NBER WORKING PAPER SERIES

PORTFOLIO DIVERSIFICATION AND CITY AGGLOMERATION

William N. Goetzmann
Massimo Massa
Andrei Simonov

Portfolio Diversification and City Agglomeration
William N. Goetzmann, Massimo Massa, and Andrei Simonov
NBER Working Paper No. 10343
March 2004
JEL No. G11, G14

ABSTRACT

We relate the degree of investor portfolio focus to the broader urban economic context of the household. Using a detailed panel of investors in Sweden over the period 1995 to 2000, we find that the level of investor diversification, as measured by number of stocks in the portfolio and by the average correlation among holdings, is partially explained by city industrial characteristics. We find that rural portfolios are more diversified than urban portfolios and that portfolio diversification is characterized by factors associated with urban growth. We consider a number of theories to explain investor focus, including behavioral biases, real and perceived informational advantage, local social competition and hedging of non-tradable risk. We find little evidence to support social and hedging motives to explain the lack of portfolio diversification, and some evidence in favor of perceived informational advantage in an urban setting. We attribute this evidence as support for the broader "knowledge-spillover" processes documented in the recent urban economics literature. Portfolio effects may be added to the list of factors that define and differentiate urbanism.

William N. Goetzmann
School of Management
Yale University
Box 208200
New Haven, CT 06520-8200
and NBER
william.goetzmann@yale.edu

Massimo Massa
INSEAD

Andrei Simonov
Stockholm School of Economics

Introduction

Urban economics in recent years has taught us a lot about the economic function of cities, from their special capacity as engines of enterprise, to the role they play in stimulating the acquisition of labor skills and knowledge. For example, by bringing firms from a variety of industries together in a single location, cities present a wider array of opportunities for employment, even while attracting a motivated labor force to compete for these jobs. By the same token, by bringing firms within a single industry together, cites have been shown to compound the potential for technological development. Papers over the past decade by Edward Glaeser and co-authors in particular have explored the role that geographical concentration plays in innovation and the transfer of knowledge.[2] Testing whether industrial concentration or diversification are determinants of economic growth, these papers emphasize the benefits of knowledge spillover in an urban setting and the role of density in skill development. For example, Glaeser *et al*. (1992) show that industrial competition rather than concentration is a determining factor in growth, and Glaeser and Maré (2001) find empirical evidence suggesting that the development of human capital within cities is the major vector of growth. Taken together, this work and the work of others in the field suggest that skill development and knowledge transfer may well be the key determinant of urbanism.

The economic influence of cities on both enterprise and human capital is so important that it is natural to ask whether its effects can be documented in the financial portion of the household investment portfolio. More broadly, urbanism itself might be viewed as having an impact on investing. The cross-sectional differences afforded by urban vs. rural locations as well as the cross-sectional differences among cities themselves should allow us to test some basic theories about how and why people invest in financial assets. These theories can be broadly divided into those concerned with risk and return.

Risk

Neo-classical portfolio theory emphasizes the risk-reducing role of the investment portfolio. Ross's (1978) K-fund separation theory, for instance, posits that all investors will hold some combination of diversified investment funds – at least in a frictionless world where all

---

[2] See, for example, Ellison and Glaeser, 1999, Glaeser *et al*. 1992, 1995, 1997 and 2002, Glaeser and Mare, 2001.

1

assets are traded and returns correspond to a quite general set of distributions. With non-traded human capital, and frictions like the high costs of re-location or search costs for selling a home, answers to the household portfolio problem are more complicated. The financial portion of the investment portfolio becomes a potential tool for hedging non-financial income risk.[3]

Besides standard macro and micro economic risks, recent research has also posited a geographically based "social risk." Bakshi and Chen (1996) for example, build an economy in which investors are motivated by the social status of wealth – presumably with respect to their geographical neighbors.[4] DeMarzo, Kaniel and Kremer (2002) term this regional factor in the investment motive "district risk." Thus, cities can be seen not only as a technology for diversifying human capital risk, but also as a medium for stimulating social competition with investment implications. These two effects suggest that the financial diversification of urban investors is likely to be different from that of rural investors. By the same token, under these various hedging stories, cities that have more focused economies might be expected to have investors whose portfolios effectively hedge out the focus of the city economy.

Return

The other side of investment is expected return. People invest in financial assets to make money, not simply to insure themselves against shocks to labor income or other sources of uncertainty. Under the assumption of efficient markets, speculation in individual securities is expected to yield no higher rate of return than provided by a diversified portfolio. However, with asymmetric information, diversification is not necessarily an optimal investment strategy. Recent behavioral studies have suggested that the average speculator in stocks does not profit by trading on his ideas (Barber and Odean, 2000). However, this does not appear to prevent the majority of investors from focusing their stock portfolios into only a few, correlated securities (cf. Goetzmann and Kumar, 2002).

One potential reason for this portfolio concentration is the existence of actual or perceived informational advantage. Given that knowledge spillover is a major characteristic of cities, one might expect to find portfolio concentration in urban, rather than rural settings.

---

[3] Recent work on the financial portfolio as hedging instrument includes, DeMarzo, Kaniel and Kremer, 2002, Heaton and Lucas, 2000, a&b, and Vissing-Jorgensen, 2002a, as well as a broader literature on precautionary savings, including Carroll and Samwick, 1997 and Carroll and Kimball, 2001.

[4] This is similar to the idea of "keeping up with the Joneses" (Ryder and Heal, 1973 and Gali, 1994).

Because much business is still largely conducted through personal interaction, cities continue to be settings for the creation and propagation of value-relevant information and skills. This is not to say that knowledge is not created in rural settings, but density of enterprise and a critical level of personal interaction are needed for diffusion and utilization of that knowledge.   Thus, investment opportunities may be expected to appear more frequently in an urban environment, and to yield economic returns to those who seize them.   In   this   context,   the   investment portfolio might be thought of as capital for exploiting short-lived investment ideas. A caveat to "don't put all of your eggs in one basket" is that you may know something special about one particular basket. Hence, we might expect urban households with access to knowledge spillover to use their investment portfolio to take advantage of this knowledge.

It is important to note that, for most investors, these investment opportunities created in an "information-dense" business environment might be illusory. Investors with access to business information flow about a particular company may not know whether the current security price already reflects this new information or not. In fact, they may be subject to behavioral biases such as over-confidence or mistaken self-attribution that would lead them to believe incorrectly that they have superior investment knowledge. Thus, unless investment activity is completely characterized by unbiased, rational expectations about returns, we may not find that urban investors perform better on average than rural investors. Since most people believe themselves to have above average skill, we might expect to find more focused, speculative portfolios among investors regularly exposed to information – whether they are truly mediocre investors or not.

Another feature we might expect from informed -- or seemingly informed -- trading on urban knowledge spill-over is investment in local industry. Zhu (2002) documents the propensity of speculators to trade in stocks close to where they live and finds little evidence that trading in localized shares yields higher profits. Academic interest in geographically proximate investing has been growing, however. Coval and Moskowitz (2001) find evidence of positive returns to institutional trading close to home.   Ivkovic and Weisbrenner (2003) report that investor holdings in local stocks outperform distant holdings. The debate about whether local investors actually profit from geographical information diffusion is still not settled, however, most researchers agree that investors at least   tend to behave as if they   believed they had superior information about local firms.

Contribution of the Current Work

A major challenge to exploring the relationship between investment portfolios and geographical location has been the lack of comprehensive and detailed data. Do investors hedge human capital risks with financial investments? Are urban portfolios more diversified than rural portfolios? These are empirical questions that rely upon data which, to date, have been difficult to obtain. Testing requires not only a detailed knowledge of the assets in the financial portfolio but also information about employment, housing, geographical location and even demographic characteristics. If people are optimizing over all the relevant economic factors in their lives, how would we know without having an embarrassing amount of personal detail to study?

In fact, in this paper we have access to a very detailed geo-coded panel dataset over the period 1995 to 2000 for a large sample of households in Sweden. Constructed from several sources, this dataset allows us to explore the geographical determinants of investment. We seek to address the question of how cities affect household risk and return, and how these effects are manifested in the choice of a diversified vs. a focused investment portfolio. We use the panel data to test a number of implications about both the risk and return stories presented above. Our findings suggest that, in general, city investors are more focused (less-diversified) and that the tendency to focus the portfolio increases with the degree to which the city itself is industry-focused. Our results are consistent with the hypothesis of real or perceived local knowledge spillover. We test for "district effects" by considering whether portfolio focus depends upon a city's relative economic prosperity, and find little support for this theory. On balance, our findings suggest that the composition of investor stock portfolios does not reflect hedging motives, but rather reflects expectations about superior information-motivated return.

Our results provide some potentially useful evidence on theories of agglomeration. Cities appear to enhance risk-taking behavior based on the economic opportunities they offer – or appear to offer -- their inhabitants. Moreover, we find that the very factors economists have associated with urban growth -- particularly industry scale, diversity and concentration, are associated with investor risk-taking.

The remainder of the paper is structured as follows. In the next section, we discuss the literature related to our work with particular focus on the various theories that might explain

investor focus, and the testable implications given our dataset.   Section 3 discusses the data, Section 4 describes the construction of the variables. Section 5 describes the estimation procedure. Section 6 reports the empirical results. A brief conclusion follows.

2. Background and Testable Hypotheses

2.1 Some alternative approaches to portfolio under-diversification

One rational explanation for apparently under-diversified investor portfolios is a hedging argument.   Merton (1971) for example, shows that, in the presence of non-financial income risk investors will hold the market portfolio and some additional position constructed to hedge the changes in the stochastic opportunity set or other sources of uncertainty. For example, an investor subject to income risk may try to diversify it away by holding, in addition to the market portfolio, a portfolio negatively correlated to his labor income.   The net overall financial portfolio may appear undiversified, when in fact the investor has reduced the overall risk of his portfolio.   The hedging motive predicts that   under-diversification should be related to non-investment income risks. We are able to test this with the current dataset.

A second explanation for portfolio focus is   motivated by asymmetric information. Investors may hold focused portfolios because they either possess value-relevant information, or believe they possess value relevant information. Superior information about a few securities implies a portfolio concentrated in those securities (Grinold and Kahn, 1999). Even the mistaken belief by the investor that information in his possession can provide superior returns will motivate a deviation from a diversified portfolio.

A third explanation relies on prospect theory and, in general, on the preference for skewness. Shore and White (2002) and Polkovnichenko (2003) point out those investors with a taste for low-probability high-stakes gambles will invest only in the very few stocks from which they expect higher returns, forfeiting the benefits of diversification in return for a chance at great wealth. Perhaps the investment accounts of individuals in our sample are, in effect, gambling accounts which lose their attractiveness to investors once they become too diversified to provide returns well in excess of the market.   Strange as the suggestion may seem, this might be rational for some form of utility function.

A fourth explanation for under-diversification is bounded rationality. Goetzmann and Kumar (2002) find that less sophisticated investors are also less diversified. Barber and Odean (2003) find that investors tend to trade more frequently in stocks in the news – suggesting either non-trivial information costs to investment research, or uncritical use of the current news. The direct implication of bounded-rationality theories is a strong correlation between familiarity and investment in stocks (Merton, 1987 and Shapiro, 2000). In fact, Grinblatt and Keloharju (2001) document just such a familiarity bias in Finland using a similarly geo-coded investor dataset. If investor interest drops off with the distance fro the company, individual portfolios will tend to be regionally focused, and thus exposed to more volatility than they should be. This is a question we can put to the data.

A fifth explanation is a theory of social competition. The theory of district risk argues that investors -- competing for local resources within their district -- have utility that depends on both their own wealth as well as the aggregate district wealth (DeMarzo, Kaniel and Kremer, 2002). In their model, investors, in order to keep up with their neighbors, invest in the same stocks in which their neighbors invest, resulting in an un-diversified portfolio.[5] The district risk theory postulates that the higher the growth of wealth of the district, the higher the local (consumption) prices, the more investors would concentrate on the same stocks. This suggests a strong positive correlation between the growth and change in wealth of the local district and the degree of under-diversification. It is important to note that this theory assumes away the mere possibility of migration across districts, while we will argue that this is one of the main components of the story. Our approach to testing this is to examine the correlations between the economic performance of the district and the degree of diversification of the investors within it.

It is important to notice 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. Moreover, in general professional proximity (that is the degree to which a stock is close to the profession of the investor) does not induce investors to tilt their portfolio toward stocks of the company they work for or of companies belonging to the same industry. This suggests that investor loyalty and own company ownership do not provide a valid alternative explanation.

---

[5] This argument does not quite explain focused portfolios, because one equilibrium is that all investors in a district might hold the market. However, if keeping up with the Joneses depends upon seeking to keep up with the most successful investor in the district, the extreme portfolio in a population is *de facto* undiversified.

2.2 The city agglomeration approach

As emphasized above, the economists have provided considerable evidence suggesting that knowledge creation and knowledge spillover are the main determinants of city agglomeration and development (Glaeser, *et al*., 1992, Glaeser, *et al*., 2002, Glaeser and Mare, 2001, Ellison and Glaeser, 1997, Glaeser, *et al.,* 1995). Knowledge creation and spillover are defined in terms of professional specialization and proximity. Agglomeration economies, characterized by dynamic local externalities, increase the returns to investment in high labor-capital intensive technologies and induce specialization.

A major question addressed by urban economists is whether intra-industry or inter-industry spillover is more important. Glaeser *et al.* (1992) characterize the former as a Marshall, Arrow and Romer [MAR] model. Along these lines, Porter (1990) argues that the main determinant of growth is specialization. Local monopoly increases growth as it allows externalities to be internalized. On the other hand, Jacobs (1969) links city growth to the existence of various and proximate industries, and in tests of these off-setting propositions, Glaeser *et al.,* 1992 find that "important knowledge transfers come from outside the core industry". They find that cross-industry spillover as opposed to within-industry spillover is the main driver to city growth.

The archtypal MAR region is Silicon Valley. Concentration of the high-tech industry around San Jose, California generates knowledge spillover between firms in the same industry. Jacobs's model represents the development of industries that grow out of specialization of a particular process. The example he gives is the brassiere industry, which evolved from the dressmakers' industry as opposed to the lingerie industry. An additional feature of the Jacobs model is that it suggests that local competition induces knowledge creation and innovation. The given example is the Italian ceramics industry, in which technical and aesthetic innovation is driven by intense, local competition among many similar firms.

These theories disagree in terms of the form of the market most conducive to knowledge creation and specialization (i.e., competition as opposed to local monopoly). However, they do agree on relating growth to the process of knowledge creation and specialization. While existing theories of urban economics focus on development of skills and stimulus for technological innovation, they apply equally well to the propagation of investment opportunities and ideas.

Workers in both diversified and concentrated urban settings are likely to be exposed to value-relevant information. If this information is about publicly traded firms, then we would expect knowledge spillover to stimulate investment activity. Whether this spill-over results in a portfolio with many or few stocks is somewhat ambiguous. If the city employment is concentrated in a few major firms, we might expect to find local investor portfolios highly concentrated. However, if a city has a number of publicly traded companies and there is knowledge transfer among them, then local speculators might take a number of positions at once.

2.3 A simple model of specialization

An important consideration in the question of portfolio focus and knowledge spill-over is that employment specialization itself has the potential to affect investment behavior. Investment activities compete for the attention of the worker. As specialization increases, and labor market competition rises, the time investors allocate to search for new stocks may decrease, even as the value-relevance – or apparent value-relevance – of acquired information increases. In a MAR setting, for example, in which an investor acquires knowledge of the particular sector/industry, this implies a lower search cost for the stocks of that industry, but a higher information search cost for stocks outside of the industry.

To examine this issue in a bit more detail, we consider a search framework (Carlson and McAfee, 1983, Hortacsu and Syverson, 2002). We assume that investors have a certain wealth (*W*) to invest and a search cost (*x*). The search cost may be alternatively seen as a discrete amount that has to be paid to enter a specific market.[6] Investors are risk averse and are aware that different stocks have different risk-return profiles, but are uncertain about the asset that represents the best next addition to the portfolio. Each stock has a specific risk-return profile (*R*) that is a function of the stock expected return, correlation and volatility. If we also assume that investors have identical perceptions of the risk-returns profiles of the different stocks, it can be shown (Hortacsu and Syverson, 2002) that an investor with *W* to invest and a search cost x will adopt the following investment rule: keep on investigating one additional stock if the cost of searching is less than the benefit, that is,

$$\int_0^{R^*} \left(\left[EU\left(R^*, W\right) \right] - \left[EU\left(R^i, W\right) \right] \right) dF > x \qquad (1)$$

---

[6] See, for example, Ukhov (2002) for a model with a fixed investor cost of market entry.

otherwise stop and invest in that stock. The intuition behind equation 1 is the existence of discreteness in the cost of investment that effectively segments the market and generates a discontinuity in the investment process. This approach assumes that the degree of portfolio diversification is mainly a function of the search cost to invest in a specific stock. Equation 1 provides a stylized representation that contains all the elements of portfolio choice, but it is completely unlike the neo-classical portfolio problem, in that past choices are not re-evaluated in the context of current ones. It has the feature of habit-formation in which the existing portfolio is the acquired habit.

Investment in stocks provides three advantages: the first is that a search may turn up a genuine, undervalued security. While insider trading laws may prevent the exploitation of this information, trading in the stocks of related companies exposed to such information may not be prohibited (Tookes, 2003). The second is the reduction of the overall portfolio risk due to diversification. Therefore, if the investor is risk averse, he has an incentive to keep on searching and not to invest only in a single stock. A risk neutral investor will only care about the return payoff and will concentrate on one stock – it is well-known that the maximum expected return portfolio is a composed of a single security, however it is costly to continue to search for that highest-returning security.

What is exactly the search cost? It is a function of the actual purchase of information as well as the time spent doing so. The latter increases with alternative opportunities the investor faces and, therefore, with the reward/opportunity ratio provided by his professional activity. Therefore, the more professionally involved the investor is, the less time he will have to devote to portfolio allocation. We can think of this in terms of bounded rationality (Simon, 1986, Sargent, 1993) or limited processing capacity (Sims, 2000). Building on Sims' results in terms of limited processing capacity, Peng and Xiong (2002) show how limited capacity, defined in terms of time to process information, affects portfolio choice.

The search cost is negatively related to the investor's ability to get information. Therefore, in a business setting in which an investor is naturally exposed to certain kind of information, professional, geographical and informational proximity are by-products of labor and thus reduce the search cost. Even in a situation in which there is no excess return expected from an investment, if the search cost is discrete and high enough, an investor will forfeit the benefits of broader diversification and hold the few stocks related to his job – in a sense, free-riding on

professional knowledge spillover. It thus follows that the degree of portfolio diversification is a direct function of the search cost. This simple model is consistent with existing empirical evidence showing that investors tend to invest in stocks of companies headquartered close to where they live (Coval and Moskowitz, 1999, 2001, Hau, 2001) or of the country they come from (Bhattacharya and Groznik, 2001). Proximity may be seen as a proxy for lower search costs.

The main implication of this simple model is to suggest portfolio diversification may be related to an investor's professional life. This allows us to link portfolio diversification to the process of city agglomeration and industrial specialization. We find evidence that the same factors that drive agglomeration are those that affect the professional lives of investors, and by extension, their search cost.

3 The data

In order to examine the relationship between urbanism and diversification, we require a set of variables that proxy for knowledge spillover and local prosperity, as well as a set of variables to test competing theories. To test hypotheses about hedging, we require a measure of non-financial income risk by the investor, measures of the level and risk of non-financial income, as well as proxies for investor borrowing constraints. Theories related to information spillover and bounded rationality require proxies for the degree of professional and geographical proximity of the investor. The district risk theory requires a measure of local prosperity as well as investor specific wealth and sources of income. Finally, a fairly general set of individual investor control variables are needed. These will allow us to separately identify the impact of belonging to a particular district from the investor's specific characteristics.

We collect data from different sources. For each investor we have detailed information about his or her individual holdings of stocks (broken down at the stock level), mutual funds, bank accounts, real estate and other types of wealth. Fiscal authorities provide us with information on the different sources of investor income, as well as demographic and family characteristics. This information is 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. We now explain the data sources more in detail.

3.1 Individual stockholding

We use the data on individual shareholders collected by Vardepapperscentralen (VPC), the Security Register Center. The data contain both stockholding held directly and in a street name, including holdings of US-listed ADRs. In addition, SIS Ägarservice AB collects information on ultimate owners of shares held via trusts, foreign holding companies and alike (for details see Sundin and Sundquist, 2002). Our data cover the period 1995-2000. Overall, the records provide information about the owners of 98% of the market capitalization of publicly traded Swedish companies. For the median company, we have information about 97.9% of the equity, and in the worst case we have information on 81.6% of market capitalization of the company. The data provided by SIS Ägarservice AB were linked by Statistics Sweden with the LINDA dataset described below.

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, representative panel of households for the population over the period 1960 to 2000. For each year, information on all family members of the sampled individuals is added to the data set. 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 which, in general, is lacking in surveys based on different cohorts polled over time.

The variables include individual characteristics (gender, 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

11

contributions, taxes paid and taxable wealth. A detailed description of the dataset is provided by Edin and Fredriksson, (2000) and is available on the web site http://linda.nek.uu.se/.

The tax aspect deserves more detailed discussion. In Sweden, in addition to 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 the tax-assessed value 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-assessed value provided by LINDA. We evaluate it at current prices by using the average ratio of market value to tax-assessed value that is provided for each year and county by the Swedish Office of Statistics. There is no estimate of the market value of privately held companies. However, the data contains an indicator variable for owners of privately held companies and entrepreneurs who file their business tax return along with their personal tax return. For the privately held unlimited liability companies the value of the assets is included in the tax return. For the privately held unlimited liability companies that are not listed, the value of assets held is generally missing. However, 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 we use contains 1,757,406 observations. In addition, we also use 1990-1994 data from LINDA in the implementation of the Carroll 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 for each year, members in household, adults in household, age of the oldest member of household, percent of the sample with secondary and higher education, percent of immigrants) as well the proxies for diversification (D_1, D_2, D_3 and D_4). Panels B report the characteristics of the local districts in terms of the main variables we will focus on (i.e., number of active enterprises, competition, specialization, diversity,...). Panels C and D report, respectively, the age and gender distribution of the sample and their wealth and income characteristics.

3.3 Firm-level information and other data

For individual security returns (including dividends) and the overall market index (SIX market index), we use the SIX Trust Database. For information on 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 municipality location of the plant. The consumer confidence index is provided by Statistics Sweden. Geographical coordinates are supplied by the Swedish Postal Service and contain latitude and longitude of Swedish Postal Offices.

4 Construction of variables

4.1 Measures of portfolio diversification

We consider four measures of portfolio diversification. These are derived from Goetzmann and Kumar (2002). We refer to their paper for a proper description and the rationale of their use. The first measure (D_1) is simply the number of positions in the portfolio (i.e., $D_1 = N$). The second (D_2) is constructed as:

$$D_2 = -\left[\frac{1}{N} + (1 - \frac{1}{N}) * \overline{Corr}\right]$$

where N is the number of positions and $\overline{Corr}$ is the average correlation of the stocks in the portfolio. It can be shown that D_2 is just a normalized version of the portfolio variance multiplied by –1. It increases with the degree of diversification of the financial portfolio. Diversification increases when the variance is reduced. The variance can be reduced either by increasing the number of shares in the portfolio (N), or by selecting the stocks so as to reduce the average correlation among the stocks in the portfolio ($\overline{Corr}$). In the limit as N goes to infinity the portfolio variance converges to the covariance among the stocks in the portfolio.

The third and fourth measures (D_3 and D_4) are constructed as:

$$D_3 = -\sum_{i=1}^{N}\left(w_i - w_{mkt_3}\right)^2, \quad \text{and} \quad D_4 = -\sum_{i=1}^{N}\left(w_i - w_{mkt_4}\right)^2 \quad \text{for each } i = 1, ..., N,$$

where w_i is the weight of the stock in the portfolio of the investor and w_{mkt} is the weight that the same stock would have in the market portfolio. They differ in the way w_{mkt} is constructed. In the case of D_3, w_{mkt3} is constructed by using the overall capitalization of the company, while in the case of D_4, w_{mkt4} is constructed by using just the free float.[7] These measures express diversification in terms of divergence of the financial portfolio from the market portfolio of the investor. Higher diversification implies a divergence closer to zero. Both D_3, and D_4 are defined so that as diversification decreases, the measures become more negative. They will be very useful in terms of comparison with our measure of active hedging of non-financial income risk. We will indeed see how the distance from the market portfolio relates to the desire to hedge non-financial income risk. All our measures of portfolio versification increase with the degree of diversification of the portfolio that is "a higher value of these variables is indicative of a higher level of diversification".

4.2 Measures of professional specialization

To proxy for professional specialization, we use the measures similar to those developed by Glaeser *et al*. (1992): specialization, competition, uniformity and diversity. We create a specialization variable as ratio of the share of the top five industries in local employment to the share of these same five industries in national employment. Thus, the denominator of this variable differs across municipalities. This provides a measure of specialization at the municipality level. Competition is captured by the number of firms per employee in a municipality relative to the number of firms per employee in Sweden. As Glaeser *et al*. (1992) point out, this is not distinguishable from a relative scale variable, but should be correlated to the degree of firm competition for employees. Our measure of diversity is somewhat different from that used by Glaeser *et al*. (1992), since their study is conducted on industry-level data. We proxy for diversity using the negative of the share of the top 5 industries in municipal employment. Cites with more diversity in employment beyond the top five industries by employment will have a higher value. Glaeser *et al*. (1992), using a sample of US cities, between 1956 and 1987, find that specialization and diversity reduce city growth, while

---

[7] This is all the more relevant in a country like Sweden where a sizable fraction of the companies' shares are in the hands of a restricted number of shareholders.

competition increases it. In order to assess how these variables affect city growth in our sample, we estimate an analogous regression[8] for Sweden. Our estimation of this relationship for Sweden (not reported) suggests that all three variables are associated with the growth rate of the city.

We also construct variables that proxy for the degree of "isolation" of the local district. These are meant to describe how isolated the district is with respect to major financial and information centers. These measures are: the index of rural areas, the distance from the closest civilian airport and the population density. The index of rural areas is the Urban Code as reported by Statistics Sweden. It ranges from 1 in the case of a Metropolitan area to 9 in the case of the countryside. The distance to the closest civilian airport is measured as the logarithm of the distance between airport and central post office in the district. Population density is the population (in tens of thousands) per square kilometer.

4.3 Measures of local prosperity

We consider six variables that measure the degree of prosperity of the district in a given year: the percentage of new start-ups, the percentage of bankruptcies, the number of enterprises active in the district, the employment growth and some measures of profitability of the companies operating in the district (i.e., the profitability of sales and the return on capital). These variables are constructed at the firm level and aggregated at the district level. As an alternative measure of local prosperity we also considered local consumption price indexes. This would be the variable required by the district risk hypothesis. However, this variable is very highly correlated to the other variables of local prosperity. We therefore omit it in the reported specification.

4.4 Measures of financial and municipal services availability

The degree of financial sophistication and bank coverage of a district provide a good proxy for both the availability of financial information and the vitality of the local economy (Garmaise and Moskowitz, 2003). In order to proxy for the availability of financial services, we include the degree of bank coverage. It is constructed as follows. We first identify the list of

---

[8] While for Glaeser et al. (1992) the unit of measure is the industry-level data, we focus on municipality-level data. These results, that replicate Table 3 in Glaeser *et al*. (1992) using this alternative specification are available upon request from the authors

credit institutions as reported by the Finance Inspection (Swedish equivalent of SEC) and then we calculate, for each municipality, the number of branches that each institution has for a given year. The resulting variable "Bank Coverage" is the logarithm of 1 + the number of branches per district. To proxy for the degree of availability of municipal services, we use the percentage of the population that is employed by the municipality.

4.5 Non-financial risk variables.

We define as non-financial risk variables those that allow us to test for portfolio choice in the presence of non-financial income risk (cf. Heaton and Lucas, 2000). We consider three variables. The first variable an the index of investor hedging (Γ_i). It measures the extent to which the investor's portfolio differs from the market portfolio in terms of correlation with investor's non-financial risk. It is constructed as:

$$\Gamma_i = corr(Y_i, r_m) - corr(Y_i, r_{i,port}),$$

where r_m is the return on the market portfolio, $r_{i,port}$ is the return on the financial portfolio of the *ith* investor and Y_i is the investor non-financial expected income. In the Appendix we provide a detailed description of how this variable is constructed. Γ_i proxies for the change in correlation between financial and non-financial risk induced by investor's portfolio choice and quantifies the extent to which the investor deviates from a passive strategy. It is positive in the case of active hedging and captures the contribution of portfolio choice to the reduction of the overall investor's risk.

The second variable is a measure of borrowing constraints. Standard portfolio theory links portfolio choice and market participation to the existence of borrowing constraints. Our proxy for these is the ratio of investor debt to total assets. It is constructed at the investor level at time *t*. The third proxy for non-financial risk is the percentage of the population in the district that is enrolled in welfare programs. It proxies for the risk of unemployment that comes from the fact of living in a depressed area. It is worth stressing that this is above and beyond the risk of unemployment that each individual investor perceives. The latter is constructed for each investor and included among the control variables (see below).

4.6 Measures of familiarity

The index of familiarity (Ψ_i) represents the degree of "professional" or "geographical proximity" to a particular stock. In the case of professional proximity, the index 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, for an investor working in the mining sector who holds the stock of a mining firm, the dummy would be equal to 1.

In the case of geographical proximity, we use the proximity between the residence of the investor and the place where the company is located. We consider 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 headquarter. Given that the results do not differ and the variables are highly collinear, we report only the first specification. These measures are analogous to those proposed by Coval and Moskowitz (1999, 2001) in a study of geographical preferences in mutual fund investment. The greater the value of the variable, the closer the investor is to the stock. These measures are constructed at the stock level and then aggregated at the investor level, across all the stocks of his portfolio, weighting them by their share in the portfolio. This procedure delivers three measures of familiarity for each investor and time.

4.7 Control variables

We consider the following sets of control variables: measures of income and wealth, contemporaneous gain/loss variables, demographic variables, momentum variables, and macroeconomic and social variables.

The measures of wealth include the overall level of wealth of the investor and a decomposition by components. Overall wealth is defined as the sum of financial and real estate wealth. The measures of income variance include the variance of labor and entrepreneurial income of the investor and the correlations between them and financial and real estate income. In order to make the results comparable with the standard literature on portfolio choice in the presence of non-financial income risk, we construct measures of the permanent (expected) non-

financial income following the approach of Carrol and Samwick (1997) and Vissing-Jorgensen (2001). In the Appendix we provide a brief description of the methodology. We consider non-financial income to be labor income and entrepreneurial income. We also construct 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, geographic and professional dummies.

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

The momentum variables include the return of the portfolio of the investor and of the market portfolio in the previous 12 months. These variables are meant to control for the possibility that the change in the degree of portfolio diversification or portfolio choice (i.e., familiarity bias) is due to momentum, that is, changes to the variation in the value of the stock market or in the value of the portfolio holdings.

The demographic variables include: the level of education of the investor, 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-Jorgensen, 2002) which find a non-linear relationship between age and the degree of stock market participation. We also include 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.

To control for the effect of the local economy, we also include a Stockholm and an immigration dummy. The Stockholm dummy takes the value of 1 if the investor lives in the capital and 0 otherwise. The immigration dummy takes the value 0 if all the members of the household are native Swedes, and 1 if at least one member of household immigrated from

abroad. Furthermore, we construct a variable to proxy for the ability of the investor in his occupation. This is based on the difference between his income and the average income of his profession. The assumption is that the higher the income of the investor relative to the average income of the other investors in the same area, the higher his ability should be.

We also consider macroeconomic and social variables. The macroeconomic variable is the Index of Consumer Confidence. The social variables include the local tax rate and the percentage of foreign-born households. The local tax rate controls for disparities across districts merely due to tax treatment. The percentage of foreign-born in the population of the district controls for the assistances provided to the immigrants. It can be particularly relevant in some areas of the country. We also include a variable that reports the number of injured in auto accidents per km of the roads, as reported by the Ministry of Industry, Employment and Communications. This allows us to control for some outside background risk.

5 Econometric methodology

We concentrate on the following specification:

$$Y_{it} = a + \beta A_{it} + ? C_{it}^{Y} + e_{it}, \qquad (2)$$

where Y_{it} is alternatively one of our measures of portfolio diversification or one of our proxies of familiarity (i.e., professional and geographical proximity). The matrix A_{it} contains the sets of variables we are interested in (i.e., proxies of professional specialization, local prosperity, local welfare assistance and degree of isolation). The matrix C_{it}^{Y} contains all the control variables (i.e., income, wealth, momentum, demographic and macroeconomic variables).

The econometric estimation of equation (2) has to account for the selection bias due to the fact that we do not observe the investment decision of investors who do not participate in the financial market. Given that the participation decision is endogenous, the standard estimates of equation (2) would be biased (Maddala, 1983, Nijman and Verbeek, 1996). To address this issue we use a Heckman (1979) two-stage procedure and separately estimate the factors associated with holding stocks and what influences his choices of assets. The decision to enter the market can be represented as:

$$P_{it} = a + \beta A_{it} + ? C_{it}^{P} + e_{it}, \qquad (3)$$

where P_{it} is a dummy that takes the value of 1 if the investor participates in the financial market and zero otherwise, $\mathbf{A_{it}}$ is defined as before, while $\mathbf{C_{it}^P}$ contains the vector of control variables.[9] The probability that the investor enters the financial market (P_{it}) is modeled as a normal c.d.f., defined on an expanded dataset that contains both households who hold financial assets and households who do not. The expanded dataset includes the totality of the households tracked over time over each of the sample years 1995 through 2000, regardless of whether they invested in the stock market. It totals 1,757,406 households-year observations.

From the estimation of equation (3) we derive a variable (λ_{it}) that is employed in the second stage to control for the selection bias (see Heckman, 1979). The significance of the estimate of λ_{it} provides a test of the null of no sample selection bias. The results show that in all the specifications λ_{it} is always strongly significant, suggesting that self-selection is indeed important in the sample. We will therefore estimate:

$$Y_{it} = \mathbf{a} + \mathbf{\beta A_{it}} + \mathbf{?C_{it}^Y} + \mathbf{dl}_{it} + \mathbf{e}_{it}, \qquad (4)$$

Given that equation (3) is just an auxiliary regression only needed for the proper estimation of the second stage, but out of the scope of this paper, for brevity we will not report the results and we will focus on specification (4). A few results are, however, worth mentioning. Professional specialization always increases stock market participation. This result holds across all the different specifications and for both the low wealth and the high wealth investors. Local isolation reduces stock market participation, while local prosperity has an ambiguous impact. Higher company profitability reduces stock market participation, while an increase in the number of companies raises stock market participation. This is consistent with the fact that if the industrial area is thriving, local investors are more likely to invest in their own business and to start their own enterprises and ventures than to invest in the stock market. On the contrary, an increase in the number of companies, by providing better job opportunities and higher

---

[9] $\mathbf{C_{it}^P}$ differs from $\mathbf{C_{it}^Y}$ only in terms of some variables that provide the identification restriction in the Heckman specification. $\mathbf{C_{it}^P}$ also contains time dummies, macro-regions, industry dummies and the correlations between non-financial income and the market portfolio. That is, the correlations between labor income and entrepreneurial labor income and the market portfolio and the correlation between the investor's real estate and the market portfolio. $\mathbf{C_{it}^Y}$ also contains the prior 12 month returns and volatility of the investor's portfolio, his prior capital gains and losses separately considered, and his tax rate. The main implicit assumption is that the participation decision is a function of the market portfolio (i.e., its correlations), while the portfolio decision is a function of the individual stocks.

employment raises stock market participation. This is controlling for individual wealth (both financial and real estate) and for the level and volatility of income (labor and entrepreneurial).

Considering the potential endogeneity of some of the explanatory variables, we use an instrumental variable estimation.[10] In order to control for heteroskedasticity, we correct the standard errors in the second stage regression. We therefore use a two stage least squares with a consistent variance-covariance matrix. We employ data disaggregated at the individual investor level.

6. Main results

We proceed in two steps. First, we consider the determinants of portfolio diversification. This allows us to directly compare evidence in favor of the different theories described above. This first part of the analysis, while it addresses the question of *why* investors might hold few stocks in their portfolio, however, does not explain *which* stocks they invest in. Therefore, the second part of the analysis addresses the issue of stock-selection. That is, we relate the drive to invest in local or familiar stocks to the degree of professional specialization and city agglomeration.

6.1 Determinants of portfolio under-diversification

We regress our measures of portfolio diversification on the proxies for local professional specialization, local prosperity, financial and municipal service availability and the non-financial risk variables, as well as a set of control variables meant to control for investors' idiosyncratic wealth, income and demographic characteristics and for geographic and country shocks. The results are reported in Tables 2-6. In particular, Table 2 contains the main specification for the 4 measures of portfolio diversification (respectively for the different measures of portfolio diversification D_1, D_2, D_3, and D_4) for the entire sample. Tables 3-6 report the main specifications for the different measures of portfolio diversification, with the sample broken down into low-wealth and wealthy investors. For each class of investors we consider four alternative specifications differing on the basis of the control variables that are used. We will mostly focus on the complete specification with all controls -- Specification 1 in the tables. The

---

[10] Lagged variables and a set of demographic variables, industry and time dummies are used as instruments.

results broadly support the city agglomeration theory and the limited information theory, while they fail to support the standard portfolio theory and the district risk theory.

The first finding is a significant negative correlation between portfolio diversification and our urban measures of professional specialization. This holds overall and for the different classes of investors. Industrial specialization and competition negatively relate to portfolio diversification. That is, the more concentrated in a particular area the industry is, the higher the degree of local competition for employees and the higher the labor concentration, the lower is the degree of portfolio diversification. This result is robust across all specifications and for different measures of portfolio diversification for low-wealth investors. It also holds for high-wealth investors when we use D_3 and D_4 as proxies for portfolio diversification, while it is not always significant when we use D_1 and D_2. This may be due to the fact that, since high-wealth investors are more financially sophisticated, crude measures of diversification such as D_1 and D_2 are not able to capture the extent of the investor's portfolio policies. The effect of industry diversity is consistently negative for the low-wealth investors for D_1 and D_2 and insignificant otherwise. This lack of significance may be due to the relatively high correlation between this variable and the specialization variable. The results on specialization, competition and diversity are striking in that they are the same factors related in previous research to knowledge spillover and urban growth.

As the tables indicate, diversification decreases significantly with the growth of competition, specialization and diversity. Thus, a one standard deviation increase in our measure of competition leads to a decrease in diversification of 2.2%, 0.5%, 1.7% and 2.0% for the measures of diversification D_1, D_2, D_3, D_4, respectively. Similarly, one standard deviation increase in specialization leads to a 1.1%, 0.2%, 1.6% and 1.5% for the corresponding measures of diversification. The effect of diversity seems to be smaller at less than 1%.

It is interesting to compare these findings to the results from the first stage: professional specialization increases stock market participation, but reduces portfolio diversification. This is consistent with the spill-over hypothesis. In particular, it suggests that the very process of city growth generates more – perhaps illusory - investment opportunities for the investors who rush to invest into them. This simultaneously increases stock market participation and portfolio concentration.

The results about the measures of local prosperity are also interesting. Local prosperity may induce two effects: on the one hand it increases investors' wealth and therefore financial sophistication. On the other hand, it raises the incentive to invest in local stocks. The latter may be due to an incentive to catch-up with the neighbors as the local-district theory would suggest, or to an incentive to increase the stakes in locally successful enterprises. This would be particularly true for the variables representing the number of local start-ups, local bankruptcies and local sale profitability. These variables proxy for the existence of successful local companies in which it is worth investing. A typical example would be an area like Silicon Valley, where we would expect most of the local investors to be highly invested on successful local companies. In this case, we would expect a negative relationship between local prosperity and diversification.

The results hold overall and for the different classes of investors and are particularly strong for the low-wealth investors. A one standard deviation increase in the number of startups (employment growth) leads to a 5.6% (1%) decrease in average correlation. The effect is noticeably stronger for wealthy households. Thus, a one standard deviation increase in bankruptcies rate leads to a 4.0% decrease in average correlation (D_2) for the low wealth households, and a 4.6% decrease for the high wealth households.

The findings show that portfolio diversification increases with the prosperity of the local area. There is, indeed, a positive relationship between portfolio diversification and employment growth, the number of new start-ups, the measures of profitability of local establishment (Sales Profitability and Return on Capital) and the number of active enterprises. An increase in bankruptcies on the other hand reduces diversification. It is also worth remembering that if we consider local consumption prices as a proxy for local prosperity there is a strong positive correlation between prosperity and portfolio diversification. The results hold overall as well as for different classes of investors and are particularly strong for low-wealth investors.

These findings do not support the district risk theory, which postulates a negative relationship between local prosperity and portfolio diversification. However, they support the information story. Indeed, it is likely that more information is generated in more prosperous districts. For example, newspapers may be established, brokers and financial analysts may flock in, banks may increase the number of their branches and the number of financial services provided. This would reduce the search cost and therefore increase portfolio diversification.

A further element to support the limited information theory is the strong, positive and statistically significant relationship between portfolio diversification and availability of financial services. That is, the higher the degree of bank coverage, the more investors are likely to diversify their portfolios. This holds across all the measures of diversification and for both classes of investors. A one standard deviation increase in the number of banking branches in a municipality leads to an average diversification increase between 1.4 and 2.0%. Moreover, the effect is stronger for high-wealth households. Thus, the number of assets increases 3.5% for high wealth households vs. 0.8% for low-wealth households.

It is worth stressing that in these specifications we are also controlling for the wealth and income of the investors. Therefore, these effects are "district effects" that act over and above the individual wealth and income effects. Analogously, our proxy for the availability of public services (the percentage of municipal employees) is mostly[11] positively related to portfolio diversification.

What do these results tell us in terms of standard portfolio theory? They suggest that diversification increases at the very time when either non-financial income risk decreases (reduction in the number of bankruptcies, increase in growth and profitability) or its effects are alleviated by the existence of public services or of a financial network. This runs against standard portfolio theory. Indeed, this theory would suggest that portfolio diversification should be greater when non-financial risk is higher. Areas characterized by a high employment growth rate, high start-up rate, and low bankruptcy rate should be areas characterized by low non-financial income risk and should therefore display lower financial diversification. Indeed, if the local economy is prosperous and growing, the risk of unemployment and well as the risk of a wage reduction should be lower. Therefore, portfolio diversification should be lower. The findings point in the opposite direction.

The percentage of people on welfare assistance impacts investors differently, depending on their level of wealth. It is always negative and statistically significant for the high-wealth investors and either not significant (for D_1 and D_2) or positive for the low-wealth investors (for D_3 and D_4). This provides some partial evidence of an impact of non financial risk for low-wealth investors.

---

[11] This holds for D_2, D_3 and D_4, for both classes of investors. However, it is mostly insignificant in the case of D_1.

These findings provide some preliminary and mixed evidence on standard portfolio theory in the presence of background risk. In order to further examine this issue, we need to consider the relationship between portfolio diversification and non-financial income risk hedging. This is represented by the correlation between our measures of portfolio diversification and Γ_i. The results are very striking. They show no correlation between hedging and portfolio diversification for the low-wealth investors and a negative correlation for the high wealth households and for the entire sample. These results are very robust across different specifications and alternative measures of portfolio diversification. This suggests that the portfolio under-diversification is not due to the need to hedge non-financial income risk. Quite the contrary, for the high wealth investors the negative correlation suggests familiarity-based investment or knowledge spill-over as an explanation for under-diversification.

Therefore, the relative lack of diversification in cities may not be attributed to the fact that cities allow investors to diversify away non-financial income risk by providing them with alternative occupational and professional possibilities. A possible rationalization of this is the fact that investors simply do not use their financial portfolio to hedge against non-financial income risk. In this case, the process of knowledge creation that takes place in the city just reduces the cost of search and increases portfolio diversification. This would be supported by recent findings showing that optimally choose not to hedge but invest in familiar stocks (Massa and Simonov, 2002).

It is interesting to note the role played by borrowing constraints. They do not affect the low-wealth investors, while they do increase portfolio diversification for the high-wealth ones. At the aggregate level they are positively related to D_1 and D_2 and not related to the other proxies of diversification. This apparently counterintuitive result may be explained as follows. For the low-wealth investors, borrowing constraints mostly affect the decision to participate in the stock market. And indeed, from the estimation of equation 3)[12], we find that the borrowing constraints negatively affect stock market participation for the low-wealth investors. For the high-wealth investors, instead, the borrowing constraints, while not sufficient to prevent stock market participation, induce the investors to reduce the financial exposure of their portfolio. This additional prudence increases the degree of financial diversification.

---

[12] Not reported, but available upon request.

To this point, both the limited information theory and the city agglomeration ones seem to be supported by the data. However, if we consider variables that proxy for the "degree of isolation", we find that, *all else equal*, portfolio diversification increases with isolation. That is, the further away an investor is from an urban center, the lower his connection to the rest of the world is and the less densely populated the district where he lives is, the more diversified his portfolio. In particular, one standard deviation increase in population density leads to a decrease of 2.46% in the number of risky assets held (D_1), a 1% increase in average correlation between assets (D_2) and a 7.5-8% increase in concentration as measured by D_3 and D_4. These results are contrary to what the limited information theory would predict[13], while they are consistent with the city-agglomeration and knowledge spill-over explanations.

To sum up, the local district affects the investors in two ways, on the one hand, the process of professional specialization that takes place in the city reduces the drive to diversification. On the other hand, the very same of process of knowledge creation reduces the search costs and may also increase portfolio diversification. We now move on to directly study portfolio choice.

6.2 Familiarity and city agglomeration

We regress our measures of familiarity – i.e., professional and geographical proximity – on the proxies for local professional specialization, local prosperity, financial and municipal service availability and the non-financial risk variables, as well as a set of control variables meant to control for investors' idiosyncratic wealth, income and demographic characteristics and for geographic and country shocks.

As before, we consider a specification based on the aggregated sample and one where the sample has been broken down into high and low wealth investors. The results are reported in Tables 7-9. In particular, Table 7 contains the main specification for both geographical and professional proximity for the entire sample. Tables 8 and 9 report the estimates for the low and high wealth investors, for geographical and professional proximity respectively. For each class of investors we consider four alternative specifications differing on the basis of the control

---

[13] They would actually lend some support to standard portfolio theory. Indeed, the more isolated the investor is, the higher is the probability that he will not be able to hedge away his income or professional idiosyncratic shock and, therefore, the more diversified he should be.

variables that are used. The results show that that the same variables that determine the degree of portfolio under-diversification are also those which affect investors' familiarity bias.

As a preliminary check, we consider the relationship between familiarity and hedging. Our working hypothesis is that familiarity should be negatively related to hedging as professional specialization induces investment in stocks that are professionally and geographically closer. This induces a behavior opposite to that required by hedging. And indeed, the relationship between our measures of familiarity and the indexes of non-financial risk show, that, as expected, hedging is negatively related to familiarity. That is, the more the investors choose familiar stocks, the more they tilt their financial portfolio away from the optimal composition required to hedge non-financial income risk. This suggests that familiarity is not due to hedging.

Let us now consider our measures of professional specialization and city agglomeration. The main finding is a significant positive correlation between both measures of familiarity and our measures of professional specialization. This relationship is strong, statistically significant and holds across all the specifications, for the overall sample and for both classes of investors in the case of competition and specialization. It is less significant for specialization. This may be due, as it was the case in the previous section, to the relatively high correlation between this variable and the diversity variable. These results suggest that the higher the degree of professional specialization of the district, the more the investors of such a district tend to invest in familiar stocks. This holds for both geographical and professional proximity. In particular, one standard deviation increase in competition (diversity) results in the portfolio being 1.9 (1.7) times "closer" to the investor as measured by value-weighted portfolio proximity. It also results in increase of probability of investors choosing the professionally related stocks by 2.1% (6.0%).

If we then consider the variables that proxy for the degree of isolation, we find that, all else equal, the more isolated the investors are, the more they tend to invest in stocks professionally and geographically close to them. That is, the further away an investor is from a big center, the less connected he is to the rest of the world and the less densely populated the district where he lives is, the more he will invest in nearby stocks. Living in a municipality with a population density one standard deviation away from the national mean results in the portfolio being 1.6 times closer to the investor. These results hold across investors (overall and for the different classes) and for different specifications.

The findings on the measures of local prosperity are also interesting. Investment in familiar stocks increases with the number of new start-ups and the local sales profitability, while it decreases with the number of local bankruptcies, number of active enterprises and return on capital. Employment growth reduces familiarity for the low-wealth investors, while, for the high-wealth investors, it reduces familiarity for the case of geographical proximity and seems to increase it for professional proximity. While in general higher prosperity seems to reduce the impact of familiarity, three measures of prosperity – new start-ups, bankruptcies and sales profitability – seem to go in the opposite direction and increase familiarity.[14] These variables are proxies for the "entrepreneurial dimension" of the district. As we mentioned before, a local investor in Silicon Valley would tend to invest in companies that are geographically and professionally close to him due to the fact that he is likely to be linked to these start-ups and to their profitability. Therefore, in general, local prosperity by increasing wealth and, therefore, financial sophistication reduces the impact of familiarity. However, if local prosperity is related to the local entrepreneurial dimension shared by the investor, the mere need to be part of such a dimension will induce the investor to invest in such shares. Finally, it is worth noting that the availability of financial services (bank coverage) in general reduces familiarity[15]. Indeed, the availability of financial services is a way of increasing financial sophistication and therefore reducing the local bias.

These results are consistent with the previous findings on portfolio under-diversification and provide a direct link between the familiarity/limited information story and the city agglomeration story. They suggest that the tension between professional specialization and prosperity also affects the exposure to the familiarity bias. The more professionally specialized the investors are and the more numerous the local investment opportunities where they are locally involved (i.e., local entrepreneurial dimension), the more they will invest in closely related stocks. In contrast, the richer and more prosperous the area where they live, the greater and the easier the access to information, the lower is the search cost and therefore the lower is the impact of the familiarity bias.

---

[14] The only exception are the high-wealth investors for the case of professional proximity.

[15] This is the case for both professional and geographical proximity for the low-wealth investors and for professional proximity for the high-wealth investors. Instead, it decreases familiarity for the high-wealth investors in the case of geographical proximity. For the latter case, we do not have a fully satisfactory explanation. We may argue that a better local financial coverage provide them with a quick way to act on the basis of the inside knowledge provided by proximity.

Conclusion

In this paper we address the puzzle of portfolio under-diversification from a broader perspective, linking it to theories of labor and the urban economy. In particular, we argue that portfolio under-diversification is closely related to factors linked to knowledge creation and spill-over in the urban environment. We showed that the processes of professional specialization and knowledge spillover that characterizes city agglomeration on the one hand reduce the availability of time to collect and analyze financial information and by the same token increases the relative information that an investor has with respect to stocks closer – professionally or geographically – to him. The net effect of these two factors appears to be that investors in an information-rich urban professional environment   -- educated though they may be -- appear to diversify less.

Appendix: Construction of income-related variables

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

$$E(?_{it}|?_{it\text{-}1}, X_{it\text{-}1}), Var(?_{it}|?_{it\text{-}1}, X_{it\text{-}1}) \text{ and } ?_{it},$$

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

$$\ln?_{it} = p_{it}+?_{it},$$

where:

$$p_{it} = g_{it}+p_{it\text{-}1}+?_{it}, \ ?_{it}\sim N(0,s_{\epsilon t}{}^2), \ ?_{it}\sim N(0,s_{?i}{}^2),$$

and

$$cov(?_{it}, ?_{is}) = 0, cov(?_{it}, ?_{is}) = 0, cov(?_{it}, ?_{is}) = 0 \text{ for each t, s.}$$

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

$$\ln?_{it}\text{-}\ln?_{it\text{-}1} = p_{it}\text{-}p_{it\text{-}1}+?_{it}\text{-}?_{it\text{-}1} = g_{it}+ ?_{it}\text{-}?_{it\text{-}1}+?_{it} ,$$

or

$$\ln?_{it} = \ln?_{it\text{-}1} + g_{it}+ ?_{it} + ?_{it}\text{-} ?_{it\text{-}1}$$

This implies:

$$E(?_{it}|?_{it\text{-}1}, X_{it\text{-}1}) = \ln?_{it\text{-}1} + g_{it} = ?_{i,t\text{-}1} G_{it} \exp\{J_{it} /2\}$$
$$Var(?_{it}|?_{it\text{-}1}, X_{it\text{-}1}) = (?_{i,t\text{-}1} G_{it})^2 \exp(J_{it}) \{ \exp(J_{it}) \text{-} 1\},$$

where:

$$G_{it} = \exp(g_{it}), J_{it} = s_{?i}{}^2+2s_{?i}{}^2$$

and $X_{i,t\text{-}1}$ is the set of variables usable to predict g_{it}.

In order to estimate $E(?_{it}|?_{it\text{-}1}, X_{it\text{-}1})$ and $Var(?_{it}|?_{it\text{-}1}, X_{it\text{-}1})$, we use data for the period 1990-2000, with a 5-year lagged rolling window. Following the Carrol and Samwick (1997) and Vissing-Jørgensen (2001) methodologies, we regress $\ln?_{it} - \ln?_{it\text{-}1}$ on the set of explanatory

variables $X_{i,t-1}$ and use the predicted values of such a regression as an estimate of g_t and the residuals as an estimate of $?_{it}+?_{it}-?_{it-1}$. The correlation between financial and non-financial income ($?_{it}$) is constructed as the conditional correlation between shocks to log non-financial income ($?_{it}+?_{it}-?_{it-1}$) and the log gross stock returns (i.e., $\ln(1+R_t)$). We use rolling five year windows to estimate the parameters.

The set of variables contained in $X_{i,t-1}$ are: demographic variables (secondary education, higher education, age, age squared, marriage status, size of the household, number of adults belonging to the household), changes in the demographic variables, industry dummies for the company the investor is working for (e.g., oil industry), dummies for the type of profession of the investor (e.g., doctor), emigration status. Following Vissing-Jørgensen (2002a), given the potential inaccuracy of estimates based on few observations, we calculate the correlation over the entire sample.

References

Arrow, Kenneth J., 1962, "The Economic Implications of Learning by Doing," *Review of Economic Studies*, 29(3), June, pp.155-73.

Bakshi, Gurdip. and Zhiwu Chen, 1996, "The Spirit of Capitalism and Stock Market Prices," *American Economic Review*, 86(1), March, pp. 133-157.

Barber, Brad M. and Terrance Odean, 2000, "Trading is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors," *Journal of Finance*, 55(2), April, pp. 773-806.

Barber, Brad M.. and Terrance Odean, 2003, "All that Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors." Working Paper, University of California, Berkeley.

Barry, Christopher B. and Stephen J. Brown, 1985, "Differential Information and Security Market Equilibrium," *Journal of Financial and Quantitative Analysis*, 20(4), December, pp. 407-422.

Brav, Alon. and John B. Heaton, 2002, "Competing Theories of Financial Anomalies," *The Review of Financial Studies*, 15, 575-606.

Benartzi, Shlomo. and Richard H. Thaler, 1995, "Myopic Loss Aversion and the Equity Premium Puzzle," *Quarterly Journal of Economics*, 110(1), February, pp.73-92.

Benartzi, Shlomo., 2001, "Excessive Extrapolation and the Allocation of 401(k) Accounts to Company Stock," Journal of Finance, 56(5), October, pp. 1747-1764.

Bhattacharya, Utpal. and Peter Groznik, 2001, "Melting Pot or Salad Bowl: Some Evidence from US Investments Abroad," Working Paper, Indiana and MIT.

Carlson, John A. and R. Preston MacAfee, 1983, R.P., "Discrete Price Equilibrium Dispersion," *Journal of Political Economy*, 91(3), June, pp. 480-493.

Carroll, Christopher. D. and Andrew A., Samwick, 1997, "The Nature of Precautionary Wealth," *Journal of Monetary Economics*, 40(1), pp. 41-71.

Carroll, Christopher .D. and Miles S. Kimball., 2001, "Liquidity Constraints and Precautionary Saving," NBER Working Paper 8496.

Coval, Joshua. D. and Tobias J. Moskowitz., 1999, "Home Bias at Home: Local Equity Preference in Domestic Portfolios," *Journal of Finance*, 54(6), December, pp. 2045-2073.

Coval, J. D. and Tobias J. Moskowitz, 2001, "The Geography of Investment: Informed Trading and Asset Prices," *Journal of Political Economy*, 109, pp. 811-841.

DeMarzo, Peter M., Ronald Kaniel and Ilan Kremer., 2002, "District Effects and Externalities in Portfolio Choice, "Working paper, Stanford and Duke Universities.

Edin, P.A., and Fredriksson, P., 2000, "LINDA: Longitudinal Individual Data for Sweden 1960-1997" Working Paper 2000:19, Department of Economics, Uppsala University.

Ellison, Glenn and Glaeser, Edward L., 1999, "The Geographic Concentration of Industry: does Natural Advantage Explain Agglomeration?" *American Economic Review*, 89(2), May, pp. 311-16

Englund, Peter, John M. Quigley, ., and Christian Redfearn, 1998, "Improved Price Indexes for Real Estate: Measuring the Course of Swedish Housing Prices," *Journal of Urban Economics* 44, pp. 171-196.

Feenberg, Daniel R. and James M., 2000, "The Income and Tax Share of Very High-Income Households, 1960-1995" *American Economic Review*, 90, May, pp. 264-270.

Gal,i Jordi, 1994, Keeping Up With the Joneses : Consumption Externalities, Portfolio Choice, and Asset Prices, *Journal of Money, Credit, and Banking* 26(1), pp. 1-8.

Garmaise, Mark and Tobias J. Moskowitz. 2003, "More Banks Less Crime. The Real and Social Effects of Bank Competition," Working Paper, The Wharton School and University of Chicago.

Glaeser Edward L., Hedi D. Kallal, , Jose A. Scheinkman and Andrei Shleifer, 1992, "Growth in Cities," *The Journal of Political Economy*, 100(6), December, pp. 1126-1151.

Glaeser Edward L Jose A. Scheinkman and Andrei Shleifer, 1995, "Economic Growth in a Cross-Section of Cities,"*Journal of Monetary Economics*, 36(1), August, pp. 117-143.

Glaeser, Edward L., Jed Kolko and Albert Saiz, 2002, "Consumer City," Harvard Institute of Economic Research Working Papers.

Glaeser, Edward L and David C. Mare, 2001, "Cities and Skills," *Journal of Labor Economics*, 19(2), April, pp. 316-42.

Glaeser, Edward L. 1997, "Learning in Cities," *Journal of Urban Economics*, 46(2) September, pp. 254-277.

Goetzmann, William N. and Alok Kumar, 2002, "Equity Portfolio Diversification," National Bureau of Economic Research Working Paper # 8686.

Grinblatt, Mark and Matti Keloharju,, 2000, "The Investment Behavior and Performance of Various Investor Types: A Study of Finland's Unique Data Set," *Journal of Financial Economics*, 55(1), January, pp. 43-67.

Grinblatt, Mark and Matti Keloharju, 2001, "How Distance, Language, and Culture Influence Stockholdings and Trades," *Journal of Finance*, 56(3), June, pp. 1053-1073.

Grinblatt, M. and Keloharju, M., 2001, "What Makes Investors Trade?," *Journal of Finance*, 56(2), April, pp. 589-616.

Grinold, Richard C. and Ronald N. Kahn, 1999, *Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Selecting Superior Returns and Controlling Risk*, McGraw-Hill.

Guiso, Luigi and Tullio Jappelli, 2002, "Household Portfolios in Italy," in Luigi Guiso, Michael Halaiassos and Tullio Jappelli (eds.) *Households Portfolios*, MIT Press, Cambridge, pp. 250-289.

Haliassos, Michael. and Christis Hassapis, 2002, "Borrowing Constraints, Portfolio Behavior and Precautionary Motives" in Kontoghiorghes, E., B. Rustem and S. Siokos (Eds.), *Computational Methods in Decision-making, Economics and Finance (Volume 1: Optimization Models)*, Kluwer Publishers Applied Optimization Series, vol. 74, 2002, pp. 183-210.

Hau, Harald, 2001, "Location Matters: An Examination of Trading Profits," *Journal of Finance*, 56(5), October, pp. 1959-1983.

Heaton, John B. and Deborah Lucas, 2000, "Portfolio Choice and Asset Prices: The Importance of Entrepreneurial Risk," *Journal of Finance*, 55(3), pp. 1163-1198.

Heaton, John B. and Deborah Lucas, 2000, "Portfolio Choice in the Presence of Background Risk," *Economic Journal*, 110(460), January, pp. 1-26.

Heckman, James., 1979, "Sample Selection Bias as a Specification Error," Econometrica 47(1), January, pp. 153-161.

Hortacsu, Ali and Chad Syverson, 2002, Search Costs, Product Differentation, and Welfare Effects of Entry: A Case Study of S&P 500 Index Funds, forthcoming, *The Quarterly Journal of Economics*.

Ivkovic, Zoran and Scott J. Weisbenner, 2003, "Local Does as Local Is: Information Content of the Geography of Individual Investors' Common Stock Investments," Working Paper, University of Illinois.

Huberman, Gur, 2001, "Familiarity Breeds Investment, "Review of Financial Studies, 14(3), Fall, pp. 659-680.

Jagannathan, Ravi and Kocherlakota, N. R., 1996, Why Should Older People Invest Less in Stocks Than Younger People?; Federal Reserve Bank of Minneapolis Quarterly Review, 20(3), Summer, pp. 11-23.

Jacobs, Jane, 1969, *The Economy of Cities*, New York, Vintage.

Marshall, Alfred, 1890, *Principles of Economics*, London, Macmillan.

Merton, Robert C., 1971, "Optimum Consumption and Portfolio Rules in a Continuous-Time Model," Journal of Economic Theory, 3(4), December, pp. 247-413.

Merton, R.C., 1987, "A Simple Model of Capital Market Equilibrium with Incomplete Information," Journal of Finance, 42(3), July, pp. 483-510.

Pastor, Lubos. and Pietro Veronesi, 2002, "Stock Valuation and Learning about Profitability," *Journal of Finance*, 58(5), October, pp. 1749-89.

Peng, Lin and Wei Xiong, 2002, "Time to Digest and Volatility Dynamics," Working Paper, Duke and Princeton Universities.

Porter, Michael E., 1990, *The Competitive Advantage of Nations*, New York, Free Press.

Poterba, James M. and Andrew A. Samwick,. 1997, "Household Portfolio Allocation Over the Life Cycle," NBER Working Paper: 6185.

Ross, Stephen A., 1978, "Mutual Fund Separation in Financial Theory: the Separating Distributions," *Journal of Economic Theory* 17(2), April, pp. 254-286.

Ryder Harl E. and Geoffrey M. Heal, 1973, "Optimal Growth with Intertemporally Dependent Preferences," *The Review of Economic Studies,* 40(1), January, pp. 1-31.

Shapiro, Alexander, 2002, "The Investor Recognition Hypothesis in a Dynamic General Equilibrium: Theory and Evidence," *Review of Financial Studies*, 15(1), Spring, pp. 97-141.

Shore, Stephen H. and Joshua S. White, 2002," External Habit Formation and the Home Bias Puzzle," Working Paper, Harvard University.

Sargent, Thomas J., 1993, *Bounded Rationality in Macroeconomics*, Oxford University Press, Clarendon Press.

Simon, Herbert A, 1986, *Theories of Bounded Rationality*, University of Minnesota Press, Minneapolis.

Sims, Christopher, 2000, "Implications of Rational Inattention," Working Paper, Princeton University.

Sundin, Anneli and Sven-Ivan Sundqvist, 2002, "Owners and Power in Sweden's Listed Companies," SIS Ägarservice AB, 1986-2002.

Tookes, Heather E., 2003, "Cross-Sectional Implications of Insider Trading: Insider Gains and Product Market Dominance," Cornell University Johnson School of Business, Working Paper.

Vissing-Jorgensen, Annette, 2002, "Towards an Explanation of Household Portfolio Choice Heterogeneity: Non-Financial Income and Participation Cost Structures," NBER Working Paper No. w8884.

Vissing-Jorgensen, A., 2002, "Limited Stock Market Participation and the Elasticity of Intertemporal Substitution," *Journal of Political Economy*, 110(4), August, pp. 825-53.

Zhu, Ning., 2003, The Local Bias of Individual Investors," Working Paper, Yale School of Management.

Table 1: Descriptive statistics

This table contains the descriptive statistics of the sample. Panel A reports the general demographic characteristics (number of households for each year, members in household, adults in household, age of the oldest member of household, percentage of the sample with secondary and higher education, percentage of immigrants). Panel B describes local characteristics. We used 289 municipalities for 1999-2000 and 288 for 1995-1998. We also report descriptive statistics for our measures of diversification $D_1 = N$, $D_2 =- 1/N -(1-1/N) \overline{Corr}$,

$$D_3 = -\sum_{i=1}^{N}(w_i - w_{mkt})^2, \text{ and } D_4 = -\sum_{i=1}^{N}(w_i - w_{ffl})^2,$$ where N is number of positions in the portfolio, $\overline{Corr}$

is average correlation of stocks in the portfolio, w_i is the weight of the stock in the portfolio of the investor, w_{mkt} is the weight that the same stock would have in the market portfolio, and w_{ffl} is the weight that the same stock would have in the free float portfolio. *Specialization* is the share of the top five industries in local employment to the share of the top five industries in national employment (source: Statistics Sweden). *Competition* is the number of firm per employee incorporated in a municipality relative to the number of firms per employee in Sweden (sources: Statistics Sweden, MM Partners). *Diversity* is the negative of the share of the top 5 industries in municipal employment (source: Statistics Sweden). *Active enterprises* is number (in thousands) of active firms in municipality in a given year (source: MM Partners). *Start-ups* and *Bankruptcies* measure percent of failed (ceased business activities) and started firms in municipality in a given year as percentage of active firms (source: MM Partners*). Profitability* and *Return on Capital* are sales-weighted average of profitability and return on capital of the enterprises in municipality (source: MM Partners). Employment growth is growth of employment in municipality w.r.t. 1985 (source: Statistics Sweden).*Index of Rural Areas* is a code that is set by Statistics Sweden from 1 (Metropolitan area) to 9 (countryside). *Distance from Airport* is measured as logarithm of the distance between closest civilian airport and central post office in municipality (sources: Cartesia Informationsteknik AB, Swedish Civil Aviation Board). *Population Density* is population (in tens of thousands) per square kilometer (source: Statistics Sweden). *Bank Coverage* is logarithm of number of credit institutions' branches in municipality (source: MM Partners). *Percent of population on welfare assistance* and *Percent of municipal employees* are provided by Statistics Sweden. Panel C reports the age and gender distribution of the sample. Panel D reports the percentage of the households paying wealth tax, having labor income, having entrepreneurial income and having real estate wealth. We report mean, standard deviation, median and inter-quartile range (IQR). They have been calculated over the whole sample (i.e., across-investors and time). 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 statistics (Mean, Standard Deviation, IQR, Maximum) of, respectively, the value wealth, labor and entrepreneurial income gross yearly income) and real estate. All monetary values are in Swedish krowns (SEK).

Panel A: General demographic characteristics

Variable	Mean	Median	Std.Dev.	IQR	Maximum
Number of households	292,901	291,913	647	686	293,320
# of members in household	2.67	2.00	1.51	3.00	16.00
# of adults in household	1.77	2.00	0.69	1.00	9.00
Age of oldest household member	49.28	47	17	24	107
% with secondary education	43.5%	43.5%	0.6%	0.5%	44.3%
% with higher education	31.4%	31.2%	1.4%	1.4%	33.7%
% of immigrants	16.4%	16.3%	2.7%	4.6%	19.3%
D_1	1.63	1.00	1.78	0.00	68.00
D_2	-0.87	-1.00	0.23	0.00	0.33
D_3	-0.98	-0.88	0.60	0.61	-3.1E-7
D_4	-0.96	-0.95	0.50	0.61	-1.6E-7

Panel B: Characteristics of the local district

Variable	Mean	Median	Std. Dev.	I. Q. R.	Maximum
Active Enterprises	0.795	0.239	1.947	0.406	12.589
Competition	1.934	1.874	0.466	0.708	3.778
Specialization	1.618	1.519	0.463	0.389	6.521
Diversity	-0.230	-0.230	0.023	0.024	-0.114
Bankruptcies	0.717	0.746	1.457	0.299	13.750
Start-ups	0.501	0.512	1.427	0.263	9.750
Sale Profitability	0.103	0.091	0.154	0.040	2.746
Return on Capital	0.046	0.001	0.107	0.008	1.128
Employment Growth	-0.094	-0.097	0.118	0.133	0.426
% of Population on Welfare Asst.	0.075	0.073	0.031	0.038	0.161
% of Municipal Employees	0.065	0.065	0.011	0.015	0.104
Index of Rural Areas	3.544	3.000	2.104	3.000	9.000
Distance from Airport	8.425	8.909	1.538	2.622	10.874
Population Density	0.609	0.078	1.159	0.516	3.971
Bank Coverage	1.768	1.792	0.711	0.811	5.501

Panel C: 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 D: Wealth and income characteristics of the 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
Real Estate Holders	54.6%	449,400	387,000	348,736	340,000	78,140,000
Labor Income Earners	100.0%	321,489	287,722	237,526	276,190	43,445,271
Entrepr. Income Earners	9.8%	88,114	43,268	172,565	111,726	7,320,000

Table 2: Measures of diversification for overall sample

We report the results for the full specification where the dependent variables are measures of diversification $D_1 = N$, $D_2 = -1/N - (1-1/N) \overline{Corr}$, $D_3 = -\sum_{i=1}^{N}(w_i - w_{mkt})^2$, and $D_4 = -\sum_{i=1}^{N}(w_i - w_{ffl})^2$, where N is number of positions in the portfolio, $\overline{Corr}$ is average correlation of stocks in the portfolio, w_i is the weight of the stock in the portfolio of the investor, w_{mkt} is the weight that the same stock would have in the market portfolio, and w_{ffl} is the weight that the same stock would have in the free float portfolio. The main variables are as described in Table 1, while the control variables are described in the text. We also control in each specification for consumer confidence and local tax rate (not reported). t-statistics are reported in parentheses. We also report the *Adjusted R^2*. All the coefficients are multiplied by 100.

Variable	D_1 Coeff.	t-stat	D_2 Coeff.	t-stat	D_3 Coeff.	t-stat	D_4 Coeff.	t-stat
Local Professional Specialization								
Competition	-7.50	(-7.18)	-0.88	(-7.14)	-1.84	(-6.57)	-2.17	(-9.39)
Specialization	-3.80	(-3.30)	-0.35	(-2.57)	-1.70	(-5.50)	-1.61	(-6.34)
Diversity	-58.80	(-2.81)	-8.69	(-3.52)	10.62	(1.90)	3.30	(0.72)
Degree of Local Isolation								
Population Density	-22.96	(-1.25)	-6.35	(-2.86)	-39.06	(-7.81)	-41.87	(-10.21)
Index of Rural Areas	1.90	(5.63)	0.14	(3.61)	0.49	(5.49)	0.39	(5.36)
Distance from Airport	2.70	(7.15)	0.25	(5.55)	0.47	(4.59)	0.40	(4.72)
Local Prosperity								
Bankruptcies	-41.10	(-27.00)	-4.33	(-24.48)	-2.42	(-6.16)	-3.65	(-11.26)
Start-ups	37.20	(25.56)	3.94	(23.17)	2.09	(5.50)	3.27	(10.42)
Sale Profitability	17.10	(6.01)	1.95	(5.82)	1.07	(1.41)	0.20	(0.32)
Return on Capital	27.09	(4.63)	5.51	(7.91)	17.40	(10.75)	17.50	(13.13)
Employment Growth	66.10	(13.65)	8.72	(15.40)	6.61	(5.19)	11.42	(10.88)
Active Enterprises	-1.89	(-5.85)	-0.10	(-2.85)	0.39	(4.74)	0.29	(4.10)
Financial Services Availability								
Bank Coverage	4.50	(6.34)	0.77	(9.14)	1.64	(8.52)	1.53	(9.69)
% of public employees	-101.70	(-1.86)	8.63	(1.31)	178.14	(11.55)	160.81	(12.72)
Non-Financial Risk and Financial Constraints								
Active Hedging Index	-10.01	(-11.80)	-1.53	(-14.14)	-1.43	(-5.29)	-1.50	(-6.81)
Borrowing Constraints	0.13	(2.58)	0.02	(3.49)	-0.01	(-0.34)	0.00	(-0.11)
% of population on welfare asst.	74.90	(3.59)	-0.72	(-0.29)	-65.38	(-11.58)	-62.79	(-13.53)
Wealth Variables	Yes		Yes		Yes		Yes	
Income Variables	Yes		Yes		Yes		Yes	
Momentum Variables	Yes		Yes		Yes		Yes	
Demographic Variables	Yes		Yes		Yes		Yes	
Macro and Social Variables	Yes		Yes		Yes		Yes	
Lambda	-251.10	(-37.46)	-22.58	(-29.09)	-7.04	(-4.10)	-15.38	(-10.84)
Constant	668.08	(37.27)	-46.83	(-22.14)	-69.64	(-14.49)	-58.96	(-14.92)
Adj R2	0.057		0.064		0.115		0.178	

Table 3: Dependent variable D_1

We report the results for the specification where the dependent variable $D_1 = N$ where N is number of positions in the portfolio. We report results for Low and High-wealth households. In each case, four different specifications are reported. The main variables are as described in Table 1, while the control variables are described in the text. We also control in each specification for consumer confidence and local tax rate (not reported). T-statistics is reported in parentheses. We also report adjusted R^2. All the coefficients are multiplied by 100.

	Low-wealth households								High-wealth households							
	(1)		(2)		(3)		(4)		(1)		(2)		(3)		(4)	
Variable	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Local Professional Specialization																
Competition	-2.61	(-4.97)	-1.72	(-3.40)	-1.68	(-3.36)	-2.36	(-4.64)	-6.38	(-1.93)	-1.38	(-0.52)	-7.34	(-2.04)	-2.03	(-0.78)
Specialization	-1.65	(-2.83)	-1.52	(-2.67)	-1.38	(-2.44)	-1.44	(-2.51)	-7.11	(-1.96)	-4.11	(-1.41)	-7.6	(-1.92)	-4.52	(-1.57)
Diversity	-31.2	(-2.96)	-21.55	(-2.10)	-21.65	(-2.12)	-29.98	(-2.89)	-71.54	(-1.08)	-21.72	(-0.41)	-100.53	(-1.40)	-40.22	(-0.77)
Degree of Local Isolation																
Population Density	-0.82	(-0.08)	7.15	(0.75)	5.44	(0.57)	0.99	(0.10)	20	(0.41)	20	(0.53)	10	(0.22)	10	(0.29)
Index of Rural Areas	0.52	(3.16)	0.22	(1.36)	0.13	(0.83)	0.32	(1.99)	-1.89	(-1.85)	3.19	(3.86)	1.15	(1.04)	3.37	(4.13)
Distance from Airport	0.76	(4.04)	0.56	(3.05)	0.48	(2.67)	0.67	(3.66)	5.84	(4.48)	2.15	(2.12)	7.03	(4.97)	1.63	(1.66)
Local Prosperity																
Bankruptcies	-9.82	(-13.05)	-6.16	(-9.53)	-5.4	(-10.29)	-8.2	(-14.47)	-52.6	(-9.87)	-18.72	(-5.15)	-63.96	(-11.14)	-17.67	(-5.58)
Start-ups	8.78	(12.18)	5.53	(8.74)	4.83	(9.04)	7.33	(12.88)	50.57	(9.69)	18.25	(5.08)	61.45	(10.93)	17.15	(5.43)
Sales Profitability	3.24	(2.29)	1.67	(1.21)	1.14	(0.84)	2.35	(1.71)	27.9	(2.97)	1.2	(0.16)	36.93	(3.63)	-1.14	(-0.16)
Return on Capital	18.4	(6.02)	17.7	(5.91)	17.7	(5.92)	18	(5.95)	25	(1.37)	28.1	(1.86)	24.9	(1.26)	27.8	(1.86)
Employment Growth	24.87	(10.17)	20.83	(8.99)	18.91	(8.31)	22.94	(10.00)	120.56	(7.22)	47.41	(3.83)	142.93	(7.89)	37.25	(3.20)
Active Enterprises	0.55	(3.32)	0.70	(4.37)	0.76	(4.79)	0.67	(4.10)	-4.10	(-4.45)	-2.80	(-3.86)	-4.60	(-4.59)	-2.70	(-3.66)
Financial and Municipal Services Availability																
Bank Coverage	1.70	(4.68)	1.52	(4.28)	1.44	(4.08)	1.58	(4.41)	7.95	(3.49)	7.85	(4.26)	8.36	(3.38)	8.03	(4.41)
% of Municipal Employees	46.19	(1.63)	44.55	(1.60)	51.07	(1.84)	43.37	(1.54)	-333.14	(-1.88)	-278.93	(-1.85)	-359.25	(-1.89)	-239.05	(-1.60)
Non-financial Risk and Financial Constraints																
Active Hedging Index	0.98	(1.76)	1.17	(1.10)	1.04	(1.76)	1.16	(1.68)	-17.56	(-8.58)	-16.16	(-8.07)	-17.88	(-8.63)	-16.2	(-8.11)
Borrowing Constraints	0.03	(1.27)	0.03	(1.09)	0.02	(1.00)	0.03	(1.26)	69.88	(5.32)	58.11	(5.23)	62.35	(4.47)	61.15	(5.71)
% of Population on Welfare Asst.	22.46	(2.13)	-0.57	(-0.06)	-3.53	(-0.36)	19.55	(1.94)	-106.3	(-1.66)	-233.28	(-4.52)	-68.93	(-1.00)	-240.11	(-4.72)
Control Variables																
Wealth Variables	Yes		Yes		No		Yes		Yes		Yes		No		Yes	
Income Variables	Yes		No		Yes		Yes		Yes		No		Yes		Yes	
Momentum Variables	Yes		Yes		No		Yes		Yes		Yes		No		Yes	
Demographic Variables	Yes		Yes		Yes		No		Yes		Yes		Yes		No	
Macro and Social Variables	Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Lambda	-50.31	(-14.94)	-27.54	(-11.25)	-22.72	(-26.96)	-40.01	(-30.31)	-413.48	(-14.37)	-149.41	(-9.34)	-497.48	(-16.27)	-131.43	(-15.68)
Constant	221.12	(23.70)	172.53	(23.89)	153.43	(29.43)	200.77	(40.09)	621.25	(15.23)	419.17	(13.32)	729.77	(17.41)	297.77	(13.27)
Adj R2	0.037		0.033		0.035		0.035		0.053		0.047		0.052		0.048	

Table 4: Dependent variable D_2

We report the results for the specification where the dependent variable is the measure of diversification $D_2 = -1/N -(1-1/N)\,\overline{Corr}$ where N is number of positions in the portfolio and $\overline{Corr}$ is average correlation of stocks in the portfolio. The notations are as in Table 2. All the coefficients are multiplied by 100.

Variable	Low-wealth households (1) Coeff.	t-stat	(2) Coeff.	t-stat	(3) Coeff.	t-stat	(4) Coeff.	t-stat	High-wealth households (1) Coeff.	t-stat	(2) Coeff.	t-stat	(3) Coeff.	t-stat	(4) Coeff.	t-stat
Local Professional Specialization																
Competition	-0.82	(-7.24)	-0.59	(-5.46)	-0.54	(-5.14)	-0.72	(-6.63)	-0.49	(-1.77)	-0.27	(-1.04)	-0.52	(-1.85)	-0.43	(-1.63)
Specialization	-0.3	(-2.37)	-0.26	(-2.18)	-0.22	(-1.82)	-0.22	(-1.80)	-0.3	(-0.98)	-0.2	(-0.68)	-0.32	(-1.02)	-0.28	(-0.96)
Diversity	-7.24	(-3.18)	-4.73	(-2.16)	-4.4	(-2.04)	-6.53	(-2.95)	-9.95	(-1.79)	-6.91	(-1.32)	-10.89	(-1.92)	-6.68	(-1.25)
Degree of Local Isolation																
Population Density	-6.22	(-2.95)	-4.36	(-2.15)	-4.34	(-2.16)	-5.71	(-2.78)	-3.23	(-0.69)	-3.58	(-0.80)	-3.51	(-0.73)	-4.72	(-1.03)
Index of Rural Areas	0.18	(5.02)	0.1	(2.99)	0.06	(1.89)	0.11	(3.21)	-0.16	(-1.90)	0.2	(2.40)	0.14	(1.57)	0.18	(2.12)
Distance from Airport	0.16	(3.81)	0.1	(2.55)	0.07	(1.88)	0.12	(3.03)	0.35	(3.18)	0.22	(2.14)	0.39	(3.48)	0.26	(2.55)
Local Prosperity																
Bankruptcies	-2.88	(-17.66)	-1.91	(-13.83)	-1.57	(-14.09)	-2.28	(-18.83)	-3.28	(-7.39)	-2.12	(-5.91)	-3.66	(-8.16)	-2.84	(-8.77)
Start-ups	2.59	(16.58)	1.72	(12.78)	1.41	(12.49)	2.05	(16.87)	3.18	(7.30)	2.07	(5.84)	3.54	(8.05)	2.75	(8.52)
Sales Profitability	1.27	(4.16)	0.84	(2.87)	0.65	(2.25)	0.94	(3.20)	1.7	(2.14)	0.75	(1.02)	2	(2.47)	1.29	(1.75)
Return on Capital	4.7	(7.09)	4.5	(7.08)	4.4	(7.02)	4.5	(7.02)	5	(3.20)	5.3	(3.57)	5	(3.14)	5	(3.30)
Employment Growth	6.35	(12.01)	5.19	(10.52)	4.58	(9.50)	5.59	(11.43)	8.69	(6.22)	6.12	(5.02)	9.45	(6.64)	6.98	(5.87)
Active Enterprises	0.11	(3.18)	0.16	(4.54)	0.18	(5.25)	0.16	(4.47)	-0.31	(-3.99)	-0.27	(-3.76)	-0.33	(-4.12)	-0.28	(-3.84)
Financial and Municipal Services Availability																
Bank Coverage	0.58	(7.36)	0.54	(7.08)	0.5	(6.69)	0.55	(7.23)	0.76	(3.96)	0.79	(4.32)	0.78	(3.96)	0.91	(4.91)
% of Municipal Employees	23.75	(3.90)	24.25	(4.09)	25.17	(4.27)	23.23	(3.89)	2.59	(0.17)	6	(0.40)	1.69	(0.11)	5.98	(0.39)
Non-financial Risk and Financial Constraints																
Active Hedging Index	-0.04	(-0.37)	0.01	(0.05)	-0.03	(-0.22)	0.03	(0.23)	-2.58	(-12.87)	-2.52	(-12.63)	-2.59	(-12.91)	-2.41	(-12.08)
Borrowing Constraints	0.01	(2.79)	0.01	(2.56)	0.01	(2.39)	0.01	(2.72)	12.84	(11.14)	12.73	(11.57)	12.55	(10.81)	15.7	(14.45)
% of Population on Welfare Asst.	4.14	(1.82)	-1.95	(-0.90)	-3.55	(-1.69)	2.15	(1.00)	-17.19	(-3.17)	-22.07	(-4.33)	-15.87	(-2.87)	-21.67	(-4.18)
Control Variables																
Wealth Variables	Yes		Yes		No		Yes		Yes		Yes		No		Yes	
Income Variables	Yes		No		Yes		Yes		Yes		No		Yes		Yes	
Momentum Variables	Yes		Yes		No		Yes		Yes		Yes		No		Yes	
Demographic Variables	Yes		Yes		Yes		No		Yes		Yes		Yes		No	
Macro and Social Variables	Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Lambda	-13.42	(-18.38)	-7.33	(-14.04)	-5.22	(-29.26)	-9.56	(-34.00)	-21.75	(-8.91)	-12.63	(-8.01)	-24.6	(-10.07)	-17.13	(-20.05)
Constant	-64.29	(-31.92)	-77.97	(-50.71)	-84.33	(-76.38)	-73.39	(-68.82)	-45.27	(-12.53)	-53.12	(-17.05)	-41.44	(-11.72)	-65.15	(-28.48)
Adj R2	0.0333		0.0288		0.0310		0.0290		0.0400		0.0366		0.0399		0.0357	

Table 5: Dependent variable D_3

We report the results for the specification where the dependent variable is the measure of diversification $D_3 = -\sum_{i=1}^{N}(w_i - w_{mkt})^2$, where w_i is the weight of the stock in the portfolio of the investor and w_{mkt} is the weight that the same stock would have in the market portfolio. The notations are as in Table 2. All the coefficients are multiplied by 100.

	Low-wealth households								High-wealth households							
	(1)		(2)		(3)		(4)		(1)		(2)		(3)		(4)	
Variable	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Local Professional Specialization																
Competition	-2.28	(-9.44)	-2.31	(-9.69)	-2.04	(-8.64)	-2.62	(-10.91)	-0.78	(-1.02)	-0.76	(-1.00)	-0.48	(-0.61)	-1.42	(-1.88)
Specialization	-1.17	(-4.38)	-1.22	(-4.56)	-1.1	(-4.10)	-1.18	(-4.34)	-1.98	(-2.37)	-1.97	(-2.35)	-1.84	(-2.11)	-2.75	(-3.30)
Diversity	-4.46	(-0.92)	-4.1	(-0.85)	-6.72	(-1.40)	-0.39	(-0.08)	23.37	(1.53)	23.75	(1.56)	31.61	(1.99)	11.42	(0.75)
Degree of Local Isolation																
Population Density	-42	(-9.30)	-43.74	(-9.71)	-41	(-9.14)	-44	(-9.62)	-44	(-3.36)	-45	(-3.45)	-41	(-3.06)	-45	(-3.47)
Index of Rural Areas	-0.08	(-1.07)	-0.01	(-0.19)	0.17	(2.32)	0.07	(0.92)	1.34	(5.62)	1.33	(5.61)	1.55	(6.29)	1.03	(4.39)
Distance from Airport	0.48	(5.53)	0.48	(5.58)	0.41	(4.80)	0.58	(6.72)	-0.01	(-0.03)	-0.06	(-0.22)	-0.35	(-1.14)	0.64	(2.24)
Local Prosperity																
Bankruptcies	-1.65	(-4.80)	-1.87	(-6.16)	-0.57	(-2.31)	-3.2	(-11.94)	1.29	(1.06)	1.57	(1.51)	4.54	(3.63)	-4.89	(-5.36)
Start-ups	1.42	(4.28)	1.61	(5.42)	0.45	(1.78)	2.79	(10.37)	-1.62	(-1.36)	-1.89	(-1.84)	-4.73	(-3.86)	4.32	(4.75)
Sales Profitability	1.08	(1.66)	1.17	(1.81)	0.56	(0.88)	1.98	(3.05)	-2.92	(-1.33)	-3.16	(-1.48)	-5.51	(-2.43)	2.37	(1.13)
Return on Capital	15.7	(11.06)	16.1	(11.34)	15.6	(11.02)	15.9	(11.11)	19	(4.38)	19.4	(4.46)	19	(4.23)	19	(4.39)
Employment Growth	6.95	(6.19)	6.76	(6.19)	5.44	(5.07)	10.09	(9.31)	-6.12	(-1.60)	-7.05	(-1.99)	-12.49	(-3.15)	8.03	(2.39)
Active Enterprises	0.55	(7.27)	0.54	(7.13)	0.61	(8.06)	0.47	(6.16)	0.41	(1.91)	0.42	(1.98)	0.55	(2.48)	0.13	(0.61)
Financial and Municipal Services Availability																
Bank Coverage	1.52	(9.12)	1.62	(9.69)	1.45	(8.69)	1.57	(9.27)	1.42	(2.69)	1.44	(2.73)	1.29	(2.34)	1.85	(3.52)
% of Municipal Employees	166.27	(12.62)	171.62	(13.03)	168.29	(12.79)	161.26	(12.12)	249.3	(5.73)	252.46	(5.79)	256.8	(5.74)	235.39	(5.41)
Non-financial Risk and Financial Constraints																
Active Hedging Index	-0.30	(-1.13)	-0.27	(-1.02)	-0.28	(-1.07)	-0.27	(-1.04)	-2.15	(-3.65)	-2.15	(-3.65)	-2.06	(-3.47)	-2.15	(-3.66)
Borrowing Constraints	-0.01	(-0.82)	-0.01	(-0.74)	-0.01	(-0.97)	-0.01	(-0.56)	13.12	(4.05)	12.91	(4.01)	15.62	(4.74)	21.55	(6.92)
% of Population on Welfare Asst.	43.51	(8.93)	43.1	(9.02)	50.89	(10.87)	32.91	(6.90)	-92.89	(-6.20)	-93.62	(-6.30)	-104.01	(-6.71)	-78.97	(-5.36)
Control Variables																
Wealth Variables	Yes		Yes		No		Yes		Yes		Yes		No		Yes	
Income Variables	Yes		No		Yes		Yes		Yes		No		Yes		Yes	
Momentum Variables	Yes		Yes		No		Yes		Yes		Yes		No		Yes	
Demographic Variables	Yes		Yes		Yes		No		Yes		Yes		Yes		No	
Macro and Social Variables	Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Lambda	-5.94	(-3.86)	-6.93	(-6.02)	0.78	(1.99)	-16.33	(-26.18)	24.63	(3.65)	27.31	(5.95)	48.71	(7.10)	-23.08	(-9.51)
Constant	-84.45	(-19.68)	-85.68	(-25.17)	-102.72	(-41.65)	-67.86	(-28.66)	1.41	(0.14)	-3.77	(-0.41)	-31.11	(-3.08)	-27.3	(-4.20)
Adj R2	0.226		0.225		0.226		0.226		0.042		0.015		0.040		0.038	

Table 6: Dependent variable D_4

We report the results for the specification where the dependent variable is the measure of diversification $D_4 = -\sum_{i=1}^{N}(w_i - w_{ffl})^2$, where w_i is the weight of the stock in the portfolio of the investor and w_{ffl} is the weight that the same stock would have in the free float portfolio. The notations are as in Table 2. All the coefficients are multiplied by 100.

| | Low-wealth households | | | | | | | | High-wealth households | | | | | | | |
| | (1) | | (2) | | (3) | | (4) | | (1) | | (2) | | (3) | | (4) | |
Variable	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Local Professional Specialization																
Competition	-2.28	(-11.67)	-2.06	(-10.74)	-2.00	(-10.54)	-2.53	(-12.92)	-2.20	(-3.58)	-1.65	(-2.73)	-1.71	(-2.79)	-2.14	(-3.50)
Specialization	-1.20	(-5.52)	-1.18	(-5.46)	-1.12	(-5.19)	-1.23	(-5.54)	-2.24	(-3.31)	-1.83	(-2.74)	-1.87	(-2.78)	-2.47	(-3.66)
Diversity	-0.25	(-0.06)	-2.01	(-0.52)	-2.62	(-0.68)	3.58	(0.89)	-2.66	(-0.22)	13.97	(1.15)	15.55	(1.27)	5.44	(0.44)
Degree of Local Isolation																
Population Density	-41.00	(-11.13)	-40.00	(-10.95)	-39.00	(-10.84)	-42.00	(-11.26)	-50.20	(-4.80)	-51.00	(-4.92)	-47.00	(-4.53)	-50.00	(-4.76)
Index of Rural Areas	-0.06	(-1.00)	-0.14	(-1.33)	0.18	(3.04)	0.04	(0.72)	0.91	(4.75)	1.10	(5.78)	1.19	(6.25)	0.90	(4.71)
Distance from Airport	0.38	(5.34)	0.29	(4.20)	0.29	(4.24)	0.46	(6.43)	0.50	(2.06)	-0.02	(-0.09)	-0.04	(-0.17)	0.51	(2.19)
Local Prosperity																
Bankruptcies	-2.32	(-8.29)	-1.17	(-4.80)	-0.98	(-4.91)	-3.48	(-15.84)	-5.10	(-5.19)	-0.62	(-0.75)	-0.07	(-0.08)	-5.27	(-7.12)
Start-ups	2.00	(7.43)	0.97	(4.07)	0.80	(3.93)	3.03	(13.76)	4.79	(4.97)	0.50	(0.61)	-0.03	(-0.03)	4.97	(6.73)
Sales Profitability	0.90	(1.70)	0.36	(0.69)	0.26	(0.50)	1.55	(2.91)	-1.98	(-1.12)	-5.57	(-3.26)	-5.96	(-3.39)	-1.62	(-0.95)
Return on Capital	15.10	(13.15)	15.10	(13.29)	14.90	(13.13)	15.20	(13.06)	18.80	(5.35)	19.40	(5.58)	18.90	(5.41)	19.00	(5.39)
Employment Growth	9.66	(10.60)	7.99	(9.13)	7.83	(9.07)	11.90	(13.40)	11.62	(3.75)	1.61	(0.57)	1.17	(0.38)	12.74	(4.68)
Active Enterprises	0.45	(7.33)	0.51	(8.31)	0.52	(8.58)	0.39	(6.28)	0.10	(0.57)	0.29	(1.73)	0.32	(1.90)	0.08	(0.45)
Financial and Municipal Services Availability																
Bank Coverage	1.45	(10.69)	1.42	(10.62)	1.36	(10.21)	1.48	(10.69)	1.07	(2.51)	1	(2.37)	0.93	(2.20)	1.21	(2.83)
% of Municipal Employees	156.24	(14.66)	160.34	(15.18)	158.11	(14.96)	152.45	(14.08)	213.15	(6.06)	225.09	(6.45)	222.95	(6.37)	209.83	(5.96)
Non-financial Risk and Financial Constraints																
Active Hedging Index	0.07	(0.30)	0.08	(0.37)	0.08	(0.39)	0.04	(0.20)	-2.4	(-5.05)	-2.27	(-4.79)	-2.27	(-4.78)	-2.44	(-5.13)
Borrowing Constraints	-0.01	(-0.88)	-0.01	(-1.01)	-0.01	(-1.10)	-0.01	(-0.68)	13.22	(5.06)	11.16	(4.33)	12.65	(4.90)	14.87	(5.90)
% of Population on Welfare Asst.	44.99	(11.39)	51.68	(13.49)	53.33	(14.18)	37.01	(9.50)	-71.56	(-5.91)	-86.49	(-7.29)	-87.65	(-7.29)	-74.61	(-6.25)
Control Variables																
Wealth Variables	Yes		Yes		No		Yes		Yes		Yes		No		Yes	
Income Variables	Yes		No		Yes		Yes		Yes		No		Yes		Yes	
Momentum Variables	Yes		Yes		No		Yes		Yes		Yes		No		Yes	
Demographic Variables	Yes		Yes		Yes		No		Yes		Yes		Yes		No	
Macro and Social Variables	Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Lambda	-11.6	(-9.27)	-4.07	(-4.41)	-3.23	(-10.09)	-19.36	(-37.93)	-21.76	(-3.99)	13.54	(3.70)	16.48	(3.10)	-23.2	(-11.79)
Constant	-74.61	(-21.42)	-94.23	(-34.53)	-95.93	(-48.43)	-61.35	(-31.71)	18.51	(2.25)	-13.27	(-1.82)	-16.92	(-2.14)	-29.6	(-5.62)
Adj R2	0.313		0.312		0.313		0.313		0.067		0.066		0.066		0.064	

Table 7: Measures of proximity for overall sample

We report the results for the specification where the dependent variables are our measures of geographical and professional proximity. Geographical proximity is measured as value-weighted average of inverse of the logarithm 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. Professional proximity for each position it takes value 1 if any of the household members are employed in the same industry of the company they are investing in. For investor the measure is computed as value-weighted average over all position in household portfolio. The notations are as in Table 2.

Variable	Geographical proximity		Professional proximity	
	Coeff.	t-stat	Coeff.	t-stat
Local Professional Specialization				
Competition	0.577	(34.18)	0.046	(24.24)
Specialization	0.001	(58.84)	-3.295	(-1.54)
Diversity	10.426	(33.36)	2.491	(9.58)
Degree of Local Isolation				
Population Density	-1.284	(-4.24)	-22.200	(-10.90)
Index of Rural Areas	0.057	(9.95)	0.088	(3.68)
Distance from Airport	0.033	(5.49)	0.588	(8.32)
Local Prosperity				
Bankruptcies	-4.342	(-17.11)	-0.111	(-10.28)
Start-ups	5.146	(25.09)	0.112	(19.63)
Sales Profitability	-0.110	(-2.41)	0.187	(17.40)
Return on Capital	-0.037	(-38.01)	-0.002	(-19.97)
Employment Growth	-1.226	(-15.98)	0.026	(8.02)
Active Enterprises	0.604	(29.03)	-0.020	(-26.55)
Financial and Municipal Services Availability				
Bank Coverage	0.052	(4.68)	-0.022	(-17.72)
% of Municipal Employees	17.411	(19.04)	0.111	(2.99)
Background Risk and Financial Constraints				
Active Hedging Index	-1.427	(-91.97)	-0.043	(-35.45)
Borrowing Constraints	-0.003	(-0.40)	0.000	(0.92)
% of Population on Welfare Asst.	-19.304	(-50.67)	0.278	(17.56)
Wealth Variables	Yes		Yes	
Income Variables	Yes		Yes	
Momentum Variables	Yes		Yes	
Demographic Variables	Yes		Yes	
Macro and Social Variables	Yes		Yes	
Lambda	-2.235	(-19.93)	-0.167	(-4.64)
Constant	-10.217	(-34.80)	0.269	(22.39)
Adj R2	0.0792		0.0094	

Table 8: Dependent variable geographical proximity

We report the results for the specification where the dependent variable is our measure of geographical proximity. It is measured as value-weighted average of inverse of the logarithm 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 notations are as in Table 2. Coefficients for Specialization are multiplied by 100.

Variable	Low-wealth (1) Coeff.	t-stat	(2) Coeff.	t-stat	(3) Coeff.	t-stat	(4) Coeff.	t-stat	High-wealth (1) Coeff.	t-stat	(2) Coeff.	t-stat	(3) Coeff.	t-stat	(4) Coeff.	t-stat
Local Professional Specialization																
Competition	0.751	(32.91)	0.654	(30.63)	0.751	(32.91)	0.778	(38.49)	0.059	(0.53)	0.233	(11.43)	0.056	(0.50)	0.247	(11.90)
Specialization	0.076	(33.32)	0.042	(18.95)	0.076	(33.31)	0.065	(30.47)	0.032	(4.33)	0.041	(21.15)	0.031	(4.22)	0.046	(23.15)
Diversity	12.703	(29.92)	12.289	(30.64)	12.702	(29.92)	12.006	(31.37)	15.593	(7.40)	3.056	(7.98)	15.613	(7.40)	2.839	(7.30)
Degree of Local Isolation																
Population Density	-2.425	(-5.60)	0.493	(1.21)	-2.311	(-5.60)	-0.510	(-1.24)	-4.089	(-2.26)	0.160	(0.47)	-4.139	(-2.26)	-0.089	(-0.31)
Index of Rural Areas	0.066	(8.67)	0.055	(7.95)	0.066	(8.66)	0.021	(3.24)	0.379	(10.33)	0.029	(4.36)	0.380	(10.34)	0.015	(2.21)
Distance from Airport	0.072	(8.84)	0.091	(11.74)	0.072	(8.84)	0.078	(10.62)	0.144	(3.31)	0.047	(5.98)	0.144	(3.32)	0.040	(5.04)
Local Prosperity																
Bankruptcies	-6.051	(-17.50)	-6.020	(-21.85)	-6.050	(-17.50)	-5.632	(-24.26)	-5.001	(-24.58)	-1.533	(-53.03)	-5.013	(-24.62)	-1.572	(-62.26)
Start-ups	6.838	(24.31)	6.691	(25.56)	6.837	(24.30)	6.304	(27.62)	6.417	(32.50)	1.950	(68.48)	6.428	(32.54)	2.014	(80.01)
Sales Profitability	1.860	(30.34)	4.265	(73.93)	1.861	(30.36)	2.978	(54.41)	-3.006	(-9.53)	-1.085	(-18.96)	-2.914	(-9.24)	-2.384	(-41.56)
Return on Capital	-0.055	(-41.92)	-0.059	(-46.64)	-0.055	(-41.92)	-0.057	(-47.14)	-0.008	(-1.35)	-0.003	(-2.85)	-0.008	(-1.37)	-0.004	(-3.64)
Employment Growth	-1.271	(-12.04)	-0.929	(-9.45)	-1.271	(-12.04)	-1.178	(-12.69)	2.234	(3.99)	0.061	(0.64)	2.235	(3.99)	-0.004	(-0.04)
Active Enterprises	-1.028	(-12.89)	-1.074	(-15.16)	-1.028	(-14.90)	-1.070	(-16.56)	-0.160	(-5.25)	-0.094	(-16.51)	-0.158	(-5.16)	-0.111	(-18.80)
Financial and Municipal Services Availability																
Bank Coverage	-0.035	(-2.30)	-0.084	(-5.79)	-0.035	(-2.30)	-0.117	(-8.53)	0.706	(9.45)	0.158	(11.36)	0.705	(9.44)	0.157	(11.12)
% of Municipal Employees	13.840	(11.34)	6.508	(5.57)	13.838	(11.33)	8.853	(7.87)	37.464	(6.96)	11.405	(9.82)	37.369	(6.94)	12.909	(10.81)
Non-financial Risk and Financial Constraints																
Active Hedging Index	-1.274	(-53.87)	-1.655	(-72.06)	-1.274	(-53.86)	-1.457	(-64.69)	-1.991	(-57.31)	-1.185	(-76.94)	-1.985	(-57.15)	-1.324	(-82.93)
Borrowing Constraints	-0.012	(-1.29)	-0.014	(1.05)	-0.012	(0.74)	-0.017	(0.61)	19.797	(4.98)	7.775	(9.11)	19.775	(4.97)	8.198	(9.62)
% of Population on Welfare Asst.	-16.884	(-32.71)	-12.127	(-24.67)	-16.881	(-32.71)	-12.129	(-25.77)	-53.340	(-21.11)	-16.186	(-34.46)	-53.273	(-21.07)	-17.161	(-35.89)
Control Variables																
Wealth Variables	Yes		Yes		No		Yes		Yes		Yes		No		Yes	
Income Variables	Yes		No		Yes		Yes		Yes		No		Yes		Yes	
Momentum Variables	Yes		Yes		No		Yes		Yes		Yes		No		Yes	
Demographic Variables	Yes		Yes		Yes		No		Yes		Yes		Yes		No	
Macro and Social Variables	Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Lambda	-2.767	(-16.81)	-2.086	(-19.26)	-2.767	(-16.81)	-1.068	(-20.22)	-20.716	(-18.33)	-1.150	(-8.81)	-20.729	(-18.33)	-0.944	(-14.07)
Constant	-10.204	(-23.90)	-10.121	(-33.86)	-10.204	(-23.90)	-13.112	(-70.06)	4.090	(3.51)	-6.246	(-26.89)	4.101	(3.52)	-7.650	(-45.97)
Adj R2	0.0594		0.0569		0.0594		0.0566		0.0732		0.0724		0.0732		0.0705	

Table 9: Dependent variable professional proximity

We report the results for the specification where the dependent variable is our measure of geographical proximity. For each position it takes value 1 if any of the household members are employed in the same industry of the company they are investing in. For investor the measure is computed as value-weighted average over all position in household portfolio. The notations are as in Table 2. . Coefficients for Specialization for non-wealthy households are multiplied by 100.

Variable	Low-wealth households (1) Coeff.	t-stat	(2) Coeff.	t-stat	(3) Coeff.	t-stat	(4) Coeff.	t-stat	High-wealth households (1) Coeff.	t-stat	(2) Coeff.	t-stat	(3) Coeff.	t-stat	(4) Coeff.	t-stat
Local Professional Specialization																
Competition	0.051	(14.74)	0.040	(24.48)	0.043	(24.49)	0.010	(6.18)	0.024	(9.40)	0.010	(3.57)	0.022	(13.13)	0.012	(3.06)
Specialization	-0.001	(-0.28)	-0.001	(-0.86)	-0.001	(-0.72)	0.001	(1.25)	-5.101	(-0.78)	-8.184	(-1.29)	-6.138	(-1.51)	0.608	(2.76)
Diversity	0.162	(2.48)	0.177	(5.71)	0.175	(5.34)	0.120	(3.95)	1.455	(4.62)	1.789	(5.36)	2.275	(11.28)	0.253	(3.57)
Degree of Local Isolation																
Population Density	-0.209	(-2.36)	-0.195	(-5.09)	-0.200	(-5.48)	-0.785	(-4.80)	-1.982	(-2.61)	-1.990	(-2.80)	-1.785	(-3.92)	-1.345	(-1.17)
Index of Rural Areas	-0.270	(-0.23)	1.476	(2.71)	1.027	(1.75)	-0.020	(-0.38)	0.140	(1.63)	0.583	(6.45)	0.319	(1.99)	-0.042	(-0.35)
Distance from Airport	0.271	(0.49)	0.050	(0.87)	0.045	(0.74)	0.043	(0.76)	0.242	(2.47)	0.389	(3.72)	0.091	(1.45)	0.431	(3.07)
Local Prosperity																
Bankruptcies	-0.108	(-20.31)	-0.101	(-43.94)	-0.102	(-37.91)	-0.055	(-31.27)	0.042	(9.08)	0.022	(5.09)	0.022	(7.56)	0.019	(4.27)
Start -ups	0.113	(22.42)	0.109	(49.50)	0.110	(43.10)	0.059	(33.13)	-0.023	(-5.04)	0.006	(1.55)	-0.002	(-0.57)	-0.012	(-2.66)
Sales Profitability	-0.010	(-1.10)	0.052	(11.73)	0.041	(8.71)	-0.005	(-1.05)	0.213	(28.56)	0.273	(35.12)	0.306	(64.11)	-0.099	(-9.56)
Return on Capital	-0.001	(-5.31)	-0.001	(-10.18)	-0.001	(-9.92)	-0.001	(-7.63)	-0.001	(-10.36)	-0.001	(-7.47)	-0.002	(-21.01)	0.000	(0.20)
Employment Growth	-0.041	(-2.55)	-0.028	(-3.72)	-0.032	(-3.89)	-0.020	(-2.76)	0.017	(1.39)	0.051	(3.95)	-0.004	(-0.50)	0.029	(1.74)
Active Enterprises	-0.020	(-15.81)	-0.019	(-31.48)	-0.020	(-31.54)	-0.010	(-26.74)	-0.020	(-22.69)	-0.009	(-11.16)	-0.020	(-50.73)	0.002	(1.58)
Financial and Municipal Services Availability																
Bank Coverage	-0.006	(-2.71)	-0.004	(-3.91)	-0.005	(-4.43)	-0.004	(-3.55)	-0.020	(-11.38)	-0.009	(-5.00)	-0.018	(-16.52)	-0.015	(-5.86)
% of Municipal Employees	0.001	(0.01)	0.075	(0.84)	0.089	(0.94)	0.024	(0.27)	0.310	(2.13)	0.292	(1.91)	0.280	(3.05)	0.181	(0.83)
Non-financial Risk and Financial Constraints																
Active Hedging Index	-0.080	(-21.16)	-0.064	(-35.78)	-0.060	(-31.44)	-0.043	(-24.02)	-0.052	(-26.65)	-0.066	(-34.87)	-0.058	(-48.72)	-0.025	(-8.52)
Borrowing Constraints	0.000	(0.14)	0.000	(0.13)	0.000	(0.09)	0.000	(-0.09)	0.483	(4.46)	0.655	(5.84)	0.384	(5.63)	0.442	(2.84)
% of Population on Welfare Asst.	-0.525	(-6.66)	-0.444	(-11.76)	-0.438	(-11.01)	-0.048	(-1.30)	0.084	(1.44)	-0.225	(-3.61)	-0.059	(-1.58)	0.728	(8.49)
Control Variables																
Wealth Variables	Yes		Yes		No		Yes		Yes		Yes		No		Yes	
Income Variables	Yes		No		Yes		Yes		Yes		No		Yes		Yes	
Momentum Variables	Yes		Yes		No		Yes		Yes		Yes		No		Yes	
Demographic Variables	Yes		Yes		Yes		No		Yes		Yes		Yes		No	
Macro and Social Variables	Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Lambda	-0.088	(-3.56)	-0.102	(-10.41)	-0.092	(-7.39)	0.012	(2.81)	-0.101	(-3.72)	-0.279	(-12.93)	-0.125	(-7.20)	0.042	(3.49)
Constant	0.130	(1.99)	0.171	(6.37)	0.134	(4.10)	-0.080	(-4.86)	-0.013	(-0.41)	0.201	(6.57)	0.000	(0.02)	-0.139	(-4.73)
Adj R2	0.0058		0.0053		0.0057		0.0035		0.0202		0.0193		0.0196		0.0112	