, P., 2001, Melting Pot or Salad Bowl: Some Evidence from US Investments Abroad, Mimeo.

[13] Brennan, M.J., 1997, The Role of Learning in Dynamic Portfolio Decisions, European Finance Review, 1, 295-306.

[14] Campbell, J.C., 2000, Asset Pricing at the Millennium, The Journal of Finance, 40, 1515-1567.

[15] Carroll, C. D. and Samwick, A. A., 1997, The Nature of Precautionary Wealth, Journal of Monetary Economics, 40, 41-71.

[16] Carroll, C.D. and Kimball, M.S., 2001, Liquidity Constraints and Precautionary Saving, NBER Working Paper 8496.

[17] Coval, J. D. and Moskowitz, T. J., 1999, Home Bias at Home: Local Equity Preference in Domestic Portfolios, Journal of Finance, 54, 2045-2073.

[18] Coval, J. D. and Moskowitz, T. J., 2001, The Geography of Investment: Informed Trading and Asset Prices, Journal of Political Economy, 109, 811-841.

[19] Edin, P.A., and Fredriksson, P., 2000, LINDA: Longitudinal Individual Data for Sweden 1960-1997, Uppsala University Working Paper.

[20] Englund, P., Quigley, J.M., and Redfearn, C., 1998, Improved Price Indexes for Real Estate: Measuring the Course of Swedish Housing Prices, Journal of Urban Economics 44, 171-196.

[21] Feenberg, D. R. and Poterba, J. M., 2000, The Income and Tax Share of Very High-Income Households, 1960-1995, American Economic Review, 90, 264-270 .

[22] Frieder, L. and Subrahmanyan, A., 2002, Brand Perceptions and the Market for Common Stocks, Mimeo.

[23] Genesove, D. and Mayer, C., 2001, Loss Aversion and Seller Behavior: Evidence from the Housing Market, Quarterly Journal of Economics, 116, 1233-1260.

[24] Gervais, S., Heaton, J.B. and Odean, T., 2000, Capital Budgeting in the Presence of Managerial Overconfidence and Optimism", Mimeo.

[25] Grinblatt, M. and Keloharju,, 2000, The Investment Behavior and Performance of Various Investor Types: A Study of Finland's Unique Data Set, Journal of Financial Economics, 55, 43-67.

[26] Grinblatt, M. and Keloharju, M., 2001, How Distance, Language, and Culture Influence Stockholdings and Trades, Journal of Finance, 56, 1053-1073.

[27] Grinblatt, M. and Keloharju, M., 2001, What Makes Investors Trade?, Journal of Finance, 56, 589-616.

[28] Guiso, L. and Paiella, M., 2002, Risk Aversion, Wealth and Background Risk, CEPR Discussion Paper 2728.

[29] Hau, H., 2001, Location Matters: An Examination of Trading Profits, Journal of Finance, 56, 1959-1983.

[30] Heaton, J. and Lucas, D., 2000, Portfolio Choice and Asset Prices: The Importance of Entrepreneurial Risk, Journal of Finance, 55, 1163-1198.

[31] Heaton, J. and Lucas, D., 2000, Portfolio Choice in the Presence of Background Risk, Economic Journal, 110, 1-26.

[32] Heckman, J., 1979, Sample Selection Bias as a Specification Error, Econometrica 47, 153-161.

[33] Huberman, G., 2001, Familiarity Breeds Investment, Review of Financial Studies, 14, 659-680.

[34] Ioannides, Yannis M., 1989, Housing, Other Real Estate, and Wealth Portfolios: An Empirical Investigation Based on the 1983 Survey of Consumer Finances, Regional Science and Urban Economics, 19, 259-280.

[35] Kiviet, J.F., 1995, On Bias, Inconsistency and Efficiency of Various Estimators in Dynamic Panel Data Models. Journal of Econometrics 68, 53-78.

[36] Merton, R.C., 1987, A Simple Model of Capital Market Equilibrium with Incomplete Information, Journal of Finance, 42, 483-510.

[37] Odean, T., 1998, Are Investors Reluctant to Realize Their Losses?, Journal of Finance, 53, 1775-1798.

[38] Odean, T., 1999, Do Investors Trade Too Much?, American Economic Review, 89, 1279-1298

[39] Pastor, L. and Veronesi, P., 2002, Stock Valuation and Learning about Profitability, Mimeo.

[40] Peress, J., 2002, Wealth, Information Acquisition and Portfolio Choice, Mimeo.

[41] Poterba, J. M. and Samwick, A.A., 1997, Household Portfolio Allocation Over the Life Cycle, NBER Working Paper: 6185.

[42] Shapiro, A., 2002, The Investor Recognition Hypothesis in a Dynamic General Equilibrium: Theory and Evidence, Review of Financial Studies, 15, 97-141.

[43] Sundin A. and Sundqvist, S.I., 2002, "Owners and Power in Sweden's Listed Companies," SIS Ägarservice AB, 1986-2002.

[44] Vissing-Jørgensen, A., 2002, Towards an Explanation of Household Portfolio Choice Heterogeneity: Non-Financial Income and Participation Cost Structures, Mimeo.

[45] Vissing-Jørgensen, A., 2002, Limited Stock Market Participation and the Equity Premium Puzzle, Mimeo.

[46] Xia, Y., 2001, Learning about Predictability: The Effects of Parameter Uncertainty on Dynamic Asset Allocation, Journal of Finance, 56, 205-246.

Table 1: Descriptive Statistics of the sample

This table contains the descriptive statistics of the sample. Panel A reports the general demographic characteristics (number of households for each year, members in household, adults in household, age of the oldest member of household, percent of the sample with secondary and higher education, percent of immigrants. Panel B reports the regional distribution of the sample based on NUTS2 classification. Panel C reports the age and gender distribution of the sample. Panel D reports the percentage of the households holding risky securities, having real estate holdings, having debt, and being involved in entrepreneurial activities. Panel E reports the financial characteristics of households in SEK. Where applicable, we report mean, standard deviation, median and inter-quartile range (IQR).

Panel A: General demographic characteristics

Variable		*Years*				
		1995	*1996*	*1997*	*1998*	*1999*
Number of households		292396	297832	298479	299053	299842
# of members in household						
	Mean	2.70	2.66	2.64	2.62	2.62
	Std. Dev.	1.51	1.51	1.51	1.51	1.51
	Median	2	2	2	2	2
	I.Q.R	3	3	3	3	3
# of adults in household						
	Mean	1.78	1.78	1.76	1.76	1.76
	Std. Dev.	0.69	0.69	0.69	0.69	0.70
	Median	2	2	2	2	2
	I.Q.R	1	1	1	1	1
Age of oldest household member						
	Mean	48.90	49.05	49.08	48.25	48.33
	Std. Dev.	17.17	17.15	17.13	18.29	18.34
	Median	46	47	47	46	46
	I.Q.R	24	24	24	25	24
% of the sample with secondary education		42.40%	41.63%	41.07%	40.73%	40.38%
% of the sample with higher education		28.86%	28.96%	29.31%	29.80%	30.31%
% of immigrants		13.22%	13.37%	13.51%	17.41%	17.80%

Panel B: Distribution of the sample over geographical regions

Regions	*Years*				
	1995	*1996*	*1997*	*1998*	*1999*
Stockholm	18.74%	18.62%	18.72%	19.15%	19.33%
Eastern middle Sweden	16.27%	16.00%	15.86%	15.99%	15.99%
Småland and the islands	8.60%	8.44%	8.55%	8.62%	8.60%
South Sweden	13.73%	13.53%	13.47%	13.62%	13.65%
West Sweden	19.14%	18.86%	18.57%	18.88%	18.90%
Northern Middle Sweden	9.46%	9.26%	9.11%	9.26%	9.20%
Middle Northern Sweden	4.24%	4.14%	4.07%	4.15%	4.11%
Upper Northern Sweden	9.82%	11.15%	11.65%	10.33%	10.22%

Panel C: Age and gender distribution of the sample

Age	*Males*	*Females*
0-9	9.13%	8.73%
10-19	9.03%	8.47%
20-29	4.82%	4.88%
30-39	7.05%	8.16%
40-49	7.35%	7.42%
50-59	5.85%	5.28%
60-69	3.10%	3.04%
70-79	2.38%	2.57%
80+	1.14%	1.61%
Total	49.85%	50.16%

Panel D: Participation rates

	Years				
Type of financial asset	*1995*	*1996*	*1997*	*1998*	*1999*
Security Market	5.1%	5.5%	20.0%	21.1%	16.4%
Real Estate Market	55.1%	54.1%	53.5%	53.8%	54.1%
Bank Loans	7.0%	9.6%	9.9%	9.6%	10.5%

Table 2: Descriptive statistics of profit measures.

This table reports the descriptive statistics for three measures of profit: wealth-based, securities' capital gains/losses based and real estate capital gains and losses-based. The first measure is defined as $\Pi_{Wt}=(Wealth_t/Wealth_{t-1}-1)$ and capital gains/losses measures are defined as $\Pi_{F,R\ t}=(CAPITAL\ GAINS_t-\ CAPITAL\ LOSSES_t)/\ Wealth_{t-1}$, where capital gains are the ones for securities (real estate) correspondingly. We report statistics for both raw measures and risk-adjusted measures. Panel B reports the results of t-test, Wilcoxon and Kolmogorov-Smirnov tests of profit measures equality for high-wealth vs. low-wealth households. . Panel C reports the results of t-test, Wilcoxon and Kolmogorov-Smirnov tests of profit measures equality for liquid vs. illiquid high-wealth households.

Panel A: Descriptive statistics of profit measures

	Low-wealth				*Wealthy*			
Variable	*Mean*	*Median*	*StdDev*	*I.Q.R.*	*Mean*	*Median*	*StdDev*	*I.Q.R.*
Raw Profits								
$\Pi_{F\ t}$	0.007	0.000	0.033	0.000	0.008	0.000	0.029	0.000
$\Pi_{R\ t}$	0.018	0.000	0.083	0.003	0.026	0.002	0.067	0.024
Π_{Wt}	0.094	0.050	0.415	0.162	0.123	0.087	0.259	0.204
Risk-adjusted profits								
$\Pi_{F\ t}$	0.037	0.000	0.236	0.002	0.069	0.000	0.335	0.002
$\Pi_{R\ t}$	0.161	0.000	0.921	0.014	0.350	0.016	0.981	0.258
Π_{Wt}	0.792	0.214	4.228	0.750	1.556	0.519	3.840	2.188

Panel B: Tests of differences between high- and low-wealth households

	t-test		*Wilcoxon Test*		*Kolmogorov-Smirnov test*	
Variable	*t-Value*	*Prob>|t|*	*Z*	*Prob>|Z|*	*KSa*	*Prob>KSa*
Raw Profits						
$\Pi_{F\ t}$	4.25	<.0001	4.78	<.0001	5.47	<.0001
$\Pi_{R\ t}$	16.16	<.0001	75.66	<.0001	43.51	<.0001
Π_{Wt}	13.65	<.0001	41.59	<.0001	21.58	<.0001
Risk-adjusted profits						
$\Pi_{F\ t}$	15.96	<.0001	4.13	<.0001	9.27	<.0001
$\Pi_{R\ t}$	29.99	<.0001	84.25	<.0001	43.67	<.0001
Π_{Wt}	29.36	<.0001	59.60	<.0001	33.70	<.0001

Panel C: Tests of differences between liquid and illiquid households

	t-test		*Wilcoxon Test*		*Kolmogorov-Smirnov test*	
Variable	*t-Value*	*Prob>|t|*	*Z*	*Prob>|Z|*	*KSa*	*Prob>KSa*
Raw Profits						
$\Pi_{F\ t}$	8.91	<.0001	7.97	<.0001	4.09	<.0001
$\Pi_{R\ t}$	-6.02	<.0001	14.11	<.0001	8.24	<.0001
Π_{Wt}	-1.55	0.1205	1.96	0.0501	3.76	<.0001
Risk-adjusted profits						
$\Pi_{F\ t}$	3.98	<.0001	5.06	<.0001	4.09	<.0001
$\Pi_{R\ t}$	-6.93	<.0001	14.94	<.0001	8.23	<.0001
Π_{Wt}	-0.81	0.4166	4.85	<.0001	4.15	<.0001

Table 3: Risk taking and prior gains/losses.

This table reports estimates of the determinants of risk taking. We report the results for the overall sample, the samples of the low- and high wealth investors and the samples of liquid and illiquid investors. The dependent variable is the percentage change in holdings of risky assets (stocks, mutual funds) with respect to the holdings at the beginning of the period. We define three measures of profitability, based on overall wealth and capital gains/losses from financial securities and real estate. $\Delta^{+}OWealth$ is defined as $Max(0,Wealth_t/Wealth_{t-1}-1)$, while $\Delta^{-}OWealth$ is defined as $Max(0,1-Wealth_t/Wealth_{t-1})$. Capital gains/losses are the capital gains and losses standardized by the overall level of wealth at the beginning of the period. We consider 6 groups of control variables: profit variables, income variables, borrowing constraint variables, demographic variables, geographic variables and macro-economic variables. The *profit variables* are the contemporaneous capital gains/loss and changes in overall wealth. Capital gains/losses include the capital gains/losses from the financial market, as well as the capital gains and losses from the real estate. They are standardized by the overall level of wealth at the beginning of the period. The *income variables* include the mean and variance of both labor and entrepreneurial income estimated using the methodology of Carrol and Samwick (1997) and Vissing-Jorgensen (2001), the *Return and the Volatility of the Portfolio* of the investor in the previous 12 months. The *borrowing constraint variables* are the *Debt/Wealth Ratio* and the *Debt/Income Ratio.* The former proxies for the leverage of the household, whereas the latter for the interest payment coverage. The *demographic variables* include: *Secondary Education* and *Higher Education.* These are dummies that take value 1 if the highest level of education in the household is, respectively, secondary education and University or higher education and 0 otherwise. *Ability* proxies for the individual abilities of the members of the household. The *Number of Adults* in the household and *Size of Household* are, respectively, the number of adults (18 year old and older) and the number of family members in household. *Age* and *Age2* are, respectively, the age and square of the age of oldest member of household. *Immigrant Status* is a dummy that takes value 1 if at least one household member immigrated to Sweden and zero otherwise. The *geographic variables* are a *Stockholm Dummy,* that is, a dummy that takes value 1 if household lives in Stockholm and zero otherwise, a set of eleven industry dummies and eight regional dummies. The *macro-economic variables* include an index of *Consumer Confidence* that represents the year-by-year change in the Swedish consumer confidence index (provided by Statistics Sweden) and the *Return on the Market* in the previous 12 months. For the specification based on the overall sample, we include a dummy that takes the value 1 if the investor is high-wealth and zero otherwise. Their coefficients are not reported. We also include λ, ("Heckman labda") , that is the variables that controls for the selection bias. Panel A reports the result for "full" specification. Panels B reports the specification without contemporaneous profits. Panel C reports the specification without demographic variables. We report the results for estimates performed on the sample of participants with Heckman correction. Estimates were performed using 2SLS. The *t-statistics* are reported in parenthesis.

Panel A

Variables	*Total Sample*		*Low-wealth*		*High-wealth*		*Liquid*		*Illiquid*	
	Estimate	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estima*	*t-stat*	*Estimate*	*t-stat*
Intercept	0.120	(2.62)	-0.478	(-7.02)	-0.547	(-2.15)	0.331	(1.44)	0.058	(0.11)
Δ^{+}OWealth	0.261	(3.16)	0.187	(11.44)	0.330	(6.47)	0.343	(7.62)	0.266	(2.37)
Δ^{-}OWealth	-0.051	(-2.16)	-0.123	(-3.67)	-0.152	(-2.26)	-0.532	(-2.91)	-0.003	(-0.01)
Security Capital Gains	0.174	(2.76)	2.179	(24.49)	4.169	(10.00)	2.817	(7.95)	0.972	(0.86)
Security Capital Losses	-0.791	(-1.99)	-0.332	(-2.58)	-5.065	(-2.64)	-2.992	(-2.25)	-22.508	(-1.49)
Real Estate Capital Gains	0.312	(7.48)	0.490	(7.36)	0.905	(5.36)	0.823	(5.15)	1.216	(4.12)
Real Estate Capital Losses	-0.179	(-2.63)	-1.704	(-11.50)	-11.008	(-8.27)	-5.434	(-4.18)	0.543	(0.22)
Control Variables										
Profit Variables	yes		yes		yes		yes		yes	
Income Variables	yes		yes		yes		yes		yes	
Borr. Constraint Variables	yes		yes		yes		yes		yes	
Demographic Variables	yes		yes		yes		yes		yes	
Geographic Variables	yes		yes		yes		yes		yes	
Macro-economic Variables	yes		yes		yes		yes		yes	
λ	-0.129	(-17.71)	-0.138	(-3.30)	-0.183	(-4.38)	0.055	(3.29)	-0.349	(-5.32)
Adjusted R^2	*0.0509*		*0.0550*		*0.0428*		*0.0446*		*0.0416*	
Number of Observations	*110399*		*76899*		*35000*		*19314*		*7093*	

Panel B

Variables	*Total Sample*		*Low-wealth*		*High-wealth*		*Liquid*		*Illiquid*	
	Estimate	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estima*	*t-stat*	*Estimate*	*t-stat*
Intercept	0.260	(25.97)	0.189	(11.74)	0.710	(22.98)	0.546	(16.65)	0.612	(10.14)
Δ^{+}OWealth	0.056	(2.33)	0.402	(11.06)	0.404	(9.07)	0.344	(7.95)	0.283	(2.63)
Δ^{-}OWealth	-0.006	(-2.55)	-0.178	(-10.22)	-0.079	(-2.51)	-0.506	(-2.85)	-0.145	(-0.55)
Security Capital Gains	0.428	(32.47)	1.629	(75.50)	0.302	(5.19)	0.195	(3.15)	0.881	(7.95)
Security Capital Losses	-0.186	(-8.31)	0.806	(22.79)	-2.314	(-20.22)	-1.822	(-16.07)	7.761	(26.96)
Real Estate Capital Gains	0.175	(2.78)	2.418	(24.92)	3.960	(10.79)	2.670	(7.75)	1.151	(1.05)
Real Estate Capital Losses	-0.822	(-2.00)	5.287	(1.45)	-2.651	(-1.97)	-2.789	(-2.20)	-16.886	(-1.16)
Control Variables										
Profit Variables	yes		yes		yes		yes		yes	
Income Variables	yes		yes		yes		yes		yes	
Borr. Constraint Variables	yes		yes		yes		yes		yes	
Demographic Variables	no		no		no		no		no	
Geographic Variables	yes		yes		yes		yes		yes	
Macro-economic Variables	yes		yes		yes		yes		yes	
λ	-0.133	(-19.46)	-0.085	(-7.81)	-0.164	(-4.66)	0.014	(3.36)	-0.357	(-5.76)
Adjusted R^2	*0.0501*		*0.0532*		*0.0422*		*0.0437*		*0.0422*	
Number of Observations	*110399*		*76899*		*35000*		*19314*		*7093*	

Panel C

Variables	Total Sample		Low-wealth		High-wealth		Liquid		Illiquid	
	Estimate	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estima*	*t-stat*	*Estimate*	*t-stat*
Intercept	0.312	(6.84)	-0.107	(-2.30)	-0.110	(-0.71)	0.106	(0.61)	-0.203	(-0.50)
Δ^{+}OWealth	0.214	(9.19)	0.183	(8.23)	0.418	(13.48)	0.371	(10.94)	0.497	(5.86)
Δ^{-}OWealth	-0.031	(-2.91)	-0.045	(-4.03)	-0.152	(-2.88)	-0.493	(-3.59)	-0.298	(-1.44)
Security Capital Gains	0.225	(3.61)	0.160	(2.67)	0.566	(2.24)	0.710	(2.66)	0.428	(0.50)
Security Capital Losses	-0.544	(-2.64)	-0.415	(-2.10)	-2.695	(-2.85)	-2.557	(-1.41)	-22.927	(-2.00)
Real Estate Capital Gains	0.333	(8.05)	0.312	(6.93)	0.376	(3.68)	0.350	(2.92)	0.402	(1.82)
Real Estate Capital Losses	-0.051	(-0.46)	-0.011	(-0.11)	0.723	(0.90)	0.920	(0.95)	-0.424	(-0.24)
Control Variables										
Profit Variables	no		no		no		no		no	
Income Variables	ves		ves		ves		ves		ves	
Borr. Constraint Variables	ves		ves		ves		ves		ves	
Demographic Variables	ves		ves		ves		ves		ves	
Geographic Variables	ves		ves		ves		ves		ves	
Macro-economic Variables	ves		ves		ves		ves		ves	
λ	-0.169	(-23.37)	-0.103	(-14.18)	-0.161	(-6.46)	-0.138	(-4.42)	-0.272	(-5.62)
Adjusted R^2	*0.0379*		*0.0378*		*0.0408*		*0.0434*		*0.0381*	
Number of Observations	*110399*		*76899*		*35000*		*19314*		*7093*	

Table 4: Stock-picking

This table reports estimates for the demand of individual risky assets. Demand is represented as a share of the risky portfolio. We consider the case where risky assets are only stocks (Panel A) and the case where the risky assets are both stocks and mutual funds (Panel B). We report the results for the overall sample, the sample of low-wealth households, the sample of high-wealth households and the samples of liquid and illiquid households. The measures of familiarity are*: Professional Proximity*, that is, a dummy that takes the value 1 if at least one household member is employed in the same industry in which the company is active and 0 otherwise. Comparison is done using 1-digit SNI92 code, *Geographical Proximity*, that is, the logarithm of the inverse of the distance (in kilometres) between the home of the investor and the closest company plant, *Time since Incorporation*, that is, the logarithm of the years since when the firm was initially registered, and *Time since IPO*, that is, the logarithm of the number of years since when the company was first listed on the exchange, holding period, that is, the number of years investor held the stock in his portfolio. The control variables are defined as in Table 4. We also include some company-specific characteristics such as: the *dividend-yield*, *book-to-market*, *size and glamour*. Book value is defined in terms of the previous fiscal year. The *market value* is defined at the last trading day of the year. *Size* is the logarithm of market capitalization. *Glamour* is a dummy that takes value 1 if the company is part of the high tech industry and 0 otherwise. We also include firm industry dummies and stock listing dummy. The firm industry dummies control for the industry the company is in, while the stock listing pertains to whether the company is o-listed (see data description in the text). We also include among the control variables the lagged value (as of year-1) of the dependent variable, the number of individual securities held in the portfolio and the return and the volatility in the previous 2 years of the stock (mutual fund) whose demand we estimate. The estimations are done using two-step Heckman procedure. The second stage of the Heckman estimation is performed using 2SLS. Lags of income, wealth and firm-level variables and demographic variables are used as instruments. *t-statistics* are reported in parenthesis.

Panel A: Stocks only

	Total Sample		Low-wealth		High-wealth		Liquid		Illiquid	
Variables	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*
Intercept	0.538	(13.84)	0.182	(3.97)	-1.607	(-74.80)	-1.697	(-34.22)	-0.029	(-0.27)
Professional Proximity	0.085	(5.21)	0.042	(6.40)	-0.021	(-2.91)	-0.032	(-2.03)	0.050	(1.98)
Geographical Proximity	0.018	(10.09)	0.008	(3.39)	-0.011	(-11.66)	-0.024	(-11.17)	0.065	(14.46)
Time since Incorporation	0.108	(8.75)	0.243	(16.98)	-0.009	(-1.32)	-0.261	(-17.95)	0.148	(4.47)
Time since IPO	0.176	(20.17)	0.266	(25.76)	-0.070	(-14.69)	-0.331	(-31.04)	-0.014	(-0.60)
Holding Period	-0.024	(-12.72)	-0.009	(-3.47)	-0.007	(-1.80)	0.032	(5.55)	-0.010	(-2.07)
Control Variables										
Profit Variables	yes		yes		yes		yes		yes	
Income Variables	yes		yes		yes		yes		yes	
Borr. Constr. Variables	yes		yes		yes		yes		yes	
Demographic Variables	yes		yes		yes		yes		yes	
Geographic Variables	yes		yes		yes		yes		yes	
Macro-econ. Variables	yes		yes		yes		yes		yes	
λ	0.133	(35.40)	0.093	(23.87)	0.294	(75.33)	0.201	(20.75)	0.122	(8.31)
Adjusted R^2	*0.4235*		*0.3562*		*0.4661*		*0.401*		*0.375*	
Number of Observations	*250558*		*124471*		*126087*		*31933*		*9119*	

Panel B: Stocks and Mutual Funds

	Total Sample		Low-wealth		High-wealth		Liquid		Illiquid	
Variables	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*
Intercept	5.307	(99.29)	6.476	(73.63)	-0.550	(-28.89)	-1.348	(-43.09)	-0.011	(-0.13)
Professional Proximity	-0.004	(-0.15)	0.169	(2.65)	-0.017	(-2.44)	-0.041	(-3.54)	0.174	(5.41)
Geographical Proximity	0.141	(71.83)	0.191	(57.34)	0.049	(72.78)	-0.037	(-29.74)	0.125	(45.55)
Time since Incorporation	1.746	(92.77)	1.643	(53.93)	-0.087	(-13.89)	-0.012	(-1.25)	0.274	(10.01)
Time since IPO	0.615	(55.49)	0.630	(34.11)	-0.141	(-33.77)	-0.140	(-22.34)	0.039	(2.13)
Holding Period	-0.029	(-10.19)	-0.146	(-27.53)	-0.003	(-1.50)	0.015	(10.54)	-0.066	(-15.37)
Control Variables										
Profit Variables	yes		yes		yes		yes		yes	
Income Variables	yes		yes		yes		yes		yes	
Borr. Constr. Variables	yes		yes		yes		yes		yes	
Demographic Variables	yes		yes		yes		yes		yes	
Geographic Variables	yes		yes		yes		yes		yes	
Macro-econ. Variables	yes		yes		yes		yes		yes	
λ	0.076	(12.68)	0.088	(9.72)	0.197	(47.09)	0.218	(30.93)	0.133	(8.63)
Adjusted R^2	*0.537*		*0.578*		*0.4889*		*0.4383*		*0.572*	
Number of Observations	*382314*		*227053*		*155461*		*40186*		*14398*	

Table 5: Differential effect of wealth and liquidity

We report the results of the estimation of the specifications reported in Table 5, augmented by interactive dummies to separate high-wealth and low-wealth households (columns 1 and 2) as well as liquid and illiquid households (columns 3 and 4). All the variables are defined as in table 5. We also include the level of the familiarity variables (i.e., *Professional Proximity*, *Geographical Proximity*, *Time since Incorporation*, *Time since IPO* and *Holding Period*). We consider both the case of investment in stocks and the case of investment in both stocks and mutual funds.

	Stocks Only				*Stocks and Mutual Funds*			
	Total Sample		*High-wealth*		*Total Sample*		*High-wealth*	
Variables	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*
Intercept	-0.037	(-0.67)	-0.950	(-40.16)	7.953	(133.82)	-0.457	(-20.74)
Professional Proximity	0.071	(2.49)	0.038	(4.12)	0.015	(0.27)	0.010	(3.18)
Professional Proximity Dummy	0.075	(1.38)	-0.125	(-6.96)	-0.073	(-1.14)	-0.140	(-8.19)
Geographical Proximity	0.022	(5.50)	0.004	(3.90)	0.163	(68.44)	0.029	(28.71)
Geographical Proximity Dummy	-0.011	(-3.35)	-0.032	(-17.34)	-0.013	(-5.52)	-0.083	(-49.80)
Time since Incorporation	0.097	(5.06)	0.358	(45.63)	2.018	(90.72)	0.420	(56.64)
Time since Incorporation Dummy	0.047	(2.54)	-1.710	(-209.07)	0.225	(12.08)	-0.642	(-82.82)
Time since IPO	-0.013	(-0.87)	1.891	(223.17)	0.934	(62.17)	0.427	(52.77)
Time since IPO Dummy	-0.074	(-3.85)	-0.673	(-114.76)	-0.333	(-16.04)	-0.166	(-32.43)
Holding Period	-0.046	(-11.87)	-0.056	(-48.95)	-0.132	(-30.07)	-0.112	(-104.06)
Holding Period Dummy	0.022	(4.46)	0.211	(98.59)	0.242	(4.29)	0.435	(219.28)
Control Variables								
Profit Variables	yes		yes		yes		yes	
Income Variables	yes		yes		yes		yes	
Borr. Constr. Variables	yes		yes		yes		yes	
Demographic Variables	yes		yes		yes		yes	
Geographic Variables	yes		yes		yes		yes	
Macro-econ. Variables	yes		yes		yes		yes	
λ	0.073	(13.42)	0.294	(67.39)	0.053	(7.77)	0.226	(48.46)
Adjusted R^2	0.553		0.4666		0.4436		0.4806	
Number of Observations	250558		126087		382314		155461	