View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Research Papers in Economics

A RESEARCH REPORT FROM STOCKHOLM INSTITUTE FOR FINANCIAL RESEARCH

Hedging, Familiarity and Portfolio Choice

MASSIMO MASSA

ANDREI SIMONOV

NO 21 — MAY 2004

Stockholm Institute for Financial Research

Stockholm Institute for Financial Research (SIFR) is a private and independent non-profit organization established at the initiative of members of the financial industry and actors from the academic arena. SIFR was launched in January 2001 and is situated in the center of Stockholm. Magnus Dahlquist serves as director of the Institute. The mission of SIFR is to:

- Conduct and stimulate high quality research on issues in financial economics, where there are promising prospects for practical applications,
- Disseminate research results through publications, seminars, conferences, and other meetings, and
- Establish a natural channel of communication about research issues in finance between the academic world and the financial sector.

The activities of SIFR are supported by a foundation based on donations from Swedish financial institutions. Major contributions have been made by: AFA, Alecta, Alfred Berg, AMF Pension, Brummer & Partners, Carnegie, Handelsbanken, Kapitalmarknadsgruppen, Länsförsäkringar, Nordea, Svenska Fondhandlareföreningen, and Östgöta Enskilda Bank.

Sveriges Riksbank funds a position as visiting professor at SIFR.

SIFR also gratefully acknowledges research grants received from Bankforskningsinstitutet, Föreningsbankens Forskningsstiftelse, Jan Wallanders och Tom Hedelius Stiftelse, Riksbankens Jubileumsfond, and Torsten och Ragnar Söderbergs stiftelser.

Stockholm Institute for Financial Research, Saltmätargatan 19A 11, SE-113 59 Stockholm, Sweden
Phone: +46-8-728 51 20, Fax: +46-8-728 51 30, E-mail: info@sifr.org, Web: www.sifr.org

Hedging, Familiarity and Portfolio Choice

Massimo Massa and Andrei Simonov

Stockholm Institute
for Financial Research

Hedging, familiarity and portfolio choice*.

Massimo Massa

INSEAD

Andrei Simonov

Stockholm School of Economics

March 1, 2004

Abstract

We exploit the restrictions of intertemporal portfolio choice in the presence of non-financial income risk to design and implement tests of hedging that use the information contained in the actual portfolio of the investor. We use a unique dataset of Swedish investors with information broken down at the investor level and into various components of wealth, investor income, tax positions and investor demographic characteristics. Portfolio holdings are identified at the stock level. We show that investors do not engage in hedging, but invest in stocks closely related to their non-financial income. We explain this with familiarity, that is the tendency to concentrate holdings in stocks with which the investor is familiar in terms of geographical or professional proximity or that he has held for a long period. We show that familiarity is not a behavioral bias, but is information-driven. Familiarity-based investment allows investors to earn higher returns than they would have otherwise earned if they had hedged.

JEL classification: **G11,G14.**

Keywords: **Asset pricing, portfolio decision, hedging.**

*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 J.Campbell, J.Cocco, B.Dumas, F.Gomes, J.Heaton, S.Livingston, M.Lettau, A.Lynch, P.Maenhout, T.Moskowitz, J.Shanken, P.Sodini, P.Veronesi, L.Viceira, A.Vissing-Jorgenson, M.Weber and the participants of the Summer Financial Markets Symposium at Gerzensee, 2003 WFA meeting and the NBER Asset Pricing Summer Institute. 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. We also thank Jean Cropper for editorial assistance. All the remaining errors are ours.

1 Introduction

One of the main questions in finance is what determines portfolio choice. The theoretical literature proposes hedging as one important driving motive: investors hold risky financial assets in order to offset their non-financial income risk. This stands in stark contrast to anecdotal evidence suggesting that investors, far from selecting an optimal portfolio, pick individual stocks on the basis of heuristics and stock familiarity. The lack of good-quality data on stock holdings, broken down at the investor level, and the scarcity of information about investors' overall assets, wealth and income have made it almost impossible to test the competing explanations. We bridge this gap by using portfolio data to assess the extent to which investors actively hedge non-financial income risk. Our contribution is along four dimensions.

First, we design and implement a first micro-based test of hedging. As we will explain in more detail in the following section, the standard tests of hedging - based on the relationship between the investment in risky assets and the correlation between investor non-financial income and the market portfolio - do not have power and provide indeterminate results if investors can choose among many risky assets and do not hold the market portfolio. Our test - based on *multi-asset* intertemporal portfolio choice in the presence of non-financial income risk - does not face this problem as it directly exploits data on the actual portfolio composition of the investor. Information detailed at the stock level allows us to identify the degree to which an investor tilts his portfolio away from the market portfolio and to use this tilt to test for hedging. We show that investors, in general, do not deliberately hedge. Quite the contrary, they tilt their portfolios toward stocks that are more correlated with their non-financial income. We also report evidence of how demographic, professional and wealth heterogeneity affect this tilt.

The second main contribution is the analysis of what induces investors to behave in a way opposite to the one posited by hedging. We document that this is directly related to "familiarity", that is, the tilt to invest in stocks that are geographically and professionally close to the investor or that have been held for a longer period. We show that the effect of familiarity is strong enough to more than offset the hedging motive, inducing an overall portfolio choice skewed toward familiar stocks. Moreover, familiarity affects the overall risk-taking of the investors

Third, we investigate the nature of familiarity and we show that familiarity-driven investment is a rational response to information constraints as opposed to a behavioral heuristic. In particular, we identify classes of investors differentially informed and subject to different types of "familiarity shocks" and trace how sensitive their portfolio choice is to familiarity. We show that the sensitivity changes with the degree of informativeness of the investors and with their exposure to familiarity shocks. In particular, familiarity mostly affects the less informed investors. When an investor has been subject in the immediate past to a shock affecting the source of familiarity (change of profession, relocation, change of employment status), the sensitivity of the investor to familiarity decreases.

Finally, we document how different strategies - i.e., hedging and familiarity-based investment - differ in terms of their profitability/costs. The portfolios of investors who hedge are characterized by lower returns than those of familiarity-based investors. This is consistent with an information-based nature of familiarity.

Our approach has also two additional merits. It is, to our knowledge, the first test of hedging that properly accounts for the lack of financial diversification. Indeed, it explicitly controls for the fact that investors hold a very undiversified portfolio. Moreover, it distinguishes hedging from other portfolio-choice motives - such as speculative investment ("myopic portfolio choice") - and controls for spurious cross-sectional correlation - such as the one between non-financial income and overall stock returns that naturally exists before any portfolio choice is made.

Our results are robust and improve over the existing literature thanks to the use of a new and unique dataset that combines, for the first time, individual portfolio holding data with comprehensive information on *all* the components of non-financial household wealth. We are able to inspect the *individual components of the investor's overall portfolio and relate them to his non-financial income.*

The dataset contains a representative sample of the Swedish population and has information on the wealth of the investors, broken down into their components (cash, equity holdings, mutual funds, real estate, loans, bonds and other assets). We also have available the income, wealth and the tax position of the investors as well as detailed information on their demographic and employment characteristics. The richness of our data allows us to overcome some of the main limitations that have affected the empirical studies of portfolio choice: the lack of information on individual portfolio composition, the use of aggregated data, the problem of inference based on survey data without a proper panel structure and the lack of information about real estate.

We improve in all the aforementioned dimensions. That is, we use *individual portfolio data* to quantify the extent to which investors hedge. We exploit information on demographic and employment characteristics of the investors as well as information on their level and sources of wealth and income to study the *heterogeneity in investor behavior*. We use, for the first time, a panel in which the *same investors are traced over time*, in terms of both portfolio choices and income, wealth, demographic and occupational characteristics. The *complete panel dimension* also allows us to control for past portfolio choices and return-related strategies such as trend-chasing, momentum, etc. Finally, we use a very broad and comprehensive measure of total wealth that also properly accounts for *real estate.* This allows us to explicitly consider the correlation between real estate and other sources of income of the investor.

The remainder of the paper is articulated as follows. In Section 2, we relate our contribution to the existing literature and describe our approach and contribution in detail. In Sections 3 and 4, we describe the datasets we use and the construction of the variables. In Section 5, we discuss the way we identify the informed investors and define the main econometric issues. In Section 6, we report the main findings. A brief conclusion follows.

2 Relation to the previous literature and our approach

2.1 The link to the previous literature

There exists a vast theoretical literature that analyzes portfolio choice in the presence of non-financial income. Non-financial income is modelled as "endowed exposure" that affects the desired financial exposure (Campbell, 2000). Non-financial income may include labor income (Campbell, 2000, Davis and Willen, 2000a, 2000b, Haliassos and Michaelides, 2002, Haliassos and Hassapis, 2003) and entrepreneurial income (Polkovnichencko, 1999, Heaton and Lucas, 2000b). Models may rely on the standard intertemporal Merton framework (Telmer, 1993, Heaton and Lucas, 1997, Koo, 1998, Viceira, 2001, Michaelides, 2001) as well as include limited horizon (Dammon *et al.*, 2001) and life-cycle considerations (Campbell *et al.*, 2001, Cocco, 2001, Cocco *et al.*, 2002, Gomes and Michaelides, 2002, Hu, 2001, Storesletten *et al.*, 2001).

There are different channels through which non-financial income may affect portfolio choice. Let us start by assuming that there is *only one risky asset* (i.e., the market portfolio). If non-financial income is perceived as riskless, it should induce HARA investors to increase their investment in risky assets (Merton, 1971, Bodie *et al.*, 1992, Jagannathan and Kocherlakota, 1996, Heaton and Lucas, 1997). This has been defined as the "level effect". If, however, non-financial income is perceived as risky, it should also affect the portfolio choice by "changing people's tolerance for stock market risk" (Heaton and Lucas, 2000a).

If non-financial income shocks are uncorrelated with the return of the risky asset,[1] the variance of non-financial income generates a crowding-out effect ("variance effect") on stockholding. The intuition is that, in the absence of correlation, the variance of non-financial income reduces the investment in the risky asset. This effect is then magnified (reduced) in the case in which these shocks are positively (negatively) correlated. In particular, a positive correlation between labor income shocks and the stock return reduces the investment in the risky asset, while a negative correlation increases it (Viceira, 2001). We define this

[1] The case of zero correlation corresponds to the situation in which non-financial income risk is just background risk.

latter effect, relating the investment decision to the correlation between financial risk and non-financial risk, as the "correlation effect".[2] More formally, this can be written as:

$$h = \beta_1 Var_y + \gamma_1 Corr_{y,m} + \boldsymbol{\delta_1}\mathbf{F_1}, \tag{1}$$

where Var_y represents the non-financial income risk, $Corr_{y,m}$ is the correlation between non-financial income and the single risky asset - proxied by the market portfolio - and h is the investment in the risky asset. $\mathbf{F_1}$ is a vector of control variables. All the *single-asset* models (from the infinite-horizon models to the life-cycle models) deliver the same set of restrictions between the share of wealth invested in the risky asset and non-financial risk. These restrictions can be summarized as: $\beta_1 < 0$ and $\gamma_1 < 0$. The latter ("covariance effect") provides a direct test of hedging.

At the empirical level, however, a definitive assessment of the way investors react to non-financial risk has, until now, eluded the literature. The results have been partial and often contradictory. For example, Vissing-Jørgensen (2002a) shows evidence of a level and variance effect of labor income on portfolio choice, but finds no evidence that investment in risky assets is affected by the correlation between labor income and the return on risky assets (*proxied by the market portfolio*). Heaton and Lucas (2000a) report considerable heterogeneity in exposure to non-financial income risk and show that households with greater exposure tend to hold a smaller share of stocks in their portfolio. Heaton and Lucas (2000b) show the existence of a significant positive correlation between equity returns (*proxied by the market portfolio*) and the income of self-employed persons, Cocco *et al.* (1999) find a very low or nil correlation between labor income and the market portfolio and Campbell *et al.* (1999) find a significant and positive correlation between labor income and the market portfolio lagged one year.

We argue that the lack of an unequivocal answer is due to the data used and the testing methodology that is embedded in equation 1. The literature has typically not used individual portfolio data to test for hedging. It has, instead, employed the market return as a proxy for the returns on the financial portfolio the investor may use to hedge his non-financial risk. Moreover, all the empirical tests have been based on theoretical restrictions defined in a *single-asset framework where the market portfolio is the only asset investors can choose.* This approach has two main limitations.

- First, in terms of data, the market portfolio is a very poor proxy for the assets the investor may use to hedge his non-financial income risk. It implicitly assumes that the investor holds a well diversified portfolio. This is not an innocuous assumption as, in fact,

[2] In the case where the correlation between non-financial and financial income risk is zero, non-financial income risk is in general defined as "Pure background risk". Background risk reduces risk taking for investors with DARA utility (Pratt and Zeckhauser, 1997, Kimball, 1993, Gollier and Pratt, 1996) and lowers the investment in risky assets (Elmendorf and Kimball, 1999, Koo, 1995, Guiso, et al., 1996).

investors concentrate their holdings in very few stocks. This has been clearly documented, for example, by Goetzmann and Kumar (2001) who report a very low degree of diversification, for a representative sample of individual accounts held with a large US broker.[3] In Sweden, the average investor holds 2.66 stocks, the low wealth investor 1.92 and the high wealth investor 3.6. On average 36% of the investor population holds stocks only (28% and 53% for low wealth and high wealth investors respectively), 57% holds only mutual funds (66% and 34% for low wealth and high wealth investors respectively) and 43% holds both stocks and mutual funds (34% and 66% for low wealth and high wealth investors respectively).

If investors do not hold the market portfolio, a more direct approach should be based on the analysis of the investor's actual portfolio. However, since the original studies of Lease, Lewellen and Schlarbaum (1974), the use of data disaggregated at the stock level in the analysis of portfolio investment has petered out.[4] A simple specification based on actual portfolio data - *but still framed in terms of a single asset model* - would be:

$$h = \beta_2 Var_y * sign(Corr_{y,p}) + \gamma_2 Corr_{y,p} + \boldsymbol{\delta}_2 \mathbf{F}_2, \tag{2}$$

where $Corr_{y,p}$ is the correlation between the non-financial income of the investor and his financial income (i.e., return on his portfolio) and $\mathbf{F}_2$ is a vector of control variables. This specification provides two testable restrictions. First, if investors want to hedge, we expect a positive correlation between non-financial and financial income to be associated with a reduction of the investment in risky assets and a negative correlation to be associated with an increase of it. Therefore, hedging requires that $\gamma_2 < 0$ if $Corr_{y,p} > 0$ and $\gamma_2 > 0$ if $Corr_{y,p} < 0$.

The second restriction deals with the variance effect. In the case of stochastic non-financial income, the variance of non-financial income reduces the investment in risky assets only if the correlation between financial and non-financial income is positive or close to zero (Heaton and Lucas, 1997, Cocco *et al.*, 1999, Viceira, 2001). A negative correlation, instead, makes it more attractive to invest in risky assets as it reduces the risk of the overall portfolio (Haliassos, 2002). The incentive is stronger the higher the non-financial income risk is. That is, if the correlation between non-financial and financial income is negative, investing in risky

[3]Also, Grinblatt and Kelhorju (2000, 2001a, 2001b), using Finnish data, show that the average investor holds a very undiversified portfolio.

[4]Only recently, Barber and Odean (2000, 2001, 2002), Goetzmann and Kumar (2001) and Odean (1998, 1999) have used information disaggregated at the stock level to study the influence of behavioral biases. However, they have information only on a subset of the entire stock-portfolio of the investors and do not have non-financial income variables.

Grinblatt and Kelloarji (2000, 2001a, 2001b) study investors' trading behavior with a dataset that contains, for the first time, the entire stock holdings of the investors. However, they focus their analysis on issues such as geographical preferences and momentum trading, without considering the overall dimension of the portfolio problem. Nor do they consider the correlation with the other sources of wealth. In fact, in all these studies, information on other sources of income of the investor - i.e., labor income, entrepreneurial income - is not available. This makes it impossible to study the correlation between financial and non-financial income.

assets decreases the overall risk and therefore the investment in risky assets should increase with the variance of non-financial income. This implies that the test of hedging is a joint test that requires a negative correlation between the investor's financial and non-financial income and a positive correlation between the investment in risky assets and the variance of non-financial income. The multiplication of non financial risk by the term $sign(Corr_{y,m})$ allows us to account for it. The higher the non-financial risk, the more it is worth investing in risky assets if this allows to diversification (i.e., $Corr_{y,p} < 0$) and the less it is worth investing if this instead increases the overall risk (i.e., $Corr_{y,p} > 0$). That is, if investors perceive the investment in financial assets as a way of hedging non-financial income, we expect $\boldsymbol{\beta_2 < 0}$.

- However, even this test would face a second and more structural limitation. It effectively assumes a single asset framework. In fact, if there are *multiple risky assets,* the investor may hedge by increasing his holdings of the assets that are negatively related to his non-financial income, as well as by reducing his holdings of the assets that are positively related to his non-financial income. For example, if an investor has available two stocks, one positively correlated to his non-financial income and one negatively correlated to it, he may either buy the stock that is negatively correlated with his non-financial income or sell the one that is positively related to it. Therefore, the decision to hedge may actually involve either an *increase* in risky assets (that are negatively correlated with non-financial income) or a *reduction* in risky assets (that are positively correlated with non-financial income).

Another way of seeing this is that a negative correlation between the share of the portfolio invested in risky assets and the correlation between financial and non-financial income is consistent with *both* investors buying risky assets to hedge and investors selling risky assets to enact a familiarity-based strategy. Notice that we cannot anymore make an unconditional statement *on whether investors hedge.* We can only say that, *if investors hedge*, they do so by increasing the investment in risky assets. This implies that, if investors hold more than one risky asset, the standard tests of hedging which relate holdings of risky assets to the non-financial risk of the investor, are indeterminate and do not have any power.

One way out is to redefine the tests by directly exploiting the information contained in the actual portfolio of the investor to construct an aggregate statistic that captures in an unequivocal way the decision to hedge. This statistic can be based on the way investors tilt their portfolios away from the market portfolio. Hedging requires investors to induce a more negative correlation between their non-financial and financial income than they would get by directly investing in the market portfolio. We will exploit this intuition in formulating our test of hedging. In particular, we will use a standard intertemporal model with multiple assets (Merton, 1971), [5] and derive the testable predictions that directly relate a measure

[5]In a multiple-asset framework, no life-cycle model exists that studies the portfolio decision with multiple risky assets. The only exception is Gomes and Michaelides (2002), where they consider two risky assets with

of the "tilt to hedge", based on the *actual* portfolio composition, to non-financial income risk. Before moving to it, it is worth mentioning two other problems that the estimation of equation 1 has faced and upon which we improve.

The first problem is the definition of the share of risky assets out of total wealth. Theoretical models in general provide restrictions based on the share of risky assets over *total wealth rather than just financial wealth*. However, total wealth has in general not been available. In particular, one dimension of investor's choice has often been ignored: *real estate*. This is all the more important considering the percentage of the investor's overall wealth that is tied up in real estate. No direct estimate of the trade-off between real estate and portfolio composition with multiple risky assets and a proper control for other sources of non-financial (labor and entrepreneurial) income has been attempted.[6] We will be able to explicitly control for real estate and return profile.

The second problem is the reliance on survey data with scarce panel dimension. The empirical literature focusing on the relationship between portfolio choice and income risk has used survey data that do not trace *the same investor* over time.[7] The lack of a proper panel dimension makes it difficult to identify the unexpected income shocks and to address issues of spurious correlation. Moreover, it does not control for the impact of past portfolio performance on the investor's choice. Therefore, momentum and trend-chasing motivations, and income effects become observationally equivalent. This may not be a big problem in a single-asset framework where a change in stock prices is directly related to overall income, wealth and consumption, but it becomes problematic in a multi-asset framework. We address these issues by using *actual data with a proper panel structure* that traces the same investor and his investment in different assets over time.

2.2 A new test of hedging

The availability of data broken down at the portfolio level allows us to devise a new test for hedging. This can be broken into two hypotheses. Let us consider them separately.

H1: The direction of the tilt in the portfolio risk profile.

The risk profile of the financial portfolio should be tilted toward assets with a negative correlation with the non-financial income of the investor and away from assets with a positive correlation.

different correlations. However, given the different focus, they do not report/conduct any of the comparative statics that would guide us.

[6]It has been shown that homeowners have riskier (financial) portfolios than renters, that house-price crowd-outs stockholdings and that leverage is positively related to stockholdings (Cocco, 2001, Yao and Zhang, 2001, Wu, 2002). There is also some evidence of causality running from the stock market to the real estate market, (Okunev, Wilson and Zurbruegg, 2000).

[7]The only exception is Vissing-Jorgenson (2000a, 2000b). However, the time series dimension for investors' holding information is only of two years.

The intuition is that holding financial assets allows the investor to hedge only if such assets are negatively related to his non-financial risk.[8] Ideally, if the investor wants to hedge, he will choose stocks that are negatively related to his non-financial risk. This will induce a negative correlation between the investor's financial and non-financial income.

However, a possible criticism of quantifying hedging on the basis of the actual correlation between financial and non-financial income is that this measure may be affected by the preexisting correlation between stocks and an investor non-financial income. For example, it is possible that the investor's income is negatively related to the average stock available on the market. In this case, a negative correlation between the non-financial income of the investor and his portfolio would not be evidence of deliberate hedging if such a correlation *is less negative than that between the investor's non-financial income and the market portfolio*. That is, the investor would actually be increasing the exposure to his non-financial risk by deliberately holding stocks more related to his non-financial income.

We therefore need a measure that proxies for hedging or the extent to which investors actively pursue a negative correlation between financial and non-financial income that differs from the one embedded in the correlation between the investor's non-financial income and the market. We will call this measure "index of hedging". It quantifies the extent to which the investor's portfolio differs from the market portfolio in terms of correlation (covariance) with his non-financial risk. In the following we will consider indexes of hedging built using differences in correlations and indexes of hedging built using differences in covariances.[9] We consider two alternative indexes of hedging:

$$\Gamma = Corr_{y,m} - Corr_{y,p} \text{ and } \Delta = Cov_{y,m} - Cov_{y,p}. \tag{3}$$

These measures track, at the investor level, the difference between two correlations (covariances). The first is the correlation (covariance) between his non-financial income and the overall *market* portfolio ($Corr_{y,m}$ or $Cov_{y,m}$). The second is the correlation (covariance) between his non-financial income and *his* financial portfolio ($Corr_{y,p}$ or $Cov_{y,p}$). The correlation (covariance) between the investor's non-financial income and the market portfolio represents the extent to which holding the market portfolio would help the investor diversify away his non-financial risk. It is a benchmark that can be used to assess the actual strategy of the investor. Γ (Δ) captures the contribution of the portfolio choice to the reduction of the investor's overall risk. It is positive in the case of hedging, so that restriction H1 requires:

[8] We are considering hedging and not mere portfolio diversification. This is achievable by increasing the number of assets in which to invest.

[9] The advantage of using the correlation is that it is measure-free, that is, it is a standardized variable that is not affected by the size of the investment. This is particularly useful if we want to assess the impact of hedging on portfolio choices over time and across investors. On the other hand, the standard intertemporal portfolio model with multiple assets provides sharper restrictions when these are cast in terms of covariances (i.e., restriction H2). In the Appendix, we will provide a more elaborate discussion of this point.

$\Gamma > 0$ ($\Delta > 0$). We will construct indexes of hedging for both sources of non-financial risk: labor and entrepreneurial risk.[10]

H2.A: The determinants of the tilt in the portfolio risk profile.

The tilt in the risk profile of the financial portfolio should be positively related to the variance of the non-financial income and to the covariance between the different sources of non-financial income.

If we assume that the investor has two sources of non-financial income (Y_z and Y_x, or labor and entrepreneurial income), and a level of wealth W, this restriction can be expressed as:

$$\Delta_z = \frac{Y_z}{W} Var_{Y_z} + \frac{Y_x}{W} Cov_{Y_z,Y_x} + \frac{1}{W}\Theta_z \tag{4}$$

where $\Theta_z = -(Y_z + Y_x)\sum_{j=1}^{n} \Omega_{S_j} * Cov_{S_j,Y_z}$ and $\Omega_{S_j} = \frac{(\mu_{s_j} - r)'}{(1-\gamma)\sigma^2_{s_j}}$. S_j, μ_{s_j} and $\sigma^2_{s_j}$ are, respectively, the price, the mean and the variance of the *jth* risky asset. In the Appendix we report a detailed derivation of equation 4. The intuition is the following. The investor will actively hedge more (i.e., will increase Δ_z or actively tilt his portfolio toward assets negatively correlated to his non-financial income), the higher the risk of the non-financial income (Var_{Y_z}) and the higher the covariance between his non-financial sources of income (Cov_{Y_z,Y_x}). Indeed, if the other non-financial sources of income are negatively related to one another, they should already provide a hedge. This should reduce the demand for financial hedging.

Hedging is also related to the assets' mean/variance ratios. An asset that is positively correlated to the non-financial income of the investor (Cov_{S_j,Y_z}), but has a high expected mean/variance ratio ($\Omega_{S_j} = \frac{Mean/Variance\ Ratio}{(1-\gamma)}$), will reduce hedging. Indeed, in this case, hedging would be expensive as it requires to forgo the gains implied by the high mean/variance ratio. We can recast the testable restriction H2.A. as:

$$\Delta_z = \beta_3 \frac{Y_z}{W} Var_{Y_z} + \gamma_3 \frac{Y_x}{W} Cov_{Y_z,Y_x} + \zeta_3 \frac{1}{W}\Theta_z + \boldsymbol{\delta}_3 \mathbf{F}_3, \tag{5}$$

where $\mathbf{F}_3$ is a vector of control variables. Hedging requires that $\beta_3 > 0$, $\gamma_3 > 0$ and $\zeta_3 > 0$.

The alternative hypothesis posits that the investor buys more of the stocks that covary with his non-financial sources of income. We will rationalize this alternative behavior in terms of "familiarity". What is familiarity? Huberman (2001) argues that there is a "general tendency of people to have concentrated portfolios, ...to hold their own company' s stock in their

[10] The first index is the difference between the correlation (covariance) of the investor's labor income and the market and the correlation (covariance) of his labor income and his financial portfolio. The second one is the difference between the correlation (covariance) of the investor's entrepreneurial income and the market and the correlation (covariance) of his entrepreneurial income and his financial portfolio.

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". In general familiarity is defined in terms of professional or geographical proximity to the stocks. For example, 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). Also, 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). Alternatively, Frieder and Subrahmanyam (2002) report that individual investors tend to hold disproportionate amounts of stocks with high brand recognition. Can the tilt toward stocks that covary with the investor's non-financial income be explained in terms of a tilt toward "familiar" stocks? This question can be addressed by testing whether the deviation from the optimal portfolio in a direction opposite to the one of hedging can be explained by the decision to invest in familiar stocks.

H2.B: The role of familiarity.

If familiarity affects investor portfolio choice, we expect the index of familiarity to be negatively related to the portfolio tilt.

This can be expressed by expanding the specification 5:

$$\Delta_z = \beta_3 \frac{Y_z}{W} Var_{Y_z} + \gamma_3 \frac{Y_x}{W} Cov_{Y_z,Y_x} + \zeta_3 \frac{1}{W} \Theta_z + \boldsymbol{\delta}_3 \mathbf{F}_3 + \nu_3 \Psi, \tag{6}$$

where Ψ is the index of familiarity as defined before. If it induces investors to reduce hedging, we expect $\nu_3 < 0$. Therefore, the testing of restrictions H2.A and H2.B can be reduced to testing specification 6.

2.3 The nature of familiarity

Familiarity may be due either to some behavioral heuristics or to better information on the particular stock.[11] 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 than information that is blended in the background. We will define this hypothesis, entirely grounded on behavioral heuristics, as "pure familiarity".

[11] A recent paper (DeMarzo, Kaniel and Kremers, 2002) provides an alternative modelization that rationally explains investors' bent for familiarity. In this context, familiarity arises out of investors' desire to hedge "relative consumption" vis-a-vis the neighbors.

The alternative approach is the "information-based familiarity". This states that "investors buy and hold only those securities about which they have enough information" (Merton, 1987, Shapiro, 2002). That is, investors are either not aware of all the stocks or do not know them well enough to be willing to invest in them. Information about a stock affects the investment decision by altering the perceived expected pay-off in a rational portfolio decision.[12]

What differentiates the two hypotheses is the relationship to information. In the case of pure familiarity, investors, even if they have available more information, erroneously rely on what is often mentioned or is closer to them (i.e., geographically or professionally) because this seems more salient and relevant. In the case of information-based familiarity, instead, familiarity is a way to cope with limited information. Therefore, in the case of pure familiarity, access to more information should not alter an investor's bias, while in the case of information-based familiarity more information would reduce the bias. In the limit, an informed investor would not rely on "cheap" public familiarity-based information such as the one related to professional or geographical proximity.

The analysis is complicated by the confounding effect of income and wealth shocks as well as by specific individual characteristics. For instance, let us consider the standard test of the impact of familiarity on investment. If the investor is subject to the shocks of the geographical area in which 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. Therefore, controlling for income and wealth effects is of paramount importance. To assess the nature of familiarity, we rely on the following tests.

H3: The nature of familiarity.

If familiarity (i.e., Ψ) proxies for information, the sensitivity of the investors to familiarity should differ with their degree of informativeness as well as with their exposure to familiarity shocks. That is, referring to equation 6, we have:

$$H_0 : |\nu_{3,\text{high info}}| = |\nu_{3,\text{low info}}| \,;\; H_a : |\nu_{3,\text{high info}}| \neq |\nu_{3,\text{low info}}| \tag{7}$$

in the case of different degrees of informativeness and

$$H_0 : |\nu_{3,\text{fam. shocked}}| = |\nu_{3,\text{non fam. shocked}}| \,;\; H_a : |\nu_{3,\text{fam. shocked}}| \neq |\nu_{3,\text{non-fam. shocked}}| \tag{8}$$

in the case of familiarity shocks.

[12] In particular, the active purchase of information on the stocks held in the portfolio reduces the sensitivity to risk (Bawa and Brown, 1984, 1985, Pastor and Veronesi, 2002) and increases the propensity to invest in such an asset.

The subscripts "high info" and "low info" identify high informed and low informed investors while the subscripts "fam. shocked" and "non-fam. shocked" identify investors who have experienced a "familiarity shock". The null of no change in the case of pure familiarity is tested against the alternative of information-based familiarity. If familiarity is information-based, it should differ across investors with different degrees of informativeness (restriction 7) and should change for investors subject to familiarity shocks (restriction 8).

A few words about the familiarity shocks. These are events that change investor's proximity to the stocks. If familiarity is information-based, a change in the proximity to the stock should affect its impact on investor behavior. For example, let us consider an investor living close to an IBM factory who moves to a location far from IBM and close to a CISCO subsidiary. If familiarity (i.e., geographical proximity) is a behavioral bias, the investor should switch from IBM to CISCO, but his overall tendency to invest in closer stocks should not change. If, however, familiarity is information-based, the mere process of moving should affect the amount of information the investor derives from his geographical proximity to such stocks. Indeed, it will take time for the investor to adjust to the new source of information on CISCO that his new geographical location entails and, in the meantime, the sensitivity to familiarity should drop.

Therefore, if proximity to a stock is a way of using cheap information, the impact of familiarity should be different and lower if the investor has recently been subject to a shock that has affected his source of familiarity - e.g., unemployment, professional and location shocks. On the contrary, if familiarity is a behavioral heuristics, prior changes in proximity to the stock should not matter.

To test restrictions 7 and 8, we exploit the richness of our dataset and focus on cross-sectional differences across investors. We will expand more on this later on. It is worth noting that this analysis, disaggregated at the investor level, is itself an innovation with respect to the standard literature, where most of the existing analyses are carried out at the aggregate level and where investors' professional or wealth heterogeneity as well as exposure to shocks are largely ignored.[13]

2.4 Implications for investment and profitability

The direct implications of the previous hypotheses manifests itself in the investor's risk-taking decision as well as in the profitability of its portfolio. The effect on risk taking can be represented as:

$$h = \beta_4 \Gamma + \nu_4 \Psi + \boldsymbol{\delta}_4 \mathbf{F}_4, \tag{9}$$

[13]Only Davis and Willen (2000a, 2000b) have recently considered data disaggregated by occupations and shown that while at the aggregate level, there is no correlation between income innovations and equity returns, at a disaggregated level a "portfolio formed on firm size is significantly correlated with income innovations for several occupations, and so are selected industry-level equity portfolios".

where h is the investment in the risky asset, Γ and Δ are our indexes of hedging and familiarity respectively and $\mathbf{F}_4$ is a vector of control variables. If investors buy stocks to hedge and their risk-taking decision is not affected by familiarity, we expect: $\beta_4 > 0$ and $\nu_4 = 0$, while if investors do not buy to hedge and are affected by familiarity, we expect: $\beta_4 < 0$ and $\nu_4 > 0$.

The decision to implement hedging as opposed to familiarity-driven strategies should also have direct implications in terms of profitability. In particular, we expect that if familiarity is information-based, conditioning on the level of private information of the investor (e.g., proxied by wealth or liquidity of the portfolio) it should increase the profitability, while hedging, being a sort of costly insurance, should reduce them. We can therefore test:

$$\Pi = \beta_5 \mathbf{\Gamma} + \boldsymbol{\gamma}_5 \mathbf{\Psi}_t + \boldsymbol{\delta}_5 \mathbf{F}_5, \tag{10}$$

where Π represents the financial profits of the investor. We expect that $\beta_5 < 0$ and $\boldsymbol{\gamma}_5 > 0$.

3 Data description

We use Swedish data. Sweden provides a very good experiment as, contrary to common belief, it has a flexible labor market similar to that of the US in terms of companies' freedom to hire and fire. The termination notice is the shortest one among all the European countries (including the UK).[14] Moreover, unemployment benefits are phased out over time and terminated after 6 months. This makes non-financial income risk-hedging more relevant. Data has been collected from different sources. For each investor, we have detailed information on his individual holdings of stocks (broken down at the stock level), the holdings of mutual funds[15], bank accounts, real estate and other types of wealth. We also have available information on the different sources of income 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 as to construct a time series of investment and income for each investor. For each stock, we have detailed information on the company and the price, volume and volatility at which it trades. We also use aggregated data on Swedish macro-economic conditions and on the indexes of the real estate market. Let us look at the sources in more detail.

3.1 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

[14] The termination notice in Sweden is 30 days, compared with, for example, 90 days in the UK and France and 60 days in Belgium.

[15] We have the aggregate value of the money invested in mutual funds. For the purpose of this study, we consider mutual funds as risky assets analogous to stocks and we proxy their return with the market index. Our results are robust to the way we treat the mutual funds. That is, re-estimating our specifications excluding the mutual funds from the set of risky assets we get results consistent with those that are reported.

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 the like (for details, see Sundin and Sundquist, 2002).

Our data cover the period 1995-2000. Overall, the records provide information about the owners of 98% of the market capitalization of publicly traded Swedish companies.[16] The data provided by SIS Ägarservice AB were linked by Statistics Sweden with the LINDA dataset described below.

3.2 LINDA

LINDA (Longitudinal INdividual DAtaset for Sweden) is a register-based longitudinal data set and is a joint endeavor 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 of the population during the period 1966 to 2000. For each year, information on all family members of the sampled individuals are added to the dataset. 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/. We do not have information on the implicit claims on retirement benefits through state provided pensions. However, it is worth mentioning that the level of these benefits (just like in most European countries) is directly related to the salary level. Therefore, by including the level of the non-financial income (wage, salary, etc.) we are implicitly and partially controlling for them.

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

[16]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.

of real estate, market value of publicly listed securities, balance of bank accounts and fair value of valuable possessions (including jewelry, cars, antiques, etc.). For the purpose 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 Statistics Sweden.[17] For the privately held unlimited liability companies, the value of the assets is included in the household's tax return. There is no estimate of market value of privately held limited liability companies that are not listed. However, the data contain an indicator variable for owners of privately held companies. The size of the group is rather small (1.74%-1.91% of the sample depending on a year) and is unlikely to affect our estimates in a significant way. Moreover, for the members of the wealthiest 5,000 families, we have been able to reconstruct their values and to correctly impute it by using information from SIS Ägarservice AB (Sundin and Sundquist, 2002).

The combined LINDA/Shareholding dataset covers the period 1995-2000. The overall sample contains 1,807,602 observations. However, only 1,757,406 observations were used.[18] In addition, we also use 1990-1994 data from LINDA in the implementation of the Carrol and Samwick (1996) procedure to construct the moments of conditional non-financial income. In Table 1, we report some descriptive statistics. In particular, Panel A contains the general demographic characteristics (number of households, members in household, age of the oldest member of household, percentage of the sample with secondary and higher education, etc.). Panels B and C report, respectively, the age and gender distribution of the sample and their wealth and income characteristics, defined in terms of wealth, real estate, labor and entrepreneurial income.

One point that is worth stressing is the fact that we use data from the stock market bubble period (1995-2000). This might affect the results on hedging as investors may be overall less cautious in their investment strategies, jumping on the bandwagon of global euphoria of the period. This would induce them to hedge less and allocate their investment toward the risky assets where they think to have better information (i.e., "closer stocks"). However, the availability of a proper panel structure helps us along this dimension in many different ways. First, the data on non-financial income span an entire decade (1990-2000) that covers a recession (1990-1994), a recovery (1994-1996), a boom (1997-1999) and the burst of the bubble (2000). This implies that our measures of permanent non-financial income, its volatility and its correlation to financial income are scarcely affected by the bubble.

[17] 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.

[18] We excluded observations for households that were in the sample for less than three years and households with the oldest member being younger than 18 years old. Also, it is worth noting that we define as shareholder anyone who has more than SEK 2,000 worth of stock (that is US$ 200). This is the definition used by Statistics Sweden.

Second, even if financial data (investor holdings and portfolios) are determined during a bubble period, two caveats apply. First, the bubble in Sweden did not affect all the stocks in the same way and we are dealing with disaggregated data and information detailed at the stock level. Second, the availability of stock level information allows us to construct measures of past performance and volatility of the investor portfolio. These are the "momentum/stock performance variables". They are meant to capture the shift in an investor's portfolio due to stock market changes or to stock-tracking, momentum or performance-chasing activity of the investor. They should, as least partially, control for trend-chasing, momentum investing and short-term strategies induced by the bubble. These make our results more robust than the equivalent ones based on US data, for which the sample is shorter (in general 1 or, at best, 2 non consecutive years) and with no panel structure.

3.3 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 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 the 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.[19] 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).

4 Construction of variables

4.1 Income-related variables

Following the standard approach, we specify investors' portfolio policies in terms of their permanent income, that is the conditional moments the long-term income, as in Heaton and Lucas (2000b). As an additional robustness check, we also replicate our results by using the actual income. Given that the results are consistent, we will report only those based on permanent income. The other results are available upon request.

In order to construct proxies for permanent non-financial income, its variance and its correlation to financial and real estate income, we use the approach of Carrol and Samwick (1997) and Vissing-Jørgensen (2002a). We consider as non-financial income: labor income

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

and entrepreneurial income. In particular, we define the conditional moments of the long-term investor's non-financial income:

$$E(Y_t|Y_{t-1}, X_{t-1}) \text{ and } Var(Y_t|Y_{t-1}, X_{t-1}), \tag{11}$$

where Y_t is the non-financial income of investor i at time t and X_{t-1} are the variables that can be used to predict income next period. We assume that non-financial income follows:

$$\ln Y_t = p_t + \varepsilon_t, \tag{12}$$

$$\begin{aligned} \text{where, } p_t &= g_t + p_{t-1} + \eta_t,\ \varepsilon_t \sim N(0, \sigma_\varepsilon^2),\ \eta_t \sim N(0, \sigma_\eta^2), \\ \text{and } \operatorname{cov}(\varepsilon_t,\ \varepsilon_s) &= 0,\ \operatorname{cov}(\eta_t,\ \eta_s) = 0,\ \text{for each } t \neq s \text{ and } \operatorname{cov}(\varepsilon_t,\ \eta_s) = 0 \text{ for each } t, s. \end{aligned}$$

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

$$\ln Y_t - \ln Y_{t-1} = p_t - p_{t-1} + \varepsilon_t - \varepsilon_{t-1} = g_t + \varepsilon_t - \varepsilon_{t-1} + \eta_t \tag{13}$$

$$\text{or } \ln Y_t = \ln Y_{t-1} + g_t + \varepsilon_t - \varepsilon_{t-1} + \eta_t. \tag{14}$$

This implies:

$$\begin{cases} E(Y_t|Y_{t-1}, X_{t-1}) = Y_{t-1} G_t \exp\{0.5 J_t\} \\ Var(Y_t|Y_{t-1}, X_{t-1}) = J_t = (Y_{t-1} G_t)^2 \exp(J_t)\{\exp(J_t) - 1\}, \end{cases} \tag{15}$$

where $G_t = \exp(g_t)$, $J_t = \sigma_\eta^2 + 2\sigma_\varepsilon^2$ and X_{t-1} is the set of variables usable to predict g_t.

In order to estimate $E(Y_t|Y_{t-1}, X_{t-1})$ and $Var(Y_t|Y_{t-1}, X_{t-1})$, we use income data for the period 1990-2000, with a 5-year rolling window, based on the previous 5 years of data. Following Carrol and Samwick (1997) and Vissing-Jørgensen (2002a) methodology, we regress $\ln Y_t - \ln Y_{t-1}$ on the set of explanatory variables X_{t-1} and use the predicted values of such a regression as an estimate of g_t and the residuals as an estimate of $\eta_t + \varepsilon_t - \varepsilon_{t-1}$.[20] We then use the sample variance to construct $\sigma_\eta^2 + 2\sigma_\varepsilon^2$. $E(Y_t|Y_{t-1}, X_{t-1})$ and $Var(Y_t|Y_{t-1}, X_{t-1})$ are our measures of income and variance of income (i.e., Y and Var_y). We report descriptive statistics of them in Table 1, Panel D. We also construct a measure of the conditional correlation between shocks to log non-financial income ($\eta_t + \varepsilon_t - \varepsilon_{t-1}$) and the log gross stock returns (i.e., $ln(1 + R_t)$). Following Vissing-Jørgensen (2002a), given the potential inaccuracy of estimates based on few observations, we calculate the correlation over the entire sample.

[20]The set of variables contained in X_{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), immigration status.

We consider two financial returns (R_t), the return on the market portfolio (R_{mt}) and the return on the portfolio of the investor ($R_{port,t}$). We analogously construct the covariances. These correlations and covariances correspond to the $Corr_{y,m}$, $Corr_{y,p}$, $Cov_{y,m}$ and $Cov_{y,p}$ we defined before, and that we will use in the estimations. Descriptive statistics of it are plotted using frequency diagrams of different correlation coefficients in Figures 1 and 2 and are reported in Table 3.

Also, as a robustness check, following Campbell *et al.*, (1999), we consider the correlation between non-financial income and stock returns lagged one year. In line with the literature, the correlation increases. However, the results of the estimates of the main specifications do not differ from those reported.[21]

4.2 Indexes of familiarity

We need a measure that captures the extent to which an investor tilts his portfolio toward assets with which he is more familiar. We will call this "index of familiarity". We consider three indexes 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 company whose stock is 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 of an investor working in the mining sector holding a stock of a mining company, the dummy would be equal to 1.

The second measure is related to "geographical proximity", that is the proximity 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 logarithm of the inverse of 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 logarithm of the inverse of the distance between the ZIP 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 put 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 is the investor to the stock.

Finally, we may argue that investors are more likely to be informed about the 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 actively purchasing information about it. In other words, stockholding may proxy for selective attention and active purchase of (private) information. We therefore construct a variable that proxies for "holding period", based on the time a stock

[21] They are available upon request from the authors.

entered the investor's portfolio.

These measures are constructed at the stock level. They are then aggregated across all the stocks of the investor, and weighted by their share in the portfolio. This procedure delivers three measures of familiarity for each investor and time t.

4.3 Control variables

We consider five types of control variables: measures of income and wealth, demographic variables, professional ability and risk, momentum/stock performance variables and residual control variables.

The *measures of income and wealth* contain the vector of the wealth of the each investor at time t, broken down into its individual components (i.e., financial, real estate and other), as well as measures of income (i.e., labor and entrepreneurial) and overall (i.e., financial and non-financial) capital gains and losses of the each investor at t. We also include the correlation between non-financial income (both labor income and entrepreneurial income) and real estate.

The *demographic variables* include: the profession of the investor, his level of education, broken down into high-school and university level, the age of the oldest 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, 2002a) which find a non-linear relationship between age and the degree of stock market participation.

We also construct variables to account for the *professional ability and risk* of the investor. A first variable proxies 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. A second variable is a measure of *unemployment risk* that proxies for the probability of being unemployed in the following year. It is the one-year-ahead forecast of a linear probability model where the unemployment status (i.e., 1 if unemployed and zero otherwise) is regressed on demographic variables, measures of income and wealth and regional, geographical and professional dummies.

The *momentum/stock performance variables* are meant to capture the shift in the investor's portfolio due to stock market changes or to stock-tracking, momentum or performance-chasing activity of the investor. They are the return and volatility on the investor's portfolio in the previous twelve months.

The *residual control variables* include standardized levels of debt for the investor (ratio of investor debt to total income and ratio of investor debt to total wealth), the return and volatility on the market portfolio in the previous twelve months, 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. The debt ratios may be considered as proxies for borrowing constraints. Indeed, Hayashi (1985) and Zeldes (1989) define as liquidity constrained the households with low savings or low financial assets. Given that we can directly observe the debt, we can use it as a proxy. We also consider 8 geographical areas and 11 industries.[22]

It is worth noting that a significant fraction of Swedish households is employed by the government (either national or municipal). Their incomes are much less variable, and the risk of unemployment dramatically different. Therefore, their income should be significantly less correlated with the return on the stock market. This may affect the results. We address this issue in two ways. First, we also include among the control variables a dummy that accounts for the profession of the investor. Being a public employee is just one of the 11 industries. This directly controls for the fact that the risk of unemployment may drastically differ in the case of public employee. Second, as an additional robustness check, we also perform our analysis on two separate subsamples: government employees and non-government employees. The results (not reported but available upon request) are consistent with those reported.

Finally, we also include a Stockholm dummy and a dummy that controls for the immigration status. The Stockholm dummy takes the value of 1 if the investor lives in the capital, and 0 otherwise. The immigration status is a dummy that takes the value 0 if all the members of the household are native Swedes, and 1 if at least one member of household immigrated.[23] All monetary variables (level and variance of non-financial income, wealth, ...) except Capital Gains/Losses, have been transformed into logarithms.

5 Identification and econometric issues

5.1 Identification of informed investors

In order to identify the informed investors, we use investor 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 us start by considering wealth.

Rational theories have a role for wealth. Higher wealth may relax informational constraints and make it easier to purchase more information. If we assume a standard information technology, the wealthy investor would be willing to spend more to purchase information on a particular stock than a less wealthy investor, because the relative cost of investing in information decreases with the level of wealth (Calvet *et al.*, 2000, Peress, 2002). A wealthier

[22]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.

[23]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 those reported. These results are available upon request from the authors.

investor, having the resources to consider a wider menu of assets, would be less dependent on "cheap" publicly available information.[24] Therefore, if familiarity is a proxy for cheap information, the demand of stocks of wealthier investors should be less sensitive to it. In general, an informed investor is less influenced by public sources of information (i.e., familiarity) as he can rely on his "private" one.

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.[25] Indeed, it makes sense to assume that if we are really dealing with human biases, saliency and behavioral heuristics should equally affect wealthy and non wealthy investors. For example, in the case where familiarity rests on geographical proximity, a limited information story posits that investors are more likely to invest in stocks located near them simply because geographical proximity provides a cheap way of acquiring information. A behavioral story, on the other hand, postulates that geographical proximity is relevant as it makes more "salient/available" the characteristics of the stocks that are located closer. A change in wealth would not necessarily affect this. Therefore, wealth provides a good starting point to distinguish these theories.

Let us 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 net financial wealth of the investor and negatively related to the share of real estate in his 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 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.[26]

The positive mapping between information and the degree of liquidity of the investor's portfolio suggests that we can use the ratio of liquid over illiquid assets as a proxy for

[24] 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.

[25] Some experiments have shown that biases decrease when the amount at stake is relatively bigger. However, an increase in the wealth of the investor, even if it may lead to higher investment in risky assets, is very different from a change in the amount at stake. Indeed, an increase in the wealth of the investor may actually lead to a reduction in the stake invested (relatively to investor's wealth).

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

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 three 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 wealth tax. We define as low-wealth investors all the others. 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.

A way of assessing the quality of our identification is to look at the profits made by the different classes of investors. In the last section, we will see that high wealth investors consistently make more profits than the low wealth ones and the liquid investors more than the less liquid ones. 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. We refer the reader to the last section for the detailed analysis.

5.2 Econometric Issues

We now move on to the econometric issues. We assume that the investment decision takes place in two steps: first, the investor decides whether to enter the stock market (stocks, mutual funds),[27] and then he selects which asset to buy. The decision to enter the market can be described as:

$$P_t^* = \alpha_0 + \boldsymbol{\beta}_0 \mathbf{X}_{0t} + \varepsilon_{0,t}, \tag{16}$$

where P_t^* is a latent unobservable variable, and $\mathbf{X}_{0t}$ is a set of variables that explain stock market participation. We cannot observe P_t^* directly, but we can observe a dummy (P_t) that takes the value 1 if the investor participates in the financial market, and zero otherwise. That is,

$$P_t = 1 \text{ if } P_t^* > 0 \text{ and } P_t = 0 \text{ if } P_t^* \leq 0. \tag{17}$$

[27] We will, in general, refer to stocks and mutual funds. However, it is worth noting that we also include among the risky assets warrants, convertible and risky bonds. These, however, represent a very tiny fraction of the actual holdings.

We therefore rewrite equation 16 as:

$$P_t = \alpha_0 + \boldsymbol{\beta}_0 X_{0,t} + \varepsilon_{0,t}, \tag{18}$$

where P_t is the observed probability of market participation (i.e., $P_t = 1$ if $P_t^* > 0$). Equation 18 is the selection equation. The probability that the investor enters the financial market is modeled as a normal c.d.f. In order to estimate this probability, we need to consider a bigger dataset based on the whole sample universe: i.e., both the households that hold financial assets and those that do not. The variables $X_{0,t}$ include the correlation between the different sources of non-financial income and the market portfolio, the volatility of the sources of non financial income and all the aforementioned control variables (i.e., measures of income and wealth, measures of income, demographic variables, professional ability and risk of the investor, the momentum/stock performance variables and the residual control variables).

The second stage deals with portfolio choice. For expositional purposes, let us define C_t as the generic dependent variable (i.e.,h in the case of specifications 1, 2, 9, Δ in the case of specification 6 and Π in the case of specification 10) and $X_{1,t}$ is a generic vector of explanatory variables as defined in the previous restrictions.[28] We therefore have:

$$C_t = \alpha_1 + \boldsymbol{\beta}\mathbf{X}_{1,t} + \varepsilon_{1,t}, \tag{19}$$

The identification restrictions require us to use control variables in the first stage that do not appear in the second stage.[29] We assume the following error correlation structure:

$$\begin{pmatrix} \varepsilon_{0,t} \\ \varepsilon_{1,t} \end{pmatrix} \sim N \begin{pmatrix} 0, & \sigma_0^2 & \sigma_{01} \\ 0, & \sigma_{01} & \sigma_1^2 \end{pmatrix}. \tag{20}$$

We use Heckman's (1979) two-stage procedure. In the first stage, we estimate stock market participation. In the second stage, we include a variable that accounts for the possibility of selection bias at the first stage. This variable is defined as $\lambda_{i,t}$ ("Heckman's lambda") and controls for the problem of omission of variables due to self-selection. We therefore estimate:

$$C_t = \alpha_1 + \boldsymbol{\beta}\mathbf{X}_{1,t} + \theta_1\lambda_t + \varepsilon_{1,t}. \tag{21}$$

The significance of the values of θ_1 provides a test of the null of no sample selection bias. We will see that, in all the specifications, a high degree of significance of θ_1 suggests that self-selection is indeed important in the sample. Specification 21 is estimated by using two-stage least squares with consistent variance-covariance matrix. We perform the analysis at household level.

[28]In the specifications where the dependent variable is the investment in risky assets we include among the explanatory variables the lagged dependent variable to account for possible feedback effects from past values of the dependent variable.

[29]These variables include the set of time and industry dummies as well as the correlations between the market portfolio and investors' sources of non-financial income (i.e., labor income, entrepreneurial income and real estate).

Some of the explanatory variables (i.e., the proxies for hedging $\mathbf{Corr}_{\mathsf{Y},p,t}$ and $\mathbf{\Gamma}_t$) are affected by the investor's choice and are therefore endogenous. Moreover, non-financial income itself may be endogenous (Moskowitz and Vissing-Jørgenson, 2002, Gentry and Hubbard, 2002). To address this issue, we pursue a two-pronged approach. First, we use an instrumental variable methodology (Vissing-Jørgensen, 2002a). We instrument the potentially endogenous variables (measures of hedging and familiarity, non-financial sources of income, lagged dependent variables, etc.) using as instruments a combination of strictly exogenous variables (i.e., demographic variables, industry and time dummies) and the lagged values of the main variables in the different specifications.[30] Alternatively, we modify the estimation of the second stage of Heckman's procedure and perform a robustness check based on a the estimation of a system of simultaneous equations. That is, we re-estimate equation 21 as part of a two-equation system where also the proxies for hedging are jointly determined. The results (not reported) do not differ from those derived from the instrumental variable estimation (reported).

6 Main findings: hedging versus familiarity

We proceed as follows. First, we consider the determinants of portfolio choice and test whether investors select their portfolio so as to hedge non-financial income risk. Then, we provide evidence of familiarity, we analyze it and show its information-based nature. Finally, we analyze the implications of hedging and familiarity-driven investment both in terms of risk taking and in terms of costs and benefits of the two strategies for the investors.

6.1 Portfolio choice: a link to the literature

In order to directly relate to the existing literature, we start by replicating the test of Vissing-Jorgenson (2002a and 2002b) on risk-taking and non-financial income. We focus on restriction 1, estimated conditional on market participation, and estimate:

$$h_t = \alpha_1 + \boldsymbol{\beta}_1 \mathbf{Var}_{\mathsf{Y},\mathsf{t}} + \boldsymbol{\gamma}_1 \mathbf{Corr}_{\mathsf{Y},\mathsf{m},t} + \boldsymbol{\delta}_1 \mathbf{F}_{1,t} + \theta_1 \lambda_t + \mu_1 h_{t-1} + \varepsilon_{1,t}. \tag{22}$$

As in the traditional literature, we consider a specification based on the percentage value of the investment in risky assets (stocks and mutual funds) over overall wealth (Risky Share) and a specification based on the dollar value of the investment (Risky Value). The results are reported in Table 2, Panel A. The findings on hedging are similar with those of Vissing-Jørgensen (2002a). We find that $\boldsymbol{\beta}_1$ is negative, but not always significant, while $\boldsymbol{\gamma}_1$ is mostly

[30] The endogeneity issue further complicates the task of finding proper instrumental variables, as only strictly exogenous variables or predetermined ones can be used. As Arellano (1989) and Kiviet (1995) showed, lagged values represent predetermined variables, uncorrelated with the residuals, whereas the demographic variables are strictly exogenous in the Granger-Sims sense. The Adjusted RSquares of the first stage regressions range between 25% and 68%.

negative and significant. These results appear robust across specifications and for different levels of wealth. They provide weak evidence in favor of hedging. Let us now move to portfolio data and consider restriction 2. We estimate:

$$h_t = \alpha_2 + \boldsymbol{\beta}_2 \mathbf{Var}_{\mathsf{Y,t}} * \mathbf{sign}(\mathbf{Corr}_{\mathsf{Y,p,t}}) + \boldsymbol{\gamma}_2 \mathbf{Corr}_{\mathsf{Y,p},t} + \boldsymbol{\delta}_2 \mathbf{F}_{2,t} + \theta_2 \lambda_t + \mu_2 h_{t-1} + \varepsilon_{2,t}. \quad (23)$$

We recall that hedging requires that $\boldsymbol{\beta}_2 < \mathbf{0}$ and $\boldsymbol{\gamma}_2 < \mathbf{0}$ if $\mathbf{Corr}_{y,p} > \mathbf{0}$ and $\boldsymbol{\gamma}_2 > \mathbf{0}$ if $\mathbf{Corr}_{y,p} < \mathbf{0}$. As before, we consider a specification based on the percentage value of the investment in risky assets (stocks and mutual funds) over overall wealth (Risky Share) and a specification based on the dollar value of the investment (Risky Value). The results are reported in Table 2, Panel B. Notice that now we use the investor portfolio as opposed to the market portfolio and that, as required by the restriction, we separately consider the case where the correlation between financial and non-financial income is positive from the case where it is negative.

As expected, now the results change. They strongly reject the hypothesis that investors buy to hedge. Indeed, $\boldsymbol{\gamma}_2$ is always positive in the case of a positive correlation, and negative or not significant in the case of a negative correlation. This holds for both labor income and entrepreneurial income and also for both the specification based on the percentage investment and the one based on the dollar value of the holdings. Another way of reading these results is that the investors who invest more in risky assets are those who have chosen a portfolio composition more positively related to their non-financial income. This also provides a first evidence of familiarity.

The results are robust across investors, regardless of their wealth level. The only noticeable difference is that the relationship between the correlation of non-financial income and financial income and the investment in risky assets is stronger for low-wealth investors. This may suggest that low-wealth investors are more affected by familiarity than the high-wealth investors are. In the following section, we will see that this is indeed the case.

It is also worth noting that there is an overall negative correlation between the variance of non-financial income and the investment in risky assets. This holds for both labor and entrepreneurial income and is robust across specifications (i.e., both in the case of absolute holdings and in the case of share in the portfolio). However, if we split the sample according to the level of wealth, there is a strong heterogeneity. Indeed, the variance of non-financial income increases risk-taking for the high-wealth investors and decreases it for the low-wealth ones. This may suggest that investing in risky assets is perceived as a risky strategy for the low-wealth investors. Low-wealth investors perceive that investment in the stock market increases their overall risk and therefore invest less, the higher their non-financial risk is. In contrast, the high-wealth investors, with access to professional advice and better information, may perceive the investment in risky assets as a way of diversifying their non-financial income risk.

6.2 Portfolio choice: the risk profile of the financial portfolio

6.2.1 Hypothesis H.1: The tilt in the risk profile.

How do we reconcile the previous results with the literature? The inconsistency lies in the fact that the standard tests do not properly account for the fact that investors tilt their portfolio away from the market portfolio. This affects all the standard tests based on the market portfolio. We now analyze this tilt more in detail. We start by studying the correlation (covariance) between the portfolio of the investor and his non-financial income. The test of hypothesis H.1. can be seen as a preliminary evidence of investor behavior.

An overall view is provided by Figures 1 and 2. They display the different correlations between investors' real estate, financial portfolio and non-financial income (i.e., labor and entrepreneurial income). They are constructed in the following way. First, we calculate for each investor the correlation of the measure of income, as defined in Section 4.1, with real estate and the returns of the stocks held in the portfolio. Then, we average these correlations across investors. In a similar fashion, we also construct a measure of correlation between investors' non-financial income and the return the value-weighted Swedish stock market index (SIX Index). There are some interesting points worth noting.

The first thing to note is that there is a high fraction of investors who display a negative correlation between financial and non-financial income. This holds in the case of both the correlation between non-financial income and the market portfolio and of the correlation between non-financial income and the investor's portfolio. This finding is also robust for both entrepreneurial income and labor income.

The second point to note is the strong bi-modal shape of the relationship between non-financial income and portfolio return. This suggests a heterogeneity across investors in their reaction to fluctuations in entrepreneurial income, possibly due to wealth effects. We will investigate this later on.

The most interesting point is that the distribution based on the correlation between non-financial income and the investor portfolio is more skewed to the right. That is, investor portfolios are more positively related to investor income than the market portfolio is. This may be due either to deliberate behavior or to spurious correlation.[31]

To more properly quantify these claims, we consider some descriptive statistics. In Table 3, we report descriptive statistics on the average value of the correlations between non-financial income and either investor portfolios or aggregate stock market and tests of the difference between them. We also report our indexes of hedging (i.e., Γ and Δ).

[31] For instance, Sweden underwent a massive restructuring in some sectors in the 1990s, at the very time the stock market was booming. An investor working in one of these sectors affected by such a restructuring would display a negative correlation induced by exogenous reasons.

The comparison of the correlation with the portfolio and the correlation with the market confirms the tilt away from hedging. Indeed, while the correlation of labor income with the market is negative (-0.01 for the low-wealth investors and -0.04 for the high-wealth investors), the correlation of the labor income of the investor with his portfolio is positive (0.03 for the low-wealth investors and 0.05 for the high-wealth investors). In other words, investors construct their portfolios so as to actually turn their natural negative correlation with the market into a positive one.

The behavior is analogous for the case of entrepreneurial risk. Investors reduce the negative correlation that would otherwise exist between their non-financial income and the market. That is, the correlation of entrepreneurial income with the investor portfolio is less negative than the one with the market. In particular, it drops from -0.07 to -0.01 for the low-wealth investors and from -0.07 to -0.06 for the high-wealth investors. Both in the case of labor income and in the case of entrepreneurial income the differences between statistics based on the market portfolio and statistics based on the investor portfolio are reflected in the indexes of hedging (Γ and Δ).[32]

These results point in the direction opposite to hedging and show in an intuitive way how investors move away from the market portfolio. In the next section we will use a more formal test based on a model of portfolio choice.

6.2.2 Hypothesis H.2: Determinants of the tilt in the risk profile.

We now focus on restriction H.2., estimated conditional on market participation. The general equation that describes the decision to invest in risky financial assets is:

$$\Delta_{z,t} = \alpha_3 + \beta_3 \frac{Y_{z,t}}{W} Var_{Y_z,t} + \gamma_3 \frac{Y_{x,t}}{W} Cov_{Y_z,Y_x,t} + \zeta_3 \frac{1}{W} \Theta_{z,t} + \boldsymbol{\delta}_3 \mathbf{F}_{3,t} + \boldsymbol{\nu}_3 \boldsymbol{\Psi}_t + \theta_3 \lambda_t + \varepsilon_{3,t}. \quad (24)$$

We have different options for the definition of non-financial income. We can aggregate the different sources of non-financial income of the investor and assume that he just hedges their aggregated value. Alternatively, we can analyze them separately. Finally, we can consider investors who have both sources of income separately from those who have just labor income.[33] We proceed as follows. We start by focusing on the separate sources of income, as restriction 6 would require and test hedging of labor income and hedging of entrepreneurial income separately. Then, as a robustness check, we also consider the case in

[32] Another point that is worth noting is the low average correlation between real estate and non-financial income. The wide dispersion of behavior suggests that some investors hedge their non-financial income with real estate, while others react to an increase in real estate wealth by raising their investment in risky financial assets. The latter effect is consistent with analogous results in the US (Case, Quigley and Shiller, 2001). Finally, it is worth noting that entrepreneurial income seems to be negatively related to real estate. Evidence on the correlation between real estate returns and entrepreneurial income has not yet been properly documented in the literature.

[33] We do not consider the case in which investors have only entrepreneurial income as the sample is very small.

which the sources of income are aggregated. Finally, we turn to the investors who have only labor income. Given that all the results are consistent, we will not report the last ones. All the specifications include the five sets of control variables: measures of income and wealth, demographic variables, professional ability and risk, momentum/stock performance variables and residual control variables.

We consider two main specifications. The first (Specification I) does not include the measures of familiarity ($\mathbf{\Psi}_t$). Hedging requires that $\beta_3 > 0$, $\gamma_3 > 0$ and $\zeta_3 > 0$. The second specification (Specification II) includes also the measures of familiarity. Familiarity requires that $\boldsymbol{\nu}_3 > 0$. The results are displayed in Table 4, Panel A for the case of labor income risk, Panel B for the case of entrepreneurial income risk and Panel C for the case of aggregated total income risk.

The results reject the hypothesis of hedging. For the overall sample, all three coefficients, β_3, γ_3 and ζ_3, are always negative. If we consider labor income, we see that β_3 is always negative and significant for both high- and low-wealth investors. This is a particularly strong rejection of hedging, as β_3 represents the very coefficient that captures the direct impact of the riskiness of the *zth* source of income on the portfolio tilt to hedge it. In the case of entrepreneurial income, β_3 is still negative and significant for the low-wealth investors and becomes positive and significant for the high-wealth investors. Also the coefficients γ_3 and ζ_3 do not support the hypothesis of hedging. In particular, the coefficient ζ_3, is always negative for both types of non-financial income and across different specifications, while γ_3 is either negative or not significant. These results are robust across specifications and for different definitions of non-financial income - i.e., aggregate, labor and entrepreneurial income considered separately. They also hold for the case in which we focus on investors who have just labor income.

These results suggest that we can reject the hypothesis of hedging, overall, for different classes of investors and for both labor and entrepreneurial income. Investors characterized by a negative correlation between their labor income and the market tend to increase their loadings on risky assets that are more closely correlated to their income than the market as a whole. In particular, in the case of the low-wealth investors, it seems that they systematically act in a fashion opposite to that required by hedging. This is more consistent with the alternative hypothesis of familiarity.

Let us now consider familiarity. The results show that familiarity affects investors' decision to hedge labor income as well as entrepreneurial income. In particular, there is a strong negative correlation between active hedging and two measures of familiarity: geographical proximity and the holding period-based. This holds regardless of the level of wealth of the investors [34] and across specifications. These findings confirm the intuition that investors

[34]The only exception is the case of low-wealth investors who are not affected by geographical proximity in

deliberately behave in a way opposite to that required by hedging because they want to invest in stocks with which they are familiar. This induces a more positive correlation with their non-financial income.

Professional proximity is not related to hedging labor income risk. This finding suggests that the type of familiarity that affects investors and induces them to deviate from hedging is not related to the company/sector they work for. In other words, the decision not to hedge is a deliberate one, not due to employment constraints such as belonging to a stock plan that invests in the company's stocks.[35] Professional proximity is, rather, negatively related to the hedging of entrepreneurial income for the case of the low-wealth investors. This may be due to the existence of borrowing constraints and limited capital that induce the low-wealth investors to focus on their main business: i.e., entrepreneurial activity.[36]

It is interesting to note that θ_3 is always significant and positive in the case of labor income and significant and negative in the case of entrepreneurial income. This suggests that the selection bias due to endogenous stock market participation is significant. In particular, the omission of λ would mean that the least squares methodology overestimates the value of the relationship between the hedging index and the non-financial income variables in the case of labor income hedging and underestimates it in the case of entrepreneurial income hedging. In other words, the people who participate are more likely to hedge labor income and less likely to hedge entrepreneurial income. This bias is properly accounted for by including λ_t.

Finally, two points are worth stressing. First, our results are robust to the way we treat the mutual funds. To guarantee this, we re-estimated our specifications excluding the mutual funds from the set of risky assets and the results are consistent with those that are reported.

Second, our results are not affected by the share of the labor force employed with the government. Indeed, a significant fraction of Swedish households is employed by the government (either national or municipal). This suggests that their incomes are much less variable, and the risk of unemployment different. Moreover, their income should be significantly less correlated with the return on the stock market. Therefore, as an additional robustness check, we also performed the analysis on two separate subsamples: government employees; and non-government employees.[37] The results are reported in Table 5, Panels A for the case of the

the case of entrepreneurial income.

[35]It is important to note that, in Sweden, investors are not forced or induced to own company stocks. This is not only confirmed by casual evidence collected by the authors, but also by the data. Indeed, the fact that professional proximity is mostly not significant indicates that investors do not tilt their portfolio toward stocks of the company they work for or of companies belonging to the same industry.

[36]Also, age increases hedging for the young people and decreases it for the old ones. For the case of the low-wealth investors, the decision to hedge is also a function of the funds available. More funds available seem to increase hedging labor income for the low-wealth investors. In particular, an increase in wealth (both financial and real estate wealth) and capital gains induce the low-wealth investors to hedge more labor income, while the high-wealth investors do not seem to be affected by the level of financial wealth or capital gains and reduce their hedging of both labor income as their real-estate wealth increases.

[37]We define as public sector households the ones whose share of income coming from occupations with SNI

private-sector and Panel B for the case of the public sector. The findings are consistent with the previous results. In particular, investors employed in the private sector do not hedge and the decision to invest in assets with payoffs closely related to their income is strongly affected by familiarity. Public employee, instead, seem to be less affected by familiarity. Indeed, both low- and high-wealth investors do not seem to be affected by any proximity variable.

6.3 H3: The nature of familiarity

We now study the nature of familiarity by investigating in more detail the heterogeneity across investors. For brevity, we directly focus on the relationship 24 by including interactive dummies to account for cross-sectional differences across investors.[38] Recall that if familiarity is information-driven, we expect it to differ across differentially informed investors and to change after familiarity shocks. As explained in Section 5.1, we use the wealth of the investor and the degree of liquidity of his portfolio to identify the informed investors. We estimate:

$$\Delta_{z,t} = \alpha_3+\beta_3\frac{Y_{z,t}}{W}Var_{Y_z,t}+\gamma_3\frac{Y_{x,t}}{W}Cov_{Y_z,Y_x,t}+\zeta_3\frac{1}{W}\Theta_{z,t}+\boldsymbol{\nu}_3\boldsymbol{\Psi}(1-\xi_1)+\boldsymbol{\pi}_3\boldsymbol{\Psi}\xi+\boldsymbol{\delta}_3\mathbf{F}_{3,t}+\theta_3\lambda_t+\varepsilon_{3,t}. \tag{25}$$

In the case of differences in wealth or liquidity, ξ is a continuous variable based on the amount of wealth or liquidity of the investor and $\xi_1 = 0$. In the case of familiarity shocks, instead, ξ is a dummy that takes the value 1 if in the previous three years the investor has been subject to a "familiarity shock" and zero otherwise and $\xi_1 = \xi$.

We define as familiarity shocks events that change the investor's proximity to the stocks. Unemployment shocks and professional change shocks are natural candidates as they represent a structural break in the professional life of the investor. In the case of unemployment shocks, ξ is a dummy that takes the value 1 if the investor has been unemployed at least once in the previous three years and zero otherwise. In the case of professional change shocks, ξ is a dummy that takes the value 1 if the investor has changed his profession and the current profession differs from the previous one and zero otherwise.

Given that our familiarity variables are related to the geographical location of the investor, another potential candidate is the change in the location of the investor. This proxies for the "relocation shocks". In this case, $\boldsymbol{\xi}$ is a dummy that takes the value 1 if the investor has moved and changed county (municipality) at least once in the previous three years and the current address differs from the previous one and zero otherwise.

We first focus on how the impact of familiarity differs across differentially informed investors. We use the level of wealth and the liquidity of the portfolio as a proxy for the degree

codes between 75000 and 91999 is in excess of 50%. This provides approximately 192,000 investors belonging to the private sector and 58,000 belonging to the public sector. We focus only on labor income as entrepreneurial income is very rare among public sector employees.

[38]The results for the estimation of the differential behavior in the case of specification 26 are not reported but available upon request from the authors.

of informativeness. The results are reported in Table 6, Panels A and B. In this case $\boldsymbol{\nu}_3$ represents the sensitivity to familiarity of investors in general and $\boldsymbol{\pi}_3$ proxies for the differential effect of being more informed (i.e., wealthier or more liquid). The first thing to note is that the impact of familiarity on the decision not to hedge is decreasing in wealth. This holds for the case of geographic, professional and holding-period based familiarity. There is evidence of a statistical difference between the two classes of investors. This suggests that wealthier investors rely less on the information based on professional proximity because they can afford better quality information.

If, then, we break the sample of the high-wealth investors down into liquid and illiquid ones, we find another strong heterogeneity. The illiquid investors are more strongly affected by familiarity than the liquid investors. This holds for both geographical and holding-period based proximity and for the different sources of income - labor, entrepreneurial and total income. Professional proximity is insignificant both for the level and for the interactive dummy. As in the case of wealth, there is a statistical difference between the two classes of investors. These results are consistent with our identification of high-wealth and liquid investors as informed ones and illiquid investors as uninformed.

Let us now consider the familiarity shocks. In this case $\boldsymbol{\nu}_3$ represents the sensitivity to familiarity of investors who have not been affected by shocks, while $\boldsymbol{\pi}_3$ represents the sensitivity to familiarity of investors who have been affected by shocks. As we mentioned before, we focus on three types of shocks: unemployment shocks, changes of profession shocks and relocation shocks. Given that these shocks are likely to be correlated - i.e., a change in profession may induce the investor to move to another town - we expect each shock to affect all the three sources of familiarity and not just the one to which the shock appears to be more related (i.e., the relocation shock should be mostly related to geographical based familiarity). The results are reported in Table 6, Panel C, for the case of shocks due to unemployment, Panel D, for the case of shocks due to professional change and Panel E, for the case of shocks due to relocation.

The results show that familiarity shocks in general change investor's sensitivity to familiarity. (Unreported) tests show that the values of the coefficients $\boldsymbol{\nu}_3$ and $\boldsymbol{\pi}_3$ are statistical different from one another. In particular, an investor who has changed profession or has been unemployed recently is less subject to a professional bias. This is expected as such an investor has not yet had the time to absorb the information embedded with his new profession. Analogously, an investor who has changed his address and has moved is less subject to geographical proximity than an investor who has not moved. This is consistent with the intuition that relocation affects the availability of geographically-based information. After a move it takes time for the investor to accumulate the same information on closer stocks that he used to have before the move. In many cases the shocks not only reduce the impact of familiarity, but also seem to increase hedging. This is consistent with the fact that shocks increase the informational uncertainty of the investor, increasing his desire to hedge.

These findings support the idea that the impact of familiarity depends on the degree of informativeness of the investor - i.e., more informed investors are less affected by familiarity - and on its stability over time - i.e., familiarity shocks modify the impact of familiarity on investor behavior. This suggests that the investment choice is driven by the availability of information, and that familiarity is a substitute for better information. Its importance decreases when the investor has access to more information or if there is a perturbation to his source of familiarity.

6.4 Some effects of the two strategies

We now consider how these two strategies (hedging and familiarity) affect the investor decision to take risk and his profitability. We start by focusing on risk taking.

6.4.1 Risk taking

We test how risk taking is affected by the hedging/familiarity strategy and whether the decision to invest in "familiar" stocks can explain the investment in risky assets and the tendency to invest in stocks positively correlated to non-financial income. We start by considering restriction 9 and estimate:

$$h_t = \alpha_4 + \boldsymbol{\beta}_4 \boldsymbol{\Gamma}_t + \gamma_4 \boldsymbol{\Psi} + \boldsymbol{\delta}_4 \mathbf{F}_{4,t} + \theta_4 \lambda_t + \mu_4 h_{t-1} + \varepsilon_{4,t}. \tag{26}$$

If risk taking is determined by hedging, we expect $\boldsymbol{\beta}_4 > 0$ and $\gamma_4 = 0$. In the case of familiarity, we expect $\boldsymbol{\beta}_4 < 0$ and $\gamma_4 > 0$. As in the previous tests, we consider both a specification based on the percentage investment (Risky Share) and one based on the dollar value of the investment (Risky Value).[39]

The results are reported in Table 7. They show that $\boldsymbol{\beta}_4$ is always significantly negative, while γ_4 is most of the time significantly positive, with the single exception of professional proximity for the low-wealth investors. These results are consistent with the previous findings. Risk taking is affected by familiarity. For the overall sample, investors do not hedge but tend to buy stocks that are either geographically or professionally close to them, or that they have held for a long period. This also holds across the different specifications. There is, however, a strong heterogeneity across investors, as the impact of familiarity is stronger for the low-wealth investors, while it is weaker and disappears for the high-wealth investors. The coefficients of our indexes of familiarity decrease with the level of wealth for both geographical proximity (from 0.11 to 0.02 for the percentage investment and from 0.79 to 0.61 for the dollar

[39] It is worth remembering that we use instrumental variables to control for the the endgeneity of some of the explanatory variables and in this case, in particular, of $\boldsymbol{\Gamma}$.

value of the investment) and holding period proximity (from 0.1 to 0.0001 for the percentage investment and from 0.75 to 0.20 for the dollar value of the investment). Conversely, it seems that professional proximity affects more the high-wealth investors.

6.4.2 Cost of hedging/familiarity

If familiarity is a proxy for information, can we quantify its cost/benefits if compared to a hedging strategy? We consider two measures of "profits": the financial gains/losses in the year standardized by the value of the risky assets at the beginning of the year and the change in wealth standardized by the value of wealth at the beginning of the year. Financial gains/losses include the realized capital gains/losses and the dividends.

Both measures may be subject to criticism and must be taken with a pinch of salt. Financial gains/losses represent an imprecise measure if investors do not turn over their positions regularly. A change in wealth does not account properly for the saving decisions of the investors. Given that both biases may be related to the income of the investor, we include both the level and the variance of investors' income and wealth among the control variables.

The fact that gains and losses are directly reported by the financial intermediary through which the transaction is executed reduces the potential bias due to the underreporting of capital gains for tax purposes. Other biases may be the result of the "lock-in effect" that generates different incentives to sell the stocks for investors with different marginal tax rates, and to clientele effects. However, for all these biases (underreporting, lock-in and clientele) Yitzhaki (1987) shows that "the likely effects of these biases is to cause an underestimate of the observed differences in rates of return among income classes." Therefore, the bias acts *against* the possibility of actually finding a statistically significant difference between classes.

Descriptive statistics of the profits and their differences between groups of investors and tests of the differences are reported in Table 8, Panels A and B. Bootstrapping has been employed to assess the robustness of the tests. We consider both profits based on the changes of wealth (Π_w) and profits based on financial gains/losses (Π_F). As expected, the high-wealth investors make more profits than the low-wealth ones. This holds for the statistics based on the mean as well as those based on the median. In Table 8, Panel C, we compare the profits of the investors who hedge (i.e., $\mathbf{\Gamma} > \mathbf{0}$) and those who do not hedge (i.e., $\mathbf{\Gamma} < \mathbf{0}$). Hedgers, in general, earn lower profits than non-hedgers. This suggests that hedging is indeed expensive, as we would expect it to be in equilibrium. Alternatively, this also means that familiarity provides the low-wealth investors with a cheap source of information.

Note that this comparison is *separately* done for both the high-wealth and the low-wealth investors, and therefore it is not inconsistent with the fact that the high-wealth investors make more profits than do the low-wealth ones. It is also interesting to note that the cost of hedging (i.e., the difference in profits between the hedgers and the non-hedgers) is higher for the low-wealth investors than for the high-wealth ones. This fits with the intuition that high-wealth investors have access to better financial services and advice.

In order to assess the relationship between profits and hedging/familiarity strategies, we estimate:

$$\Pi_t = \alpha_5 + \beta_5 \boldsymbol{\Gamma} + \boldsymbol{\gamma}_5 \boldsymbol{\Psi}_t + \boldsymbol{\delta}_5 \mathbf{F}_t + \theta_5 \lambda_t + \varepsilon_{5,t}, \tag{27}$$

where Π_t are the profits realized at t by the investor. The results are reported in Table 9, Panel A, for the case of profits based on changes of wealth (Π_w) and Panel B, for the case of profits based on financial gains/losses (Π_F). Notice that the fact that we control for the level of wealth (both in the definition of the sample and by including it among the control variables) allows us to condition on the level of private information of the investor. We expect that $\beta_5 < 0$ and $\boldsymbol{\gamma}_5 > 0$.

The results show that there is a negative or non-significant correlation between hedging and profits (β_5) and a positive correlation between profits and geographical proximity and holding period proximity ($\boldsymbol{\gamma}_5$). This holds, regardless of the level of wealth of the investors, across specifications and for different classes of investors. However, the impact of familiarity is stronger for low-wealth investors. Professional proximity, on the other hand, does not seem to be related to profits. It is interesting to note that the sources of familiarity that are mostly related to the low-wealth investors (i.e., geographical proximity and holding period) are those which are related to profits. Conversely, professional proximity, which is more related to the behavior of the high-wealth investors, is not related to profits.

These findings are consistent with the intuition that familiarity is, for the low-wealth investors, a proxy for cheap information. This allows the low-wealth investors to have higher profits. This is not the case for the high-wealth investors who, presumably, have access to better private information. Hedging, instead, appears to be expensive.

6.5 Main implications

These results suggest that investors, in general, do not hedge and for those who do, hedging is costly. What is the relevance of this phenomenon? A random investor does not hedge, but does the wealth-weighted investor hedge? That is, how important do we think the small investors are? In our sample, the high-wealth investors represent a mere 8% of the sample, but on a value basis they represent 73.42% of the assets. Moreover, the low-wealth and high-wealth investors are almost evenly split between high (greater than 0.2) and low (less than 0.2) correlation with the market.

To address this issue, we focus on labor income. In Figure 3, we report the distribution of the shareholdings of the different classes of investors broken down according to investors' measure of hedging (Γ_1). Each histogram represents the value of the stock market held by the investors with a specific tendency to hedge/go for familiar stocks. Investors are grouped into the ones who invest in "close" and "distant" stocks (right and left columns, correspondingly). The distributions are reported separately over sub-samples of low-wealth and high-wealth investors (Figures 3a and 3b, correspondingly) and for the overall sample (Figure 3c). We use five ranges for Γ_1: from -2 to -1.2, from -1.2 to -0.4, from –0.4 to 0.4, from 0.4 to 1.2 and from 1.2 to 2.

The first thing to note is that, among the high-wealth investors, the distribution is slightly skewed toward investors who hedge. In contrast, for the low-wealth investors, the distribution is strongly skewed toward familiarity-driven investors. How does this aggregate? In aggregate, the fact that the high-wealth investors own more than 73% of the assets offsets the bent toward familiarity of the low-wealth investors. This implies that, even if the low-wealth investors represents 92% of the population, their impact at the aggregate level is very tiny. The wealth-weighed investor is less subject to familiarity and behaves more in line with what theory would dictate. This has implications in terms of the standard theory that assumes that prices are set by "smart rational" investors and also validates the standard theoretical models that generate a scarce impact coming from individual heterogeneity (Heaton and Lucas, 1997).

However, this does not provide evidence at the individual stock level. If familiarity hardly aggregates into a factor that affects prices, it may still explain cross-sectional heterogeneity among stocks and phenomena such as fads, bubbles, momentum and value premia in all those instances where individual stock values diverge from their fundamentals. A more detailed analysis of this point is outside the scope of the present paper. It is interesting to note that the scarce or null impact on prices may also explain why familiarity-driven gains may accrue and sustain in equilibrium.

7 Conclusion

We studied the question of whether investors use their investment in financial assets to hedge their non-financial income. We used a new approach based on the inspection of the relationship between investor non-financial and financial income and a portfolio-based index of hedging.

We provided evidence that investors do not engage in hedging, but seem to deliberately tilt their portfolio toward stocks that are most closely related to them. We rationalized this in terms of "familiarity" and we provided evidence of it. We showed a high degree of heterogeneity across investors and we identified differences in terms of wealth and liquidity of their overall portfolio. We then investigated the nature of familiarity and showed that it is information-based. Familiarity-driven behavior is a way of conditioning on a cheap source of information for the investors.

We used a unique and new dataset where all the sources of income, as well as the asset holdings, are broken down at the investor level.

These results challenge the standard portfolio theory and, at the same time, generate evidence to extend it, by shedding new insights on investor behavior. The identification of the determinant of familiarity has practical relevance, given that the information-based familiarity hypothesis and the pure familiarity hypothesis 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, on the other hand, 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 increases. Therefore, processes such as globalization and financial integration, by increasing information, should reduce the impact of familiarity on investors' choices, and therefore on asset prices.

8 Appendix

In this Appendix we develop the testable restrictions on the tilt in the portfolio. We rely on the standard literature on portfolio choice with multiple assets and, in particular, Tepla (2001). Let us consider an economy with n risky securities denoted by S and a riskless asset B. The riskless asset earns and instantaneous interest rate $r > 0$, while the risky securities follow a geometric Brownian motion, such that:

$$dB = rBdt \tag{28}$$

$$\mathbf{dS} = \mathbf{I}_s\boldsymbol{\mu}dt + \mathbf{I}_s\boldsymbol{\Sigma}\mathbf{dw}, \tag{29}$$

where, $\mathbf{w}$ is a n-dimensional standard Brownian motion, $\mathbf{I}_s$ is an n-dimensional diagonal matrix with the risky securities prices as entries, $\boldsymbol{\mu}$ is an n-dimensional vector of mean returns and $\boldsymbol{\Sigma}$ is the matrix of diffusion coefficients. We assume $\boldsymbol{\Sigma}$ to be diagonal, that is markets are complete and each asset loads only on a specific source of uncertainty. The investor has other non-financial sources of income ($\mathbf{Y}$):

$$\mathbf{dY} = \mathbf{I}_y\mathbf{a}dt + \mathbf{I}_y\mathbf{s}\mathbf{dw}, \tag{30}$$

where $\mathbf{I}_y$ is an n-dimensional diagonal matrix with the value of the income source as entries, $\mathbf{a}$ is an n-dimensional drift vector and $\mathbf{s}$ is the matrix of diffusion coefficients. We assume that to rule out arbitrage opportunities, $(\mathbf{a} - r) = \mathbf{s}\boldsymbol{\Sigma}'(\boldsymbol{\Sigma}\boldsymbol{\Sigma}')^{-1}(\boldsymbol{\mu} - r)$. The representative investor maximizes utility of terminal wealth ($W(T)$),[40] where the wealth follows:

$$dW = [\boldsymbol{\theta}'(\boldsymbol{\mu} - r) + Wr]dt + \boldsymbol{\theta}\boldsymbol{\Sigma}\mathbf{dw} \tag{31}$$

The representative investor is endowed with a HARA utility function

$$U = \frac{1-\gamma}{\gamma}(\frac{y}{1-\gamma})^{\gamma} \tag{32}$$

of terminal $y = W + \mathbf{Y}'\mathbf{e}$, where $\mathbf{e}$ is a vector of ones. Let us define $cov_{\mathsf{Y},p}$ as the covariance between the return on the financial portfolio of the investor and the rate of change of his non-financial income and $cov_{\mathsf{Y},m}$ as the covariance between the return on the market portfolio and the rate of change of the investor's non-financial income. The measure of deviation from the market is:

$$\Delta = cov_{\mathsf{Y},m} - cov_{\mathsf{Y},p}. \tag{33}$$

The covariances between the non-financial risk and, respectively, the market portfolio and the investor's actual portfolio are:

$$cov_{\mathsf{Y},m} = \boldsymbol{\theta}_m \boldsymbol{\Sigma}\mathbf{s}' \text{ and } cov_{\mathsf{Y},p} = \boldsymbol{\theta}_p \boldsymbol{\Sigma}\mathbf{s}', \tag{34}$$

where $\boldsymbol{\theta}'_m$ and $\boldsymbol{\theta}'_p$ are, respectively, the vectors of the proportion of wealth invested in risky assets for the market portfolio in the absence of non-financial risk and for the investor's portfolio in the presence of non-financial risk. They are:

$$\boldsymbol{\theta}'_p = \frac{1}{W}\boldsymbol{\Sigma}^{-2}\left\{\frac{W + \mathbf{Y}'\mathbf{e}}{1-\gamma}(\boldsymbol{\mu} - r) - \boldsymbol{\Sigma}\mathbf{s}\mathbf{Y}\right\} \text{ and } \boldsymbol{\theta}'_m = \frac{1}{W}\boldsymbol{\Sigma}^{-2}\left\{\frac{W}{1-\gamma}(\boldsymbol{\mu} - r)\right\} \tag{35}$$

Therefore, equation 33 can be rewritten as:

$$\boldsymbol{\Delta}{*}W = \mathbf{Y}'\left[\mathbf{s}\mathbf{s}' - \mathbf{e}\frac{(\boldsymbol{\mu} - r)'}{(1-\gamma)}\boldsymbol{\Sigma}^{-2}\boldsymbol{\Sigma}\mathbf{s}'\right]. \tag{36}$$

Let us assume that there are two non-financial sources of income: x and z. The deviation from the market to hedge income y is:

$$\Delta_z * W = Y_z Var_{Y_z} + Y_x Cov_{Y_z,Y_x} - (Y_z + Y_x)\sum_{j=1}^{n}\Omega_{S_j} * Cov_{S_j,Y_z}, \tag{37}$$

where $\Omega_{S_j} = \frac{(\boldsymbol{\mu}_{s_j} - r)'}{(1-\gamma)\boldsymbol{\sigma}^2_{s_j}}$ and μ_{s_j} and $\sigma^2_{s_j}$ are, respectively, the mean and the variance of the *jth* risky asset.

[40] Alternatively, we can consider the case of the labor income flow. In this case, the investor does not maximize the utility of terminal wealth, but $MaxE[\int_0^T U(C(s))ds]$,s.t.$E[\int_0^T \xi_s C(s)ds] \leq$ $\xi_0 W_0 + E[\int_0^T \xi_s Y(s)ds]$, where ξ_s is the state price density. The result is analogous to the one we report in equation 35, but with labor income defined from time t to T. That is, the residual working life of the invetsor. $\boldsymbol{\theta}'_p = \frac{1}{W}\boldsymbol{\Sigma}^{-2}\left\{\frac{W + \mathsf{Y}(\mathsf{T}-\mathsf{t})^{\theta}\mathsf{e}}{1-\gamma}(\boldsymbol{\mu} - r) \;- \boldsymbol{\Sigma}\mathbf{s}\mathbf{Y}(\mathbf{T} - \mathbf{t})\right\}$. This is effectively what we construct when we use the Carrol and Samwick and Vissing-Jorgenson methodology to define permanent income.

References

[1] Arellano, M., 1989, Testing for Autocorrelation in Dynamic Random Effects Models. Review of Economic Studies, 57 127-134.

[2] 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.

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

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

[5] Bertaut C. and Haliassos M., 1997, Precautionary Portfolio Behavior from a Life-cycle Perspective, Journal of Economic Dynamics and Control , 21, pp. 1511-42.

[6] Bodie, Z., Merton, R.C. and Samuleson, W.C., 1992, Labor Supply Flexibility and Portfolio Choice in a Life Cycle Model, Journal of Economi Dynamics and Control, 16, 427-449.

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

[8] Calvet, L., Gonzalez-Eiras and Sodini, P., 2002, Financial Innovation, Stock Market Participation and Asset prices, SSE/WFI Working Paper 464.

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

[10] Carroll, C.D. and Samwick, A.A., 1996, The Nature of Precautionary Wealth, Journal of Monetary Economics, 40, 41-71.

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

[12] Case, K.E., Quigley, J.M. and Shiller, R., 2001, Comparing Wealth Effects: the Stock Market versus the Housing Market, UC Berkeley, Working Papers.

[13] Cocco, J. F., 2001, Hedging House Price Risk With Incomplete Markets Mimeo.

[14] Campbell, J.Y., Cocco, J.F., Gomes, F.J., Maenhout, P.J., and Viceira, 1997, Stock Market Mean Reversion and the Optimal Equity Allocation of a Long-Lived Investor, Mimeo.

[15] Campbell, J.Y., Cocco, J.F., Gomes, F.J., Maenhout, P.J., and Viceira, 1998, Investing Retirement Wealth: A Life-Cycle Model, Mimeo.

[16] Cocco, J.F., Gomes, F.J. and Maenhout, P.J., 1999, Consumption and Portfolio Choice over the Life-Cycle, Mimeo.

[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] Cuoco, D., 1997, Optimal Consumption and Equilibrium Prices with Portfolio Constraints and Stochastic Income, Journal of Economic Theory, 72, 33-73.

[20] Dammon, R., Spatt, C. and H. Zhang, 2001, Optimal Consumption and Investment with Capital Gains Taxes, Review of Financial Studies, 14, 583-616.

[21] Davis, S. and Willen, P., 2000, Occupation-Level Income Shocks and Asset Returns Mimeo.

[22] Davis, S. and Willen, P., 2000, Using Financial Assets to Hedge Labor Income Risk: Estimating the Benefits, Mimeo.

[23] DeMarzo, P., Kaniel, R. and Kremers, I., 2002, Diversification as a Public Good: Community Effects in Portfolio Choice, Mimeo.

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

[25] Elmendorf, D.W. and Kimball, M.S., 2000, Taxation of Labour Income and the Demand for Risky Assets, International Economic Review, 3, 801-834.

[26] 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.

[27] 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.

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

[29] Gentry, W.M., and R.G. Hubbard, 2002, Entrepreneurship and Household Savings, Mimeo.

[30] Gollier, C. and Pratt, J.W., 1996, Risk Vulnerability and the Tempering Effect of Background Risk, Econometrica, 64, 1109-1123.

[31] Gomes, F. and Michaelides, A., 2002, Portfolio Choice with Habit Formation: A Life-Cycle Model with Uniinsuable Labor Income Risk, Mimeo.

[32] Gomes, F. and Michaelides, A., 2002, Lyfe-Cycle Asset Allocation: A Model with Borrowing Constraints, Uninsurable Labor INcome Risk and Stock MArket Participation Costs, Mimeo.

[33] Goetzmann, W.N. and Kumar, A., 2001, Equity Portfolio Diversification, NBER Working Paper, W8686.

[34] 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.

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

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

[37] Grossman, S.J. and Laroque, G., 1990, Asset Pricing and Optimal Portfolio Choice in the Presence of Illiquid Durable Consumption Goods, Econometrica, 58, 25-51.

[38] Guiso, L. and Jappelli, T., 1998, Background Uncertainty and the Demand for Insurance against Insurable Risks, Geneva Papers on Risk and Insurance Theory, 23, 7-27.

[39] Guiso, L. and Jappelli, T., 2002, Household Portfolios in Italy, in Households Portfolios, edited by Guiso, Jappelli and Haliassos, MIT Press, 250-289.

[40] Guiso, L., Jappelli, T. and Terlizzese, D., 1996, Income Risk, Borrowing Constraints, and Portfolio Choice, American Economic Review, 86, 158-172.

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

[42] Haliassos, M. and Bertaut, C., 1995, Why Do So Few Hold Stocks?" The Economic Journal, 105, pp. 1110-29.

[43] Haliassos, M., 2002, Stockholding: Recent Lessons from Theory and Computations, Mimeo.

[44] Haliassos, M. and Hassapis, A., 2003, Borrowing Constrains, Portfolio Behavior and Precautionary Motives, forthcoming in Computational Methods in Decision Making, Economics and Finance, Kluwer Academic Publishers.

[45] Haliassos, M. and Hassapis, A., 2001, Non-expected Utility, Saving, and Portfolios, The Economic Journal , pp. 69-102.

[46] Haliassos, M. and Michaelides, A., 2002, Calibration and Computation of Household Portfolio Models, Cambridge, MIT Press, pp. 55-102.

[47] Haliassos, M. and Michaelides, A., 2003, Portfolio Choice and Liquidity Constraints, Forthcoming in International Economic Review

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

[49] Heaton, J. and Lucas, D., 1997, Market Frictions, Savings Behavior, and Portfolio Choice, Macroeconomic Dynamics, 1, 76-101.

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

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

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

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

[54] Ioannides, Y.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.

[55] Kimball, M.S., 1993, Standard Risk Aversion, Econometrica, 61, 3, 589-611.

[56] Koo, H.K., 1995, Consumption and Portfolio Selection with Labor Income I: Evaluation of Human Capital", Mimeo

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

[58] Jagannathan, R. and Kocherlakota, N. R., 1996, Why Should Older People Invest Less in Stocks Than Younger People?; Federal Reserve Bank of Minneapolis Quarterly Review, v. 20, iss. 3, pp. 11-23.

[59] Lease, R.C., Lewellen, W.G. and Schlarbaum, G.G., 1974, The Individual Investor: Attributes and Attitudes, Journal of Finance, 29, 413-433.

[60] Maddala, G.S., 1983, Limited-Dependent and Qualitative Variables in Econometrics. Cambridge: Cambridge University Press.

[61] Merton, R.C., 1971, Optimum Consumption and Portfolio Rules in a Continuous-Time Model, Journal of Economic Theory, 3, 247-413.

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

[63] Moskowitz, T. and A. Vissing-Jorgenson, 2002, "The Returns to Entrepreneurial Investment: A Private Equity Premium Puzzle?", American Economic Review, 2002.

[64] Nijman, T. and Verbeek, M., 1996, Incomplete Panels and Selection Bias, The Econometrics of Panel Data: A Handbook of the Theory with Applications, 449-490.

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

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

[67] Okunev, J., Wilson, P. and Zurbruegg, R., 2000, The Causal Relationship between Real Estate and Stock Markets, Journal of Real Estate Finance and Economics, 21, 251-61.

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

[69] Polkovnichenko, V., 2002, Limited Stock Market Participation and the Equity Premium, Mimeo.

[70] Poterba, J.M., 2000, Stock Market Wealth and Consumption, Journal of Economic Perspectives, 14, 99-118.

[71] Poterba, J.M., Venti, S.F. and Wise, D.A., 2000, Saver Behavior and 401(k) Retirement Wealth, American Economic Review, 90, 297-302.

[72] Poterba, J. M., 2001, Demographic Structure and Asset Returns, Review of Economics and Statistics, 83, 565-584.

[73] Pratt, J.W., and Zeckhauser, R.J., 1987, Proper Risk Aversion, Econometrica, 55, 1, 143-154.

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

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

[76] Telmer, C. I., 1993, Asset Pricing Puzzles and Incomplete Markets, Journal of Finance, 48, 1803-1832.

[77] Tepla, L., 2001, Optimal Investment with Minimum Performance Constraints, Journal of Economic Dynamics & Control, 25, 1629-1645.

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

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

[80] Viceira, L.M., 2001, Optimal Portfolio Choice for Long-Horizon Investors with Non-Tradable Labor Income, forthcoming Journal of Finance.

[81] Yao, R. and H. Zhang, 2002, Optimal Consumption and Portfolio Choices with Risky Housing and Stochastic Labor Income, Working Paper, UNC.

[82] Yitzhaki, S., 1987, The Reationship between Return and Income, The Quarterly Journal of Economics, 78-95.

Table 1: Descriptive Statistics of the sample

This table contains the descriptive statistics of the sample. In Panel A we report the number of households for each year and the descriptive statistics of household characteristics, for the full sample and for the sub-sample of stock market participants. We report the logarithm of financial and real estate wealth, capital gains and losses (standardized using mean and standard deviation of capital gains/losses in overall sample), dummies for secondary education (defined as 1 if the person' highest level of education is either completed or uncompleted high school, and 0 otherwise), dummies for high education (defined as 1 if the person has some post-high school education, and 0 otherwise), an immigration dummy (it takes the value 0 if all the members of the household are native Swedes, and 1 if at least one member of household immigrated), a Stockholm dummy (it is defined as 1 if the household resides in the capital, and 0 otherwise), the size of household and the age of its oldest member. We also report variables to account for the professional ability and unemployment risk of the investor. The first variable is based on the difference between his income and the average income of his profession (normalized by the standard deviation of the income in the profession). The second variable is the one-year-ahead forecast of a linear probability model where the unemployment status (i.e., 1 if unemployed and zero otherwise) is regressed on demographic variables, measures of income and wealth and regional, geographical and professional dummies. For stock market participants we also report the yearly portfolio return. Panel B presents the age and gender distribution of the sample. We report Mean, Median, Standard Deviation, Inter-Quartile Range (IQR) and Maximum value. They have been calculated over the whole sample (i.e., across-investors and time). Panel C reports the percentage of the households paying wealth tax, having labor income, having entrepreneurial income and having real estate wealth. The column "Representation in the sample" reports the fraction of households in the sample who pay wealth tax, earn labor or entrepreneurial income or hold real estate wealth. The other columns report the statistics (Mean, Median, Standard Deviation, IQR, Maximum) of, respectively, the value of wealth, labor and entrepreneurial income (gross yearly income) and real estate. Panel D reports the estimates for conditional moments of labor and entrepreneurial income estimated using Carroll and Samwick (1996) methodology described in Section 4.1. Otherwise noted, all monetary values are in Swedish Krones.

Panel A: Descriptive statistics of households

Variable	*Mean*	*Median*	*Std.Dev.*	*IQR*	*Maximum*
Number of households	292,901	291,913	647	686	293,320
		Full Sample			
Financial Wealth	2.98	4.26	2.35	4.96	8.88
Real Estate Wealth	2.12	5.04	3.76	7.62	7.89
Capital Gains/Losses	0.00	-0.02	1.00	0.00	1035.35
Secondary Education	0.43	0.00	0.46	1.00	1.00
Higher Education	0.31	0.00	0.50	1.00	1.00
Ability	0.05	0.00	0.58	0.31	4.81
Size of Household	2.67	2.00	1.51	3.00	16.00
Immigration Status	0.14	0.00	0.35	0.00	1.00
Age	49	47	17	24	107
Unemployment Risk	0.14	0.15	0.10	0.15	1.04
Stockholm Dummy	0.20	0.00	0.40	0.00	1.00
		Sample of Stock Market Participants			
RetPortfolio	0.56	0.23	1.52	0.85	16.11
Financial Wealth	4.51	4.95	1.47	0.94	8.88
Real Estate Wealth	3.87	5.58	3.24	1.05	7.54
Capital Gains/Losses	0.07	-0.02	2.62	0.02	1035.35
Secondary Education	0.35	0.00	0.44	1.00	1.00
Higher Education	0.47	0.00	0.48	1.00	1.00
Ability	0.08	1.00	0.44	0.29	3.97
Size of Household	2.78	2.00	1.35	2.00	14.00
Immigration Status	0.11	0.00	0.32	0.00	1.00
Age	54	53	15	22	102
Unemployment Risk	0.12	0.13	0.10	0.17	0.83
Stockholm Dummy	0.22	0.00	0.42	0.00	1.00

Panel B: Age and gender distribution of the sample

Age	*Males*	*Females*	*Age of oldest household member*
0-19	18.2%	17.2%	0.5%
20-29	4.8%	4.9%	10.7%
30-39	7.1%	8.2%	21.7%
40-49	7.4%	7.4%	23.6%
50-59	5.9%	5.3%	17.9%
60+	6.6%	7.2%	25.8%
Total	49.9%	50.2%	100%

Panel C: Wealth and income characteristics of households

Variable	Representation in the sample	Mean	Median	Std.Dev.	I. Q. R.	Maximum
Wealth-tax Payers	7.9%	359,592	102,700	2,648,521	353,400	1,023,147,857
Labor Inc. Earners	100.0%	321,489	287,722	237,526	276,190	43,445,271
Entrepr. Inc. Earners	9.8%	88,114	43,268	172,565	111,726	7,320,000
Real Estate Holders	54.6%	449,400	387,000	348,736	340,000	78,140,000

Panel D: Conditional moments of income characteristics of households

Variable	Mean	Median	Std.Dev.	IQR	Maximum
Total Sample					
Conditional Mean of Labor Income ('000 SEK)	420	332	503	352	52,238
Conditional Variance of Labor Income ('000,000 SEK^2)	212,459	3,609	5,491,396	24,285	3,293,430,136
Conditional Mean of Entrepreneurial Income ('000 SEK)	954	211	2284	964	233,433
Conditional Variance of Entr. Income ('000,000 SEK^2)	1,973,182	258	51,928,730	8,638	5,985,515,790
Sample of Non-Wealthy Households					
Conditional Mean of Labor Income ('000 SEK)	383	317	348	329	46,687
Conditional Variance of Labor Income ('000,000 SEK^2)	164,040	3,206	4,851,850	20,942	3,293,430,136
Conditional Mean of Entrepreneurial Income ('000 SEK)	1,069	253	2,393	1,331	2,334,33
Conditional Variance of Entr. Income ('000,000 SEK^2)	1,187,598	84	25,523,650	5,965	2,539,552,010
Sample of Wealthy Households					
Conditional Mean of Labor Income ('000 SEK)	677	462	1,037	517	52,238
Conditional Variance of Labor Income ('000,000 SEK^2)	535,052	8195	8,608,691	50,228	2,745,839,821
Conditional Mean of Entrepreneurial Income ('000 SEK)	431	101	1605	300	67,474
Conditional Variance of Entr. Income ('000,000 SEK^2)	5,522,516	2145	5,985,515,821	31,084	5,985,515,790

Table 2: Portfolio choice: investment in risky assets

We report the estimates of the determinants of the investment in risky financial assets. The dependent variable is either the percentage of the investment in risky assets over overall wealth (Risky Share) or the dollar value of the investment (Risky Value). Panel A reports a specification analogous to the one estimated in the literature based on the use of the market portfolio as a proxy for the investor portfolio, while Panel B reports the results for a specification where the actual investor portfolio is employed. In Panel B, positive and negative correlations ("+" and"-") are separately considered. We consider types of control variables: *measures of income and wealth, demographic variables, professional ability and risk, momentum/stock performance variables and residual control variables.* The *measures of income and wealth* contain the vector of the wealth of the ith investor at time t, broken down into its individual components (i.e., logarithm of financial wealth, logarithm of real estate wealth and other), as well as measures of income (i.e., labor and entrepreneurial) and overall (i.e., financial and non-financial) capital gains and losses of the ith investor at t. We also include the correlation between non-financial income (both labor income and entrepreneurial income) and real estate. All monetary variables (level and variance of non-financial income, wealth,…) except *Capital Gains/Losses,* have been transformed into logarithms. The *demographic variables* include: the profession of the investor, his level of education, broken down into high-school and university level, the age of the oldest member of the family of the investor and its value squared. We also construct variables to account for the professional ability and risk of the investor. The first variable is based on the difference between his income and the average income of his profession. The second variable is the one-year-ahead forecast of a linear probability model where the unemployment status (i.e., 1 if unemployed and zero otherwise) is regressed on demographic variables, measures of income and wealth and regional, geographical and professional dummies. The *momentum/stock performance variables* are the return and volatility on the investor's portfolio in the previous twelve months. The *residual control variables* include the standardized levels of debt for the investor (ratio of investor debt to total income and ratio of investor debt to total wealth), the return and volatility on the market portfolio in the previous twelve months, 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 also include 8 geographical areas and 10 industries. We also include a Stockholm dummy and a dummy that controls for the immigration status. The Stockholm dummy takes the value of 1 if the investor lives in the capital and 0 otherwise. The immigration status is a dummy that takes the value 0 if all the members of the household are native Swedes, and 1 if at least one member of household immigrated. *Age2* is divided by 1000. Estimates are performed using 2SLS.

Panel A: Investor's portfolio proxied by the market portfolio

	All households				Low-wealth households				High-wealth households			
	Risky Share		Risky Value		Risky Share		Risky Value		Risky Share		Risky Value	
Variable	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*
Labor Income (Variance)	0.04	1.67	0.02	1.44	-0.02	-3.91	0.01	1.33	-0.01	-1.27	0.01	1.01
Entrepr. Income (Variance)	0.01	0.67	-0.01	-1.44	0.04	1.58	0.00	-0.27	0.00	-1.23	-0.13	-7.70
Corr(Labor Inc., Market)	-0.13	-4.34	-0.92	-3.56	-0.04	-0.40	-0.77	-3.02	-0.38	-2.28	0.28	1.56
Corr(Entr.Inc., Market)	-0.04	-5.34	-0.09	-2.01	-0.15	-0.45	-0.10	-2.41	-0.01	-2.45	0.29	1.01
Control Variables												
Intercept	0.77	12.53	5.27	9.78	1.93	8.39	5.98	13.24	0.25	8.18	6.17	17.39
Labor Income (Level)	-0.06	-4.01	-0.18	-1.95	-0.02	-1.53	-0.23	-2.78	-0.02	-6.01	-0.03	-1.47
Entrepr. Income (Level)	0.06	7.49	0.11	1.98	0.02	1.54	0.28	5.58	0.00	-3.26	-0.02	-1.45
Corr(Labor Inc., Real Estate)	-0.07	-5.28	-0.11	-1.20	-0.02	-0.66	0.00	-0.01	-0.02	-4.65	-0.16	-3.08
Corr(Entr. Inc., Real Estate)	0.17	3.56	-0.11	-0.53	0.02	0.59	-0.10	-0.64	0.00	-1.51	-0.10	-2.18
RetPortfolio	0.00	-0.41	0.00	-1.24	0.00	-0.72	0.00	-1.79	0.00	0.50	0.00	-0.53
Financial Wealth	-0.08	-13.49	0.36	11.70	-0.23	-9.75	0.00	0.04	0.00	-0.53	0.12	4.57
Real Estate Wealth	-0.07	-8.79	0.01	0.45	-0.06	-4.95	-0.09	-3.54	-0.01	-3.40	-0.01	-0.29
Capital Gains/Losses	0.16	17.94	0.02	3.25	0.71	8.07	0.01	2.13	0.03	10.69	0.73	32.92
Secondary Education	-0.01	-2.82	-0.17	-4.29	-0.03	-2.05	-0.18	-4.95	-0.01	-5.13	-0.04	-1.62
Higher Education	-0.03	-5.07	-0.23	-4.29	-0.05	-2.60	-0.26	-5.63	-0.01	-4.37	-0.01	-0.23
Ability	-0.01	-2.47	-0.15	-3.84	0.01	0.77	-0.11	-2.59	-0.01	-6.44	-0.06	-4.09
Size of Household	0.01	1.08	-0.10	-3.21	0.00	0.11	0.06	2.10	0.00	-0.93	-0.04	-1.29
Immigration Status	-0.01	-3.24	0.12	3.29	0.02	1.42	0.02	0.67	0.00	1.24	0.08	3.81
Age	0.01	4.04	-0.02	-2.09	0.00	-0.55	0.02	2.62	0.00	-1.07	-0.03	-2.44
Age2	-0.01	-4.32	0.01	1.30	0.00	1.54	-0.02	-2.58	0.00	0.51	0.03	2.44
Stockholm Dummy	-0.02	-5.45	0.01	0.27	0.00	-0.37	-0.07	-2.31	-0.01	-3.41	0.03	2.17
Lagged Dependent Variable	0.10	9.96	0.04	24.48	0.07	3.98	0.03	18.60	0.66	56.25	0.06	47.36
Lambda	-0.10	-5.86	-0.40	-3.37	-0.21	-3.76	-0.69	-6.23	-0.02	-2.18	-0.22	-2.05
Residual Control Variables	Yes		Yes		Yes		Yes		Yes		Yes	
Adj R^2	0.54		0.20		0.61		0.09		0.49		0.17	

Panel B: Actual market portfolio

	All households				***Low-wealth households***				***High-wealth households***			
	Risky Share		*Risky Value*		*Risky Share*		*Risky Value*		*Risky Share*		*Risky Value*	
Variable	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*
$Var(Y_L)$*sign[$Corr(Y_L$,m)]	-0.04	-7.88	-0.33	-9.80	-0.03	-7.25	-0.19	-10.53	0.08	4.11	1.28	3.99
$Var(Y_E)$*sign[$Corr(Y_E$,m)]	-0.02	-9.05	-0.18	-11.88	-0.02	-7.25	-0.10	-8.51	0.02	1.80	0.33	2.39
Corr(Labor Inc., Portfolio) +	0.69	8.11	3.48	5.57	0.37	3.13	2.73	4.95	0.52	2.00	9.52	1.96
Corr(Labor Inc., Portfolio) -	0.12	1.01	1.41	1.51	0.25	1.48	0.54	0.67	0.18	0.32	3.43	0.23
Corr(Entr. Inc., Portfolio) +	1.67	5.56	15.99	7.33	2.88	8.18	14.21	7.36	0.15	0.37	-0.18	-0.03
Corr(Entr. Inc., Portfolio) -	0.09	0.39	1.38	0.85	0.58	1.69	2.06	1.13	-0.42	-1.39	-6.73	-1.37
Control Variables												
Intercept	-0.27	-3.27	-0.75	-1.12	0.26	2.74	2.65	5.43	-0.38	-1.46	-6.23	-1.21
Labor Income (Level)	-0.13	-6.21	-1.02	-6.86	-0.11	-4.22	-0.95	-9.16	0.05	2.91	0.97	3.59
Entrepr. Income (Level)	0.11	10.13	0.85	11.46	0.15	9.74	0.59	9.92	-0.00	-0.08	-0.08	-0.61
Corr(Labor Inc.. Real Estate)	-0.08	-4.66	-0.57	-5.08	-0.16	-6.32	-0.38	-3.25	0.05	1.99	1.42	3.11
Corr(Entr. Inc., Real Estate)	0.34	8.12	2.81	9.56	0.21	3.58	1.39	5.00	0.10	1.82	1.66	1.66
RetPortfolio	-0.01	-1.77	-0.01	-1.31	0.00	1.48	-0.00	-0.63	0.00	0.06	0.01	0.37
Financial Wealth	-0.02	-7.49	0.17	7.66	-0.04	-11.82	0.08	5.17	-0.01	-1.94	0.10	2.18
Real Estate Wealth	-0.04	-3.13	0.02	3.63	-0.05	-2.07	0.01	1.52	-0.01	-3.77	-0.03	-1.24
Capital Gains/Losses	0.01	3.73	0.02	3.17	0.01	2.56	0.01	3.19	0.00	0.28	0.02	1.55
Secondary Education	0.02	2.20	0.04	0.73	-0.03	-2.81	-0.14	-3.21	0.02	1.10	0.59	1.60
Higher Education	0.03	4.85	0.20	4.76	0.00	0.00	-0.03	-0.81	0.02	1.08	0.35	1.31
Ability	0.06	8.86	0.59	12.23	0.05	6.28	0.31	8.32	0.02	1.30	0.11	0.69
Size of Household	-0.42	-7.87	-5.27	-14.22	-0.65	-7.44	-3.70	-9.42	0.08	0.66	3.97	2.05
Immigration Status	-0.01	-3.82	0.03	2.09	-0.02	-6.04	0.02	1.55	0.01	2.07	0.19	2.83
Age	0.01	5.38	0.03	5.42	0.01	4.88	0.01	2.11	0.01	2.11	0.12	2.74
Age2	-0.03	-5.71	-0.19	-4.31	-0.05	-6.21	-0.07	-2.06	-0.03	-1.91	-0.72	-2.55
Unemployment Risk	-0.42	-7.87	-5.27	-14.22	-0.65	-7.44	-3.70	-9.42	0.08	0.66	3.97	2.05
Stockholm Dummy	-0.03	-7.68	-0.30	-11.45	-0.05	-8.11	-0.19	-7.02	-0.05	-1.40	-0.89	-1.54
Lagged Dependent Variable	0.12	25.32	-0.02	-3.40	0.15	10.36	0.01	2.24	0.69	6.16	0.17	1.51
Lambda	0.18	9.47	0.92	6.43	0.06	2.94	0.31	2.98	0.11	1.39	2.23	1.63
Residual Control Variables	Yes		Yes		Yes		Yes		Yes		Yes	
Adj R^2	0.54		0.22		0.61		0.10		0.50		0.18	

Table 3: Correlation and indexes of hedging
(Tests of restriction H2.A)

This table reports statistics of the correlations of non-financial income (i.e., labor income and entrepreneurial income) with financial returns (i.e., portfolio returns and overall stock market returns) and real estate return. The non-financial income components are estimated by using the Carrol and Samwick (1997) and Vissing-Jorgensen (2002) methodology reported in the text. We report the descriptive statistics separately for high-wealth and low-wealth households as well as our indexes of active hedging (i.e., Γ_1 and Γ_2, where Γ_1 = Corr(Labor Inc., Market) -Corr(Labor Inc., Portfolio) and Γ_2 = Corr(Ent. Inc., Market) -Corr(Ent. Inc., Portfolio)). We report the the results of the *t-tests* for each group (where we test for mean = 0 hypothesis) as well as the difference between high- and low-wealth households. The values of Δ_1 and Δ_2 are multiplied by 10,000.

	Low Wealth					*High Wealth*					*T-test of the difference*	
Variable	*Mean*	*Median*	*Std.Dev.*	*I. Q. R.*	*T-test Mean=0*	*Mean*	*Median*	*Std.Dev.*	*I. Q. R.*	*T-test Mean=0*	*T-Stat*	*p-value*
Corr(Labor Inc., Market)	-0.01	0.00	0.31	0.41	-4.86	-0.04	-0.05	0.31	0.41	-16.51	10.82	<0.0001
Corr(Labor Inc., Portfolio)	0.03	0.09	0.79	1.77	2.57	0.05	0.10	0.68	1.10	7.44	-2.32	0.020
Corr(Labor. Inc., Real Estate)	-0.06	-0.09	0.47	0.76	-27.81	-0.13	-0.19	0.46	0.74	-39.51	16.71	<0.0001
Corr(Entrepr. Inc., Market)	-0.07	-0.14	0.61	0.90	-11.01	-0.07	-0.14	0.59	0.90	-9.14	0.09	0.926
Corr(Entrepr. Inc., Portfolio)	-0.01	0.00	0.84	1.88	-0.29	-0.06	-0.14	0.77	1.60	-3.99	2.10	0.036
Corr(Entrepr. Inc., Real Estate)	-0.18	-0.32	0.64	0.98	-25.24	-0.18	-0.21	0.62	0.86	-21.33	0.40	0.686
Γ_1	-0.03	-0.03	0.82	1.34	-2.31	-0.08	-0.11	0.71	1.07	-10.42	3.36	0.001
Γ_2	-0.07	-0.00	0.88	1.39	-2.71	-0.09	-0.03	0.78	1.10	-4.05	0.70	0.483
Δ_1	-0.34	-0.00	4.18	0.85	-7.07	-0.49	-0.07	3.96	0.74	-12.06	2.47	0.014
Δ_2	-0.34	-0.00	2.50	0.49	-0.99	-0.41	-0.00	10.43	0.39	-2.99	0.16	0.866

Table 4: Determinants of the tilt in the risk profile.

This table reports estimates of the determinants of Δ_1 (Panel A), Δ_2 (Panel B), and Δ_T (Panel C) where Δ_1 = Cov(Labor Inc., Market) - Cov(Labor Inc., Portfolio), Δ_2 = Cov(Entrepreneurial Inc., Market) - Cov(Entrepreneurial Inc., Portfolio), and Δ_T = Cov(Total Income, Market) - Cov(Total Income, Portfolio). Θ_L is $-(Y_L+Y_E)\Sigma_j SR_j*Cov(Ret_j, Y_L)$. It is defined in the text (see Appendix), while $Cov(Y_E, Y_L)$ is the covariance between labor and entrepreneurial income. Familiarity variables are *Professional Proximity, Geographical Proximity* and *Holding Period.* They are constructed as follows. For each stock in the portfolio we identify three measures of familiarity and then we aggregate them for each investor on the basis of his portfolio composition (i.e., using as weights the value of the portfolio holding). The first measure is *Professional Proximity.* It is a dummy taking the value 1 if the investor's profession is in the same area of activity as the company whose stock is 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 of an investor working in the mining sector holding a stock of a mining company, the dummy would be equal to 1. The second measure is *Geographical Proximity*, that is the proximity between the residence of the investor and the place where the company is located. In particular, we use the logarithm of the inverse of 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. The greater the value of the variable, the closer the investor is to the stock. Finally, we construct a variable that proxies for *Holding Period*, based on the time a stock entered the investor's portfolio. Each index of familiarity is constructed weighting the measures for each investor on the basis of his portfolio composition. The control variables are defined as in Table 2.

Panel A: Determinants of Δ_1 *(Hedging labor income risk)*

	All households				Low-wealth households				High-wealth households			
	I		II		I		II		I		II	
Variable	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*
$Y_L Var(Y_L)$	-0.26	-10.85	-0.25	-10.76	-0.34	-16.59	-0.33	-15.51	-9.88	-3.26	-10.12	-3.02
$Y_E Cov(Y_E, Y_L)$	-1.48	-16.67	-1.38	-14.95	-1.56	-19.73	-1.40	-16.52	-4.83	-1.93	-3.83	-1.31
Θ_L	-0.84	-12.27	-0.76	-10.70	-0.92	-14.85	-0.80	-11.99	-0.85	-3.79	-0.86	-3.43
Geographical proximity			-1.96	-4.68			-2.48	-4.05			-1.68	-1.88
Professional proximity			-0.61	-0.83			-1.09	-1.05			0.25	0.23
Holding Period			-6.36	-5.84			-7.58	-5.68			-4.43	-3.08
Control Variables												
Intercept	-32.66	-6.52	-44.70	-4.97	-49.58	-9.80	-72.58	-8.52	-27.97	-1.35	-18.04	-0.77
Corr(Labor Inc., Real Estate)	2.73	21.76	2.56	9.31	3.05	23.51	2.97	21.78	3.62	1.77	2.87	1.33
Corr(Entr. Inc., Real Estate)	-0.07	-0.32	-0.13	-0.64	0.04	0.19	-0.02	-0.08	-1.28	-3.39	-1.22	-2.94
RetPortfolio	0.21	2.88	0.19	2.64	0.15	2.09	0.17	2.36	0.00	-0.80	0.00	-0.79
Financial Wealth	1.25	7.83	1.50	9.09	1.79	10.79	1.94	11.16	0.48	1.45	0.25	0.67
Real Estate Wealth	0.13	4.16	0.19	5.71	0.18	6.25	0.24	7.59	-0.32	-2.62	-0.34	-2.57
Capital Gains/Losses	0.46	9.71	0.44	9.21	0.65	12.20	0.62	11.10	0.11	1.12	0.08	0.73
Secondary Education	-1.11	-3.23	-0.88	-2.57	-1.09	-2.87	-0.72	-1.80	0.08	0.06	-0.24	-0.19
Higher Education	-0.68	-1.61	-0.43	-1.01	-0.32	-0.73	-0.02	-0.04	0.63	0.42	0.22	0.14
Ability	-0.54	-2.54	-0.05	-0.21	-0.55	-2.52	0.00	0.00	-0.49	-0.54	0.00	-0.01
Size of Household	0.93	8.67	0.95	8.25	0.95	10.34	1.09	10.12	-0.05	-0.14	-0.02	-0.04
Immigration Status	-1.13	-5.18	-1.51	-6.52	-1.77	-7.57	-2.30	-9.08	1.89	2.24	2.18	2.27
Age	0.73	6.06	0.71	5.95	0.91	8.98	0.81	7.59	0.65	0.84	0.55	0.67
Age2	-0.74	-6.50	-0.72	-6.30	-0.95	-9.27	-0.85	-7.88	-0.55	-0.83	-0.50	-0.72
Unemployment Risk	-8.13	-5.04	-13.65	-6.39	-9.15	-5.28	-16.10	-7.47	5.53	0.53	-0.14	-0.01
Stockholm Dummy	-0.30	-1.92	0.03	0.18	-0.55	-3.34	-0.50	-2.59	-0.24	-0.42	0.42	0.60
Lambda	6.19	6.92	7.32	8.02	10.07	10.01	10.94	10.37	5.28	1.47	4.32	1.18
Residual Control Variables	yes		yes		yes		yes		yes		yes	
Adj R^2	0.02		0.03		0.02		0.03		0.02		0.03	

Panel B: Determinants of Δ_2 *(Hedging entrepreneurial income risk)*

	All households				Low-wealth households				High-wealth households			
	I		II		I		II		I		II	
Variable	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*
$Y_E Var(Y_E)$	-0.43	-34.05	-0.43	-33.66	-0.22	-9.06	-0.26	-7.77	1.09	9.95	0.73	9.31
$Y_L Cov(Y_E, Y_L)$	-0.49	-3.82	-0.48	-3.72	-1.34	-3.85	-1.32	-3.23	-11.17	-3.28	-7.15	-2.65
Θ_E	-0.34	-50.41	-0.34	-49.34	-0.23	-20.63	-0.23	-17.75	-0.59	-20.40	-0.62	-27.06
Geographical proximity			-2.90	-3.83			-0.82	-0.70			-5.38	-4.43
Professional proximity			-2.43	-1.65			-2.64	-2.30			0.22	0.09
Holding Period			-4.89	-2.26			-6.49	-2.27			-22.27	-5.02
Control Variables												
Intercept	37.29	3.93	91.24	5.03	35.32	3.74	-27.88	-1.33	-255.51	-3.82	43.10	1.90
Corr(Labor Inc., Real Estate)	-0.21	-0.86	-0.03	-0.10	-0.13	-0.55	-0.24	-0.84	-2.10	-2.74	-2.27	-3.58
Corr(Entr. Inc., Real Estate)	-2.06	-5.30	-2.09	-5.32	-1.61	-4.02	-1.64	-3.49	-2.82	-3.43	-3.08	-4.63
RetPortfolio	0.00	-0.00	0.00	-0.03	0.00	0.52	-0.03	-0.88	0.00	-1.00	0.00	-0.79
Financial Wealth	-0.99	-3.30	-0.88	-2.78	-0.76	-2.36	-0.96	-2.42	-1.35	-2.75	-0.23	-0.67
Real Estate Wealth	-0.35	-5.74	-0.35	-5.34	-0.09	-1.51	-0.01	-1.22	-0.66	-3.43	-0.34	-2.51
Capital Gains/Losses	0.26	2.81	0.25	2.73	0.03	0.43	0.01	-0.04	0.55	2.28	0.66	3.43
Secondary Education	-4.80	-6.97	-4.56	-6.54	-2.58	-3.38	-2.71	-3.02	-8.06	-3.48	-1.01	-0.63
Higher Education	-4.87	-6.00	-4.65	-5.65	-2.14	-2.47	-2.39	-2.34	-9.29	-3.09	0.44	0.22
Ability	-3.22	-5.43	-3.51	-5.74	-3.12	-4.60	-3.08	-3.86	-0.02	-0.02	-0.40	-0.41
Size of Household	-0.74	-3.47	-1.07	-4.65	-0.53	-3.71	-0.46	-2.69	0.44	0.94	-0.62	-1.79
Immigration Status	0.21	0.49	0.31	0.68	0.61	1.35	-0.16	-0.28	-1.19	-1.01	0.21	0.22
Age	0.53	2.31	0.62	2.64	-0.35	-1.64	-0.29	-1.15	1.69	2.21	-0.34	-0.62
Age2	-0.59	-2.69	-0.63	-2.84	0.27	1.29	0.22	0.88	-1.74	-2.56	0.10	0.21
Unemployment Risk	27.52	7.94	32.61	7.59	23.36	6.55	13.48	2.79	-9.50	-0.73	3.71	0.39
Stockholm Dummy	0.29	0.91	0.87	2.46	0.17	0.52	-0.24	-0.57	2.67	2.77	2.50	3.15
Lambda	-6.61	-3.90	-6.00	-3.37	-5.31	-2.73	-5.74	-2.43	-17.13	-2.02	11.74	2.02
Residual Control Variables	yes		yes		yes		yes		yes		Yes	
Adj R^2	0.01		0.02		0.01		0.02		0.01		0.02	

Panel C: Determinants of Δ_T *(Hedging total income risk)*

	All households				***Low-wealth households***				***High-wealth households***			
	I		**II**		**I**		**II**		**I**		**II**	
Variable	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*
$Y_T Var(Y_T)$	-0.07	-3.51	-0.12	-5.33	-0.13	-9.89	-0.16	-9.34	1.24	12.19	1.20	10.92
Θ_T	-0.37	-8.01	-0.37	-7.72	-0.10	-3.88	-0.07	-2.01	-5.15	-7.86	-5.20	-6.85
Geographical proximity			-1.26	-2.56			-1.55	-2.39			-3.44	-5.64
Professional proximity			-0.79	-0.91			-0.11	-0.10			-2.88	-2.37
Holding Period			-8.39	-6.74			-9.78	-6.93			-10.80	-4.18
Control Variables												
Intercept	-22.80	-4.19	-93.25	-9.19	-52.05	-12.01	-144.68	-16.04	-3.45	-0.25	34.39	1.50
Corr(Total Inc., Real Estate)	2.67	18.44	2.21	14.20	2.70	23.91	2.53	17.33	2.61	7.46	2.34	5.68
RetPortfolio	0.35	4.91	0.30	3.97	0.09	1.48	0.03	0.38	0.11	2.06	0.07	1.31
Financial Wealth	0.74	4.23	1.25	6.68	1.59	11.23	1.99	10.84	-0.31	-1.77	-0.34	-1.90
Real Estate Wealth	0.05	3.50	0.18	4.64	0.12	4.86	0.26	7.84	-0.11	-1.45	-0.16	-1.76
Capital Gains/Losses	0.42	7.75	0.38	6.75	0.63	13.47	0.58	9.76	0.00	0.01	-0.02	-0.22
Secondary Education	-1.29	-3.32	-0.89	-2.20	-1.99	-6.07	-0.89	-2.09	-1.88	-2.08	-1.16	-1.18
Higher Education	-0.92	-1.92	-0.50	-1.02	-1.10	-2.88	-0.07	-0.14	-2.68	-2.52	-1.77	-1.51
Ability	-0.97	-4.11	0.56	1.96	-0.48	-2.54	1.32	4.93	-0.19	-0.42	0.79	1.31
Size of Household	0.99	8.41	1.35	10.28	0.89	11.22	1.55	13.43	0.90	2.99	0.71	2.22
Immigration Status	-0.94	-3.79	-1.86	-6.91	-1.84	-9.04	-2.99	-11.01	0.30	0.66	0.82	1.66
Age	0.50	3.78	0.46	3.35	0.98	11.13	0.73	6.40	0.75	1.44	0.60	1.10
Age2	-0.53	-4.19	-0.50	-3.83	-1.02	-11.55	-0.80	-6.97	-0.71	-1.62	-0.58	-1.25
Unemployment Risk	-2.81	-1.56	-19.59	-7.98	-5.71	-3.90	-26.67	-11.67	-4.09	-0.85	-12.15	-2.02
Stockholm Dummy	-0.14	-0.79	-0.12	-0.56	-0.28	-1.95	-0.84	-4.01	-0.11	-0.30	0.91	2.17
Lambda	3.48	3.56	5.72	5.53	8.91	10.40	11.45	10.30	-4.98	-1.70	-0.94	-2.30
Residual Control Variables	yes		yes		yes		yes		yes		yes	
Adj R^2	0.01		0.02		0.01		0.02		0.01		0.02	

Table 5: Determinants of the tilt in the risk profile: public vs. private sector employees.

This table contains estimates of the determinants of the tilt for private sector (Panel A) and public sector (Panel B) employees. We define as public sector employees those whose share of income coming from occupations with SNI codes between 75000 and 91999 is in excess of 50%. The notations are as in Table 4. We consider labor income. We use the same control variables as in Table 5.

Panel A: Determinants of Δ_1 *for private sector employees*

	All households				*Low-wealth households*				*High-wealth households*			
	I		**II**		**I**		**II**		**I**		**II**	
Variable	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*
$Y_L Var(Y_L)$	-0.16	-6.44	-0.17	-6.65	-0.22	-11.49	-0.22	-11.16	-7.00	-3.10	-7.12	-3.12
$Y_E Cov(Y_E, Y_L)$	-1.55	-13.31	-1.49	-12.49	-1.48	-17.37	-1.36	-15.10	-2.38	-1.13	-2.20	-0.94
Θ_L	-0.97	-10.60	-0.93	-9.83	-0.93	-13.70	-0.83	-11.61	-0.60	-3.61	-0.61	-3.59
Geographical proximity			-2.05	-4.20			-2.86	-4.87			-0.45	-1.68
Professional proximity			-0.74	-0.80			-1.36	-1.25			0.35	0.27
Holding Period			-5.77	-4.37			-7.79	-5.95			-1.29	-3.48
Control Variables	Yes		Yes		Yes		Yes		Yes		yes	
Adj R^2	0.02		0.03		0.02		0.03		0.02		0.03	

Panel B: Determinants of Δ_1 *for public sector employees*

	All households				*Low-wealth households*				*High-wealth households*			
	I		**II**		**I**		**II**		**I**		**II**	
Variable	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*
$Y_L Var(Y_L)$	-0.45	-15.90	-0.45	-15.92	-0.64	-20.91	-0.64	-20.94	5.40	1.66	-3.43	-0.76
$Y_E Cov(Y_E, Y_L)$	-0.74	-9.46	-0.74	-9.45	-0.86	-10.68	-0.86	-10.66	-2.36	-0.74	-5.12	-1.50
Θ_L	-0.18	-26.20	-0.18	-26.20	-0.19	-25.34	-0.19	-25.34	0.11	0.45	-0.52	-1.59
Geographical proximity			-0.14	-2.14			0.01	0.16			-1.85	-1.75
Professional proximity			0.36	0.26			0.58	0.31			-1.93	-0.76
Holding Period			-0.15	-1.46			-0.22	-1.76			-5.41	-1.27
Control Variables	Yes		Yes		Yes		Yes		Yes		yes	
Adj R^2	0.02		0.03		0.02		0.03		0.02		0.03	

Table 6: The nature of familiarity

We estimate:

$$_{z,t}=\alpha_3+\beta_3(Y_{z,t}/W)\ VarY_{z,t}+\gamma_3(Y_{x,t}/W)CovY_z,Y_x+\zeta_3(\Theta_{z,t}/W)+\nu_3\Psi(1-\xi_1)+\pi_3\Psi\xi+\delta_3F_{3,t}+\theta_3\lambda_t+\varepsilon_{3,t},$$

where F_3 is a vector of control variables, Ψ is a vector of proximity variables, and λ_t is Heckman' lambda. In the case of differences in wealth or liquidity, ξ is a continuous variable based on the amount of wealth or liquidity of the investor and $\xi_1=0$. In the case of familiarity shocks, ξ is a dummy that takes the value 1 if in the previous three years the investor has been subject to a "familiarity shock" and zero otherwise and $\xi_1=\xi$. In Panel A, we use the logarithm of net wealth as a proxy for wealth (i.e., $\xi_1=0$, ξ=WEALTH). The estimates are done for the whole sample. In Panel B, we define liquidity as the fraction of wealth invested in liquid assets (i.e., $\xi_1=0$, ξ=LIQUIDITY). The estimates are performed for the sample of wealthy households. We define as familiarity shocks events that change investor's proximity to the stocks. In Panel C, we use an unemployment dummy ($\xi_1=\xi$=UNEMPL) that takes the value one if the investor has been unemployed at least once in the previous three years and zero otherwise. In Panel D, we use a professional change dummy ($\xi_1=\xi$=PCHANGE) that takes the value one if the investor has changed profession at least once in the previous three years and the current profession differs from the previous one and zero otherwise. Finally, in Panel E, we use a relocation dummy ($\xi_1=\xi$=MOVER) that takes the value one if the investor has moved and changed county (municipality) at least once in the previous three years and the current address differs from the previous one and zero otherwise. In Panels C, D, and E the estimates are done for the full sample. The notations are as in Table 4. We use the same control variables as in the previous tables.

Panel A: Wealth and Tilt of the portfolio.

	Labor Income		Entrepreneurial Income		Total Income	
Variable	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*
$Y_iVar(Yi)$	-0.25	-4.48	-0.43	-30.80		
$Y_iCov(Yi, Yj)$	-1.47	-9.26	-0.40	-2.74		
Θ_I	-1.04	-8.46	-0.34	-44.77		
$Y_TVar(Y_T)$					0.05	1.03
Θ_T					-1.20	-3.50
Geographical proximity	-38.42	-2.85	-62.23	-4.02	-78.39	-7.30
Geographical proximity*WEALTH	4.52	2.57	10.22	4.21	11.97	6.89
Professional proximity	-245.45	-4.88	-364.97	-4.59	-127.34	-2.34
Professional proximity*WEALTH	39.18	4.02	57.51	4.59	12.02	2.33
Holding Period	-62.82	-2.55	-132.42	-4.14	-137.10	-6.93
Holding Period*WEALTH	6.15	2.01	21.84	4.18	15.55	6.24
Control Variables	Yes		Yes		Yes	
Adj R^2	0.02		0.01		0.02	

Panel B: Liquidity and Tilt of the portfolio.

	Labor Income		Entrepreneurial Income		Total Income	
Variable	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*
$Y_iVar(Yi)$	-13.20	-3.15	0.88	7.82		
$Y_iCov(Yi, Yj)$	-8.64	-2.57	-8.08	-2.33		
Θ_I	-1.12	-3.61	-0.64	-21.89		
$Y_TVar(Y_T)$					1.07	5.51
Θ_T					-2.97	-2.37
Geographical proximity	-7.95	-2.43	-27.55	-3.91	-20.01	-3.91
Geographical proximity*LIQUIDITY	32.62	1.96	146.08	3.14	104.80	3.37
Professional proximity	5.72	1.95	-131.10	-0.81	-272.92	-2.40
Professional proximity*LIQUIDITY	-16.57	-1.24	880.10	0.80	183.71	2.36
Holding Period	-10.01	-2.26	-63.95	-4.62	-38.45	-4.07
Holding Period*LIQUIDITY	64.51	2.04	258.81	3.08	177.34	3.10
Control Variables	Yes		Yes		Yes	
Adj R^2	0.02		0.02		0.02	

Panel C: Unemployment and Tilt of the portfolio.

Variable	Labor Income		Entrepreneurial Income		Total Income	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
$Y_i Var(Yi)$	-0.27	-10.40	-0.42	-34.44		
$Y_i Cov(Yi, Yj)$	-1.40	-14.13	-0.51	-4.09		
Θ_I	-0.78	-10.23	-0.34	-51.13		
$Y_T Var(Y_T)$					-0.09	-4.21
Θ_T					-0.28	-5.07
Geographical proximity*(1-UNEMPL)	-3.78	-7.90	-2.90	-2.67	-2.38	-11.24
Geographical proximity*UNEMPL	0.99	4.95	-0.76	-1.87	1.41	6.50
Professional proximity*(1-UNEMPL)	-2.21	-8.49	-6.49	-2.21	-3.51	-5.79
Professional proximity* UNEMPL	2.98	8.51	4.50	1.77	4.82	5.82
Holding Period*(1-UNEMPL)	-11.45	-9.03	-3.44	-3.69	-3.90	-2.52
Holding Period* UNEMPL	-1.02	-5.23	0.89	3.04	2.47	5.41
Control Variables	Yes		Yes		Yes	
Adj R^2	0.02		0.01		0.01	

Panel D: Professional Change and Tilt of the portfolio.

Variable	Labor Income		Entrepreneurial Income		Total Income	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
$Y_i Var(Yi)$	-0.27	-10.77	-0.43	-31.43		
$Y_i Cov(Yi, Yj)$	-1.40	-14.60	-0.55	-3.88		
Θ_I	-0.78	-10.55	-0.34	-46.18		
$Y_T Var(Y_T)$					-0.05	-2.33
Θ_T					-0.24	-2.78
Geographical proximity*(1-PCHANGE)	-4.38	-5.95	-4.03	-2.88	-5.50	-6.36
Geographical proximity*PCHANGE	2.33	6.13	3.06	0.81	4.53	10.21
Professional proximity*(1-PCHANGE)	-3.61	-3.44	-1.84	-1.80	-1.75	-1.97
Professional proximity* PCHANGE	13.38	4.42	3.35	1.66	14.73	4.10
Holding Period*(1-PCHANGE)	-11.93	-6.53	-6.64	-1.99	-12.08	-8.48
Holding Period* PCHANGE	1.52	4.96	8.33	1.04	2.03	5.55
Control Variables	Yes		Yes		Yes	
Adj R^2	0.02		0.01		0.01	

Panel E: Relocation and Tilt of the portfolio.

Variable	Labor Income		Entrepreneurial Income		Total Income	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
$Y_i Var(Yi)$	-0.28	-12.45	-0.43	-33.76		
$Y_i Cov(Yi, Yj)$	-1.52	-17.03	-0.55	-4.21		
Θ_I	-0.87	-12.62	-0.34	-49.61		
$Y_T Var(Y_T)$					-0.06	-2.00
Θ_T					-0.40	-8.48
Geographical proximity*(1-MOVER)	-3.91	-8.63	-3.30	-3.64	-1.37	-4.01
Geographical proximity*MOVER	0.52	1.35	17.07	3.58	-0.53	-4.24
Professional proximity*(1-MOVER)	-1.68	-2.27	-2.51	-1.67	-1.72	-2.04
Professional proximity*MOVER	45.49	4.25	8.71	1.84	17.97	1.93
Holding Period*(1-MOVER)	-8.22	-6.89	-5.47	-2.43	-2.31	-2.93
Holding Period* MOVER	2.21	4.89	35.55	3.58	-0.31	-1.22
Control Variables	Yes		Yes		Yes	
Adj R^2	0.02		0.02		0.02	

Table 7: Portfolio choice: investment in risky assets

We report the estimates of the determinants of the investment in risky financial assets. The dependent variable is either the percentage value of the investment in risky assets over overall wealth (Risky Share) or the dollar value of the investment (Risky Value). The variables are defined as in Table2. The Indexes of hedging are Γ_1 and Γ_2, where Γ_1 = Corr(Labor Inc., Market) - Corr(Labor Inc., Portfolio) and Γ_2 = Corr(Entrepreneurial Inc., Market) - Corr(Entrepreneurial Inc., Portfolio). The Indexes of Familiarity are *Professional Proximity, Geographical Proximity* and *Holding Period,* defined as before. The control variables are the same as the ones used in Table 2 plus the lagged dependent variable.

Panel A: Indexes of Hedging

	All households				***Low-wealth households***				***High-wealth households***			
	Risky Share		*Risky Value*		*Risky Share*		*Risky Value*		*Risky Share*		*Risky Value*	
Variable	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*
Γ_1	-0.06	-7.67	-0.59	-9.48	-0.05	-7.15	-0.35	-9.26	-0.01	-3.42	-0.16	-5.02
Γ_2	-2.21	-9.09	-20.73	-9.96	-2.20	-9.03	-13.05	-9.93	-0.55	-5.38	-6.47	-7.67
Labor Income (Variance)	0.08	3.50	0.01	0.07	0.02	0.96	-0.62	-5.11	0.04	1.59	0.87	4.29
Entrepr. Income (Variance)	0.12	11.29	0.66	6.80	0.17	14.58	0.52	9.19	0.01	0.69	0.22	2.47
Control Variables	Yes		Yes		Yes		Yes		Yes		Yes	
Adj R^2	0.54		0.19		0.61		0.09		0.50		0.17	

Panel B: Indexes of Hedging and Familiarity

	All households				***Low-wealth households***				***High-wealth households***			
	Risky Share		*Risky Value*		*Risky Share*		*Risky Value*		*Risky Share*		*Risky Value*	
Variable	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*
Γ_1	-0.04	-5.97	-0.36	-7.79	-0.05	-6.28	-0.35	-8.12	-0.01	-2.44	-0.07	-1.63
Γ_2	-1.59	-7.62	-13.29	-8.54	-2.15	-7.85	-13.32	-8.69	-0.52	-4.20	-4.16	-3.90
Geographical proximity	0.07	10.20	0.83	16.58	0.11	13.50	0.79	16.69	0.02	2.64	0.61	8.98
Professional proximity	0.61	5.72	3.16	3.89	0.26	1.55	0.11	0.12	0.25	2.14	2.95	2.85
Holding Period	0.04	4.47	0.59	7.90	0.11	7.99	0.75	9.83	0.00	0.08	0.20	1.58
Labor Income (Variance)	0.01	2.93	0.03	1.44	-0.01	-2.62	-0.11	-5.20	0.01	3.18	0.20	5.56
Entrepr. Income (Variance)	0.01	7.67	0.09	8.18	0.01	6.13	0.05	6.99	0.01	3.82	0.05	3.93
Control Variables	Yes		Yes		Yes		Yes		Yes		Yes	
Adj R^2	0.55		0.23		0.62		0.12		0.51		0.20	

Table 8: Profits of high-wealth and low-wealth households

We report tests of the difference of two measures of profits for high-wealth and low-wealth households. The measures are $\Pi_{Wt}=(Wealth_t/Wealth_{t-1}-1)$ and $\Pi_{Ft}=((Capital\ Gains/Losses_t+Dividends_t)/RISKY\_ASSETS_{t-1})$, where capital gains and losses are the reported realized gains/losses of the household in risky assets at year *t*. Panel A reports *t-tests*, *Wilcoxon* and *Kolmogorov-Smirnov tests* of the equality of profits between high-wealth and low-wealth households. For the *t-tests* we used *Satterwhaite* version of the *t-tests* that assumes inequality of variances in two sub-samples. In all cases equality of variances was rejected at the 1% level. In Panels B and C, households are separated into 2 groups: the "hedgers" (defined as the households with positive Γ_1) and non-hedgers (households with negative Γ_1). Panel B reports the tests of difference of profits between high- and low-wealth households after they have been separated into hedgers and non-hedgers. Panel C reports the tests of difference of profits between hedgers and non-hedgers performed separately for high-wealth and low-wealth households. The asterisks denote that the statistics are robust to bootstrapping on 0.1% ("**") and 1% ("*") level. Bootstrapping is based on 20,000 resamplings.

Panel A: Profit measures

	Mean		**Median**		**T-Test**		**Wilcoxon Test**		**Kolmogorov-Smirnov Test**	
Variable	*Low-wealth*	*High-wealth*	*Low-wealth*	*High-wealth*	*t value*	*p-value*	*Z*	*p-value*	*Ksa*	*p-value*
Π_{Ft}	0.13	0.22	0.00	0.06	41.94	<.0001**	120.39	<.0001**	62.83	<.0001**
Π_{Wt}	0.05	0.10	0.07	0.08	23.43	<.0001**	11.34	<.0001**	16.89	<.0001**

Panel B: Profit measures (low- vs. high- wealth households)

		Mean		**Median**		**T-Test**		**Wilcoxon Test**		**Kolmogorov-Smirnov Test**	
Variable	*Hedging*	*Low-wealth*	*High-wealth*	*Low-wealth*	*High-wealth*	*t value*	*p-value*	*Z*	*p-value*	*Ksa*	*p-value*
Π_{Ft}	yes	0.12	0.22	0.00	0.05	36.91	<.0001**	96.40	<.0001**	48.63	<.0001**
Π_{Ft}	no	0.19	0.23	0.05	0.07	7.44	<.0001**	31.92	<.0001**	18.66	<.0001**
Π_{Wt}	yes	0.04	0.08	0.06	0.07	17.72	<.0001**	3.56	0.0002**	14.76	<.0001**
Π_{Wt}	no	0.09	0.13	0.09	0.10	6.11	<.0001**	2.55	0.0054*	9.01	<.0001**

Panel C: Profit measures (hedgers- vs. non-hedgers)

		Mean		**Median**		**T-Test**		**Wilcoxon Test**		**Kolmogorov-Smirnov Test**	
Variable	*Wealthy*	*Hedgers*	*Non-Hedgers*	*Hedgers*	*Non-Hedgers*	*t value*	*p-value*	*Z*	*p-value*	*Ksa*	*p-value*
Π_{Ft}	yes	0.22	0.230	0.05	0.07	3.15	0.0020**	34.40	<.0001**	22.44	<.0001**
Π_{Ft}	no	0.12	0.194	0.00	0.05	19.99	<.0001**	60.75	<.0001**	32.96	<.0001**
Π_{Wt}	yes	0.08	0.091	0.07	0.10	19.48	<.0001**	24.09	<.0001**	13.62	<.0001**
Π_{Wt}	no	0.04	0.131	0.06	0.09	7.62	<.0001**	17.47	<.0001**	12.78	<.0001**

Table 9: Determinants of Profits

This table reports estimates for the two measures of profits defined in Table 8. Methodology and variables are identical to those described in Tables 2, 4 and 5. We report the results for estimates based on Heckman correction and 2SLS. All estimates are multiplied by 10. The control variables are the same as the ones the ones used in Table 2, Panel B, augmented by the variance of both labor and entrepreneurial income.

Panel A: Determinants of Π_{Ft}

	All households				***Low-wealth households***				***High-wealth households***			
	I		II		I		II		I		II	
Variable	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*
Γ_1	-0.11	-0.87	0.07	0.06	0.06	1.61	0.07	1.83	-0.03	-0.11	-0.02	-0.82
Γ_2	-0.47	-2.01	0.10	0.41	0.11	1.23	0.10	1.17	0.09	0.66	0.09	0.47
Geographical proximity			0.46	5.60			0.12	6.32			0.30	2.01
Professional proximity			-0.43	-1.86			0.13	0.56			-0.43	-1.31
Holding Period			0.26	3.88			0.04	2.36			0.17	3.94
Labor Income (Variance)	0.32	10.29	0.28	8.64	0.25	8.53	0.23	8.13	0.44	1.87	0.49	1.85
Entrepr. Income (Variance)	0.03	1.79	-0.01	-0.78	0.01	1.32	0.01	0.94	-0.02	-1.21	-0.02	-1.10
Control variables	Yes		Yes		Yes		Yes		Yes		Yes	
Adj R^2	0.01		0.01		0.01		0.01		0.01		0.01	

Panel B: Determinants of Π_{Wt}

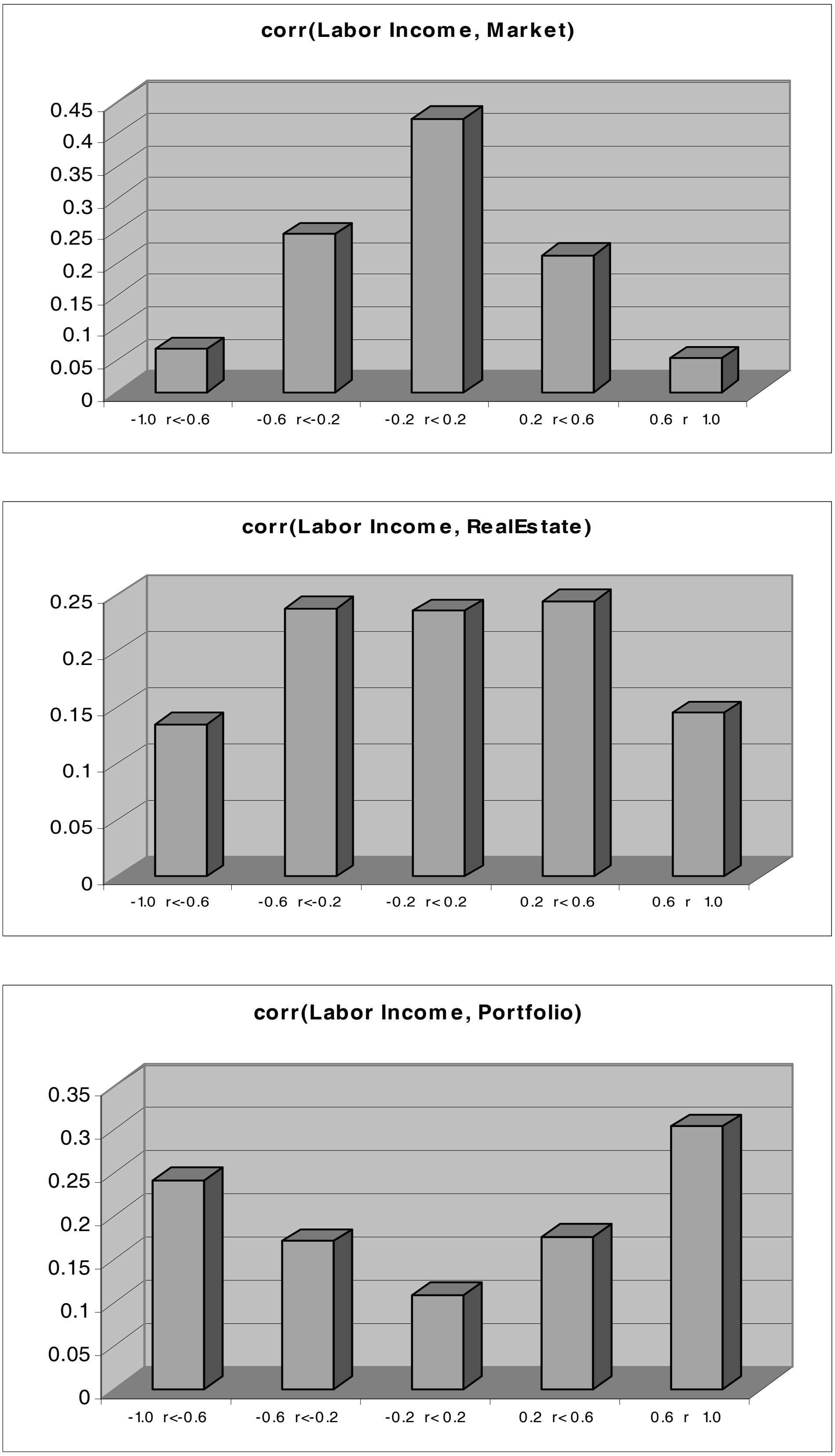
	All households				***Low-wealth households***				***High-wealth households***			
	I		II		I		II		I		II	
Variable	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*	*Estimate*	*t-stat*
Γ_1	-0.11	-3.41	-0.05	-1.43	-0.13	-2.30	-0.05	-0.93	-0.05	-2.36	-0.06	-2.35
Γ_2	0.01	0.21	0.03	0.51	-0.05	-0.45	-0.07	-0.58	-0.03	-0.76	-0.05	-1.02
Geographical proximity			0.47	4.36			0.94	5.41			0.53	5.56
Professional proximity			-0.43	-1.42			-0.74	-1.22			-0.02	-0.13
Holding Period			0.84	4.72			1.04	2.32			1.52	4.04
Labor Income (Variance)	-0.13	-4.25	-0.17	-5.04	-0.06	-1.12	-0.13	-2.49	0.03	1.22	0.11	2.59
Entrepr. Income (Variance)	-0.02	-2.93	-0.02	-3.35	-0.02	-1.86	-0.03	-2.18	-0.02	-3.06	-0.01	-1.91
Control variables	Yes		Yes		Yes		Yes		Yes		Yes	
Adj R^2	0.02		0.02		0.02		0.02		0.10		0.10	

Figure 1: Frequency plots for the correlation between labor income and overall stock market Index (SIX Index), regional real estate indices (Englund *et. al.*, 1998) and return on household portfolio of risky assets (RET_PORT). Percentage of the total sample is along the Y-axis, range of values of the correlation coefficients are along the X-axis. We use five ranges for correlations: from -1 to –0.6, from -0.6 to -0.2, from –0.2 to 0.2, from 0.2 to 0.6 and from 0.6 to 1.

Figure 2: Frequency plots for correlation between entrepreneurial income and overall stock market Index (SIX Index), regional real estate indices (Englund *et. al.*, 1998) and return on household portfolio of risky assets (RET_PORT). Percentage of the total sample is along the Y-axis, range of values of the correlation coefficients are along the X-axis. We use five ranges for correlations: from -1 to –0.6, from -0.6 to -0.2, from –0.2 to 0.2, from 0.2 to 0.6 and from 0.6 to 1.

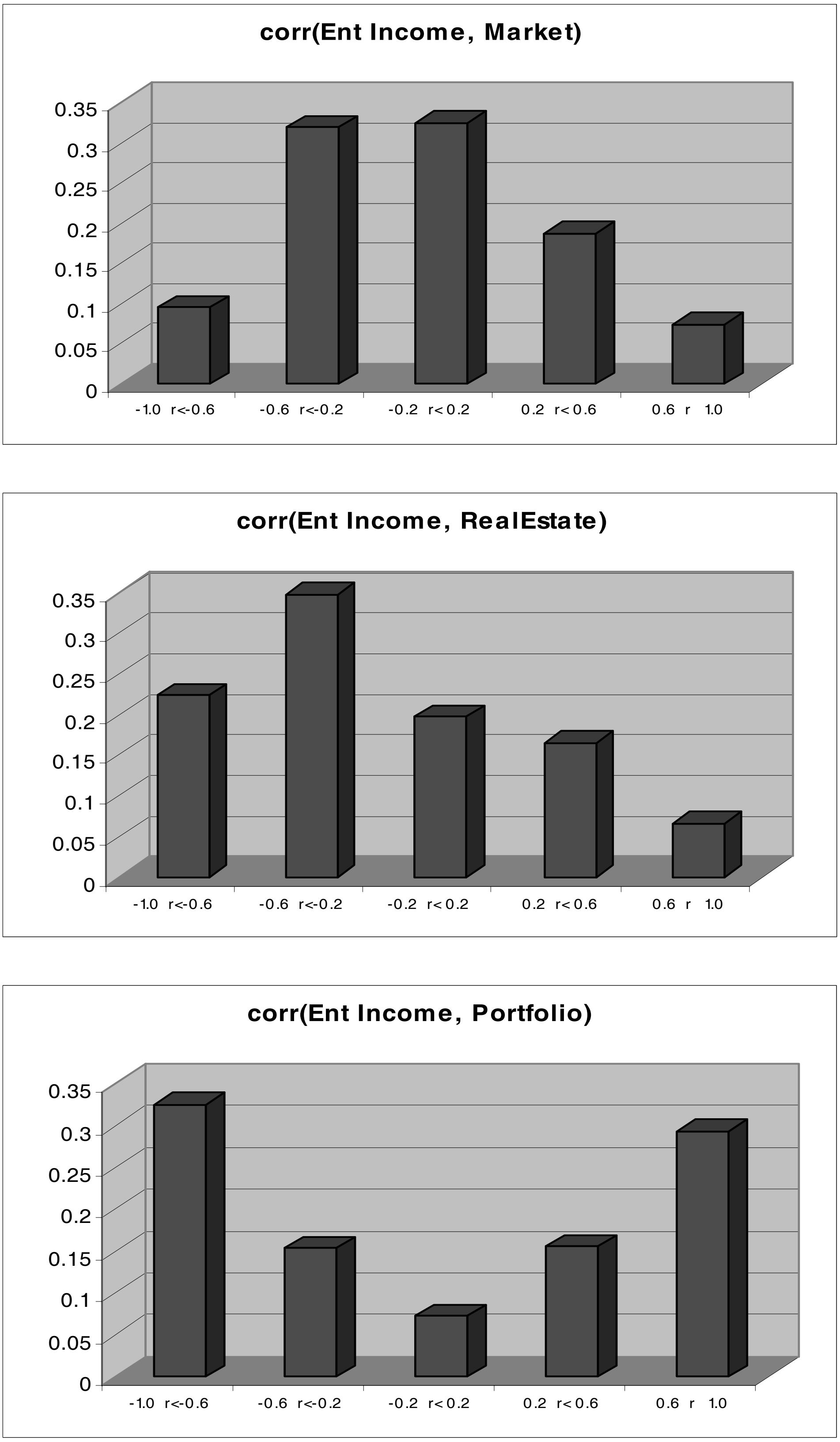
Figure 3: Distribution over index of hedging (Γ_1) *of stock holdings* of investors grouped into the ones who invest in “close” and “distant” stocks (right and left columns, correspondingly). The distributions are reported separately over sub-samples of low-wealth and high-wealth investors (Fig. 3a and 3b, correspondingly) and for overall sample (Fig.3c). We use five ranges for Γ_1 : from -2 to –1.2, from -1.2 to -0.4, from –0.4 to 0.4, from 0.4 to 1.2 and from 1.2 to 2.

Figure 1



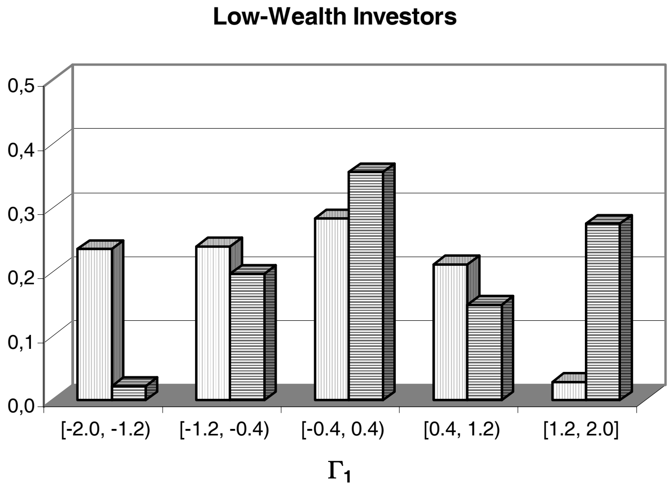
corr(Labor Income, Market)
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-1.0 r<-0.6
-0.6 r<-0.2
-0.2 r<0.2
0.2 r<0.6
0.6 r 1.0
corr(Labor Income, RealEstate)
0.25
0.2
0.15
0.1
0.05
0
-1.0 r<-0.6
-0.6 r<-0.2
-0.2 r<0.2
0.2 r<0.6
0.6 r 1.0
corr(Labor Income, Portfolio)
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-1.0 r<-0.6
-0.6 r<-0.2
-0.2 r<0.2
0.2 r<0.6
0.6 r 1.0

Figure 2

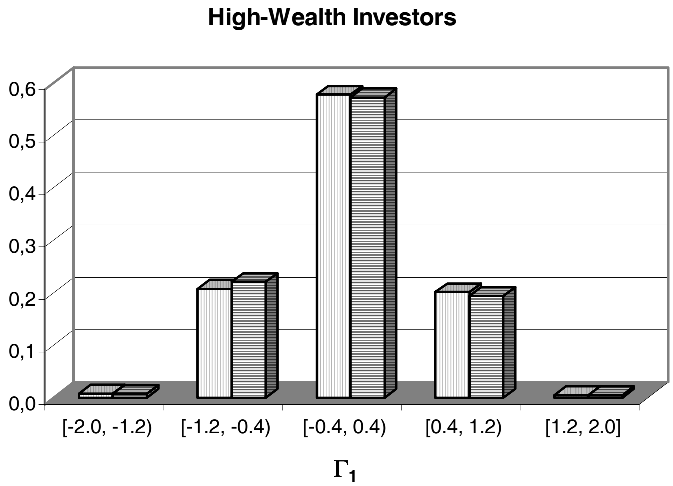


corr(Ent Income, Market)
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-1.0 r<-0.6
-0.6 r<-0.2
-0.2 r<0.2
0.2 r<0.6
0.6 r 1.0
corr(Ent Income, RealEstate)
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-1.0 r<-0.6
-0.6 r<-0.2
-0.2 r<0.2
0.2 r<0.6
0.6 r 1.0
corr(Ent Income, Portfolio)
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-1.0 r<-0.6
-0.6 r<-0.2
-0.2 r<0.2
0.2 r<0.6
0.6 r 1.0

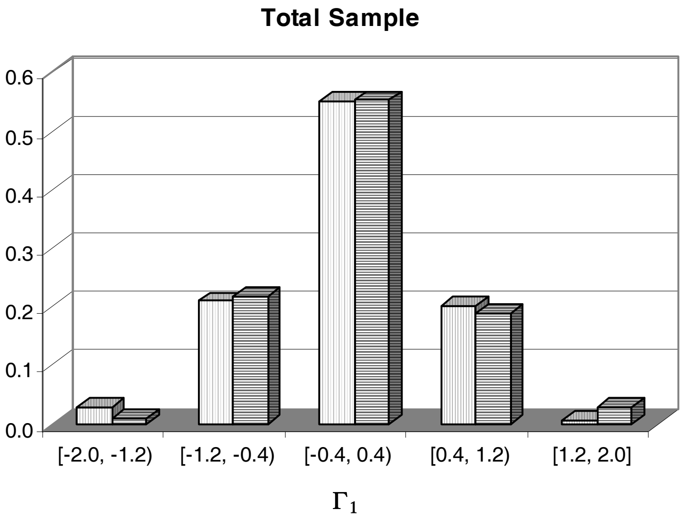
Figure 3



Low-Wealth Investors
0,5
0,4
0,3
0,2
0,1
0,0
[-2.0, -1.2)
[-1.2, -0.4)
[-0.4, 0.4)
[0.4, 1.2)
[1.2, 2.0]
Γ_1



High-Wealth Investors
0,6
0,5
0,4
0,3
0,2
0,1
0,0
[-2.0, -1.2)
[-1.2, -0.4)
[-0.4, 0.4)
[0.4, 1.2)
[1.2, 2.0]
Γ_1



Total Sample
0.6
0.5
0.4
0.3
0.2
0.1
0.0
[-2.0, -1.2)
[-1.2, -0.4)
[-0.4, 0.4)
[0.4, 1.2)
[1.2, 2.0]
Γ_1

SIFR Research Report Series

All reports can be downloaded from our website www.sifr.org, under the heading Research. Reports no. 1-15 are also available in print. In order to obtain copies of printed reports, please send your request to info@sifr.org with detailed ordering information.

1. **Foreigners' Trading and Price Effects Across Firms**
   Magnus Dahlquist and Göran Robertsson, December 2001

2. **Hedging Housing Risk**
   Peter Englund, Min Hwang, and John M. Quigley, December 2001

3. **Winner's Curse in Discriminatory Price Auctions: Evidence from the Norwegian Treasury Bill Auctions**
   Geir Høidal Bjønnes, December 2001

4. **U.S. Exchange Rates and Currency Flows**
   Dagfinn Rime, December 2001

5. **Reputation and Interdealer Trading. A Microstructure Analysis of the Treasury Bond Market**
   Massimo Massa and Andrei Simonov, December 2001

6. **Term Structures in the Office Rental Market in Stockholm**
   Åke Gunnelin and Bo Söderberg, April 2002

7. **What Factors Determine International Real Estate Security Returns?**
   Foort Hamelink and Martin Hoesli, September 2002

8. **Expropriation Risk and Return in Global Equity Markets**
   Ravi Bansal and Magnus Dahlquist, November 2002

9. **The Euro Is Good After All: Corporate Evidence**
   Arturo Bris, Yrjö Koskinen, and Mattias Nilsson, November 2002

10. **Which Investors Fear Expropriation? Evidence from Investors' Stock Picking**
    Mariassunta Giannetti and Andrei Simonov, November 2002

11. **Corporate Governance and the Home Bias**
    Magnus Dahlquist, Lee Pinkowitz, René M. Stulz, and Rohan Williamson, November 2002

12. **Implicit Forward Rents as Predictors of Future Rents**
    Peter Englund, Åke Gunnelin, Martin Hoesli, and Bo Söderberg, November 2002

13. **Accounting Anomalies and Information Uncertainty**
    Jennifer Francis, Ryan LaFond, Per Olsson, and Katherine Schipper, June 2003

14. **Characteristics, Contracts and Actions: Evidence From Venture Capitalist Analyses**
    Steven N. Kaplan and Per Strömberg, June 2003

15. **Valuing Corporate Liabilities**
Jan Ericsson and Joel Reneby, June 2003

16. **Rental Expectations and the Term Structure of Lease Rates**
Eric Clapham and Åke Gunnelin, October 2003

17. **Dealer Behavior and Trading Systems in Foreign Exchange Markets**
Geir Høidal Bjønnes and Dagfinn Rime, December 2003

18. **C-CAPM and the Cross-Section of Sharpe Ratios**
Paul Söderlind, December 2003

19. **Is there Evidence of Pessimism and Doubt in Subjective Distributions? A Comment on Abel**
Paolo Giordani and Paul Söderlind, December 2003

20. **One for the Gain, Three for the Loss**
Anders E. S. Anderson, May 2004

21. **Hedging, Familiarity and Portfolio Choice**
Massimo Massa and Andrei Simonov, May 2004

22. **The Market Pricing of Accruals Quality**
Jennifer Francis, Ryan LaFond, Per Olsson, and Katherine Schipper, May 2004

Stockholm Institute
for Financial Research