Can Individual Investors Beat the Market?

Joshua D. Coval
Harvard Business School

David Hirshleifer
University of California at Irvine

Tyler Shumway
Marriott School of Business,
Brigham Young University*

June 21, 2021

Abstract

We document persistent superior trading performance among a subset of individual investors. Investors classified in the top performance decile in the first half of our sample subsequently earn risk-adjusted returns of about 6% per year. These returns are not confined to stocks in which the investors are likely to have inside information, nor are they driven by illiquid stocks. Our results suggest that skilled individual investors exploit market inefficiencies (or perhaps conditional risk premiums) to earn abnormal profits, above and beyond any profits available from well-known strategies based on size, value, momentum, or earnings announcements. (*JEL* G11, G14, G40, G51)

*We thank Malcolm Baker, Randy Cohen, Laura Frieder, Mark Grinblatt, Mark Kamstra, Terry Odean, Jeffrey Pontiff, and Rob Stambaugh; an anonymous referee; and seminar participants at the Atlanta Federal Reserve Bank, Harvard Business School, UCLA, the University of Florida, the University of Georgia, the University of Maryland, the University of Pennsylvania, the University of Rochester, and Virginia Tech for helpful comments and discussions. We also thank Greg Allenby for his generous help with the data and Noah Stoffman for his very capable research assistance. Send correspondence to Tyler Shumway, tyler_shumway@byu.edu.

Individual investors who trade actively play in a tough league populated by highly trained professionals supported by heavy infrastructure. They do so despite the alternatives of diversifying through index funds, exchange-traded funds, or actively managed mutual funds. As such, active individual investors are generally regarded by academics and investment professionals as uninformed and perhaps even foolish market participants. We systematically evaluate whether some individual investors have investment skill. Such skill can take the form of exploiting market inefficiencies, trading on cash flow news, or capturing time-varying risk premiums.[1]

Many studies cast doubt on the ability of individual investors to beat the market. Households appear to trade too much, maintain underdiversified portfolios, buy in response to attention-grabbing news, hold onto losing positions for too long, and trade unprofitably relative to cash flow information.[2]

However, some papers suggest that not all individual investors are likely to act as "noise traders." Kelly and Tetlock (2013) show that the net buying of individual investors positively predicts cross-sectional stock returns over the next month. Korniotis and Kumar (2013) find that investors with characteristics that may be associated with higher intelligence have higher returns. Grinblatt, Keloharju, and Linnainmaa (2012) document that investors with higher IQs have better performance. And Seru, Shumway, and Stoffman (2010) show that investors seem to learn and improve as they gain experience.

But none of these studies answers the more basic question of whether individual investors are foolish to trade actively in the first place. There may exist cross-sectional dispersion in investment skill, and perhaps even some learning, both about one's ability and learning by doing, is taking place. But unless there exists a subset of investors that can expect to reliably and significantly outperform their passive alternatives, actively trading individual investors are indeed fools, and only vary in their degree of

[1]Such risk premiums can potentially provide "good deals" for some investors if they are induced by the hedging incentives of other investors who are endowed with nonmarketable assets.

[2]See, for example, Blume and Friend (1975), Ferris, Haugen, and Makhija (1988), Odean (1997, 1998, 1999), Barber and Odean (2000, 2002, 2008), Grinblatt and Keloharju (2001), Cohen, Gompers, and Vuolteenaho (2002), Goetzmann and Kumar (2008), Ivkovic, Sialm, and Weisbenner (2008), Kaniel, Saar, and Titman (2008), Hirshleifer et al. (2008), Barber, Odean, and Zhu (2009), Barber et al. (2020), and Liu, Xiong, and Xiong (2020).

foolishness.

We provide evidence that the pure noise trader view of individual investors is incomplete. We find that individual investors are highly heterogeneous in their investment skills, not just their ex post performance (as other studies have found), and that a substantial subset is able to achieve persistently superior investment performance relative to standard passive benchmarks.

Using the transaction record of a large sample of households at a major discount brokerage, we find strong evidence that individual investors who have been successful in the past tend to outperform in the future, even after controlling for firm characteristic effects or factor exposures based on size, book-to-market, momentum, and liquidity. Specifically, our results suggest that 10% to 20% of individual investors can anticipate genuinely positive alpha from their investments, with the top decile earning average abnormal returns of about 6% per annum. We also identify a subset of investors that persistently make selections that underperform. Finally, we document that valuable information can be gleaned from the trading decisions of individual investors. Portfolio strategies that mimic the selections of previously successful investors and short those of previously unsuccessful investors deliver abnormal returns of roughly 7% per annum, with modest turnover.

We document performance persistence in a variety of ways. In most of our tests, we classify investors according to their average return over either five days or their holding period during the first half of the sample. We then examine their returns over the same horizon during the second half of the sample.[3] This offers a simple and intuitive measure of whether investors are actually outperforming the appropriate benchmark in the stocks that they own over both long and short horizons. We find that trading performance is consistently cross-sectionally correlated across the two sample halves. The correlations are approximately 5% to 10%, and they differ from zero with high statistical significance.

There is some debate in the literature as to the proper way to adjust for risk in

[3]We mark open positions to market at the end of the relevant sample period, so evaluation returns for these positions are not strictly holding period returns.

evaluating investment performance. It is common to use factor benchmarks, such as the three-factor model of Fama and French (1992) or the four-factor model of Carhart (1997), to control for characteristics, such as book-to-market, size, or momentum (see, e.g., Daniel et al. 1997). In most of our tests we use both the four-factor model and the Daniel et al. (1997) characteristics-based adjustment. If these benchmarks reflect risk premiums, then deviation from the benchmark measures the abnormal performance of investors. If these benchmarks capture mispricing, then our tests describe the ability of individual investors to earn gains or losses above and beyond any profits they earn based on other well-known return predictors.

To examine economic significance, we sort investors into performance deciles based on their holding period performance in the first half of the sample. We then examine average abnormal performance by decile in the second half of the sample. Investors in the top performance decile earn characteristic-adjusted returns that are a little more than 3% larger than those of investors in the bottom decile, with high statistical significance. Since the average holding period of these investors is approximately 6 months, investors in the top decile outperform those in the bottom decile by about 6% per year on a characteristic-adjusted basis. In our holding period results the performance difference comes primarily from the top performance decile.

Finally, to investigate whether the information contained in household trading behavior offers profitable trading opportunities to those with access to this information, we construct and test the profitability of trading strategies that rely on information that would be available ex ante to someone (such as a brokerage house) that can observe trading data in real time. Specifically, the strategies mimic the portfolio holdings of previously successful investors and short the positions of previously unsuccessful investors. These strategies earn statistically significant abnormal returns of more than 7% per year. Moreover, because strategy holding periods are matched to those of the trades that are being mimicked, the strategy achieves this performance with modest turnover and therefore is likely to outperform net of reasonable estimates of transaction costs.

These findings raise the question of whether individual investors with persistent positive abnormal performance possess superior skill at identifying and exploiting market mispricing, or are exploiting inside information about fundamentals. Several of our tests provide insight about the degree to which inside information contributes to performance persistence. First, in some of our tests we only consider mid- to large-cap stocks, which presumably have less information asymmetry. Second, we examine whether abnormal performance persists when we allow each investor to only trade each stock one time. Finally, we find that some investors persistently *underperform*; it is not plausible that such investors have superior information and then trade the wrong way.[4] Most of our results are more consistent with skilled households exploiting market inefficiencies than with insiders trading on private information.

1 Performance Persistence and Individual Investors

Several studies provide evidence that different categories of individual investors have achieved different realized performance. Some evidence indicates that individual investors who trade the most underperform those who trade less (Barber and Odean 2000; Liu, Xiong, and Xiong 2020), that male investors underperform female investors (Barber and Odean 2001), that online traders underperform telephone traders (Barber and Odean 2002), that individuals with relatively concentrated portfolios outperform individuals whose portfolios are more diversified (Ivkovic, Sialm, and Weisbenner 2008), and that investors with higher IQs have better stock picking performance (Grinblatt, Keloharju, and Linnainmaa 2012). Some evidence suggests that individual investors achieve relatively higher returns when purchasing stocks from companies closer to their homes (Ivkovic and Weisbenner 2005). Finally, Kaniel, Saar, and Titman (2008) find that stocks that are heavily purchased by individual investors in one month exhibit positive excess returns in the following month, which they attribute to liquidity provision by individual investors.

[4]In contrast, the market inefficiency theory requires that some investors systematically underperform. If members of a group of investors are subject to common misperceptions, their total trading as a group will move prices in a direction adverse to their desired trades. Thus, an inefficient markets story implies not only smart investors who make money by exploiting inefficiency but also foolish investors who lose money in the processing of generating the inefficiency. This is consistent with the findings of Barber, Odean, and Zhu (2008) and Hvidkjaer (2008).

What none of these studies addresses is whether some individual investors have superior skill at analyzing public information, and are able to use this skill to make *persistent* abnormal trading profits. In a similar spirit, the issue of whether superior performance persists is a major issue in the mutual funds literature. Most studies of mutual funds find that the average fund underperforms the overall market.[5] Similarly, only limited evidence exists suggesting that those funds that outperform during one time period do so in the future.[6] Given the extensive literature addressing the question of performance persistence for mutual funds and other institutional investors, it is surprising that there has been no concerted empirical work focused directly on this issue for individual investors.

The central question we address in this paper is whether those individual investors who earn profits on their trades are merely lucky or whether some are indeed skilled. We also ask whether the most skilled investors actually earn positive profits or whether they merely earn higher returns than the less-skilled peers. We further seek to understand whether observed skill comes from short horizon or long horizon strategies, and whether skill comes from successful interpretation of public signals or from private insider signals.

We do not expect individual investors to be, on average, better skilled or informed than mutual fund managers. However, some compelling reasons point to individual investors being better positioned to *exploit* a given informational advantage. First, individual investors almost always make smaller trades than professional investors, which reduces the pressure they place on prices. This makes them far better positioned to exploit smaller or shorter-term deviations from fundamental values.

Second, individual investors are less constrained than mutual funds to track the market or a given benchmark. The theory of portfolio performance measurement

[5]See Carhart (1995), Malkiel (1995), Chevalier and Ellison (1999), and Daniel et al. (1997). The exception is Wermers (2000), who, after controlling for cash drag, finds positive average excess returns.

[6]Lehman and Modest (1987), Grinblatt and Titman (1992), Hendricks, Patel, and Zeckhauser (1993), Goetzmann and Ibbotson (1994), Brown and Goetzmann (1995), Elton, Gruber, and Blake (1996), and Wermers (1997) all document evidence of persistence in mutual fund performance. However, Carhart (1992, 1997) and Wermers (2000) find that most of the persistence can be explained by persistence in mutual fund expense ratios and momentum in stock returns. Baks, Metrick, and Wachter (2001) employ a Bayesian approach to detect managers with positive expected alphas.

shows that it is the freedom to deviate from benchmarks that allows an investor to achieve gains from stock selection. Constraints on stock selection are most extreme for index mutual funds and ETFs, which capture a substantial portion of invested wealth. But they are also present for actively managed funds.[7]

Finally, individuals are not subject to the agency problems inherent in investing other people's money.

From the standpoint of detecting performance persistence, the transaction-level data sets of individual accounts are far superior to mutual fund data, which are generally available only at a quarterly frequency. If the profit opportunities exploited by investors are transient, then tests that rely on transactions reported at a quarterly frequency are considerably disadvantaged.

2 Data

We study a data set provided by a large discount brokerage firm on the trades placed by 63,652 households in 115,856 different accounts from January 1991 through November 1996.[8] We merge the trades data with CRSP data for much of the analysis. Table 1 reports summary statistics.

For most of our tests, we restrict attention to households that trade at least 25 times in our six-year sample period. A few households hold a large number of accounts, with one household holding almost 400 different accounts. We are not certain that households with many accounts, each of which trades very little, are following simple wealth maximization strategies, so most of our tests also require each account being used to have at least 10 trades. Requiring at least 25 trades removes more than 48,000 households from consideration. However, our restrictions have the benefit of ensuring that, for each household we study, we have sufficient

[7]The personal risk borne by mutual fund managers for deviating from their performance benchmark reduces their incentives to seek superior selectivity. Empirically, Cremers and Petajisto (2009) find that those funds that deviate more from the holdings of their benchmark index achieve higher performance.

[8]While the discount brokerage data include files with information about account holdings and trades, we only use the file that records account trades. The trades file includes 126,488 accounts, of which only 115,856 accounts have at least one purchase during the sample period.

data to estimate trading profits with some accuracy. Since the 15,382 households that placed at least 25 trades during the sample period are clearly of interest, we also report statistics for this subset in Table 1. In the rest of the paper we sometimes refer to households as accounts or even as individual investors, but all of the analysis is done at the household level. Performing the analysis at the account level results in findings that are very similar to those reported.

Overall, the average household placed 27 purchases of an average size of $8,551 in 15 different companies. Not surprisingly, the median and standard deviation indicate substantial right-skewness in the number of purchases. The median household placed 11 purchase trades in 7 different companies at an average value of $4,721. For the subset of purchases that were later (at least partially) sold in the sample period, the holding period for the average household is 380 days, and for the median household it is 301 days.

The lower panel of Table 1 reports summary statistics for households that traded a minimum of 25 times during the sample period. The average active household trades 81 times with an average size of $9,738 in 42 different companies. Again, the median figures are somewhat lower, with the median account trading 51 times in 29 different companies with an average value of $5,503. Not surprisingly, the average holding period for the active accounts, 260 days, is considerably lower than the overall figure.

To address known sources of return predictability, we perform standard characteristic based and four-factor adjustments. We also investigate whether the stocks traded by successful investors differ in terms of their liquidity characteristics from those of unsuccessful investors.

Measuring the performance of investors who exhibit holding periods of various lengths and different numbers of positions is a difficult task. We measure performance at two different horizons, using both short-term and holding-period measures. Short-term performance measures have greater statistical power than holding-period measures to assess skill at exploiting short-lived investment opportunities (such as event-related opportunities). If profitable trading opportunities available to skilled individuals are short-lived, then as the holding period grows, the component of a trade's return variability that is unrelated to skill will account for an increasing fraction of total return variability. Using daily transactions data therefore allows us to test for short-lived investment opportunities with substantial statistical power. Our short horizon tests typically focus on the returns individuals obtain from their trades during the week that follows their trades.

For the short horizon factor-based risk adjustment, we estimate time-series regressions of the return on each stock net of the Treasury-bill rate on several factors: the excess of the CRSP value-weighted market over the Treasury-bill rate, a size factor, a book-to-market factor, a momentum factor, and a one-day lag of each of these factors to adjust for possible nontrading biases. We estimate these regressions using daily data for the calendar year finishing at the end of each month in the sample period, to update each stock's regression estimates each month. We take the abnormal daily return of each stock to be the sum of the regression intercept and error term, or equivalently, the realized return minus the sum of the factor loadings times the realized value of each of the factors. The size, book-to-market, and momentum factor realizations are those posted on Ken French's website.

For the short horizon characteristic-based risk adjustment, we follow a procedure similar to that of Daniel et al. (1997) (DGTW). Specifically, we rank each stock into quintiles based on its market capitalization at the end of the previous month, its book-to-market ratio based on its most recently announced book equity value (lagged by at least 60 days to ensure public availability) and its momentum status. To determine the momentum status of each stock, we sort stocks each month into deciles based on their return over the previous twelve months, excluding the most recent month. Any stock that has been in the highest decile during one of the past three months is considered a winner stock, while any stock that has been in the lowest decile is considered a loser stock. A stock appearing in both groups during the last three months is assigned to its most recent classification. Stocks that are neither losers nor winners are designated as neither, resulting in three possible momentum categories. Combining the three momentum categories with five size and five book-to-market categories results in 75 possible classifications for each stock. We calculate daily equal-weighted average returns for each of these 75 stock classifications, taking the characteristic-adjusted return of a particular stock to be its realized return minus the average return to a stock with its classification.

We also examine performance persistence using holding period returns. For positions that are open at the end of the relevant sample period, we calculate returns by marking positions to market. Using holding period returns requires adjusting for known sources of return predictability for holding periods of varying lengths. We adopt a characteristic-based method that is similar to the adjustment described above, except that for each stock position we calculate the buy-and-hold return on the characteristic matched portfolio for the position's holding period (or from the close on the purchase day to the close on the sale day). We then subtract this return from the position's return, and consider the resultant buy-and-hold return to be adjusted for known sources of predictability. Because we use buy-and-hold returns, we are confident that the long horizon results are not driven by microstructure biases.

3 Results

3.1 Return correlations

We begin with a simple correlation test for persistence in household performance in which average household returns are compared across the two sample halves. To be considered in our calculations, we require accounts to have traded at least 25 times during the first half of the sample. The six-year sample is split at the end of the third year to ensure that a roughly equivalent number of households have traded at least 25 times in both sample halves. To make sure the results are not subject to a selection bias if individuals trade more frequently after their performance is good, we place no minimum trade restriction on the second half of the sample.

We calculate the correlation across households of the performance in one half of the sample with performance in the other half. We calculate correlations of both raw and abnormal returns across the two sample halves, measuring returns at both short and long horizons. Risks are adjusted using the Carhart (1997) four-factor model, and using DGTW characteristic portfolios.

To account for the fact that average returns are calculated with varying precisions across accounts (and across sample halves), we calculate three additional return correlations. The first compares the ratio of the investor's mean return to the investor's return standard deviation (a return/risk ratio) across the two sample halves. For the second, we compute the (one-sided) p-value associated with the hypothesis that the investor's abnormal return is drawn from a t-distribution with a positive mean during the sample half. In contrast with standard deviation, the p-value should not be affected by the number of trades placed by the investor.

We then calculate the correlation in the investor's p-values across the two sample halves. In addition, as a robustness check, we also calculate the correlation in returns obtained in even and odd quarters. The first three panels of Table 2 report the results for the four short horizon return correlations.

The correlation in performance across the sample periods is consistently around 5% to 10%, and is highly statistically significant. The effects are significant for both Pearson and rank correlations and are largely invariant to whether or how we adjust for risk. The correlations are also consistently positive and significant for each of our three performance measures: the simple average returns, the return-risk ratios, and the p-values. The finding of positive correlations is also robust to splitting the sample into even and odd quarters instead of halves. Thus, the results of Table 2 provide the first evidence of short horizon persistence in the performance of individual investors.

Next, to see whether the short horizon persistence is due to an ability to time the market as a whole, we recompute our tests replacing individual stock returns with overall value-weighted market returns. This tests whether individuals consistently tend to purchase stocks before the market rises. For market timing, persistence is reduced. The p-value correlations do retain some of their statistical significance, but the magnitude of the correlation coefficients declines somewhat. Persistence in market timing is explored in more detail in Keppo, Shumway, and Weagley (2020). Overall, it appears that the persistence in the performance of individuals in our data comes primarily from stock selection rather than market timing.

In the final two panels of Table 2, we examine long horizon performance persistence. In the absence of any risk adjustment, the correlations are quite large. However, this may simply be due to heterogeneity in investors' risk-bearing styles or average holding periods. Adjusting for risk (or mispricing) using a characteristics-based approach, the correlation drops to about 9% for the p-value, which is consistent with the short horizon results reported above.

Overall, Table 2 provides fairly strong initial evidence that the performance of individual investors persists. This persistence is robust to various risk adjustments and to the return measurement horizon.

3.2 Short horizon performance classification of investors

While the above results indicate clear persistence in household performance, they do not directly convey economic magnitude, such as the difference in future returns that can be expected of households identified as among the top versus the bottom 10%. To investigate economic magnitude, we classify households according to the performance of their trades and then measure how well this classification explains the returns of the household's other trades. We report on this analysis for short horizon returns in Tables 3 through 5, and we report on similar analysis for holding-period returns in Table 7.

Since we only have six years of data, we employ a technique we call a "complementary image procedure" to maximize our power to detect persistence. In this procedure, for each trade placed by a given household, we use *all other* trades the household places in our data set to calculate the household's average return and its p-value for testing the hypothesis that the mean return of the other trades is positive. That is, to maximize the accuracy of our classification of each household, we use future trades as well as past trades to assess a household's ability at a given point in time.

Clearly, the complementary image procedure does not provide a trading strategy that would be implementable by an investor who observes household trades only as they occur. To test whether there is a profitable trading strategy based on mimicking individual household trades, we will later consider a rolling forward procedure. The purpose of the complementary image procedure is not to design a strategy for making profits, but to address the scientific question of whether some households exhibit superior skill. Although the complementary image procedure uses ex post data, it does not do so in a way that biases the measurement of households' profits. In predicting the profit for a given week, the procedure omits the profit outcome for that week from the set of data used to identify the set of smart households. Under the null hypothesis that excess returns are independent over time, this procedure should not generate excess returns.

As discussed above, corresponding to each trade of a given household is an average return of all other trades placed by this household, and a p-value that this average return is positive. We sort all trades according to their corresponding p-values into deciles, and calculate the average returns of the trades within each decile. We write the average return of household j in all trades except trade k as

$$\hat{r}_j^k = \sum_{l,l \neq k}^{n_j} \frac{r_{j,l}}{n_j - 1}, \tag{1}$$

where n_j is the number of trades placed by household j and $r_{j,l}$ is the return earned by household j on trade l during the subsequent five trading days. Using this, the average return of the trades in decile i can be expressed as

$$r_i = \frac{\sum_j \sum_k r_{j,k} I\left(\hat{r}_j^k \in [\underline{r}_i, \overline{r}_i]\right)}{\sum_j \sum_k I\left(\hat{r}_j^k \in [\underline{r}_i, \overline{r}_i]\right)}, \tag{2}$$

where $r_{j,k}$ is the return earned by household j on trade k and $I\left(\hat{r}_j^k \in [\underline{r}_i, \overline{r}_i]\right)$ is an indicator variable which is one if $\hat{r}_j^k$ is within the limits of decile i, $\underline{r}_i$ and $\overline{r}_i$.

Table 3 reports the average returns of the trades in each decile during the five days following the trades' placements. The first column of Table 3 reports the average raw return of each skill decile group. The most successful investors in all other trades outperform the least successful investors (top minus bottom decile returns) by 254 basis points per week, with a simple t-statistic of 54. The t-statistic is extremely large because we have a large number of observations and we treat each trade's return as a conditionally independent observation.

The next two columns report the average risk-adjusted performances of each of the skill groups. With both four-factor and characteristic adjusted returns, the least successful investors continue to underperform the market and the most successful investors continue to outperform. Taking the difference between the top and bottom deciles yields a spread of 2.7% per week using factor-adjusted returns and 2.5% using characteristic-adjusted returns. A spread of 2.5% per week annualizes to a spread of about 130% per year, which is substantial. These values are all very statistically significant.

In the fourth and fifth columns of Table 3, we relax the requirement that a household have at least 25 trades and allow for households with as few as five trades. The difference in risk-adjusted performance for these less frequent traders drops somewhat to about 1.2% per week, but it remains very statistically significant.

Next, we examine whether investor skill, at least for superior performers, comes from access to inside information. To do so, we rerun the previous tests removing all trades in firms in which the investor has transacted more than once. While a subset of investors may have inside information about a company or two (i.e., about their employer or friend's firm), it seems very unlikely that large numbers of investors have access to inside information across a broad set of companies.

This restriction eliminates the most plausible candidates for insider trades, and, thus, the test focuses on whether an investor has superior skill at identifying and exploiting either market mispricing or time-varying risk premiums. The results of this test appear in the final two columns of the table. In these columns, the difference in returns remains economically and statistically significant, though the magnitude of the performance differential decreases somewhat.

The results in Table 3 are very strong indicators that the performance of any given trade by any household is predictable by examining the performance of the household's other trades. However, the assumption that each trade is conditionally independent of all other trades is quite strong. To relax this assumption, we perform a very similar analysis using the first half of the sample to sort households and the second half to evaluate their performance. We report the results of this analysis in Table 4.

To form the skill decile groups in Table 4, we sort all investors by the p-value of a one-sided test that their four-factor risk-adjusted short-horizon return is positive in the first half of the sample. The first column of Table 4 reports the average raw return of each skill decile group. The most successful investors in the first half of the sample outperform the least successful investors (top minus bottom decile returns) by 54 basis points per week, with a t-statistic of 7.41.

Although the average returns in this column show a clearly increasing pattern across previous performance deciles, the relation is not monotonic. This is the case for many of the persistence results we report, and is to be expected based on sampling noise. Given the difficulty in predicting stock returns in general, it would be surprising to get an effect so strong that it would be monotonic in decile cuts.[9]

The next two columns report the average risk-adjusted performances of each of the skill groups. With both four-factor and characteristic adjusted returns, the least successful investors continue to underperform the market and the most successful investors continue to outperform. Taking the difference between the top and bottom decile alphas yields a spread of 58 basis points (bps) per week using factor-adjusted returns and 55 bps using characteristic-adjusted returns. A spread of 55 bps per week annualizes to a spread of about 28%. The t-statistics associated with these returns show high levels of significance.

The excess (negative) performance of the unskilled households is much larger than the positive performance of the high skill groups. Still, the alphas of the highest skill

[9] In some scenarios, we would expect to see skill effects concentrated in extreme quantiles. Suppose that a small fraction of individuals have genuine skill, and a moderate fraction of individuals have negative skill. This negative skill may come from herding with sentiment-driven buyers or sellers who make the stock over- or underpriced. Suppose further that many and perhaps most investors have zero ability. Then we would expect to see strong effects at the extremes, but noise and nonmonotonicity in the middle.

deciles are statistically significant and fairly large. Annualizing a return of 10.6 bps per week gives a return of about 5.5% per year. In Table 4, based on either factor or characteristic adjustment, only the top decile of investors is able to outperform significantly, whereas there is substantial and, in most cases, significant underperformance in the bottom-four deciles. This is consistent with previous findings that overall individual investors tend to underperform.

We conclude that relatively few investors are skilled and a comparatively large set of investors have negative "skill" arising from their decisions to herd in their uninformed trades (thereby pushing prices against themselves as a group). This suggests that some very useful investment information can be gleaned by an observer of individual trades about bad trades to be avoided.

The persistence of the poor performance documented in Tables 3 and 4 is notable, since it seems to indicate a special ability to underperform the market. The losses of these investors are far greater than the losses of the average individual investor documented by Odean (1999). The systematic tendency of some individuals to underperform persistently is good evidence that access to inside information is not the only source of persistent abnormal performance in the sample.

Since we calculate returns for a period that is subsequent to the day each trade is executed, the persistence of poor performance is not due to transaction costs. The bid-ask spread and price impact tend to manifest themselves on the same day as the trade. In standard microstructure settings with semistrong form efficient markets, noise traders on average lose money because of these costs. Thus, regardless of whether the investors with the worst performance in the data are viewed as noise traders, their persistent losses must be due to something beyond transaction costs.

What causes these individuals to underperform? In models of investor psychology and security prices, imperfectly rational investors, in the process of producing market mispricing, on average lose money to more sophisticated "arbitrageurs." Our findings are consistent with individual investors being a heterogeneous group that includes both naive and sophisticated investors. The evidence is consistent with systematic underperformers being individuals who tend to trade on the wrong side of market inefficiencies.

To ensure that our results are not driven by any special characteristics of frequent traders, it is useful to rerun the tests using all households that have five or more trades in the first half of the sample. The results from this test appear in the fourth and fifth columns of Table 4. The difference between the least and most successful decile average returns is about 40 bps per week, in line with the previous results.

As in our complementary image procedure results, omitting repeated trades in the same stock weakens the effect somewhat, resulting in a performance differential of about 25 bps per week. Omitting repeated trades also makes the average performance of each group negative. We note that this is not the case in the complimentary image results, where the top-two decile investor groups have large and statistically significant average returns. The negative average returns here appear to be due to sampling variability. The difference between the past successful and past unsuccessful investors remains statistically significant.

3.3 Performance after sales

Table 5 examines selling as well as buying as a potential investment skill. It reports the average five-day return of trades that have been sorted into quintiles based on past buy or sell success. We report on quintiles here just because using deciles creates a lot of numbers to report. The results only include trades of households who have placed at least 25 trades in the first half of the sample. The first two columns show decile skill groups formed according to the p-value testing whether the investor's risk-adjusted buy returns have a positive average in the first half of the sample. The second two columns are calculated using the investor's risk-adjusted sell returns, or the returns to stocks the investor has sold in the five days after the sale. A negative return after sales indicates skill, so the signs for skilled households in columns 3 and 4 follow the opposite pattern from those of columns 1 and 2. The final column considers both buy and sell returns, giving a negative sign to sell returns and a positive sign to buy returns. It requires households to place at least 40 total trades in the first half of the sample to be included.

The first column, as already reported in Tables 3 and 4, shows that as we move from lower to higher performance quintiles, greater buying skill in the first half of the sample is associated with higher buying performance in the second half of the sample. The fourth column shows that greater skill at *selling* in the first half of the sample is associated with better selling performance in the second half of the sample. The high-low difference in alphas is almost 20 bps, which is significant at the 0.1% level.

The second and third columns provide no evidence for cross-transfer in buying versus selling skills. In column 2, greater skill at buying is not associated with better selling performance in the second half of the sample. Similarly, in column 3, greater skill at selling is not associated with better subsequent buying performance. In fact, the sign of the return difference in column 3 suggests that better selling performance is related to worse buying performance. This suggests that some investors (smart bulls) are persistently skillful at identifying good stocks to buy but have no special skill at timing their exits. Other investors (smart bears) are not good at picking stocks, but are good at recognizing good times to exit their positions. Perhaps smart bulls focus their attention on screening stocks for the purpose of buy-and-hold purchases, and do not sell until liquidity considerations require a sale. Smart bears may lack skill in screening large numbers of stocks for purchases, but they may be good at analyzing carefully the few stocks that they hold, thereby acquiring skill in timing their sale.

The final column sorts investors by the p-value that corresponds to their buy return minus their sell return, and describes the returns on both buys and sells during the second half of the sample. It requires that investors have at least 40 buys and sells in the first half of the sample. This categorization reveals a strong effect of skill on subsequent performance: the difference in alphas of the high skill versus low skill investors is 31 bps, with a t-statistic of 8.22. As before, this skill is quite broad at the low end. The lowest two quintiles achieve highly significant underperformance. Outperformance is positive and significant in the highest quintile, at 14 bps.

Overall, the return differentials we identify suggest the presence of either market inefficiencies or time-varying risk premiums, that some individual investors are losing money by trading on the wrong side of such return patterns, and that a smaller set of individual investors have the skills to exploit them.

3.4 Characteristics of trades

To further explore the sources of investor skill, in Table 6 we examine some of the characteristics of the trades of successful households. The first two columns describe the transaction costs that the investors in our sample incur. The first column reports the average return from the purchase price to the closing price on the day of purchase. Our previous tests do not include these same-day returns because they calculate returns from the closing price on the day of a purchase. These same-day returns are negative on average because most households are buying at the ask price and thus paying the bid-ask spread. While this column shows the households incur transactions costs, the same-day returns of the most successful households are significantly less negative than those of the least skilled households.

The second column in Table 6 reports the average closing percentage bid-ask spread of the stocks purchased by each skill group. The most successful households tend to buy stocks with a higher bid-ask spread even though their average same-day returns are less negative than other households. Furthermore, the average same-day returns of all our household groups are not nearly as negative as the average bid-ask spreads would suggest. Apparently, our households sometimes provide liquidity and avoid paying the spread. They may also buy stocks with prices that are trending upward even during the day when they make their purchase. The overall average same-day return of about negative 28 bps is smaller than the total outperformance we document in Tables 3 and 4, which is over 200 bps for the complimentary image results and more than 50 bps for the second half of the sample results. This suggests that our performance results are not likely to be completely explained by a liquidity premium.

The third and fourth columns of Table 6 are designed to test whether trading on post earnings announcement drift (PEAD) drives the persistence that we document.

Evidence from individual investors as a whole provides surprisingly little evidence that their aggregate trading drives the underreaction to earnings (Hirshleifer et al. 2008). However, it is possible that this reflects a mixture of irrational investors who contribute to PEAD and smart investors who are exploiting PEAD. If in fact more successful investors are trading on PEAD, we should expect them to systematically purchase stocks right after earnings announcements. The third and fourth columns report the average and standard deviation of the number of calendar days between an individual's purchase date and the stock's closest earnings announcement. There is little difference in these numbers across skill deciles, and our most successful households are actually buying stocks a little further away from earnings dates. These results indicate that it is very unlikely that PEAD is a significant source of the persistence that we document.

The fifth column of Table 6 reports the average abnormal volume for each stock traded. We calculate abnormal volume by averaging the percentage difference between the stock's volume on the day of a purchase and the stock's average daily volume for the calendar year. So, for example, if an investor bought a stock on a day on which that stock's volume was 1,000 shares, and the stock's average volume over the calendar year was 500 shares a day, the abnormal volume would be (1,000 - 500)/500 = 100% for that transaction. We calculate abnormal volume to determine whether investors are trading during periods of high liquidity or disagreement (rational or otherwise). We find that abnormal volume is declining with investor ability, consistent with a view that more successful investors are less likely to trade at the same time and on the basis of the same signals as other households.

We examine the relative wealth of successful investors in the last two columns of Table 6. The sixth column reports the group median of the total account value as of the end of January 1991. We calculate this figure using holdings data. The seventh column reports the group median of each investor's average trade size, measured over the first three years of the sample. Both measures of investor wealth are roughly increasing in investor ability and suggest that successful households tend to be wealthier and potentially more sophisticated.

The results reported in Table 6 indicate that the least successful investors are particularly likely to trade at the same time as others, and have relatively little wealth in their trading accounts. The most successful investors do not appear to specialize in providing liquidity or to trade on PEAD.

3.5 Long horizon performance classification of investors

To examine long-horizon performance persistence, we rank investors into deciles by their holding period performance in the first half of the sample, and then examine the holding period performance of their trades in the second half of the sample. Our results are reported in Table 7.

The returns reported in Table 7 are buy-and-hold returns, so the return horizon associated with each position varies substantially. However, in the market-adjusted and characteristic-adjusted results, the matching portfolio returns are also buy-and-hold returns calculated over the same holding period. Therefore, if investors do not have any trading skill, the mean characteristic-adjusted return should be close to zero, regardless of the holding period.

We sort households in the first half of the sample by both their average return in excess of the market and their characteristic-adjusted returns. In the first column of the table, investors are sorted by market-adjusted returns, and the average of their market-adjusted return in the second period is reported. In all remaining columns, investors are sorted by their average characteristic-adjusted return in the first half of the sample.

The average market-adjusted returns reported in the first column are clearly increasing in the previous performance decile, though the relation is not monotonic. Since the average holding period in our sample is approximately half a year, the spread in expected returns appears to be about 9% or 10% per year. Interestingly, most of the apparent predictability in returns is positive rather than negative.

The analogous results using characteristic-adjusted returns are reported in the second column. For this and all remaining columns, accounts are sorted into deciles based on the p-value of their characteristic-adjusted returns in the first half of the sample. In the second column, average characteristic-adjusted returns in the second half of the sample are not monotonically related to previous performance decile. However, the average characteristic-adjusted performance of the five groups of households below the median is negative 17 bps. The average risk-adjusted performance of the top-two deciles is about 240 bps. The performance of the highest previous performance decile group is 3.38%, which is by far the highest of all the groups. The difference between the highest and lowest group performance is over 3%, and is highly statistically significant. Average holding periods are reported in the fifth column. They generally appear pretty close to six months, or 182 days. Given an average holding period of six months, the return of the most skilled decile corresponds to outperformance of about 6% per year.

The third column of Table 7 reports on the success of our investors in the short horizon measure used in previous tables. We construct this column by ranking investors by their five-day risk-adjusted return in the second half of the sample. We assign them a rank of 1 to 10, so the average rank is 5.5. The highest skill decile when sorted by long horizon performance has an average short-term ranking that is 1.1 points higher than the lowest skill decile. This shows a strong and significant correlation between our short-term performance measures and our long horizon measure.

The other columns in Table 7 give some details about the frequency of trade by previous performance decile and the types of stocks that investors in different deciles are buying. Looking at the trade frequency numbers, one observes that frequency does not appear to be closely related to previous performance. Comparing the number of purchases of these investors across the two sample halves, one clearly sees that previously unsuccessful investors take fewer positions in the second half of the sample, while previously successful investors take more positions. The average holding period does not seem closely related to performance, and the characteristics of the stocks that investors hold vary only a little by previous performance decile.

3.6 A short horizon trading strategy

The results thus far indicate that a subset of individual investor households have persistent ability to exploit market inefficiency or conditional risk premiums. To provide an alternative perspective on the economic magnitude of this ability, we create a trading strategy based up investor trades, and measure the performance of this strategy relative to the four-factor model.

Specifically, we construct zero-cost portfolios that go long all the trades of accounts that have performed well up to the current date and go short all the trades of investors who have performed poorly up to the current date. Since we are focusing on short-horizon returns, we sort investors according to their past performance measured over the one-week horizon following their trades. As with our earlier tests, to ensure that any short-term price pressure created by trades does not influence our results, we wait until the day after the trade is executed to begin measuring returns.

Since we only have six years of data, and much of these data are used to assess investors' performances, our power to detect abnormal returns is somewhat limited. To maximize our power to detect abnormal performance, we must accept a trade-off. If we only include trades of investors with mean returns significantly different from zero, we more reliably focus on the trades of more skilled versus less-skilled investors. However, to the extent that, at times, only a limited number of households can be classified as unusually good (or bad), such a portfolio will be highly undiversified. Since we only have about one thousand days over which to measure our strategy's expected return, such lack of diversification can result in the unexpected component of returns becoming so variable that reliable inference is impossible. Alternatively, if we are lax in our criteria for including accounts in the strategy, a larger fraction of the trades we mimic are from accounts lacking in special skill.

To strike a balance, we only consider investors who have traded at least 25 times up to the current date, and we sort them into quintile portfolios to ensure that our portfolio is diversified. Furthermore, we only measure the returns to our strategy on days when there are at least 25 stocks in the top and bottom portfolios. Specifically, we rank all investors who have traded at least 25 times up to the current date by the p-value that their abnormal return is positive. We then compute weighted average returns of all the stocks purchased during the last five days by all accounts in each of the performance quintiles. Specifically, the return to portfolio p on date t is calculated as follows:

$$r_{pt} = \sum_j \sum_i \frac{I_{ijt}}{\sum_j \sum_i I_{ijt}} r_{it}, \tag{3}$$

where I_{ijt} is an indicator variable that is one if investor j has purchased asset i in the five trading days before date t and r_{it} is the return to firm i on date t. These returns are neither equal weighted nor value weighted. They are household weighted in the sense that if k previously successful households are all holding stock i at time t (and no unsuccessful households hold it) then the weight on stock i will be k divided by the total number of positions held by all previously successful households. Since firms with larger market capitalizations will be purchased by more investors on average, this household weighting should converge to something like value weighting as the number of households gets large.

Using the strategy return defined in Equation (3), we calculate the abnormal return to the strategy that goes long the top quintile and short the bottom quintile. We use a four-factor model to adjust returns for risk. Specifically, we calculate a daily raw return from Equation (3) and then regress the difference between the top and bottom portfolio returns on the four factors and their lags. To ensure that trading in small, illiquid firms does not drive the results, we remove the smallest (by capitalization) one-third of all CRSP firms from the sample. Table 8 reports the results.

The strategy generates abnormal returns of about 17% per year, which is larger than the long-horizon average abnormal return reported previously. Both four-factor and capital asset pricing model (CAPM) alphas are very statistically significant. For the CAPM alpha, about half of the excess performance comes from the previously successful portfolio and about half comes from shorting the previously unsuccessful portfolio. For the four-factor alpha, most of the excess performance comes from the previously successful portfolio.

One important issue for these strategy returns is transactions costs. The result in Table 8 reveals potentially valuable short-term information in the trades of skilled households, but the turnover involved in implementing this strategy may make it infeasible. In part to address this issue, we consider a holding period trading strategy below.

3.7 A long horizon trading strategy

Our long horizon trading strategy is somewhat simpler than the short horizon strategy. At the end of 1993, we rank all active investors into quintiles based on their characteristic-adjusted buy-and-hold performance over the previous 3 years. For the period 1994-1996, we then mimic the portfolio holdings of the most successful quintile and short the holdings of the least successful quintile. We start the strategy with the trades of investors on January 3, 1994, ignoring holdings at the end of 1993.[10] As in the short-term strategy, we weight each stock by the number of successful or unsuccessful investors holding the stock on each day. We also evaluate the long horizon strategy with a four-factor model estimated with daily returns, and we drop stocks in the smallest tercile of the size distribution. The factor model includes three Fama-French factors and a momentum factor. Table 9 reports the long horizon strategy results.

The long horizon strategy results are reported for investors with high past performance, for those with low performance, and then results are reported for the portfolio that is long high past performance and short low past performance. The last two columns of the table show that the high minus low strategy generates daily abnormal returns of about 7% per year, which is statistically significant and similar to the outperformance reported for the short horizon strategy. In the other columns of Table 9, it is evident that the performance of the long-short strategy is driven by good average performance on the part of past successful investors.

Interestingly, the long horizon strategy requires much less portfolio turnover than

[10]The holdings of successful and unsuccessful investors at the end of 1993 appear to have alphas very close to zero.

the short horizon strategy. The total number of investor-positions over which we are averaging is 55 at the beginning of the strategy sample (there were 55 purchases on the first day) and 16,087 by the end of the sample. There are a total of 45,330 positions taken by our high and low performance investors over our three-year period. In the 735 days of trading we consider, the average number of investor positions over which we are averaging is 9,966. We have 45,330 purchases to consider and 29,243 (=45,330-16,087) sales to consider, so the average number of transactions per day is 101. If we have average transactions per day of 101 and average holdings per day of 9,966, we have an average daily turnover of 51 bps. This works out to about 128% per year, which is an upper bound on the portfolio turnover implied by the strategy. To the extent that the trades of our investors cancel each other out (e.g., past unsuccessful and successful households trade the same way, or two successful households trade in opposite ways), the implied portfolio turnover will be lower.

An annual turnover of approximately 128% per year is not uncommon for individual investors or actively managed mutual funds. Chalmers, Edelen, and Kadlec (2001) report that the mutual fund quintile with the highest turnover in their sample has average annual transaction costs of 2.44% of assets. While this is a large figure, it is much smaller than the outperformance of 17% per year implied by the strategy results in Table 8, or the outperformance of about 7% per year implied by the holding period strategy in Table 9.

Tables 8 and 9 both confirm that there are profitable trading strategies based on mimicking the trades of investors who have been successful in the past. The holding period strategy has fairly modest trading costs, both because we exclude the smallest tercile of stocks and because turnover is on the order of 100% per year. This is perhaps our strongest evidence that individual investors can beat the market in the sense of outperforming the four-factor model.

4 Conclusion

Previous literature has emphasized that on average individual investors trade badly in the sense of experiencing poor realized performance. Here, we provide evidence that some individual investors have skill, in the sense that they are persistently able to beat the market. The ability of some households at a discount brokerage to select outperforming stocks is not confined to small firms or to firms the household has traded before. Furthermore, some investors persistently and heavily *under*perform. These findings suggest that investors' persistent abnormal performance is not derived primarily from trading on inside information.

The ability of some individual investors to achieve persistent abnormal performance, at least relative to standard benchmarks, is potentially consistent with the hypothesis that the market is not completely efficient, and that some individual investors are able to exploit inefficiencies. We conjecture that some investors are good at making sense of value-relevant information (such as information about business strategy, accounting information, and qualitative discussion in the news media) that are not taken into account in the four-factor model. An interesting further question is whether large brokerage firms are aware of the value of the information contained in their customers' trades. Indeed, a number of investing websites (e.g., Marketocracy.com, Cakefinancial.com, Updown.com, and Covestor.com) are explicitly designed in varying degrees to capture, quantify, and exploit the insights of individual investors.

Investors who are classified among the top 10% (based on the performance of their trades in the first half of the sample) buy stocks that earn abnormal returns of about 12 bps during the following week. These findings are robust to different forms of risk adjustment, to the removal of small stocks from the sample, and to the removal of any firms in which the account has traded more than once. Similarly, there are also individual investors who consistently place underperforming trades. Investors classified among the bottom 10% of all investors place trades that can expect to lose 50 to 100 bps during the subsequent week. In long horizon (holding period) returns, successful investors outperform unsuccessful investors by about 6% per year.

A trading strategy that mimics the trades of successful investors earns risk-adjusted returns of about 7% per year.

Several of our test procedures are designed to minimize the possibility that survivorship bias is an important contributor to our findings. For example, we require a minimum number of trades in the first half of the sample to measure household skill or ability, but we have no required number of trades in the second half of the sample when we assess performance. We also perform correlation tests partitioning by even and odd quarters rather than sample halves. We are confident that survivorship cannot explain the persistence that we document, but survivorship could affect the magnitude of the returns achievable by mimicking successful investors.

Our findings suggest some new tools and directions for future research. Past studies have used the degree to which investors are subject to specific biases as a proxy for financial naiveté (Goetzmann and Kumar 2008; Calvet, Campbell, and Sodini 2009). Our approach suggests that the investor's past return performance can provide an alternative, comprehensive proxy for financial naiveté.

Our findings also bring a new perspective to the issue of whether on average individual investors foolishly trade too much. As discussed earlier, previous studies have shown that individual investors on average lose money in their trades. However, if investors vary widely in terms of their ability to select investments, and if they learn about and develop this ability through trading, it may, in fact, be rational for some investors to trade frequently and at a loss, in the hope of future gains.

References

Baks, K. P., A. Metrick, and J. Wachter. 2001. Should investors avoid all actively managed mutual funds? A study in Bayesian performance evaluation. *Journal of Finance* 56:45–85.

Barber, B., and T. Odean. 2000. Trading is hazardous to your wealth: The common stock investment performance of individual investors. *Journal of Finance* 55:773–806.

———. 2001. Boys will be boys: Gender, overconfidence, and common stock investment. *Quarterly Journal of Economics* 116:261–92.

———. 2002. Online investors: Do the slow die first? *Review of Financial Studies* 15:455–87.

———. 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies* 21:785–818.

Barber, B., T. Odean, and N. Zhu. 2009. Do retail trades move markets? *Review of Financial Studies* 22:151–86.

Barber, B., Y. T. Lee, Y. J. Liu, T. Odean, and K. Zhang. 2020. Learning, fast or slow. *Review of Asset Pricing Studies* 10:61–93.

Blume, M., and I. Friend. The asset structure of individual portfolios with some implications for utility functions. *Journal of Finance* 30:585–604.

Brown, S. J., and W. N. Goetzmann. 1995. Performance Persistence. *Journal of Finance* 50, 679-698.

Calvet, L. E., J. Y. Campbell, and P. Sodini. 2009. Measuring the financial sophistication of households. *American Economic Review* 99:393–8.

Carhart, M. M. 1995. Survivorship bias and mutual fund performance. PhD thesis, Graduate School of Business, University of Chicago.

———. 1997. On persistence in mutual fund performance. *Journal of Finance* 52:57–82.

Chalmers, J. M. R., R. M. Edelen, and G. B. Kadlec. 2001. Transaction-cost expenditures and the relative performance of mutual funds. Working Paper, University of Oregon.

Chevalier, J., and G. Ellison. 1999. Are some mutual fund managers better than others? Cross-sectional patterns in behavior and performance. *Journal of Finance*

54:875–900.

Cohen, R. B., P. A. Gompers, and T. Vuolteenaho. 2002. Who underreacts to cash-flow news? Evidence from trading between individuals and institutions. *Journal of Financial Economics* 66:409–62.

Cremers, K. J. M., and A. Petajisto. 2009. How active is your fund manager? A new measure that predicts performance. *Review of Financial Studies* 22:3329–65.

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers. 1997. Measuring mutual fund performance with characteristic based benchmarks. *Journal of Finance* 52:1035–58.

Elton, E. J., M. J. Gruber, and C. R. Blake. 1996. The persistence of risk-adjusted mutual fund performance. *Journal of Business* 69:133–57.

Fama, E. F., and K. R. French. 1992. The cross section of expected returns. *Journal of Finance* 47:427–65.

Ferris, S. P., R. A. Haugen and A. K. Makhija. 1988. Predicting contemporary volume with historic volume at differential price levels: Evidence supporting the disposition effect. *Journal of Finance* 43:677–97.

Grinblatt, M., and M. Keloharju. 2001. How distance, language and culture influence stockholdings and trades. *Journal of Finance* 56:1053–73.

Grinblatt, M., M. Keloharju, and J. Linnainmaa. 2012. IQ, trading behavior, and performance. *Journal of Financial Economics* 104:339–62.

Goetzmann, W. N., and R. G. Ibbotson. 1994. Do winners repeat? Patterns in mutual fund performance. *Journal of Portfolio Management* 20:9–18.

Goetzmann, W. N., and A. Kumar. 2008. Equity portfolio diversification. *Review of Finance* 12:433–63.

Grinblatt, M., and S. Titman. 1992. The persistence of mutual fund performance. *Journal of Finance* 47:1977–84.

Hendricks, D., J. Patel, and R. Zeckhauser. 1993. Hot hands in mutual funds: The persistence of performance, 1974-1988. *Journal of Finance* 48:93–130.

Hirshleifer, D., J. Myers, L. Myers, and S. H. Teoh. 2008. Do individual investors drive post-earnings announcement drift? Direct evidence from personal trades. *Accounting Review* 83:1521–50.

Hvidkjaer, S. 2008. Small trades and the cross section of stock returns. *Review of Financial Studies* 21:1123–51.

Ivkovic, Z., and S. Weisbenner. 2005. Local does as local is: Information content of the geography of individual investors' common stock investments. *Journal of Finance* 60:267–306.

Ivkovic, Z., C. Sialm, and S. Weisbenner. 2008. Portfolio concentration and the performance of individual investors. *Journal of Financial and Quantitative Analysis* 43:613–56.

Kaniel, R., G. Saar, and S. Titman. 2008. Individual investor trading and stock returns. *Journal of Finance* 63:273–310.

Kelley, E. K., and P. Tetlock. 2013. How wise are crowds? Insights from retail orders and stock returns. *Journal of Finance* 68:1229–65.

Keppo, J., T. Shumway, and D. Weagley. 2021. Are monthly market returns predictable? *Review of Asset Pricing Studies*. Advance Access published April 10, 2021, 10.1093/rapstu/raab010.

Korniotis, G. M., and A. Kumar. 2013. Do portfolio distortions reflect superior information or psychological biases? *Journal of Financial and Quantitative Analysis* 48:1–45.

Lehman, B. N., and D. Modest. 1987. Mutual fund performance evaluation: A comparison of benchmarks and a benchmark of comparisons. *Journal of Finance* 42:233–65.

Liu, H., C. Peng, W. A. Xiong and W. Xiong. 2020. Resolving the excessive trading puzzle: An integrated approach based on surveys and transactions. Working Paper, Princeton University.

Mahani, R., and D. Bernhardt. 2007. Financial speculators' underperformance: Learning, self-selection, and endogenous liquidity. *Journal of Finance* 62:1313–40.

Malkiel, B. 1995. Returns from investing in mutual funds 1971 to 1991. *Journal of Finance* 50:549–72.

Odean, T. 1998. Are investors reluctant to realize their losses? *Journal of Finance* 53:1775–98.

———. 1999. Do investors trade too much? *American Economic Review* 89:1279–98.

Seru, A., T. Shumway, and N. Stoffman. 2010. Learning by trading. *Review of Financial Studies* 23:705–39.

Wermers, R. 1997. Momentum investment strategies of mutual funds, performance persistence, and survivorship bias. Working Paper, University of Maryland.

———. 2000. Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses. *Journal of Finance* 55:1655–95.

Table 1. Summary statistics

This table reports summary statistics for the entire set of households and for the subset of households that trade at least 25 times during the sample period. The average holding period encompasses all purchases that were sold later in the data set. For consecutive buys or sells in a given company, we calculate the time between the last purchase and the first sale. The sample spans from January 1991 to December 1996.

Variable (per household)	Mean	Median	SD	Min.	Max.
Full sample ($n = 63,652$ households)					
Number of purchases	27.37	11	64.17	1	4,218
Average dollar value	8,551	4,721	16,001	2	612,668
Number of different firms purchased	15.18	7	29.02	1	1,523
Number of purchases sold later	14.88	6	34.58	1	2,216
Average holding period (days)	380.54	301	311.56	0	2,079
Number of trades $\geq$ 25 (using accounts with at least 10 trades) ($n = 15,382$ households)					
Number of purchases	81.49	51	110.53	25	3976
Average dollar value	9,738	5,503	16,035	50	692,524
Number of different firms purchased	42.27	29	47.98	1	1,523
Number of purchases sold later	38.27	24	56.05	1	2,216
Average holding period (days)	260.36	218.23	192.21	0	1,870

Table 2. Correlation tests of performance persistence

This table reports correlations of the investment performance of households across two halves of the sample. The top-four panels of the table calculate performance based on each purchased stock's return on the five trading days that follow any purchase made by a household. The bottom-two panels calculate performance based on the holding period return of each stock. The correlations in columns 1 and 2 split the sample in half at the end of the third year and calculate the correlation in performance across the two sample halves for all accounts with at least 25 trades during the first three years. The correlations in columns 3 and 4 divide all trades into those that occur during the first and third quarters of the year and those that occur during the second and fourth, using all accounts with at least 25 trades in odd quarters. The p-values are calculated using a t-distribution and a t-score corresponding to a test that average returns are positive. The four-factor risk-adjusted return correlations regress daily returns on daily realizations of the SMB, HML, RMRF and momentum factors. The DGTW characteristic-adjusted returns subtract from a given firm's daily return the daily return to the matching size, book-to-market, and momentum portfolio. The market-timing returns replace the daily risk-adjusted return of a given firm with the corresponding daily return of the value-weighted market portfolio. The characteristic-adjusted holding period returns are calculated by subtracting the return of a buy-and-hold characteristic-matched portfolio over the same holding period. All correlations are expressed as a percentage The sample sizes vary from 2,640 to 3,190. **p <.05; ***p <.01.

	Sample halves		Even and odd quarters	
Variable	Pearson	Rank order	Pearson	Rank order
Raw returns for 5 days following purchase				
Mean return	5.9**	6.7**	4.7**	9.3**
Mean ret. / SD	7.7**	7.9**	9.4**	10.3**
p-value	10.6**	11.2**	12.1**	12.0**
Four-factor risk-adjusted returns for 5 days				
Mean return	4.3**	7.3**	4.0**	8.8**
Mean ret. / SD	6.9**	7.2**	9.0**	8.8**
p-value	9.3**	9.6**	9.6**	10.0**
DGTW characteristic-adjusted returns for 5 days				
Mean return	4.9**	7.3**	4.3**	9.8**
Mean ret. / SD	6.7**	7.3**	8.9**	9.8**
p-value	9.5**	9.8**	10.6**	10.9**
Market-timing returns for 5 days				
Mean return	0.7	1.6	0.8	1.3
Mean ret. / SD	2.6	1.0	-0.8	0.6
p-value	4.1**	7.6**	4.8**	9.1**
Raw returns for holding period				
Mean return	17.1**	26.1**	–	–
Mean ret. / SD	0.1	22.1**	–	–
p-value	15.4**	26.4**	–	–
Characteristic-adjusted returns for holding period				
Mean return	6.8**	9.1**	–	–
Mean ret. / SD	0.3	8.1**	–	–
p-value	9.1**	9.6**	–	–

Table 3. Five-day returns: Complimentary image procedure

This table reports the average five-day return of trades that have been sorted into deciles according to the assessed ability of the household. Each household's ability is assessed by the average risk-adjusted return of all of its other trades during the five days after each trade is placed. The first three columns of returns only include trades of households that have placed at least 25 other trades in the sample. The first column reports the average daily raw return for each group. The next two columns report four-factor risk-adjusted and DGTW characteristic-adjusted returns. The final four columns report risk-adjusted returns for groups formed using two alternative sorting procedures. In the fourth and fifth columns, stocks are sorted according to the risk adjusted returns of the trader even if the trader had as few as five other trades in the sample. In the final two columns, each stock is only allowed to be purchased by a household once, so there are no repeat purchases of the same stock in these columns' calculations. All returns are expressed in basis points, and t-statistics are in parentheses. *$p < .1$; **$p < .05$; ***$p < .01$.

Five-day returns in basis points

Skill Decile	Raw rets Mean	Four-factor α	DGTW α	$\geq$ 5 trades Four-factor α	$\geq$ 5 trades DGTW α	No Repeats Four-factor α	No repeats DGTW α
1 (low)	-89.1	-144.6***	-136.3***	-70.7***	-67.5***	-61.8***	-59.7***
2	-37.3	-86.8***	-84.5***	-38.4***	-37.8***	-46.5***	-46.7***
3	-3.0	-52.3***	-50.4***	-26.3***	-29.0***	-22.3***	-23.9***
4	9.9	-39.0***	-40.1***	-20.3***	-22.8***	-20.7***	-21.9***
5	36.3	-12.2***	-12.1***	-11.4***	-12.9***	-5.7	-8.4**
6	49.9	5.3	3.5	7.3**	4.9	1.2	-2.1
7	76.9	31.5***	29.3***	9.5***	6.1*	-2.2	-5.3
8	93.3	55.8***	50.6***	21.9***	18.9***	5.7	2.9
9	129.9	89.6***	82.1***	32.1***	27.5***	30.7***	24.4 ***
10 (high)	165.4	126.5***	117.7***	57.1***	51.6***	35.7***	31.1***
$10 - 1$	254.5***	271.1***	254.6***	127.8***	119.2***	97.4***	90.6***
t-stat	(54.22)	(59.99)	(57.09)	(34.31)	(32.54)	(21.24)	(20.06)

Table 4. Five-day returns: Second half of the sample

This table reports the average five-day return of trades that have been sorted into deciles according to the assessed ability of the household. The first three columns of returns are for skill groups that have been formed according to the p-value that the trader's risk-adjusted returns have a positive average in the first half of the sample. These results only include trades of households that have placed at least 25 trades in the first half of the sample. The first column reports the average daily raw return for each group. The next two columns report four-factor risk-adjusted and DGTW characteristic-adjusted returns. The final four columns report risk-adjusted returns for groups formed using two alternative sorting procedures. In the fourth and fifth columns, stocks are sorted according to the risk adjusted returns of the trader even if the trader had as few as five trades in the first half of the sample. In the final two columns, each stock is only allowed to be purchased by a household once, so there are no repeat purchases of the same stock in these columns' calculations. All returns are expressed in basis points, and t-statistics are in parentheses. *$p<.1$; **$p<.05$; ***$p<.01$.

Five-day returns in basis points

Skill Decile	Raw rets Mean	Four-factor α	DGTW α	$\geq$ 5 trades Four-factor α	$\geq$ 5 trades DGTW α	No repeats Four-factor α	No repeats DGTW α
1 (low)	-6.2	-47.5***	-41.1***	-34.7***	-29.8***	-30.3***	-25.6 ***
2	16.3	-27.0***	-22.6***	-21.0***	-18.1***	-27.5***	-26.1***
3	15.4	-21.1***	-20.3***	-9.0	-7.5	-13.6***	-11.6*
4	23.1	-14.1*	-12.9*	-7.6	-6.9	-14.2***	-13.3*
5	25.2	-12.1*	-10.4	-17.6***	-15.8***	-15.8***	-11.9*
6	40.3	3.0	6.4	-3.8	-1.4	-15.3***	-13.4**
7	42.9	1.3	4.7	-6.0	-2.1	-16.8***	-13.9**
8	41.4	-0.9	4.2	-4.3	-0.2	-7.2	-6.5
9	40.2	0.2	0.5	3.4	6.0	-11.1*	-8.4
10 (high)	48.3	10.6*	14.0**	8.4*	10.3**	-2.1	-0.7
$10-1$	54.6***	58.0***	55.1***	43.1***	40.2***	28.2***	25.2***
t-stat	(7.41)	(8.24)	(7.96)	(8.39)	(7.96)	(3.94)	(3.59)

Table 5. Considering sell performance

This table reports the average five-day return of trades that have been sorted into quintiles based on past buy or sell success. Returns are calculated by cumulating the risk-adjusted returns of the stock over the five days after it was purchased. The first two columns of numbers are for skill groups that have been formed according to the p-value that the trader's risk-adjusted buy returns have a positive average in the first half of the sample. The second two columns are calculated using the trader's risk-adjusted sell returns, or the returns to stocks the investor has sold. Lower returns after sales are considered evidence of skill, so the most skilled group in columns 3 and 4 are those with the most negative returns in the five days following a sale (in the first half of the sample). The first four columns only include trades of traders that have placed at least 25 buy trades in the first half of the sample. The final column sorts traders by the p-value that corresponds to their buy return minus their sell return, and it requires that traders have at least 40 buys and sells in the first half of the sample. All returns are expressed in basis points, and both simple and portfolio t-statistics are in parentheses. *$p < .1$; **$p < .05$; ***$p < .01$.

Five-day returns in basis points

Ability	Sort by buy performance		Sort by sell performance		Sort by buy - sell
Quintile	Buy α	Sell α	Buy α	Sell α	Buy - sell α
1 (low)	-37.6***	-11.2**	-17.7***	1.1	-16.9***
2	-17.7***	-0.7	-15.8***	3.5	-11.9***
3	-3.8	5.9	-5.3	-3.7	-0.7
4	0.2	-5.9	-14.5***	-6.5	-2.4
5 (high)	6.4	-10.6**	-0.9	-18.7***	14.0***
$10 - 1$	44.0***	0.7	16.8**	-19.8***	30.9***
t-stat	(8.44)	(0.13)	(3.11)	(-3.65)	(8.22)

Table 6: Trade characteristics: Second half of the sample

This table reports on a number of characteristics of the trades placed in the second half of the sample by households sorted into skill deciles based on their performance in the first half of the sample. In the first column, the average same-day returns of purchases are listed by group. In the second column, the average closing bid-ask spread of the stocks purchased by each group is listed. The third and fourth columns contain the average number of days between a purchase and the closest earnings announcement and the standard deviation of this number of days. The fifth column contains the average abnormal volume of each stock traded on the day it is purchased. The last two columns contain proxies for investor wealth: the group median portfolio value in January of 1991 and the group median of each investor's average trade size over the first half of the sample. t-statistics are reported in parentheses.

Ability decile	Same-day return (%)	Bid-ask spread (%)	Days to closest earnings	SD of days to earnings	Daily abnormal volume (%)	Portfolio value (US$1991)	Average trade size (/)
1 (low)	-0.28	1.80	29.8	54.2	144	39,500	8,085
2	-0.20	1.75	31.9	73.2	116	41,375	7,223
3	-0.28	1.93	29.4	48.9	111	22,925	6,653
4	-0.55	2.10	31.2	57.0	104	36,540	5,930
5	-0.26	1.86	30.6	56.1	102	38,915	7,148
6	-0.29	1.93	30.2	56.7	94	41,407	6,384
7	-0.26	1.99	30.3	60.4	94	52,212	8,183
8	0.61	2.09	32.1	71.1	91	43,512	6,897
9	-0.26	2.10	31.4	63.0	97	48,068	6,450
10 (high)	-0.11	1.99	31.8	67.9	99	41,758	$8,979
$10-1$	0.17	0.19	1.9	13.7	-45	2,258	894
t-stat	(2.16)	(9.32)	(3.55)	(3.32)	(-14.15)	–	–

Table 7. Long horizon returns and characteristics: Second half of the sample

This table reports the average holding period returns and characteristics of trades in the second half of the sample that have been sorted into deciles according to the average performance of the investor in the first half of the sample. In the first column, both the average return reported and the performance measure used for sorting is the buy-and-hold return corresponding to the stock and the transactions dates of the investor minus the buy-and-hold return of the value-weighted market index over the same period. In the rest of the columns, trades are sorted by the average characteristic-adjusted return of the investor from 1990 to 1993. All returns are buy-and-hold returns and are reported as a percentage. The characteristic-adjusted returns reported in column 2 are calculated by subtracting the buy-and-hold return for a characteristic-matched portfolio from the purchased stock's return. column 3 reports the average short-term return decile ranking of each household over the second half of the sample. It varies from 1 to 10 and has an average value of 5.5. Columns 4 and 5 report the average number of positions purchased by each trader over both sample halves, and column 6 reports the average holding period (in calendar days) over the second half of the sample. The last three columns report the average size quintile, book-to-market quintile, and momentum score for each trade. The momentum score is set to one for a previous loser stock; it is set to three for a previous winner; and it is set to two for stocks that are neither winners nor losers. The penultimate row of the table reports the difference between the best and worst performance deciles, and the last row reports a t-statistic for the hypothesis that this difference is equal to zero. *p <.1; **p <.05; ***p <.01 (column 2).

Ability Decile	Mkt-adj return	Char-adj return	5-day rank	Buys 1990–1993	Buys 1994–1996	Holding period	Size quint	B/M quint	Mom score
1 (low)	0.712	0.238	5.5	77.2	115.0	225.8	2.1	3.9	2.0
2	2.018	0.254	5.4	106.1	158.9	186.5	2.1	4.1	2.0
3	1.089	-0.833**	5.6	101.3	160.3	161.9	2.0	4.1	2.0
4	0.870	-0.420	4.9	94.4	189.9	145.6	1.9	4.1	2.1
5	1.027	0.084	5.1	142.9	247.0	119.4	1.9	4.1	2.1
6	1.411	0.033	5.8	219.8	285.6	123.3	2.0	4.2	2.1
7	1.543	0.033	6.1	110.1	437.1	126.3	1.9	4.2	2.1
8	1.660	0.207	5.5	91.1	207.0	131.1	2.0	4.0	2.1
9	3.402	1.475***	5.9	91.6	188.2	150.0	2.0	4.1	2.0
10 (high)	6.045	3.379***	6.6	61.7	125.2	191.6	2.1	3.9	2.0
10 − 1	5.333	3.141	1.1	-15.5	10.2	−34.3	−0.02	−0.00	0.05
t-stat	(6.30)	(4.17)	(24.00)	(-18.05)	(6.85)	(−12.77)	(−1.30)	(−0.46)	(4.59)

Table 8. Short horizon trading strategy returns

This table reports the results of a performance regression of a short horizon trading strategy's return on the daily realizations (and lagged realizations) of four factors: the market return minus the risk-free rate (RMRF), the return of high minus low book-to-market stocks (HML), the return of small minus large stocks (SMB), and the return of a momentum portfolio that is long past winners and short past losers (MOM). Portfolios are constructed by sorting on each date accounts that have traded at least 25 times up to that date based on the p-values of their past trades. Only the largest two-thirds of all CRSP stocks are included in portfolios. When a stock is included in a portfolio, it remains in the portfolio for a calendar week. Columns 1 and 2 report the regression for the portfolio that mimics past successful traders, while columns 3 and 4 report the same regression for the portfolio that mimics past unsuccessful traders. The high minus low strategy's returns used for columns 5 and 6 are constructed by going long the trades of accounts in the top quintile and going short those of the bottom quintile. Results are only reported for days on which at least 25 stocks are in the top and bottom quintile portfolios. The sample size is 1,226 days. Intercepts are expressed as a percentage per year (i.e., multiplied by 552), and t-statistics are in parentheses.

Factor-adjusted returns (annualized daily returns)

	Strategy					
Variable	High		Low		High-low	
Intercept	9.31	16.13	-9.02	-1.44	18.33	17.57
	(2.01)	(4.13)	(-1.57)	(-0.29)	(5.54)	(5.24)
$RMRF_t$	1.44	1.37	1.52	1.39	-0.08	-0.02
	(43.85)	(35.56)	(37.36)	(27.88)	(-3.43)	(-0.56)
HML_t		0.62		0.62		-0.00
		(12.91)		(10.03)		(-0.06)
SMB_t		-0.67		-0.79		0.12
		(-12.24)		(-11.19)		(2.59)
MOM_t		0.02		0.15		-0.13
		(0.56)		(2.49)		(-3.10)
$RMRF_{t-1}$		-0.14		-0.16		0.02
		(-3.57)		(-3.22)		(0.68)
HML_{t-1}		-0.07		-0.07		0.00
		(-1.51)		(-1.22)		(0.07)
SMB_{t-1}		0.20		0.23		-0.03
		(3.62)		(3.29)		(-0.73)
MOM_{t-1}		-0.11		-0.15		0.04
		(-2.25)		(-2.45)		(1.07)

Table 9. Long horizon trading strategy returns

This table reports the results of a performance regression of a long horizon trading strategy's return on the daily realizations (and lagged realizations) of four factors: RMRF, HML, SMB, and MOM. Portfolios are constructed by sorting households with at least 25 trades at the end of 1993 by their characteristic-adjusted average holding period returns over the previous three years. Columns 1 and 2 report the regression for the portfolio that mimics past successful households, while columns 3 and 4 report the same regression for the portfolio that mimics past unsuccessful households. The high minus low strategy's returns used for columns 5 and 6 are constructed by going long the trades of accounts in the top quintile and going short those of the bottom quintile. Intercepts are expressed as percentage per year, and t-statistics are in parentheses. Each regression is estimated with 757 observations, consisting of daily returns from 1994 to 1996.

Factor-adjusted returns (annualized daily)

	Strategy					
Variable	High		Low		High-low	
Intercept	10.10	14.50	3.08	7.01	7.02	7.48
	(2.61)	(5.05)	(1.12)	(3.33)	(3.29)	(3.99)
RMRF_t	1.20	1.25	1.16	1.14	0.04	0.11
	(45.36)	(44.08)	(61.72)	(54.88)	(3.03)	(5.82)
HML_t		0.62		0.31		0.31
		(17.45)		(11.91)		(13.32)
SMB_t		-0.36		-0.40		0.04
		(-8.15)		(-12.29)		(1.32)
MOM_t		-0.15		-0.19		0.04
		(-3.94)		(-6.61)		(1.39)
RMRF_{t-1}		-0.06		-0.06		0.00
		(-2.00)		(-2.83)		(0.12)
HML_{t-1}		-0.03		-0.03		-0.01
		(-1.00)		(-1.08)		(-0.33)
SMB_{t-1}		0.06		0.06		0.00
		(1.40)		(1.83)		(0.08)
MOM_{t-1}		-0.08		-0.05		-0.04
		(-2.13)		(-1.64)		(-1.41)