School of Finance
Harvard University
Working Paper No. 04-025

Negotiation, Organization and Markets
Harvard University
Working Paper No. 02-45

Can Individual Investors Beat the Market?

Joshua D. Coval
Harvard Business School; NBER

David A. Hirshleifer
Ohio State University

Tyler Shumway
University of Michigan at Ann Arbor

Can Individual Investors Beat the Market?

Joshua D. Coval, David Hirshleifer and Tyler Shumway*

September 2005

Abstract

We document strong persistence in the performance of trades of individual investors. The correlation of the risk-adjusted performance of an individual across sample periods is about 10 percent. Investors classified in the top performance decile in the first half of our sample subsequently outperform those in the bottom decile by about 8 percent per year. Strategies long in firms purchased by previously successful investors and short in firms purchased by previously unsuccessful investors earn abnormal returns of 5 basis points per day. These returns are not confined to small stocks nor to stocks in which the investors are likely to have inside information. Our results suggest that skillful individual investors exploit market inefficiencies to earn abnormal profits, above and beyond any profits available from well-known strategies based upon size, value, or momentum.

*We thank Malcolm Baker, Randy Cohen, Laura Frieder, Mark Grinblatt, Mark Kamstra, Terry Odean, 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. Coval is at the Harvard Business School *(jcoval@hbs.edu),* Hirshleifer is at the Fisher College of Business at Ohio State University *(hirshleifer_2@cob.osu.edu),* and Shumway is at the University of Michigan Business School *(shumway@umich.edu).* Shumway acknowledges financial support from the Bank One Corporation Assistant Professorship and the NTT fellowship of the Mitsui Life Center.

Financial economists have debated the efficient market hypothesis (EMH) for decades. Most formal tests of market efficiency test whether risk-adjusted security returns are predictable by measuring the profitability of investment strategies designed to exploit potential mispricing. Although such tests are good at determining whether a particular strategy represented a potential profit opportunity, they place a heavy burden on the econometrician's ability to identify trading strategies that have power to reject the null. If the econometrician chooses strategies that are not properly designed to exploit existing mispricing, he will fail to detect market inefficiency.

Furthermore, such tests do not identify whether the proposed strategies were actually exploited, or even recognized, by sophisticated investors at the time of their trading decisions. Often econometricians employ financial, econometric, and computational databases and technology that were not readily available to investors at the time of their trades.[1] Thus, some authors have suggested that the traditional efficient market hypothesis is too strong, and have proposed milder notions of efficiency that reflect reasonable constraints on the ability of intelligent investors to process information (see, e.g., the 'adaptive efficiency' argument of Daniel and Titman (1999)). In a similar spirit, we offer an approach for evaluating market efficiency that is based upon not just whether profit opportunities are available, but whether some set of real investors have demonstrated abnormal skill in generating abnormal trading profits. Our evidence suggests that even this milder form of market efficiency is violated.

According to the EMH, investors' risk-adjusted performance should be random: unless they possess relevant private information, they should neither consistently beat the market nor should they, transaction costs aside, consistently underperform the market. If those individual investors who have performed abnormally well in the past continue to perform abnormally well in the future by an amount that is not explained by mere chance, market efficiency may be violated. Alternatively, investors whose abnormal performance persists may be exploiting superior private information about fundamentals rather than superior skill at identifying and exploiting market mispric-

[1]For example, although historical price data are in principle public information that is costlessly available to all, in reality such data are costly, as evidenced by the fees researchers pay for access to electronic databases.

ing. Therefore, in addition to testing for performance persistence, we test whether abnormal performance is persistent among mid- to large-cap stocks, which presumably have less information asymmetry. In addition, we examine whether abnormal performance persists when we exclude the few stocks in which an investor transacts frequently (and thus is more likely to possess private information), and whether abnormal performance persists even among the stocks in which an investor transacts only once. Finally, we examine whether some investors persistently *underperform*; it is not plausible that such investors have superior information and then trade the wrong way.[2]

An advantage of using individual trader performance to evaluate market efficiency is that this approach vastly expands the set of strategies indirectly being tested and the set of observable variables on which trades potentially may be conditioned. The burden is no longer on the econometrician to propose the return predictors to be used in constructing optimal trading strategies. Our approach obviates the need to run thousands of diverse tests and then debate how to discount the statistical significance of these tests to adjust for data mining.

Using the transaction record of a large sample of accounts[3] at a major discount brokerage, we conduct several tests of whether individual performance persists. We measure investor performance in two ways. In what we refer to as long horizon tests, we evaluate investors by their average holding period returns.[4] While measuring performance by holding period returns is natural and intuitive, it can cause statistical problems for some tests since the returns are generally measured for overlapping periods and are thus not independent observations.

We therefore also perform short horizon tests, in which we evaluate investors by measuring the average returns on their positions during the week after they place each

[2]The market inefficiency theory offers a simple explanation for why 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 just smart traders who make money by exploiting inefficiency, but foolish traders who lose money in the processing of generating the inefficiency.

[3]Throughout the paper, we use the terms account, investor, and trader interchangeably. However, our unit of observation is a brokerage account, which may exist for an individual or for a household.

[4]We mark open positions to market at the end of the relevant sample period, so evaluation returns are not exactly holding period returns.

purchase. To ensure that our short horizon measures of trader profits are not driven by any short term price pressure created by the trade, in most of our tests we wait a day before beginning our measurement of returns. We also measure performance as the p-value of a one-sided test of whether the mean return earned by the account over the sample half is positive, and then compute the correlation of the probability across sample halves.

There is some debate as to the proper way to adjust for risk in evaluating investment performance. It is common to use factor benchmarks such as the 3-factor model of Fama and French (1992), or to control for characteristics such as book/market, size, or momentum (see, e.g., Daniel et al (1997)). However, each of these characteristics can be a proxy for market mispricing, as can the factor loadings in factors that are generated by portfolios based upon such characteristics. Thus, such benchmarks can absorb some of the abnormal performance that the test is trying to measure (see, for example, the discussions of Loughran and Ritter (2000) and Daniel, Hirshleifer and Teoh (2001)). We use as our benchmarks the Fama French (1992) 3-Factor Model and the Daniel et al (1997) characteristics-based adjustment. If these benchmarks reflect models of risk, 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 upon other well-known return predictors.

Given our long and short horizon measures and our methods of risk adjustment, we turn to tests of whether individuals can beat the market. For our first test, to examine whether trading performance persists, we divide the sample in half and examine the correlation in the abnormal performance of an account's trades between the two halves. We find that trader performance, regardless of measurement horizon or risk adjustment, is consistently correlated across the two sample halves. The correlations are approximately 10 percent, and differ from zero with high significance. The positive correlation survives a variety of robustness checks, including comparing even and odd quarters, removing the smallest one-third of the sample of CRSP stocks, and removing an account's trade in a stock that the investor has traded previously (a possible indication that the individual has access to inside information).

Next, to examine the economic significance of short horizon performance persistence, we classify each trade of each account according to the performance of *all other* trades placed by that account. In this way, we maximize our precision in classifying an account according to the investor's degree of skill. We call this a complementary image procedure, in analogy to the phenomenon in visual perception in which a figure is identified through its contrast with a complementary background. The complementary image procedure uses ex post information, and therefore does not identify a feasible trading strategy. However, the procedure does strictly quarantine the classification stage for a trade from the performance of that trade itself. This is essential, as otherwise we would induce a mechanical bias in which trades classified as coming from skillful traders are more likely to be profitable, even if there were no true persistence. The classification for each trade is made by sorting according to the one-sided p-value in the test of the hypothesis that all other trades by that account have a positive rather than zero (possibly adjusted) mean return. Finally, we form decile portfolios and calculate the average of the returns earned on each purchase during the subsequent week.

The difference between the abnormal returns of the top and bottom decile portfolios is striking. Trades in the top decile earn between 12 and 15 basis points per day during the following week. Trades in the bottom decile lose between 11 and 12 basis points per day. The results are highly statistically significant and are invariant to using a factor- or characteristic-based risk/mispricing adjustment. The results are also robust to the removal of the smallest third of CRSP stocks and to the removal of trades in stocks traded more than once by the account. This suggests that the top traders earn economically large returns from their stock selection skills in a wide range of companies.

To examine the economic significance of long horizon persistence, we sort investors into performance deciles based upon 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 almost 4 percent larger than those of investors in the bottom decile. This excess performance is extremely statistically significant. Since the average holding period of these investors is approximately six months, investors in the top decile outperform those in the bottom decile by about 8 percent per year on a characteristic-adjusted basis.

Finally, to investigate whether the information contained in account trading behavior offers profitable trading opportunities to those with access to this information, we construct and test the profitability of both a long horizon and a short horizon trading strategy. The strategies we consider use trading data available ex ante at each point in time. For the short horizon strategy, on each date we rank all traders who have traded at least 25 times up to that point. We rank them according to the one sided p-value for testing whether the mean return is positive. Next, for each quintile of traders, we construct a portfolio consisting of all stocks purchased by traders in that quintile during the previous week, weighted by the stocks' market values. We then construct a zero-cost trading strategy that is long the portfolio of the top quintile and short the portfolio of the bottom quintile. The returns of this strategy are then benchmark-adjusted using factor and characteristic-based adjustments.

Using the one-week holding period, the short horizon strategy earns abnormal returns of 5 basis points per day, or 13.7 percent per annum. Again, the results are robust to removing small stocks and trades in stocks an account has traded previously. The strategy is also assessed using a one-day and a one-month holding period. Using the one-day holding period, the return is 7 basis points per day, whereas the daily return of one-month holding period is indistinguishable from zero, suggesting that many of the profitable trades have horizons of less than a month.

Our long horizon portfolio strategy mimics the portfolio holdings of previously successful traders and shorts the positions of previously unsuccessful traders. Like the short horizon strategy, this long horizon strategy earns statistically significant abnormal returns of 4 to 5 basis points per day. Moreover, because holding periods are matched to those of the trades that are being mimicked, this strategy achieves the outperformance with far less turnover.

The paper proceeds as follows. In Section I we describe the data. Section II reports the results of the across-sample correlation tests and the estimation of returns of trades of accounts ranked according to the performance of their other trades. Section II also discusses the results of a trading strategy designed to exploit information contained in the trades of well- and poorly-performing accounts. Section III provides some interpretation of our evidence and Section IV concludes.

I. Performance Persistence and Individual Investors

At first glance, it would seem that a search for evidence that individual investors can beat the market is not very promising. Individual traders are often regarded as at best uninformed, at worst fools. The noise trader approach to securities markets, for example, identifies individual investors as generating demands that are generally driven by liquidity or psychological considerations unrelated to the information about underlying security values (see, e.g., Black (1986), De Long et al (1990) and Lee et al (1991)). Several studies have documented the poor average performance of individual traders relative to the market and to institutional traders. For example, individual traders appear to trade too much, maintain underdiversified portfolios, and hold onto losing positions for too long.[5]

However, not all individual traders do poorly in their investments. Indeed, as Barber and Odean (2000) note, the top-performing quartile of the individual accounts in their dataset outperform the market on average by 0.5 percent per month. Ivkovich and Weisbenner (2005) find that individual investors generate relatively high returns when purchasing the stocks of companies close to their homes compared to the stocks of distant companies. Ivkovich, Sialm and Weisbenner (2005) find that individuals with relatively concentrated portfolios outperform those that are more diversified, and Kaniel Saar and Titman (2005) find evidence that stocks that are heavily purchased by individual investors in one month exhibit positive excess returns in the following

[5] See, e.g., Blume and Friend (1975), Ferris, Haugen, and Makhija (1988), Odean (1997), Odean (1998), Barber and Odean (2000), Grinblatt and Keloharju (2001). Cohen et al (2002) and Hirshleifer et al (2001) also report evidence suggesting that individual investors trade unprofitably in response to cash flow news and earnings announcements.

month. These findings raise the question of whether some individual investors have superior skill, and are able to profit thereby. Of course, given the low diversification of many of the accounts, one expects substantial cross-sectional variation in account performance by chance even if there are no differences in skill. The central question we address in this paper, therefore, is whether all individual investors who earn profits on their trades are merely lucky, or whether some are indeed skillful.

The issue of whether superior performance persistence exists has been examined most extensively for mutual funds. Most studies of mutual funds find that the abnormal performance of the average funds lags that of the overall market.[6] Similarly, only limited evidence exists suggesting that those funds that outperform can be expected to continue to do so in the future.[7]

While few would expect individual traders to be, on average, better informed than mutual fund managers, there are compelling reasons to believe that individual traders are better positioned to *exploit* a given informational advantage. First, individual traders almost always trade smaller positions than professional traders. As a result, the pressure that their trades impart on prices is likely to be much less. This makes them far better positioned to trade using strategies that exploit smaller or shorter-term deviations from fundamental values. Second, individual traders are less constrained than mutual funds to hold a diversified portfolio or to track the market or a given benchmark.

From the standpoint of the researcher seeking to detect performance persistence, the transaction-level datasets 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.

[6]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.

[7]Lehman and Modest (1987), Grinblatt and Titman (1992), Hendricks et al (1993), Goetzmann and Ibbotsen (1994), Brown and Goetzmann (1995), Elton et al (1996), and Wermers (1996) 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 et al (2001) employ a Bayesian approach to detect managers with positive expected alphas.

II. Data

This paper studies a dataset provided by a large discount brokerage firm on the trades placed by 115,856 accounts from January 1990 through November 1996.[8] Table 1 reports summary statistics. Since many of our tests focus on the 16,668 accounts that placed at least 25 trades during the sample period, we also report statistics for this subset. Overall, the average account placed 15 purchases of an average size of $8,599 in 9.2 different companies. Not surprisingly, the median and standard deviation indicate substantial right-skewness in the distribution. The median account placed six purchase trades in four different companies at an average value of $4,369. For the subset of purchases that were later (at least partially) sold in our sample period, the holding period for the average (median) account was 378.11 (293) days.

The lower panel of Table 1 reports summary statistics for accounts that traded a minimum of 25 times during the sample period. The average account traded an average size of $10,301 in 66.4 trades in 36.8 different companies. Again the median figures are somewhat lower, with the median account trading 43 times in 26 different companies at an average value of $5,675. Not surprisingly, the average holding period for the active accounts, 244.33 days, is considerably lower than the overall figure.

A key challenge to our inquiry is that we only have seven years of data, which limits our power to assess the trading skills of the individuals in our dataset. We therefore design short horizon performance procedures to maximize statistical power. To the extent that the profitable trading opportunities available to skillful individuals are short-lived, the variability of the component of a trade's return that is unrelated to skill will account for an increasing fraction of total return variability as the holding period grows. Thus, for short-lived profit opportunities (such as event-related trading), any inference about abnormal expected returns is likely to be considerably easier when the focus is on shorter horizons, which the transactions data allow. With this in mind, our short horizon tests typically focus on the returns individuals obtain

[8] While the discount brokerage data include files with information about account holdings and trades, we only use the file that records account trades. There are 126,488 accounts in the trades file, but only 115,856 accounts had at least one purchase during the sample period.

from their trades during the week that follows their trades.

An additional step we take to mitigate the inference problem is that, for many of the tests, we restrict our attention to accounts that have traded at least 25 times.This removes more than 99,000 accounts from consideration, but has the benefit of ensuring that for each account we study we have sufficient data to estimate trading profits with some accuracy.

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, and one lag of each of these factors to adjust for the possibility of non-trading 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. Both the size and the book-to-market factors are calculated by taking the equal weighted average of the top three value-weighted size and book-to-market decile portfolio returns and subtracting the average of the bottom three decile portfolio returns.

For the short horizon characteristic-based risk adjustment, we follow a procedure similar to the approach used by 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 three months. 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 that was in the 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. As a robustness check, we repeat some of our momentum calculations using stock returns over the past twelve months, exluding the most recent month, to assign stocks to the three categories. However, our results do not appear to be very sensitive to the method of adjusting for momentum. Combining our three momentum categories with five size and five book-to-market categories results in seventy-five possible classifications for each stock. We calculate daily equal-weighted average returns for each of these seventy-five 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.

While our short horizon tests come close to maximizing our statistical power in detecting performance persistence, they ignore investors' decisions to sell and thus they do not test whether investors actually beat the market during our sample period.[9] To examine whether investors translate any short horizon performance persistence into higher wealth and to examine whether our short horizon results are robust, we also examine performance persistence with holding period returns. In our holding period returns results, we calculate returns for positions that are open at the end of the relevant sample period by marking positions to market.

Adjusting holding period returns for risk or mispricing is substantially more difficult than adjusting five-day returns. We adopt a characteristic-based adjustment 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. We then subtract this return from the position's return. Since buy-and-hold returns are not simple functions of daily portfolio returns, this is a computationally cumbersome approach. However, since we use buy-and-hold returns to adjust for risk, we are confident that our long horizon results are not driven by microstructure biases like bid-ask bounce or portfolio rebalancing.

[9] Our short horizon results test whether individuals' buy trades, in combination with a mechanical rule of selling after five days, yield profits. In this sense we document that individual investors can beat the market at short horizons. It is, however, possible that individual investors who are systematically smart in their buy trades (generating positive measured profits) are systematically dumb in their sell timing, creating offsetting losses. Even if this were the case, it would not invalidate our main points: individual investors are able to trade in a way that identifies market inefficiencies, and as a result their trades contain sufficient information to beat the market.

Our short horizon tests for stock selection ability focus solely on the performance of trades that initiate or expand existing positions in companies. We ignore all sales of shares. Our rationale for this is that we expect that sales are often not strongly driven by specific analysis of (or private information about) the sold stock. Liquidity needs, or the reversing of a position taken long ago in order to diversify may motivate many sales. Such sales may also be motivated by a desire to move into other firms expected to outperform the market. Since few accounts place short-sale trades, we ignore them as well. In contrast, we regard the purchase of a particular stock (as contrasted with the alternative of investing in a mutual fund) as a relatively clear indication that the investor expects that stock to outperform the market. Of course, our long horizon tests use both purchase and sales decisions.

Using the abnormal return series, we calculate the average daily returns earned during the days that immediately follow a given purchase using four different horizons. Measuring returns at different horizons allows us to verify that our results are robust and to explore the nature of the information that successful investors incorporate into their decisions. If an account purchases shares in a company on a particular date, for our tests that use a weekly horizon, we calculate the average daily return in that company during the next five trading days. Returns calculated using a one-day horizon use only the subsequent trading day's return, whereas those using a one-month horizon use the subsequent 20 trading days' returns. Our long horizon results use the entire holding period associated with the position. We typically do not include the trading day's return in our calculations to ensure that any price pressure created by the purchase–particularly of small companies–does not distort our results in favor of the investor. This also ensures that any investor losses are not driven by same-day returns.

III. Results

A. Return Correlations

We begin with a simple correlation test for persistence in account performance in which average account 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 our sample. The seven year sample is split at the end of the fourth year to ensure that a roughly equivalent number of accounts have traded at least 25 times in both sample halves. To make sure our results are not contaminated by, for example, individuals who trade more frequently if their performance is good, we place no minimum trade restriction on the second half of the sample. We then calculate the correlation of each account's performance in one half of the sample with that account's 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 Fama and French (1992) 3-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 two additional return correlations. The first compares the ratio of the mean return to the return's standard deviation (a return/risk ratio) across the two sample halves. For the second, we compute the (one-sided) p-value associated with the t-statistic of the hypothesis that a given account's abnormal return is positive during the sample half. We then calculate the correlation in the account p-values across the two sample halves. Finally, as a robustness check, we also calculate the correlation in returns obtained in even and odd quarters. The results for the three short horizon return correlations are reported in the first three panels of Table 2.

The correlation in performance across the sample is consistently around 10 percent and is highly statistically significant. The results are significant for both Pearson and rank correlation calculations and are largely invariant to whether or how we adjust for risk.[10] 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 correlations are 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 traders.

Next, to see whether the short horizon persistence is due to trader ability (or negative ability) to time the market, we recompute our tests replacing individual stock returns with overall value-weighted market returns. This tests whether individuals consistently purchase stocks before the market rises. When individual stock returns are replaced with the overall market return, most of the evidence of performance persistence disappears. However, the p-value correlations retain most of their significance, suggesting perhaps some persistence in market timing. Overall, though, it appears that the persistence in the performance of individuals comes primarily from stock selection rather than market timing.

Last, we examine long horizon performance persistence in the last two panels of Table 2. In the absence of any risk adjustment, the correlations are quite large. However, this may simply be due to heterogeneity in investors' trading styles. Adjusting for risk (or mispricing), the correlation drops to about 11 percent for the p-value, which is consistent with the short horizon results reported above. Overall, Table 2 provides fairly strong evidence that the performance of individual investors persists. This persistence is robust to various risk adjustments and to return measurement horizon.

B. Short Horizon Performance Classification of Traders

While the above results indicate clear persistence in trader performance, they do not provide a measure of its economic magnitude. For example, what level of future returns can be expected of traders identified as among the top or bottom 10 percent? To investigate economic magnitude, we classify traders according to the performance

[10] The characteristic-adjusted returns for five days use three months of past returns to measure the momentum of each stock. Measuring momentum with returns over the past twelve months, excluding the most recent month, produces very similar results.

of their trades and then measure how well this classification explains the returns of subsequent trades. We perform this analysis for short horizon returns in Tables 3 and 4 and for long horizon returns in Table 5.

Since we only have seven years of data, and since many traders have not accumulated a sufficient number of trades to be accurately classified until fairly late in the sample, to maximize our power to identify trader ability we employ what we call the complementary image procedure. In this procedure, for each trade placed by a given trader, we use *all other* trades he has placed in our dataset to calculate his average return and the 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 the trader, we use trades placed in the future as well as those placed in the past in assessing a trader'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 individual trades only as they occur. In order to test whether there is a profitable trading strategy based upon mimicking individual investor trades, we 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 traders 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 traders' 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 traders. Under the null hypothesis that abnormal returns are unpredictable and independent over time, this procedure should not generate abnormal returns.

As discussed above, corresponding to each trade of a given trader is an average return of all other trades placed by this trader, and a p-value that this average return is positive. We then sort all trades according to the corresponding average returns and p-values, form deciles, and calculate the average returns of the trades within each decile. We write the average return of trader 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 trader j and $r_{j,l}$ is the return earned by trader 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

$$r_i = \frac{\sum_j \sum_k r_{j,k} I_{j,k}(i)}{\sum_j \sum_k I_{j,k}(i)}, \tag{2}$$

where $r_{j,k}$ is the return earned by trader j on trade k and $I_{j,k}(i)$ is an indicator variable which is one if $\hat{r}_j^k$ is within the limits of decile i.

Table 3 reports the average returns of the trades in each decile during the days following the trades' placement. Portfolios are formed according to the p-values that the *raw* returns of each trader's other trades during the five trading days after he places them are positive. We include only trades of investors who have placed at least 25 trades. In Column 1, the average raw return in investors' other trades ('sort period') is reported for each decile. The sort period average returns range from -24 basis points to $+108$ basis points per day during the week after they are placed. Since the returns are in raw terms, most deciles have positive average returns in their other trades.

Column 2 contains the average raw same-day return, earned on the trades from the time the trade is placed until the market close. We do not adjust these returns for risk since, without time-and-sales data, it is not apparent how to properly benchmark intra-day returns. For most deciles, same-day returns are on average negative. Most of the accounts appear to concede between 25 and 35 basis points on the day their trades are placed. This is likely due to the bid-ask spread component of transaction costs that the traders incur in executing their trades, as discussed in Barber and Odean (2000). Interestingly, however, the final two deciles appear to concede far less in same-day returns costs. The top decile loses only 10 basis points on the day the trade is executed, and decile nine actually earns an average of 17 basis points by the end of trading. Relative to the bottom decile, both figures are highly statistically significant. Thus, even when we focus on same-day returns, the accounts vary widely in terms of their ability to initiate positions at low cost.

Columns 3 through 8 report average daily abnormal returns for each of the portfolios during the days that follow the placement of their trades. Returns are benchmark-adjusted using the Fama-French 3-factor model. A wide difference in returns exists between the portfolios classified as having low performance and those classified with high performance. While the bottom portfolio loses between 5 and 14 basis points per day during the 5 subsequent trading days, the top portfolio averages gains of between 3 and 24 basis points. The difference between the two portfolios begins at 30 basis points and declines steadily to 8 basis points by the fifth day. Even after two weeks, a significant difference between the two portfolios remains, with the high performance portfolio outperforming the low portfolio by 6 basis points on the tenth trading day.

While the results thus far provide a strong indication of persistent differences in the ability of different individuals to select stocks, a variety of potential concerns remain. First, because we have sorted accounts according to the raw returns of their other trades, we may be sorting on their willingness to assume risk. Sorting accounts by their average risk adjusted returns rather than their raw returns should produce more powerful tests of risk adjusted performance persistence.[11]

Second, the results so far do not control for momentum. Since some have argued that momentum is a proxy for an unknown risk-factor, it is interesting to see whether our results are robust to the removal of momentum-related returns. More generally, it is useful to verify whether abnormal performance is robust to the employment of a characteristic-based benchmark that adjusts for momentum.

Third, our results may be affected by the fact that we only include accounts that have traded at least 25 times in our tests. It could be the case, for example, that investors who have done well in the past trade more often because they believe they have ability.[12] Similarly, accounts that have done poorly in the past may be more

[11]Of course, if we have errors in our risk adjustment, these will induce performance persistence when there is none. The fact that persistence exists when we sort according to raw returns helps alleviate this concern.

[12]See Barber and Odean (2000) for supportive evidence.

inclined to trade aggressively to make up for past losses.[13] This may result in a post-selection bias for persistently lucky or unlucky accounts that consequently generate the requisite 25 trades. Thus, despite a likely reduction in our power to identify trader ability correctly, it is useful to rerun our tests using all accounts that have more than one trade, sorting only on returns.

Finally, to determine whether our results are driven by trading in stocks in which investors have private information, we rerun our tests removing all trades in companies in which the investor's account has transacted more than once. It seems unlikely that many investors obtain private information sporadically in a wide range of companies. This restriction therefore focuses the test on whether an investor has superior skill at identifying and exploiting market mispricing. Each of the above tests and robustness checks are described in Table 4.

In the first four columns of Table 4, accounts are sorted by the p-value of a test that average returns adjusted for risk by the three factor model are positive. The first column reports the average excess return to each account performance decile. The portfolios differ markedly in their average excess returns. Trades placed by accounts whose other trades have average returns that are among the bottom 10 percent in performance lose 4.9 basis points per day during the next five trading days. In contrast, trades placed by accounts ranking in the top decile earn 19.4 basis points per day. When these returns are benchmark-adjusted the picture remains the same. Using the factor-based adjustment, the trades of accounts in the bottom decile lose 12 basis points per day, whereas those of accounts in the top decile earn 15 basis points per day. Both figures are significantly different from zero and their difference, 27.5 basis points, is highly significant. Because accounts are sorted according to the average risk-adjusted returns of their other trades, this average daily return differential is somewhat higher than the average over the first five trading days in Table 3, reflecting the improvement in accuracy in classifying the accounts. The characteristic-based adjustment results in a slightly lower spread of 22.8 basis points, but one that is still

[13] Coval and Shumway (2005) document such behavior among market makers in the CBOT US Treasury Bond pit.

highly statistically and economically significant.[14]

The persistence of the poor performance 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), and our characteristic-adjusted findings further indicate that this poor performance is present even after controlling for momentum. The systematic ability of some individuals to underperform indicates that access to inside information is not the primary source of abnormal performance in our sample.

Since we calculate returns for a period that is subsequent to the day each trade is executed, the persistence of poor performance cannot be due to microstructure effects. Transaction costs like the bid-ask spread or price impact tend to manifest themselves on the same day as the trade. In standard microstructure settings with semi-strong form efficient markets, liquidity traders on average lose money because of these costs. While the investors with the worst performance in our data may be liquidity 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' (see, e.g., Hirshleifer (2001) for a review of several models). Our findings are consistent with individual investors being a heterogeneous group that includes both foolish and sophisticated traders. The evidence is consistent with systematic underperformers being individuals who tend to trade on the wrong side of market inefficiencies.

Although the portfolios are constructed using ex-post data, the return differentials nonetheless raise doubts about the efficient markets hypothesis prediction that abnormal returns on an account's trades will be independent draws. Either we have adjusted for risk incorrectly and thus we have induced some cross-sectional correlation in the returns through our measurement procedure, or the accounts have significant

[14]Characteristic-adjusted returns that measure momentum with returns calculated over the past twelve months, excluding the most recent month, exhibit a slightly larger spread. We do not report these results in a table.

dispersion in their alphas.

One possibility, as mentioned earlier, is that our restriction that accounts must have traded at least 25 times over the full sample period introduces a subtle post-selection bias. To address this possibility, the fifth column of Table 4 sorts all trades of all accounts according to the average *raw* return earned on the account's other trades. We would expect this classification to be far less precise as a segregator of skillful and lucky investors. Furthermore, it results in extreme portfolios having a bias towards accounts that trade infrequently. Nevertheless, this classification produces an average return differential that is consistent with the previous findings. Accounts in the bottom decile place trades that lose 8.8 basis points per day, whereas accounts in the top decile earn 11 basis points per day during the week following their trades.

A second possibility is that frequent account trading in the same stock generates a correlation in trade returns. To control for this possibility, we reclassify accounts using only trades made in stocks they trade once during our sample. These results are reported in the final column of Table 4. Although the statistical significance declines somewhat (there are 40 percent fewer observations), the overall result is unchanged. Although traders in the bottom decile no longer perform so poorly in subsequent trades, traders in the top decile continue to place trades that earn nearly 10 basis points per day. The spread between the top and bottom deciles, 12.7 basis point per day, remains highly significant.

This finding casts further doubt upon the hypothesis that the abnormal performance we find is due to investors trading on inside information. While it is possible that a subset of the accounts have inside information about a company or two (i.e. in their employer or friend's firm), it seems doubtful that a large number of accounts have access to inside information in a broad set of companies. Finally, to see whether the results are concentrated in small, illiquid stocks, or stocks for which individual investors are likely to be insiders, we rerun our classification using only the largest two-thirds of all CRSP firms. Once again, the results (not reported here) remain essentially the same.

C. Long Horizon Performance Classification of Traders

For long horizon returns, we do not use the complementary image procedure to examine performance. Since most of the holding periods in the data overlap, it would be very difficult to calculate the statistical significance of any differences in performance across groups of traders. Instead, we simply rank investors into deciles by their long horizon performance in the first half of the sample, and then examine the long horizon performance of their trades in the second half of the sample. Our results are reported in Table 5.

All of the returns reported in Table 5 are buy-and-hold returns, so the return horizon associated with each position varies substantially. However, in the characteristic-adjusted results, the matching portfolio returns are also buy and hold returns calculated over the holding period. Therefore, if investors do not have any trading skill, the mean characteristic-adjusted return should be zero, regardless of the holding period. Average holding periods are reported in the fifth column of Table 5. Average holding periods do not appear to be monotonically related to either previous performance or to characteristic-adjusted performance reported in the second column of the table.

The average return on each position is reported, by decile, in the first column of the table, and the characteristic-adjusted return is reported in the second column. Looking at Column 2, characteristic-adjusted returns are not monotonically related to previous performance decile. However, the characteristic-adjusted performance of the lowest previous-performance group is negative and easily the lowest of all the deciles, and that of the highest previous-performance group is positive and by far the highest of all the deciles. The difference between the highest and lowest group performance is almost 4 percent, and is very statistically significant. Since the average holding period for these investors is approximately six months, this corresponds to economically significant outperformance on the order of 8 percent per year.

The other columns in Table 5 give some details about the frequency of trade by previous performance decile and the types of stocks that different decile investors are buying. Looking at the trade frequency numbers, trade frequency does not appear to be closely related to previous performance. Comparing the number of purchases of these traders across the two sample halves, it is clear that previously unsuccessful traders take fewer positions in the second half of the sample, while previously successful traders take more positions. Looking at the characteristics of the stocks that investors hold, it is not clear that any of these characteristics varies much by previous performance decile. This is again consistent with investors being able to use their trading skills to pick stocks successfully without relying on previously documented sources of returns variation like size, book-to-market, or momentum.

D. A Short Horizon Trading Strategy

The results thus far indicate that a subset of individual investors have ability of some sort. However, it is not yet clear whether these results offer a trading strategy for an observer to exploit the information contained in the accounts' trades. To investigate the real-time returns offered by individual trade information, 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 accounts that have performed poorly up to the current date. 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 seven years of data, and much of this is used to assess traders' performance, our power to detect abnormal returns is somewhat limited. To maximize our power to detect abnormal performance, there is a tradeoff. If we only include trades of accounts with mean returns significantly different from zero, we more reliably focus on the trades of more skillful versus less skillful traders. However, to the extent that, at times, only a limited number of accounts can be classified as unusually good (or bad), such a portfolio will be highly undiversified.[15] Since we only have 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 inference is impossible. Alternatively, if we are lax in our criteria for

[15]Our reliance on value-weighted portfolios makes any lack of diversification even more pronounced.

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 accounts that have traded at least 25 times up to the current date, but 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 accounts that have traded at least 25 times up to the current date by the p-value that their abnormal return is positive. We then compute value-weighted 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 i on date t is calculated as follows:

$$r_{i,t} = \sum_j \frac{MV_{j,t}}{\sum_k MV_{k,t} I_{i,k,t}} r_{j,t} I_{i,j,t}, \tag{3}$$

where $MV_{j,t}$ is the market value of firm j on date t, $r_{j,t}$ is the return to firm j on date t, and $I_{i,j,t}$ is an indicator variable which is one if an account in portfolio i has purchased firm j within the holding period preceding date t and zero otherwise.

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 4-factor model to adjust returns for risk, adding a momentum factor to the Fama French (1992) 3-factor model.[16] When we employ the 4-factor model, we calculate a *raw* return in equation (3) and then regress the difference between the top and bottom portfolio daily return on the four factors. When we benchmark-adjust using the characteristic-based adjustment, equation (3) is calculated using the individual firm characteristic-adjusted returns. The reported results focus on the 4-factor risk adjustment though highly similar results are obtained using the characteristic-based adjustment. To ensure that trading in small, illiquid firms does not drive the results, we remove the smallest (by capitalization) third of all CRSP firms from the sample. Finally, we examine the returns to the trading strategy using three portfolio formation horizons: daily, weekly, and monthly. The results are reported in Table 6.

[16] We construct our momentum factor by taking the difference of the equal-weighted average returns of the winner portfolio and the loser portfolio identified by our characteristic adjustment algorithm.

Beginning with the one-week holding period, the strategy generates abnormal returns of 5.1 basis points per day. When only a market factor is used to risk-adjust, the returns are 4.4 basis points per day. Both figures are significant at the 5 percent level but not the 1 percent level. If we measure returns at the daily horizon — that is, on the trading day following the trade placement — the effects are slightly stronger both economically and statistically. The 4-factor abnormal returns are 6.8 basis points per day and those adjusted using the market factor are 5.6 basis points per day. When we move to the one-month horizon, however, the results essentially disappear, falling below a basis point per day and losing any statistical significance. The disappearance of significance is likely due to the mismatch between the strategy's holding period and that of the trades that it mimics. As the holding period of the strategy increases, a growing fraction of the stocks owned are no longer held by the investors the strategy is designed to mimic. We address this issue in the long horizon tests reported below.

As described above, estimating the returns to a feasible trading strategy based on our data involves a careful balance. If we base our strategy on fewer traders with more extreme past performance, the variability of our results increases. If we base our strategy on more traders in an effort to reduce the variability of our results, the average performance of the strategy declines.

In the tests reported in Table 6, we form top and bottom trader-mimicking portfolios based on the top and bottom quintiles of all ranked traders. We require all ranked traders to have at least 25 previous trades and we require both the top and bottom trader mimicking portfolios to consist of at least 25 stocks on any particular day. For the returns calculated over one week, our requirements mean that out of 1,205 possible trading days, we can only evaluate the returns to our strategy on 1,072 days. If we define the top and bottom trader mimicking portfolios by taking the top and bottom deciles of ranked traders, the strategy return can only be estimated on 945 days. Using the top and bottom deciles, the one week intercepts become 4.4 ($t = 1.74$) basis points for the 4-factor model and 5.1 basis points ($t = 2.14$) for the market model. If we define the top and bottom portfolios as the top and bottom thirty percent of traders, the number of valid return days becomes 1,101. The intercepts become 3.8 basis points ($t = 2.40$) for the 4-factor model and 3.1 basis points ($t = 2.05$) for the market model. If we estimate the regressions reported in Table 5 with a weighted least squares technique that assigns weights to each observation that are proportional to the square root of the number of stocks in the top and bottom portfolios, the intercepts become 4.7 basis points ($t = 3.09$) for the 4-factor model and 4.9 basis points ($t = 3.11$) for the four-factor model.

Comparing the results in Table 6 to those in Tables 3 and 4 raises an interesting question. Why are the results so much stronger, both economically and statistically, when portfolios are formed using all available trade data? Do the additional data points really add that much power? It turns out they do. Their contribution is twofold. First, the additional data gives us far more information with which to classify traders. With the real time trading strategy, a trader can be classified using only data up to the point in time of a given trade. Thus, the accuracy with which a trader is classified improves steadily across time. Conversely, when all data are used, each trade can be classified as if it is the last. This not only allows us to rank each trader far more accurately. It also allows us to use more trades of more traders. For instance, using our minimum of 25 trades, a trader who has placed 27 trades will only have two trades considered for our real time strategy's portfolio. On the other hand, when ex-post data are used in classifying trades, all 27 can count towards portfolio return calculations.

E. A Long Horizon Trading Strategy

Our long horizon trading strategy is somewhat simpler than our short horizon strategy. At the end of 1993, we rank all traders into quintiles based on their characteristic-adjusted buy-and-hold performance over the previous 4 years. For the period 1994-1996, we then mimic the portfolio holdings of the most successful quintile and we short the holdings of the least successful quintile. Rather than value-weighting the stocks in the portfolio, we weight each stock by the number of successful or unsuccessful investors holding the stock on each day. Like the short horizon strategy, the long horizon strategy is evaluated with a four factor model estimated with daily returns. The factor model includes three Fama-French factors and a momentum factor. The long horizon strategy results are reported Table 7.

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. Looking at the last two columns of the table, the high minus low strategy generates daily abnormal returns of about 5 basis points per day, which is quite similar to the outperformance reported for the short horizon strategy. Looking at the other columns of the table, it is evident that the outperformance of the long-short strategy is driven by good average performance on the part of past successful traders.

Interestingly, the statistically significant outperformance of our long horizon strategy requires much less portfolio turnover than the outperformance of our short horizon strategy. The total number of investor-positions over which we are averaging is 144 at the beginning of the strategy sample and 16,982 by the end of the sample. There are a total of 47,511 positions taken by our high and low performance investors over our three-year period. In the 758 days of trading we consider, the average number of investor-positions over which we are averaging is 12,041. We have 47,511 purchases to consider and 30,529 (47,511 - 16,982) sales to consider, so the average number of transactions per day is 103. If we have average transactions per day of 103 and average holdings per day of 12,041, we have an average daily turnover of 86 basis points. This works out to about 215 percent 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 traders trade the same way, or two successful traders trade in opposite ways), our implied portfolio turnover will be less than 215 percent per year.

Tables 6 and 7 both confirm that it is almost certainly feasible to execute a successful trading strategy by mimicking the trades of investors that have been successful in the past. This is perhaps our strongest evidence that individual investors can beat the market.

IV. Conclusion

Recent literature has emphasized that on average individual investors are misguided in their trades. We provide evidence here that some individual investors are persistently able to beat the market. Traders that can be classified among the top 10 percent (based on the performance of their other trades) buy stocks that earn abnormal returns of between 12 and 15 basis points per day 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. Traders classified among the bottom 10 percent of all traders place trades that can expect to lose up to 12 basis points per day during the subsequent week. In long horizon (holding period) returns, successful investors outperform unsuccessful investors by about eight percent per year. A trading strategy that exploits the information in investors' trades earns risk-adjusted returns of about five basis points per day.

Our finding that some individual investors have superior investment skills, and that others systematically underperform, suggests a new perspective on the issue of whether on average individual traders foolishly trade too much. As discussed earlier, previous studies have shown that individual investors on average lose money in their trades. However, if traders 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.[17] If traders who learn that they have unusual ability move their accounts to lower-cost or higher-leveraged trading venues (e.g. options markets), evidence drawn solely from stock trades may focus on those investors who are still in the process of learning—either how to trade, or about whether they are good traders.[18]

[17] As mentioned earlier, those investors who have superior ability at *losing money* relative to the market may be individuals who are, in equilibrium, contributing to the *creation* of market inefficiencies. In principle such an investor has a clear opportunity to learn how to make abnormal profits by reversing his trading strategy.

[18] Any such tendency to change venues would mitigate the returns obtainable by mimicking the trades of smart traders in our stock-trading sample, which suggests that true skill differences may be even greater than our estimates.

Finally, this evidence does not support the efficient market hypothesis. The ability of individual traders at a discount brokerage to select outperforming companies is not confined to small firms or to only a few firms in which the traders transact frequently; and some investors persistently trade so as to underperform. 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 implies a violation of semi-strong form market efficiency. An interesting further question is whether large brokerage firms are aware of the value of the information contained in their customers' trades.

References

Baks, Klaas P., Andrew Metrick, and Jessica Wachter, 2001, "Should investors avoid all actively managed mutual funds? A study in Bayesian performance evaluation," *Journal of Finance*, 56, 45-85.

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

Black, Fischer, 1986, "Noise," *Journal of Finance*, 41, 529-543.

Blume, Marshall and Friend, Irwin, "The Asset Structure of Individual Portfolios with Some Implications for Utility Functions," *Journal of Finance*, 30, 585-604.

Brown, Stephen J., William N. Goetzmann, Roger Ibbotson, and Stephen Ross, 1992, "Survivorship bias in performance studies," *Review of Financial Studies* 5, 553-580.

Carhart, Mark M., 1995, "Survivorship Bias and Mutual Fund Performance," Ph.D. thesis. Graduate School of Business, University of Chicago.

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

Chevalier, Judith, and Glenn Ellison, 1999, "Are some mutual fund managers better than others? Cross-sectional patterns in behavior and performance," *Journal of Finance* 54, 875-900.

Cohen, Randolph B., Paul A. Gompers, and Tuomo Vuolteenaho, 2002, "Who Underreacts to Cash-Flow News? Evidence from Trading between Individuals and Institutions," *Journal of Financial Economics* 66, 409-462.

Coval, Joshua D. and Tyler Shumway, 2001, "Do behavioral biases affect prices?" *Journal of Finance* 60, 1-34.

Daniel, Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers, 1997, "Measuring mutual fund performance with characteristic based benchmarks," *Journal of Finance* 52, 1035-1058.

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam, 1998, "A theory of overconfidence, self-attribution, and security market under- and overreactions," *Journal of Finance* 53, 1839-1886.

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam, 2001, "Overconfidence, Arbitrage, and Equilibrium Asset Pricing," *Journal of Finance* 56, 921-965.

Daniel, Kent D. and Sheridan Titman, 1999 "Market Efficiency in an Irrational World," *Financial Analysts' Journal* 55(6), November/December, (1999):28-40

De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann, 1990, "Noise trader risk in financial markets," *Journal of Political Economy* 98, 703-738.

Elton, Edwin J., Martin J. Gruber, and Christopher R. Blake, 1996, "The persistence of risk-adjusted mutual fund performance," *Journal of Business* 69, 133-157.

Fama, Eugene F. and Kenneth R. French, 1992, "The Cross Section of Expected Returns," *Journal of Finance* 47, 427-465.

Ferris, S. P., R. A. Haugen and A. K. Makhija, "Predicting Contemporary Volume with Historic Volume at Differential Price Levels: Evidence Supporting the Disposition Effect," *Journal of Finance* 43(3), (1988):677-697

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

Goetzmann, William N., and Roger G. Ibbotson, 1994, "Do winners repeat? Patterns in mutual fund performance," *Journal of Portfolio Management* 20, 9-18.

Grinblatt, Mark, and Sheridan Titman, 1992, "The persistence of mutual fund performance," *Journal of Finance* 47, 1977-1984.

Hendricks, Darryl, Jayendu Patel, and Richard Zeckhauser, 1993, "Hot hands in mutual funds: The persistence of performance 1974-1988," *Journal of Finance* 48, 93-130.

Hirshleifer, David, 2001, "Investor Psychology and Asset Pricing," *Journal of Finance* 54, 4, 1533-1597.

Hirshleifer, David, James Myers, Linda Myers and Siew Hong Teoh, 2002, "Do Individual Investors Drive Post-Earnings Announcement Drift?" Working paper, Ohio State University, Fisher College of Business.

Ivkovich, Zoran, and Scott Weisbenner, 2005, "Local Does as Local Is: Information Content of the Geography of Individual Investors' Common Stock Investments," *Journal of Finance* 60, 267-306.

Ivkovich, Zoran, Clemens Sialm, and Scott Weisbenner, 2005, "Portfolio Concentration and the Performance of Individual Investors," Working paper, University of Illinois at Urbana-Champaign.

Kaniel, Ron, Gideon Saar, and Sheridan Titman, 2005, "Individual Investor Trading and Stock Returns," Working paper, New York University.

Lee, Charles, M.C., Andrei Shleifer, and Richard H. Thaler, 1991, "Investor sentiment and the closed-end mutual funds," *Journal of Finance* 46, 75-109.

Lehman, Bruce N., and David Modest, 1987, "Mutual fund performance evaluation: A comparison of benchmarks and a benchmark of comparisons," *Journal of Finance* 42, 233-265.

Loughran, Tim and Jay R. Ritter, 2000, "Uniformly Least Powerful Tests of Market Efficiency," *Journal of Financial Economics* 55, 361-389.

Malkiel, Burton, 1995, "Returns from investing in mutual funds 1971 to 1991," *Journal of Finance* 50, 549-572.

Odean, Terrance, 1998, "Are Investors Reluctant to Realize Their Losses?" *Journal of Finance* 53, 1775-1798.

Odean, Terrance, 1999, "Do Investors Trade Too Much?" *American Economic Review* 89, 1279-1298.

Wermers, Russ, 1997, "Momentum investment strategies of mutual funds, performance persistence, and survivorship bias," Working paper, University of Maryland.

Wermers, Russ, 2000, "Mutual Fund Performance: An Empirical Decomposition into Stock-Picking Talent, Style, Transactions Costs, and Expenses," *Journal of Finance* 55, 1655-1695.

Table 1: Summary Statistics

Table 1 reports summary statistics for the entire set of accounts and for the subset of accounts that have traded at least 25 times during the sample period. Average holding period is the average holding period for an account of all purchases that were sold later in the dataset. 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, 1990, to December, 1996.

Variable (per account)	Mean	Median	St. Dev.	Min.	Max.
Full Sample ($n = 115,856$)					
Number of Purchases	15.06	6	38.43	1	3,167
Average Dollar Value	8,599	4,369	28,031	0.1	6,011,360
Number of Different Firms Purchased	9.20	4	19.68	1	1,523
Number of Purchases Sold Later	9.96	4	25.03	0	2,209
Average Holding Period (days)	378.11	293	321.34	0	2,100
Accounts with at least 25 trades ($n = 16,668$)					
Number of Purchases	66.40	43	83.61	25	3167
Average Dollar Value	10,301	5,675	18,071	50	692,524
Number of Different Firms Purchased	36.83	26	41.16	1	1,523
Number of Purchases Sold Later	32.92	22	48.60	0	2,209
Average Holding Period (days)	244.33	199	192.37	0	1,870

Table 2: Correlation Tests of Performance Persistence

Table 2 reports correlations of the investment performance of accounts 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 an account. The bottom two panels calculate performance based on the holding period return of each stock. The correlations in Columns one and two split the sample in half at the end of the fourth year and calculate the correlation in performance across the two sample halves for all accounts with at least 25 trades during the first four years. The correlations in Columns three and four divide all trades into those that occur during the first and third quarter 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 3-factor risk-adjusted return correlations regress daily returns on daily realizations of the SMB, HML and RMRF 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 in percent, and statistical significance is noted by asterisks where ** indicates significantly different from zero at the one percent level and * indicates at the five percent level.

	Sample Halves		Even/Odd Quarters	
Variable	Pearson	Rank Order	Pearson	Rank Order
Raw Returns for 5 Days Following Purchase				
Mean Return	5.5**	9.1**	6.0**	7.9**
Mean Ret. / StDev.	9.7**	10.9**	9.3**	9.4**
p-value	11.4**	12.0**	9.8**	10.7**
3-Factor Risk-Adjusted Returns for 5 Days				
Mean Return	6.9**	9.2**	12.2**	10.8**
Mean Ret. / StDev.	9.0**	9.4**	11.9**	10.2**
p-value	10.1**	10.3**	9.9**	10.1**
DGTW Characteristic-Adjusted Returns for 5 Days				
Mean Return	8.9**	9.9**	2.1	3.3
Mean Ret. / StDev.	9.1**	10.9**	4.9	5.2
p-value	11.2**	11.1**	7.0*	7.1*
Market-Timing Returns for 5 Days				
Mean Return	0.9	1.2	0.6	0.9
Mean Ret. / StDev.	−1.6	0.7	5.7**	2.1
p-value	2.9	5.0**	3.9*	7.3**
Raw Returns for Holding Period				
Mean Return	38.0**	34.8**	–	–
Mean Ret. / StDev.	24.6**	25.5**	–	–
p-value	17.5**	27.0**	–	–
Characteristic-Adjusted Returns for Holding Period				
Mean Return	25.5**	11.1**	–	–
Mean Ret. / StDev.	10.5**	11.8**	–	–
p-value	11.1**	11.8**	–	–

Table 3: One-Day Portfolio Returns: Complementary Image Procedure

Table 3 reports the average daily return of trades that have been sorted into deciles according to the assessed ability of the trader. Each trader's ability is assessed by calculating the average raw returns of all of his other trades during the five trading days after he places them. These portfolios only include trades of traders that have placed at least 25 trades. Column 1 reports the average raw return of traders' other trades for each decile. Column 2 reports the average raw return earned from the time the trade is executed to the same-day close. Columns 3 through 8 report the average risk-adjusted returns of the trade on one, two, three, four, five, and ten trading days after the trade is placed. Returns are risk-adjusted using a 3-factor Fama/French model. All returns are expressed in percent and t-statistics are in parentheses.

	Return Period							
Portfolio	Sort Period	Same Day	First Day	Second Day	Third Day	Fourth Day	Fifth Day	Tenth Day
1 (low)	−0.24	−0.29	−0.06	−0.13	−0.09	−0.14	−0.05	−0.02
2	0.12	−0.35	0.00	−0.03	−0.07	−0.08	−0.05	−0.02
3	0.26	−0.26	0.03	−0.02	−0.04	−0.04	−0.03	−0.01
4	0.30	−0.33	0.02	0.00	0.00	−0.06	−0.02	0.01
5	0.45	−0.31	0.08	0.05	0.03	−0.03	−0.02	0.02
6	0.46	−0.23	0.06	0.00	0.02	−0.02	0.03	0.03
7	0.57	−0.27	0.12	0.05	0.02	−0.03	0.01	0.03
8	0.75	−0.26	0.13	0.10	0.05	0.01	0.03	0.02
9	0.73	−0.17	0.12	0.07	0.04	0.02	0.01	0.01
10 (high)	1.08	−0.10	0.24	0.12	0.10	0.06	0.03	0.04
10 − 1		0.19	0.30	0.25	0.19	0.20	0.08	0.06
		(13.7)	(20.3)	(18.3)	(14.1)	(15.3)	(6.2)	(5.0)

Table 4: Five-Day Portfolio Returns: Complementary Image Procedure

Table 4 reports the average daily return of trades that have been sorted into deciles according to the assessed ability of the trader. Returns are calculated by averaging the returns of the firm over the five days after it was purchased. The first five columns of numbers are for portfolios that have been formed according to the p-value that the trader's other trades have a positive average return. These portfolios only include trades of traders that have placed at least 25 trades. The first column reports the average daily return in excess of the risk-free rate for each portfolio. The next column reports the standard deviation of this excess return. The next two columns report 3-factor risk-adjusted and DGTW characteristic-adjusted returns. The final two columns report 3-factor risk-adjusted returns for portfolios formed using two alternative sorting procedures. In the penultimate column, stocks are sorted according to the raw returns earned by that trader in his other trades, regardless of how few trades the trader has placed. In the final column, each stock is only allowed to be purchased by an account once, so there are no repeat purchases of the same stock in this column's calculations. All returns are expressed in percent, and t-statistics are in parentheses.

5-Day Returns

	Excess Returns				All Traders	No Repeats
Portfolio	Mean	Std. Dev.	3-Factor α	DGTW α	3-Factor α	3-Factor α
1 (low)	−0.049	1.556	−0.123	−0.109	−0.088	−0.030
2	0.020	1.588	−0.034	0.002	−0.049	−0.040
3	0.036	1.593	−0.019	−0.043	−0.023	−0.031
4	0.063	1.516	−0.011	0.028	−0.014	0.011
5	0.088	1.529	0.041	−0.025	0.031	0.028
6	0.082	1.528	0.011	0.029	0.006	0.046
7	0.109	1.539	0.030	0.027	0.053	−0.025
8	0.116	1.523	0.100	0.067	0.058	0.054
9	0.162	1.482	0.053	0.039	0.045	0.004
10 (high)	0.194	1.511	0.152	0.119	0.110	0.096
10 − 1			0.275	0.228	0.198	0.127
			(21.7)	(18.0)	(17.5)	(7.0)

Table 5: Returns and Characteristics: Long Horizon Sorts

Table 5 reports the average returns and characteristics of trades that have been sorted into deciles according to the average characteristic-adjusted return of the trader from 1990 to 1993. All returns are buy-and-hold returns, and are reported in percent. 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. Columns 3 and 4 report on the average number of positions purchased by each trader over both sample halves, and Column 5 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 1 for a previous loser stock, it is set to 3 for a previous winner and it is set to 2 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 simple t-statistic for the hypothesis that this difference is equal to zero.

Holding Period Returns

Decile	Return	Char-Adj Return	Buys 1990-93	Buys 1994-96	Holding Period	Size Quint	B/M Quint	Mom Score
1 (low)	10.646	−1.492	52.1	39.2	223.4	2.15	3.95	1.94
2	11.186	0.186	57.8	48.1	191.5	2.11	4.06	2.03
3	7.982	−0.929	69.6	58.1	157.7	2.05	4.15	2.02
4	7.975	−0.167	86.2	99.8	134.2	1.93	4.13	2.08
5	7.264	0.148	125.9	132.5	117.0	1.92	4.12	2.15
6	8.101	0.639	164.2	118.1	118.1	2.02	4.12	2.11
7	6.305	−0.679	102.9	112.0	112.0	1.93	4.26	2.10
8	8.219	0.117	82.6	105.8	125.1	2.03	4.16	2.07
9	9.687	0.828	62.4	88.1	138.8	1.96	4.09	2.14
10 (high)	12.789	2.486	68.6	73.3	178.2	2.11	3.89	2.02
10 − 1	2.143	3.978	16.5	34.1	−45.2	−0.04	−0.07	0.08
	(2.47)	(4.79)	(13.27)	(22.78)	(−14.32)	(−1.87)	(−2.67)	(6.69)

Table 6: Short Horizon Trading Strategy Returns

Table 6 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. For the three holding periods, returns are measured using the first trading day (One Day), first five trading days (One Week), and first twenty trading days (One Month) after the trade is placed. The returns are then value-weighted within each portfolio. The strategy's returns are constructed by going long the trades of accounts in the top quintile and going short those of the bottom quintile on days when at least 25 stocks are in each quintile. All returns are expressed in percent, and t-statistics are in parentheses.

Factor-Adjusted Returns: High Minus Low Portfolio (Daily Returns)

	Holding Period					
Variable	One Day		One Week		One Month	
Intercept	0.0556	0.0681	0.0438	0.0510	0.0010	0.0031
	(2.49)	(2.89)	(2.25)	(2.48)	(0.09)	(0.44)
$RMRF_t$	−0.0834	−0.1499	−0.0674	−0.0551	−0.0166	−0.0142
	(−2.09)	(−2.91)	(−1.93)	(−1.23)	(−1.29)	(−0.85)
HML_t		−0.2060		−0.0333		−0.0053
		(−3.62)		(−0.71)		(−0.31)
SMB_t		−0.0249		0.0131		0.0302
		(−0.36)		(0.23)		(1.40)
MOM_t		−0.0031		−0.0371		0.0361
		(−0.08)		(−1.07)		(2.82)
$RMRF_{t-1}$		−0.0020		−0.0292		0.0015
		(−0.04)		(−0.64)		(0.09)
HML_{t-1}		0.1273		0.0594		−0.0144
		(2.27)		(1.16)		(−0.67)
SMB_{t-1}		−0.0655		0.0003		−0.0026
		(0.95)		(0.01)		(−0.16)
MOM_{t-1}		0.0654		0.0972		−0.0142
		(1.58)		(2.81)		(−1.12)
n	911	911	1072	1072	1191	1191

Table 7: Long Horizon Trading Strategy Returns

Table 7 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: 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 accounts with at least 25 trades at the end of 1993 by their characteristic-adjusted average holding period returns over the previous four years. Columns one and two report the regression for the portfolio that mimics past successful traders, while Columns three and four report the same regression for the portfolio that mimics past unsuccessful traders. The high minus low strategy's returns used for Columns five and six are constructed by going long the trades of accounts in the top quintile and going short those of the bottom quintile. All returns are expressed in percent, and t-statistics are in parentheses. Each regression is estimated with 757 observations, consisting of daily returns from 1994 to 1996.

Factor-Adjusted Returns (Daily)

	Strategy					
Variable	High		Low		High-Low	
Intercept	0.0395	0.0460	−0.0038	−0.0062	0.0433	0.0522
	(2.73)	(3.13)	(−0.34)	(−0.55)	(6.01)	(7.20)
RMRF_t	1.2543	1.3097	1.0835	1.1911	0.1708	0.1185
	(50.81)	(40.70)	(57.18)	(48.86)	(13.90)	(7.46)
HML_t		−0.0971		0.0180		−0.1151
		(−2.41)		(0.59)		(−5.78)
SMB_t		0.2427		0.2423		0.0004
		(5.65)		(7.45)		(0.02)
MOM_t		0.0451		−0.0052		0.0503
		(1.70)		(−0.26)		(3.83)
RMRF_{t-1}		−0.0757		−0.0621		−0.0135
		(−2.35)		(−2.54)		(−0.85)
HML_{t-1}		−0.0664		−0.0563		−0.0101
		(−1.66)		(−1.86)		(−0.51)
SMB_{t-1}		−0.1260		−0.0866		−0.0394
		(−2.93)		(−2.66)		(−1.86)
MOM_{t-1}		−0.0073		−0.0200		0.0127
		(−0.28)		(−1.01)		(0.99)