Mutual Fund Choices and Investor Demographics

Christopher J. Malloy,
Ning Zhu*

March 16, 2004

ABSTRACT

We test the hypothesis that investment in actively managed mutual funds is linked to investor clienteles. We document that individual investors located in less affluent, less educated, and ethnic minority neighborhoods invest more in funds with expensive load fees. They also tend to invest a disproportionate amount in funds with no minimum-balance requirements, and find fewer funds and (fewer no-load funds in particular) to invest in. Our results suggest that less sophisticated neighborhoods may be faced with an inferior investment opportunity set. We find no evidence that investors in these areas tend to trade significantly more or are more likely to redeem their fund holdings each year.

*Malloy is at London Business School; Zhu is at the Graduate School of Management, UC Davis. We thank seminar participants at London Business School for helpful comments and suggestions. Please send correspondence to: Christopher Malloy, London Business School, Sussex Place, Regent's Park, London NW1 4SA, UK, phone: 44-207-262-5050 x3278, email: cmalloy@london.edu.

Mutual Fund Choices and Investor Demographics

ABSTRACT

We test the hypothesis that investment in actively managed mutual funds is linked to investor clienteles. We document that individual investors located in less affluent, less educated, and ethnic minority neighborhoods invest more in funds with expensive load fees. They also tend to invest a disproportionate amount in funds with no minimum-balance requirements, and find fewer funds and (fewer no-load funds in particular) to invest in. Our results suggest that less sophisticated neighborhoods may be faced with an inferior investment opportunity set. We find no evidence that investors in these areas tend to trade significantly more or are more likely to redeem their fund holdings each year.

Mutual Fund Choices and Investor Demographics

We explore the puzzle of investment in actively managed mutual funds. We focus on one particular form of this investment, namely the decision to invest in funds with expensive load fees. Given that actively managed funds on average offer a negative risk-adjusted return, and that load funds perform *worse* than most other actively managed funds, this type of investment decision presents a portfolio allocation puzzle.

We test the hypothesis, developed in Gruber (1996), that the existence of investor clienteles can explain this puzzle.[1]Gruber (1996) argues that a "disadvantaged clientele" made up of unsophisticated investors will direct its money to funds based on advertising and brokerage advice, rather than based on performance. Only sophisticated investors will direct their money based on performance. As a result, we see money remain in funds that are predicted to do poorly and do in fact perform poorly. Since load fees are generally used to pay advertising costs, and since load funds have historically performed worse than other funds (see Carhart (1997)), investments into load funds provide an ideal laboratory to evaluate Gruber's (1996) hypothesis.

Using a unique dataset that merges individual accounts data from a large discount brokerage house with U.S. Census data and items from a variety of mutual fund data sources, we examine the mutual fund choices of different classes of investors. Specifically, we aggregate investment choices by zip code and use detailed Census data on the characteristics of each zip code to test the idea that investment in load funds is linked to investor clienteles. Our identifying assumption is that investors in similar neighborhoods share similar characteristics, and that the characteristics of an investor's neighborhood may influence or define (or are correlated with factors that influence or define) an investor's portfolio allocation decision.

Consistent with Gruber's (1996) conjecture, we provide evidence that individual investors located in less affluent, less educated, and ethnic minority neighborhoods invest more in mutual funds with expensive load fees. The magnitude of this effect is modest, but nontrivial. A $10,000 increase in the median income of a neighborhood corresponds to a 1.2 percent decrease in the share of mutual fund investment that is directed to load funds; our estimate of the overall dollar

[1]Nanda, Narayanan, and Warther (2000) formalize this intuition in a model in which management fees, loads, and returns are determined endogenously.

effect is \$11,200 for two zip codes with a \$10,000 difference in median income. Generalizing this figure across the population highlights the potential economic importance of our study.

We explore some explanations for our main result, and find that neighborhoods that invest more in load funds tend to invest a disproportionate amount in funds with no minimum-balance requirements. They also find fewer funds and (fewer no-load funds in particular) to invest in. Our results suggest that less sophisticated neighborhoods may simply be faced with an inferior investment opportunity set. However, we find no evidence that investors in these areas tend to trade significantly more or are more likely to redeem their fund holdings each year, casting doubt on the interpretation that fund companies optimally entice less sophisticated investors to buy expensive load funds because these are their costly clients.

The paper is organized as follows. Section I provides some background and a description of the data, while Section II presents our empirical design and results. Section III concludes.

I. Methodology

A. Background and Motivation

The empirical evidence on actively managed mutual fund performance, and load funds in particular, is not impressive. While Ippolito (1989) finds that load funds have significantly higher average returns than no-load funds, this finding is challenged in Elton, Gruber, Das, and Hlavka (1993), who find problems with the paper's methodology. Carhart (1997) finds no support for the existence of skilled or informed mutual fund managers, and provides evidence that load funds underperform no-load funds by approximately 80 basis points per year (a figure which ignores the load fees themselves). Finally, both Barber, Odean, and Zheng (2003) and Gruber (1996) find no significant difference, on average, between load and no-load performance. Thus there is little evidence that load funds, on average, outperform no-load funds.[2] And yet load funds are still widely held, as shown in Figure 1 for our sample (and in Barber, Odean, and Zheng (2003)). The

[2] Elton, Gruber, and Busse (2003) show that load index funds are clearly a dominated passive investment choice as well.

question of why investors hold what appears to be such an inferior class of investment is therefore puzzling.

We explore the notion, put forth in Gruber (1996) and developed in Nanda, Narayanan, and Warther (2000), that the existence of investor clienteles can help resolve this puzzle. Gruber (1996) argues that a group of sophisticated investors seems to recognize the fact that future performance is in part predictable from past performance, and hence directs its money to exploit this fact. By contrast, a disadvantaged clientele (composed of unsophisticated investors, institutionally disadvantaged investors, and tax disadvantaged investors) does not. In particular, unsophisticated investors direct their money based at least in part on advertising and advice from brokers.[3]

We use load funds as the testing ground to evaluate this hypothesis. As noted by Elton, Gruber, and Busse (2003) and Sirri and Tufano (1998), since load fees are used primarily for marketing purposes, the fact that load funds are held is evidence that marketing works. In this vein, we identify those investors among whom this marketing effort is successful. The paper closest in spirit to ours is Capon, Fitzsimons, and Prince (1996), who use self-reported survey data to group investors into different categories and assess their knowledge and understanding of mutual fund specifics. By contrast, we use actual transactions and holdings data, aggregated at the zip-code level, to explore the possibility that clientele effects exist at the neighborhood level.

Our results link to several strands of the literature. The importance of trying to distinguish investor types has only recently been emphasized. Barber and Odean (2001) find that men trade more actively than (and perform worse) women, and Barber and Odean (2002) show that online investors trade more actively (and perform worse) than phone-based investors. Malmendier and Shanthikumar (2003) that small, presumably naive investors do not properly account for the biased incentives of affiliated equity analysts. Our results are also consistent with anecdotal evidence that mutual fund marketing managers target (and exploit) different segments of the investor population, although we provide no direct evidence that this is the case. Zhao (2003), for

[3]Goetzmann, Greenwald, and Huberman (1992) make a similar conjecture regarding the existence of a class of naive mutual fund investors.

example, argues that brokers and financial advisors ultimately serve as the true decision makers behind investments into load funds.

We also build on a growing literature linking social networks and geography to portfolio choice. Brown, Izkovic, Smith, and Weisbenner (2003) find that individuals are influenced by the investment behavior of members of their community, and Hong and Stein (2004) present a model in which stock market participation may be influenced by social interaction. Hong, Kubik, and Stein (2002) link geographic proximity and mutual fund managers as in Coval and Moskowitz (2001), and focus on the information that investors that are close together can pass on to one another; they document a word-of-mouth effect whereby local mutual fund managers are more likely to hold a particular stock if other managers from different fund families in the same city are holding that same stock.[4]

B. Data

Our sample merges three main datasets. The first is a large dataset of individual investor transactions, provided by a major U.S. discount brokerage firm. This individual investor data contains detailed information on the positions and trades of more than 78,000 households during the period from January, 1991 to November, 1996. The dataset contains buy and sell transactions executed by all investors in a given month (in the Trade file), end-of-month portfolios of all investors (in the Positions file), and some limited demographic information–zip code, age, occupation, and income range–for a subset of investors (in the Demographics file). For 27,189 households, we have information on the residence zip code.

In an average month, households in our sample hold more than 2 billion dollars of securities in their portfolios. The average account size is $35,629 (median=$13,869). Households trade a wide variety of securities, including domestic equities, fixed income securities, mutual funds, and foreign assets. Out of about 3 million trades executed by all households, this study focuses

[4]Coval and Moskowitz (1999), Zhu (2002), Huberman (2001), and Grinblatt and Keloharju (2001) also report strong preferences for geographically local equities among investors.

on the 18 percent of all trades that took place in U.S. mutual funds. Of the 78,000 households, 32,199 (41 percent) had positions in mutual funds during at least one month.

Of the 32,199 households with positions in mutual funds the average held 3.6 funds worth $36,988 (median=2, and $12,844, respectively). For these households, 42 percent of the market value in these accounts was held in mutual funds. In aggregate, these households held 1,073 mutual funds worth $1.4 billion in December 1996.

We match each household zip code to data drawn from the 1990 U.S. Census, our second main data source. Using Census data allows us to explore a wealth of demographic variables, but of course forces us to group investors by neighborhood. We collect a wide array of demographic information for each zip code, such as median income, the fraction of the zip code's population with college education, the fraction of the zip code's population in professional occupations, and the fraction of the zip code's population in certain ethnic groups (e.g., white, Hispanic, and black).

The issue of sampling bias is important in our dataset, since we are only looking at a subset of transactions within each zip code. Our sample contains investors trading from 2748 zip codes, and the mean number of investors per zip code is 12. Using a small subsample of investors for whom we have individual income data, we can report that our investors are significantly wealthier than the median income of their neighborhood, but that this result holds for all levels of zip code income.[5]

We merge the information from our two main data sources with data drawn from the CRSP Mutual Fund Database and Morningstar. The CRSP and Morningstar databases contain detailed information on the specific mutual funds held by our individual investors. We use CRSP to collect data on fund returns, fund objective, fees, and fund turnover. We use Morningstar to collect data on minimum-balance requirements, sales channel, and fund location.

Panel A of Table I presents some summary statistics of our sample. Note that we first identify all investors who have any fund holdings during the sample, then cumulate (by zip code) all investments made into mutual funds and load mutual funds. Thus the figures here represent

[5]These results are available on request.

averages across zip codes. We see that 20.4 percent of the total fund investment in our sample is allocated to mutual funds with load fees ($PCTLOAD$). In Figure 1, we cumulate all mutual fund and load mutual fund investments in the universe of our sample by month, and see that the monthly fraction invested load funds has remained fairly steady over the past few years. The average zip code (median) income ($MEDINC$) is \$40,063.

Correlations of the main variables used in our sample are reported in Panel B of Table I. By way of preview, Panel B shows that $PCTLOAD$ is negatively correlated with $MEDINC$, $PCTCOL$ (the share of a zip code's population with college education), and $PCTWHT$ (the share of a zip code's population that is white). Not surprisingly, and consistent with well-known results, $MEDINC$ is positively correlated with $PCTCOL$, $PCTPRF$ (the share of a zip code's population in professional occupations), and $PCTWHT$.

II. Empirical Design and Results

A. Load-Fund Investment and Neighborhood Income, Education, and Ethnicity

Our first test of the hypothesis that load-fund investment is linked to investor clienteles is a cross-sectional regression of the share of total fund investment in load funds for a given zip code on the characteristics of the neighborhood of the investor (defined by zip code). These characteristics are measured as of 1990, before the start of our investor accounts data (January 1991), but our results are similar if we use data from the 2000 U.S. Census data instead. We focus on four demographic variables in an attempt to sort neighborhoods by level of sophistication: 1) $MEDINC$, the median income of the population in a zip code, 2) $PCTCOL$, the fraction of the zip code's population with college education, 3) $PCTPRF$, the fraction of the zip code's population in professional occupations, and 4) $PCTWHT$, the fraction of the zip code's population that is white. We also include a dummy variable ($METRO$) equal to one if a zip code is within 25 miles of at least one of the top twenty metropolitan areas, and zero otherwise. This variable captures the idea that metropolitan neighborhoods may contain more sophisticated residents. Finally, we include a dummy variable (*CALIF*) equal to one if a zip code is located in California, and zero otherwise; excluding this variable has virtually no effect on our results.

Our baseline cross-sectional regression specification is as follows:

$$(1)\ PCTLOAD_i = \beta_0 + \beta_1 MEDINC_i + \beta_2 PCTCOL_i + \beta_3 PCTPROF_i + \beta_4 PCTWHITE_i + \beta_5 METRO_i + \beta_6 CALIF_i + \varepsilon_i.$$

As noted above, since we aggregate investment choices by zip code, our identifying assumption is that investors in similar neighborhoods share similar characteristics, and that the characteristics of an investor's neighborhood may influence or define (or are correlated with factors that influence or define) an investor's portfolio allocation decision.

B. Baseline Results

Table II presents estimates of pooled, cross-sectional regressions of equation (1). Panel A shows that the coefficient estimate on *MEDINC* is significantly negative in univariate (t=-2.95) and multivariate (t=-2.37) specifications, indicating that more wealthy neighborhoods invest a smaller share of their total fund investment into funds with load fees. The magnitude of this effect is modest, but nontrivial. A $10,000 increase in the median income of a neighborhood corresponds to a 1.2 percent decrease in the amount invested in load funds.

Drawing from the 1995 Survey of Consumer Finances and assuming an average zip code with 10,000 in population invests 22.4 million in mutual funds (with an average load of 4.36 percent), then a 1.2 percent change in the fraction of load investment leads to a 5 basis point difference in mutual fund investment; this translates into to a $11,200 difference between two zip codes with a $10,000 difference in median income. More generally, if we assume that of the 240 million population in the U.S. that 80 million people live in poor neighborhoods and 80 million live in affluent neighborhoods, then a $10,000 difference between the two areas (the difference is surely far larger in practice) leading to a 5 basis point change in fund investment could lead to large savings for the 80 million people in poor neighborhoods investing an average of $2,200 (although this number is surely smaller for poor neighborhoods) in mutual funds.

The coefficient estimates on *PCTCOL*, *PCTWHT*, and *METRO* shown in Panel A of Table II are also negative and significant. The coefficient on *PCTPRF* is insignificant. Overall, these results depict a strong link between investor clienteles and load-fund investment. Investors from less educated, non-white, and non-metropolitan areas tend to place more of their mutual fund investments in load funds than other investors.

Although our data is a panel, and hence contains a time-series dimension that is not explicitly modelled in this specification, it turns out that cross-correlation in the residuals does not plague our results. For example, our results are robust to a Fama and MacBeth (1973) yearly specification.[6] In addition, our results are very similar if we use new purchases (i.e., trades) instead of holdings (i.e., positions) to compute the dependent variable (*PCTLOAD*).

C. Regression Broken Down by Ethnicity

Panel B of Table II reports estimates for different categories of ethnicity. Univariate regressions of *PCTLOAD* on *PCTHIS*, the fraction of a zip code's population that is Hispanic, and *PCTBLK*, the fraction of a zip code's population that is black, show that investors from neighborhoods with a higher percentage of Hispanics or blacks invest more of their fund wealth in load funds. However, this result is only significant for Hispanic neighborhoods (t=2.51). Controlling for *MEDINC*, *PCTPRF*, *METRO*, and *CALIF*, the significantly positive relation (t=2.40) between *PCTLOAD* and *PCTHIS* persists.

D. Robustness

These main results are robust to a variety of permutations. For brevity, we only report a few such checks. Table III presents results for sub-samples of our standard sample. Limiting the sample to only those zip codes that contain at least five investors from the individual accounts data, as shown in Panel A, increases the magnitude of the effect of a zip code's median income and education level (-0.015, t=-2.60 and -0.532, t=-1.94, respectively), but dampens the impact

[6] These and other untabulated results are available on request.

of a neighborhood's ethnicity.[7] When we change the dependent variable to *PCTFEES*, the share of fund investment in funds with either load fees *or* 12b-1 fees (which are generally used for marketing purposes as well), we report similar results.

Panel B presents results for a sub-sample of fund investments in equity funds only (while still subject to the five investor per zip code threshold introduced above). The results are very similar to those originally reported in Table II. The coefficient estimates for *MEDINC* and *PCTCOL* are again negative and significant (t=-1.94 and t=-2.06, respectively).

E. Regressions Using Minimum-Balance Requirements

Given the wealth of evidence that load funds underperform (or at best, do not outperform) no-load funds, the question now becomes: why do these "disadvantaged clienteles" make this seemingly sub-optimal investment decision? One possibility, which we explore in this section, is that these clienteles are constrained from investing in other funds. For example, disadvantaged clienteles may invest in load funds because these are the funds with no minimum-balance requirements.

We examine this possibility by collecting data from Morningstar on the minimum-balance requirements of all the mutual funds in our sample (as of December 1995).[8] Table IV reports the coefficient estimates of our original equation (1), except that here we replace the dependent variable *PCTLOAD* with *PCTBAL*, equal to the share of fund investment that is directed to funds with a minimum-balance requirement. As shown in Panel A, the coefficient on *MEDINC* is now positive and significant (t=2.45), indicating that more affluent neighborhoods direct more of their fund investment into funds with a minimum-balance requirement. The coefficients on *PCTCOL* and *PCTWHT* are also positive, but insignificant. Including *PCTLOAD* as an independent variable in this regression shows that the share of fund investment in load funds is strongly negatively related to the share of fund investment in funds with a balance requirement; however, controlling for *PCTLOAD* does not kill the predictive power of *MEDINC* in this regression. These results

[7]Changing this admittedly arbitrary threshold to 10 investors per zip code changes no conclusions.

[8]These requirements are fairly stable. From 1995 to 1998, only 20 percent of all funds change their minimum balance requirements. Excluding those funds that change their requirements does not affect the results presented in this section.

are consistent with the idea that disadvantaged clienteles (at least less affluent clienteles) face an inferior investment opportunity set. Much of their fund investment may go into load funds simply because these are the only funds with no minimum-balance requirements.

The last row of Panel A shows that including $PCTBAL$ as an independent variable in the original equation (1) does not kill the explanatory power of $MEDINC$ and $PCTWHT$, but does dampen the predictive ability of $PCTCOL$. The coefficient estimate on $PCTBAL$ (-0.222, t=-8.11) in this specification again indicates that the share of investment in funds with a balance requirement is strongly negatively related to the share of fund investment in load funds. Panel B shows that all of these results are robust to the introduction of a five investor per zip code threshold.

F. Number of Funds Available

Another way of investigating the limited opportunity story is to merge data from the Trade file of the individual accounts data with data from the Position file. Table V presents results from such an exercise. Specifically, we divide all zip codes from our sample into six equal categories (i.e., each category contains an approximately equal number of zip codes) according to the share of fund investment directed into load funds. Category I contains zip codes that invest very little in load funds, while Category VI contains zip codes that invest a lot in load funds. These sorts are all done using data from the Position file, which contains the end-of-month holdings of each individual investor in the sample. We then collect data on all the unique mutual fund trades (from the Trade file) for each investor in each zip code, and sum them up over each of the six pre-defined categories.

As shown in Panel A for the entire sample of funds, zip codes that invest a higher fraction of their fund wealth in load funds trade fewer unique funds overall (182 for Category VI, compared to 755 for Category I), and trade a lower fraction and number of unique no-load funds overall (35.2 percent and 64, compared to 43.0 percent and 325).[9] Meanwhile, Category I zip codes

[9] The number of load and no-load funds does not add up to the total number of funds because load information is unavailable for many funds. This should in no way bias our results, since it is unlikely that the distribution of funds

trade a far smaller fraction of load funds than Category VI zip codes (11.0 percent compared to 51.1 percent). Panel B presents similar results for a sub-sample of equity only mutual funds.

These results indicate that zip codes that invest a larger share of their fund wealth in load funds invest in fewer funds and fewer no-load funds in particular, consistent with the limited opportunity story explored in the last section. Disadvantaged clienteles, which invest disproportionately in seemingly unattractive load funds, may do so because load funds are the only funds in their investment opportunity set. For example, Panel C shows that Category VI zip codes, those with high load fund investment shares, barely ever (less than one percent of the time) trade mutual funds that have minimum balance requirements, while Category I zip codes trade minimum-balance funds 18.2 percent of the time.

G. Turnover and Likelihood of Redemption

The follow-up issue is of course *why* these disadvantaged clienteles face an inferior opportunity set. One standard explanation is that these investors are likely to have high liquidity needs of the sort modelled in Nanda, Narayanan, and Warther (2000). Since redemptions can be costly to fund managers, funds will induce these high-cost customers to buy high-cost products (i.e., funds with expensive load fees) by attaching minimum balance requirements to cheaper, no-load options.

We test the underlying assumption that load-fund investors are undesirable or more costly by examining the likelihood of redemption and portfolio turnover activity of different classes of investors. Specifically, we replace the dependent variable in equation (1) with a) $REDEEM$, the average yearly mutual fund redemption likelihood (defined as total sales divided by start-of-year position) within a given zip code, and then b) $PORTTURN$, the average monthly investor turnover of the mutual fund portion of a portfolio within a given zip code. We then run cross-sectional regressions of redemption likelihood and investor portfolio turnover on neighborhood characteristics.

for which load information is unavailable is systematically different from that of funds for which load information is available.

Table VI presents the coefficient estimates. In the regression specification with *REDEEM* as the dependent variable, none of the clientele variables are significant, and the signs on *MEDINC* and *PCTCOL* indicate that less affluent and less educated neighborhoods are *less* likely to redeem their fund holdings. The results with *PORTTURN* are similarly underwhelming; the only significant variable is *PCTPRF*, and the coefficient estimate is positive, which works in the opposite direction of the liquidity story. Untabulated results also indicate little differences in our main results when accounts are divided up into taxable and tax-deferred groups. These results do not support the contention that disadvantaged clienteles are more costly due to increased likelihood of redemption and higher portfolio turnover.

III. Conclusion

This paper explores the link between investment in actively managed mutual funds and investor clienteles. We test the hypothesis, developed in Gruber (1996), that unsophisticated investors will direct their money to funds based on advertising and brokerage advice, rather than based on performance. Since load fees are generally used to pay advertising costs, we use investments into load funds as the setting for our experiment. Consistent with Gruber's (1996) conjecture, we find that individual investors located in less affluent, less educated, and ethnic minority neighborhoods direct more of their fund wealth into funds with expensive load fees. They also tend to invest a disproportionate amount in funds with no minimum-balance requirements, and find fewer funds and (fewer no-load funds in particular) to invest in. Our results suggest that less sophisticated neighborhoods may be faced with an inferior investment opportunity set. However, we find no evidence that investors in these areas trade significantly more or are more likely to redeem their fund holdings each year.

Thus the question of why these "disadvantaged clienteles" face an inferior opportunity set remains unanswered. Do funds target and exploit unsophisticated investors, or are they simply rationally structuring and marketing their products? Anecdotal evidence suggests that mutual fund marketing managers use demographic information to attract clients and structure clients; the zip code of potential investors is a commonly used input. As such, the extent to which funds target certain investor clienteles, or alter their fees to optimally entice investment into certain products are fascinating subjects for future research.

References

Barber, Brad, and Terrance Odean, 2001, Boys will be boys: Gender, overconfidence, and common stock investment, *Quarterly Journal of Economics* 116, 261–292.

Barber, Brad, and Terrance Odean, 2002, Online investors: Do the slow die first?, *Review of Financial Studies* 15, 455–487.

Barber, Brad, Terrance Odean, and Lu Zheng, 2003, Out of sight, out of mind: The effects of expenses on mutual fund flows, Working paper, UC-Davis.

Brown, Jeffrey, Zoran Izkovic, Paul Smith, and Scott Weisbenner, 2003, The geography of stock market participation: The influence of communities and local firms, Working paper, University of Illinois at Urbana-Champaign.

Capon, Noel, Gavan Fitzsimons, and Russ Alan Prince, 1996, An individual level analysis of the mutual fund decision, *Journal of Financial Services Research* 10, 59–82.

Carhart, Mark M., 1997, On persistence in mutual fund performance, *The Journal of Finance* 52, 57–82.

Coval, Joshua D., and Tobias J. Moskowitz, 1999, Home bias at home: Local equity preference in domestic portfolios, *The Journal of Finance* 54, 2045–2073.

Coval, Joshua D., and Tobias J. Moskowitz, 2001, The geography of investment: Informed trading and asset prices, *Journal of Political Economy* 4, 811–841.

Elton, Edwin, Martin Gruber, and Jeffrey Busse, 2003, Are investors rational? Choices among index funds, *The Journal of Finance* 59, 261–288.

Elton, E., M. Gruber, S. Das, and M. Hlavka, 1993, Efficiency with costly information: a reinterpretation of evidence from managed porfolios, *Review of Financial Studies* 6, 1–22.

Fama, Eugene F., and James MacBeth, 1973, Risk, return and equilibrium: empirical tests, *Journal of Political Economy* 51, 55–84.

Goetzmann, William, Bruce Greenwald, and Gur Huberman, 1992, Market reponse to mutual fund performance, Working paper, Columbia Business School.

Grinblatt, Mark, and Matti Keloharju, 2001, Distance, language, and culture bias: The role of investor sophistication, *The Journal of Finance* 56, 1053–1073.

Gruber, Martin J., 1996, Another puzzle: The growth in actively managed mutual funds, *The Journal of Finance* 51, 783–810.

Hong, Harrison, Jeffrey D. Kubik, and Jeremy C. Stein, 2002, Thy neighbor's portfolio: word-of-mouth effects in the holdings and trades of money managers, Working paper, Stanford University.

Hong, Harrison, and Jeremy Stein, 2004, Social interaction and stock market participation, *The Journal of Finance* 59, 137–164.

Huberman, Gur, 2001, Familiarity breeds investment, *Review of Financial Studies* 14, 659–680.

Ippolito, R., 1989, Efficiency with costly information: A study of mutual fund performance, *Quarterly Journal of Economics* 104, 1–23.

Malmendier, Ulrike, and Devin Shanthikumar, 2003, Are small investors naive?, Working paper, Stanford University.

Nanda, Vikram, M. P. Narayanan, and Vincent Warther, 2000, Liquidity, investment ability, and mutual fund structure, *Journal of Financial Economics* 57, 417–443.

Sirri, Erik, and Peter Tufano, 1998, Costly search and mutual fund flows, *The Journal of Finance* 53, 1589–1622.

Zhao, Xinge, 2003, The role of brokers and financial advisors behind investments into load funds, Working paper, College of William and Mary.

Zhu, Ning, 2002, The local bias of individual investors, Working paper, Yale University.

Table I
Summary Statistics for Individual Investor Accounts and 1990 Census Data

Panel A reports summary statistics for the main variables used in the paper. $PCTLOAD$ is the fraction of mutual fund investments (in dollars) made by investors in a given zip code that are invested in funds with load fees; $PCTFEES$ is equal to the fraction that is invested in funds with load fees or 12b-1 fees. $FUNDTURN$ is the monthly turnover ratio of each mutual fund. All fee and turnover data is obtained from CRSP. $MEDINC$, $PCTCOL$, $PCTPRF$, $PCTWHT$, $PCTHIS$, and $PCTBLK$ are variables equal to the median income of the population in a zip code, the fraction of the zip code's population with college education, the fraction of the zip code's population in professional occupations, the fraction of the zip code's population that is white, the fraction of the zip code's population that is Hispanic, and the fraction of the zip code's population that is black; zip code demographic data are obtained from the 1990 U.S. Census. Panel B reports correlations for the above variables.

Panel A: Summary Statistics

Variable	No. of Obs.	Mean	Med	Std. Dev.	Min	Max
$PCTLOAD$	2748	0.204	0.000	0.329	0.000	1.000
$PCTFEES$	2748	0.278	0.099	0.352	0.000	1.000
$FUNDTURN$	2680	0.040	0.005	0.104	0.000	0.955
$MEDINC$	2748	40063.1	37507.5	15294.1	5123.0	150001.0
$PCTCOL$	2734	0.313	0.316	0.082	0.033	0.750
$PCTPRF$	2734	0.198	0.194	0.064	0.000	1.000
$PCTWHT$	2734	0.817	0.883	0.189	0.013	1.000
$PCTHIS$	2734	0.070	0.029	0.109	0.000	0.904
$PCTBLK$	2734	0.065	0.020	0.123	0.000	0.982

Panel B: Correlations

Variable	A	B	C	D	E	F	G	H	I
A=$PCTLOAD$	1	0.850	-0.005	-0.058	-0.054	-0.019	-0.061	0.056	0.041
B=$PCTFEES$		1	-0.005	-0.043	-0.041	0.001	-0.049	0.040	0.040
C=$FUNDTURN$			1	-0.015	-0.006	-0.007	0.007	-0.007	-0.003
D=$MEDINC$				1	0.522	0.448	0.262	-0.180	-0.300
E=$PCTCOL$					1	0.564	0.274	-0.207	-0.292
F=$PCTPRF$						1	0.125	-0.172	-0.082
G=$PCTWHT$							1	-0.663	-0.636
H=$PCTHIS$								1	0.004
I=$PCTBLK$									1

Table II
Regressions of Load-Fund Investment Share on Neighborhood Income, Education, and Ethnicity

Pooled, cross-sectional regressions are run using individual investor accounts data from January 1991 to November 1996. This table reports the estimated coefficients from regressions of the share of individual investors' fund investments in load funds on characteristics of the neighborhood of the investor (defined by zip code). The dependent variable *PCTLOAD* is the fraction of total mutual fund investments (in dollars) by investors in a given zip code that are invested in funds with load fees; data on loads is obtained from CRSP. *MEDINC*, *PCTCOL*, *PCTPRF*, *PCTWHT*, *PCTHIS*, and *PCTBLK* are variables equal to the median income of the population in a zip code, the fraction of the zip code's population with college education, the fraction of the zip code's population in professional occupations, the fraction of the zip code's population that is white, the fraction of the zip code's population that is Hispanic, and the fraction of the zip code's population that is black; zip code demographic data are obtained from the 1990 U.S. Census. *METRO* is a dummy variable equal to one if a zip code is within 25 miles of at least one of the top twenty metropolitan areas, and zero otherwise. *CALIF* is a dummy variable equal to one if a zip code is located in California, and zero otherwise. t-statistics, adjusted for autocorrelation and heteroskedasticity, and the number of observations (n) are in parentheses.

Panel A: Baseline Regressions

Dep. Var.	*MEDINC*	*PCTCOL*	*PCTPRF*	*PCTWHT*	*METRO*	*CALIF*
PCTLOAD	-0.012					
(n=2,747)	(-2.95)					
PCTLOAD		-0.215				
(n=2,733)		(-2.87)				
PCTLOAD			-0.078			
(n=2,733)			(-0.71)			
PCTLOAD	-0.012		0.121	-0.097	-0.051	0.002
(n=2,733)	(-2.47)		(0.95)	(-2.42)	(-2.45)	(0.10)
PCTLOAD		-0.224	0.157	-0.089	-0.044	0.007
(n=2,733)		(-2.10)	(1.11)	(-2.13)	(-2.13)	(0.10)
PCTLOAD	-0.012	-0.168	0.115	-0.087	-0.050	0.007
(n=2,733)	(-2.37)	(-1.95)	(0.91)	(-2.08)	(-2.41)	(0.10)

Panel B: Regressions Broken Down by Ethnicity

Dep. Var.	*MEDINC*	*PCTCOL*	*PCTPRF*	*PCTWHT*	*PCTHIS*	*PCTBLK*	*METRO*	*CALIF*
PCTLOAD				-0.094				
(n=2,733)				(-2.63)				
PCTLOAD					0.159			
(n=2,733)					(2.51)			
PCTLOAD						0.090		
(n=2,733)						(1.69)		
PCTLOAD	-0.012		0.147		0.175		-0.478	-0.003
(n=2,733)	(-2.63)		(1.12)		(2.40)		(-2.33)	(-0.16)

Table III
Robustness Checks

Pooled, cross-sectional regressions are run using individual investor accounts data from January 1991 to November 1996, for sub-samples of Table II. Panel A restricts the sample from Table II to only those zip codes with at least five investors, while Panel B further restricts the sample to investments in equity funds. The table reports the estimated coefficients from regressions of the share of individual investors' fund investments in load funds on characteristics of the neighborhood of the investor (defined by zip code). The dependent variable $PCTLOAD$ is the fraction of total mutual fund investments (in dollars) by investors in a given zip code that are invested in funds with load fees; data on loads is obtained from CRSP. An alternate specification uses $PCTFEES$ as the dependent variable, equal to the fraction of total mutual fund investments (in dollars) by investors in a given zip code that are invested in funds with load fees or 12b-1 fees. $MEDINC$, $PCTCOL$, $PCTPRF$, and $PCTWHT$ are variables equal to the median income of the population in a zip code, the fraction of the zip code's population with college education, the fraction of the zip code's population in professional occupations, and the fraction of the zip code's population that is white; zip code demographic data are obtained from the 1990 U.S. Census. $METRO$ is a dummy variable equal to one if a zip code is within 25 miles of at least one of the top twenty metropolitan areas, and zero otherwise. $CALIF$ is a dummy variable equal to one if a zip code is located in California, and zero otherwise. t-statistics, adjusted for autocorrelation and heteroskedasticity, and the number of observations (n) are in parentheses.

Panel A: Sub-Sample of Zip Codes with at Least Five Investors						
Dep. Var.	*MEDINC*	*PCTCOL*	*PCTPRF*	*PCTWHT*	*METRO*	*CALIF*
PCTLOAD	-0.012		0.020	-0.084	-0.050	-0.014
(n=1,795)	(-2.15)		(0.12)	(-1.58)	(-2.27)	(-0.78)
PCTLOAD		-0.432	0.113	-0.042	-0.041	-0.000
(n=1,795)		(-2.98)	(0.68)	(-0.75)	(-1.88)	(-0.01)
PCTLOAD	-0.015	-0.532	-0.038	-0.39	-0.048	0.002
(n=1,795)	(-2.60)	(-1.94)	(-0.21)	(-0.70)	(-2.17)	(0.12)
PCTFEES	-0.011		0.128	-0.096	-0.037	-0.030
(n=1,795)	(-1.85)		(0.78)	(-1.68)	(-1.48)	(-1.55)
PCTFEES		-0.439	0.232	-0.051	-0.028	-0.016
(n=1,795)		(-3.00)	(1.45)	(-0.86)	(-1.12)	(-0.77)

Panel B: Sub-Sample of Stock-Only Funds and Zip Codes with at Least Five Investors						
Dep. Var.	*MEDINC*	*PCTCOL*	*PCTPRF*	*PCTWHT*	*METRO*	*CALIF*
PCTLOAD	-0.010		0.079	-0.160	-0.087	-0.014
(n=1,608)	(-1.94)		(0.46)	(-2.45)	(-0.30)	(-0.62)
PCTLOAD		-0.325	-0.035	-0.098	-0.004	0.015
(n=1,608)		(-2.06)	(-0.21)	(-1.53)	(-0.136)	(0.67)

Table IV
Regressions Using Minimum-Balance Requirements

Pooled, cross-sectional regressions are run using individual investor accounts data from January 1991 to November 1996, and Morningstar data from 1995. This table reports the estimated coefficients from regressions of the share of individual investors' fund investments in funds with either a) a minimum balance requirement, or b) funds with load fees, on characteristics of the neighborhood of the investor (defined by zip code). *PCTBAL* is the fraction of total mutual fund investments (in dollars) by investors in a given zip code that are invested in funds with minimum balance requirements; data on minimum balance requirements is obtained from Morningstar. *PCTLOAD* is the fraction of total mutual fund investments (in dollars) by investors in a given zip code that are invested in funds with load fees; data on loads is obtained from CRSP. *MEDINC*, *PCTCOL*, *PCTPRF*, and *PCTWHT* are variables equal to the median income of the population in a zip code, the fraction of the zip code's population with college education, the fraction of the zip code's population in professional occupations, and the fraction of the zip code's population that is white; zip code demographic data are obtained from the 1990 U.S. Census. *METRO* is a dummy variable equal to one if a zip code is within 25 miles of at least one of the top twenty metropolitan areas, and zero otherwise. *CALIF* is a dummy variable equal to one if a zip code is located in California, and zero otherwise. *t*-statistics, adjusted for autocorrelation and heteroskedasticity, and the number of observations (*n*) are in parentheses.

Panel A: Baseline Regressions								
Dep. Var.	*MEDINC*	*PCTCOL*	*PCTPRF*	*PCTWHT*	*METRO*	*CALIF*	*PCTLOAD*	*PCTBAL*
PCTBAL	0.012	0.241	-0.084	0.034	0.034	-0.019		
(*n*=1,510)	(2.45)	(1.08)	(-0.62)	(0.64)	(1.24)	(-0.91)		
PCTBAL	0.011	0.218	-0.021	0.008	0.021	-0.022	-0.189	
(*n*=1,510)	(1.92)	(1.00)	(-0.16)	(0.15)	(0.76)	(-1.04)	(-8.11)	
PCTLOAD	-0.014	-0.069	0.031	-0.130	-0.064	-0.017		-0.222
(*n*=1,510)	(-1.86)	(-0.29)	(1.29)	(-2.31)	(-2.18)	(-0.75)		(-8.11)

Panel B: Sub-Sample of Zip Codes with at Least Five Investors								
Dep. Var.	*MEDINC*	*PCTCOL*	*PCTPRF*	*PCTWHT*	*METRO*	*CALIF*	*PCTLOAD*	*PCTBAL*
PCTBAL	0.020	0.224	-0.146	0.011	0.035	-0.031		
(*n*=1,115)	(2.87)	(0.66)	(-0.89)	(0.15)	(1.16)	(-1.29)		
PCTBAL	0.016	0.118	-0.104	-0.012	0.023	-0.036	-0.193	
(*n*=1,115)	(2.32)	(0.33)	(-0.65)	(-0.17)	(0.77)	(-1.49)	(-6.82)	
PCTLOAD	-0.014	-0.06	0.024	-0.155	-0.058	-0.044		-0.209
(*n*=1,510)	(-2.10)	(-0.37)	(1.47)	(-2.35)	(-1.91)	(-1.97)		(-6.85)

Table V
Mutual Fund Investment Opportunities

Using merged data from the position file of the individual accounts data and the 1990 U.S. Census, zip code areas are sorted into six categories based on the fraction of total mutual fund investments (in dollars) by investors in a given zip code that are invested in funds with load fees. Each category contains an approximately equal amount of zip codes, except in Panel C where Category I combines the bottom two categories used in Panels A and B. Using merged data from the trade file of the individual accounts data and the 1990 U.S. Census, the number of unique funds, no-load funds, and load funds traded by investors in each zip code is identified; these numbers are then summed across all zip codes within a given category. Panel B restricts the sample to stock-only funds. Panel C reports the fraction of unique funds traded in each zip code category that include minimum balance requirements.

Zip Code Share in Load Funds	No. of Funds Traded (A)	No. of No-Load Funds (B)	No. of Load Funds (C)	Share of No-Load Trades (B/A)	Share of Load Trades (C/A)
Panel A: All Mutual Funds					
Category I (Low)	755	325	83	0.430	0.110
Category II	524	167	98	0.319	0.187
Category III	1204	411	186	0.341	0.154
Category IV	980	378	181	0.386	0.185
Category V	754	294	165	0.390	0.219
Category VI (High)	182	64	93	0.352	0.511
Panel B: Stock-Only Funds					
Category I (Low)	66	55	11	0.833	0.167
Category II	96	76	20	0.792	0.208
Category III	224	165	59	0.737	0.263
Category IV	221	165	56	0.747	0.253
Category V	175	129	46	0.737	0.263
Category VI (High)	55	28	27	0.509	0.491

Panel C: Minimum Balance Requirements

Zip Code Share in Load Funds	Share of Trades in Funds with Minimum Balances
Category I (Low)	0.1819
Category II	0.1341
Category III	0.0977
Category IV	0.0842
Category V (High)	0.0082

Table VI
Investor Characteristics, Redemption Likelihood, and Individual Portfolio Turnover

Pooled, cross-sectional regressions are run using individual investor accounts data from January 1991 to November 1996. This table reports the estimated coefficients from regressions of a) *REDEEM*, the average yearly mutual fund redemption likelihood (defined as total sales divided by start-of-year position) within a given zip code, or b) *PORTTURN*, the average monthly investor turnover of the mutual fund portion of a portfolio within a given zip code, on characteristics of the neighborhood of the investor (defined by zip code). *MEDINC*, *PCTCOL*, *PCTPRF*, and *PCTWHT* are variables equal to the median income of the population in a zip code, the fraction of the zip code's population with college education, the fraction of the zip code's population in professional occupations, and the fraction of the zip code's population that is white; zip code demographic data are obtained from the 1990 U.S. Census. *METRO* is a dummy variable equal to one if a zip code is within 25 miles of at least one of the top twenty metropolitan areas, and zero otherwise. *CALIF* is a dummy variable equal to one if a zip code is located in California, and zero otherwise. t-statistics, adjusted for autocorrelation and heteroskedasticity, and the number of observations (n) are in parentheses.

Panel A: Baseline Regressions						
Dep. Var.	*MEDINC*	*PCTCOL*	*PCTPRF*	*PCTWHT*	*METRO*	*CALIF*
REDEEM	0.003	0.200	-0.011	-0.022	-0.003	-0.005
(n=1,750)	(1.58)	(1.48)	(-0.26)	(-1.33)	(-0.37)	(-0.82)
PORTTURN	-0.001	-0.147	0.104	-0.043	-0.143	0.142
(n=1,910)	(-0.26)	(-0.14)	(1.85)	(-0.20)	(-1.26)	(1.61)

Figure 1: Load Fund Investment