Do Stock Traders Learn From Experience? Evidence from an Emerging Market*

John Y. Campbell, Tarun Ramadorai, and Benjamin Ranish†

November 15, 2012

Abstract

This paper reports evidence that individual investors in Indian equities hold better performing portfolios as they become more experienced in the equity market. Related to this, Indian stocks whose individual investors have a higher average account age tend to outperform the value-weighted Indian stock market. Several standard measures of investment mistakes, including underdiversification, high turnover, and the disposition effect, also decline with account age. These mistakes become less prevalent when investors experience poor returns resulting from them, consistent with models of reinforcement learning.

*We gratefully acknowledge the National Securities Depository Limited and the Securities and Exchange Board of India for providing us with access to the data, and the Sloan Foundation for financial support. We thank Samar Banwat, Chandrashekhar Bhave, Luigi Guiso, Nagendar Parakh, Prabhakar Patil, Gagan Rai, Manoj Sathe, Ajay Shah, U.K. Sinha, Jayesh Sule, and seminar participants at NSDL, SEBI, NIPFP, and the Einaudi Institute of Economics and Finance for numerous useful comments and discussions, and Sushant Vale for excellent and dedicated research assistance.

†Campbell: Department of Economics, Littauer Center, Harvard University, Cambridge MA 02138, USA, and NBER. Email john_campbell@harvard.edu. Ramadorai: Saïd Business School, Oxford-Man Institute of Quantitative Finance, University of Oxford, Park End Street, Oxford OX1 1HP, UK, and CEPR. Email tarun.ramadorai@sbs.ox.ac.uk. Ranish: Department of Economics, Littauer Center, Harvard University, Cambridge MA 02138, USA. Email: branish@fas.harvard.edu.

1 Introduction

Equities play an important role in normative theories of household investment. Because stocks have historically offered a risk premium, households with no initial exposure to the asset class can benefit from holding at least some stocks. The optimal equity allocation depends on market conditions and many details of the household's financial situation, including the equity premium and the household's risk aversion and other risk exposures, but typical calibrations suggest it is substantial—at least for households with sufficient wealth to justify paying the fixed cost of equity market participation (Campbell and Viceira 2002, Campbell 2006, Siegel 2007).

Direct investment in stocks is not straightforward, however, and households can lose much of the benefit of stock market participation if they make some common mistakes. Three of these can be costly even in a market where all individual stocks have the same risk and the same expected return. First, *underdiversification* increases portfolio risk without increasing return (Blume and Friend 1975, Kelly 1995, Calvet et al. 2007). Second, high *turnover* of an equity portfolio leads to high trading costs (Odean 1999, Barber and Odean 2000). Third, selling stocks that have appreciated while holding those that have depreciated—a tendency known as the *disposition effect*—increases the present value of tax obligations by accelerating the realization of capital gains and deferring the realization of offsetting losses (Shefrin and Statman 1985, Odean 1998).

In a market where expected returns differ across stocks, it is also possible for households to lose by picking *underperforming stocks*. They may do this by taking risk exposures that are negatively compensated, for example by holding growth stocks in a market with a value premium, or by adopting a short-term contrarian investment strategy (perhaps driven by the disposition effect) in a market with momentum where outperforming stocks continue to outperform for a period of time. If these style tilts do not offset other risks of the household, they are welfare reducing.[1] Alternatively, households may lose by trading with informed

[1]This is true whether risk prices are driven by fundamentals or by investor sentiment (the preferences of unsophisticated investors for certain types of stocks). In a model with fundamental risks it may be more likely that households' non-equity risk exposures justify equity positions with low expected returns, but if

counterparties in a market that is not strong-form efficient and thus rewards investors with private information (Grossman and Stiglitz 1980, O'Hara 2003).

Households can limit such investment mistakes in several ways. They may hold mutual funds as a way to gain equity exposure without trading stocks directly. This, however, may result in trade-offs between reductions in household investment mistakes, the level of fees charged by intermediaries, and potential investment mistakes made by mutual fund managers. Households may also learn from observing overall patterns in the market, or from their own investment experience (Kaustia and Knüpfer 2008, Chiang et al 2011, Malmendier and Nagel 2012).[2] In this paper we report evidence that learning from experience is important. Importantly, however, we do not claim that such learning is rational; instead, it may reflect reinforcement learning, in which personal experiences are overweighted relative to broader patterns of evidence in historical data.

Our study uses data from the Indian equity market. For several reasons this is an ideal laboratory for studying learning among equity investors. First, India is an emerging market whose capitalization and investor base have been growing rapidly. In such a population of relatively young investors, rapid learning may be easier to detect than in larger and more well-established equity markets. Second, mutual funds account for a relatively small value share of Indian individuals' equity exposure, so it is meaningful to measure the diversification of directly held stock portfolios.[3] The prevalence of direct equity ownership also implies that it is more important for Indian investors to develop the skills necessary to own stocks directly than it is in a mature market with a large mutual fund share. Third, India has electronic registration of equity ownership, allowing us to track the complete ownership

this is not the case such positions still reduce household welfare just as they would in a sentiment-driven model.

[2]In related work using data on professional investors Greenwood and Nagel (2009) find that less experienced mutual fund managers act as trend-chasers during the technology bubble.

[3]In March 2010, the Security and Exchange Board of India (SEBI) reports that there are 46 million individual mutual fund accounts in India collectively worth about \$50 billion. At that time, we estimate based on our data and NSDL's market share that there were approximately 10 million individual accounts directly invested in equities collectively worth about \$150 billion. Thus, while mutual fund ownership is common, the value of mutual funds is too small for Indian equityholders to rely entirely on them for diversification.

history of listed Indian stocks over a substantial period of time. The long time dimension of our panel allows us to measure investors' performance using their realized returns, a method that is vulnerable to common shocks when applied to a short panel. Moreover, our data is monthly, and this relatively high frequency allows us to more accurately measure important determinants of performance such as momentum investing and turnover.

A limitation of our Indian data is that we have almost no information about the demographic characteristics of investors. Thus we cannot follow the strategies, common in household finance, of proxying financial sophistication using information about investors' age, education, or occupation (Calvet et al. 2007, 2009a) or survey evidence about their financial literacy (Lusardi and Mitchell 2007). Instead, we study learning by relating account age (the length of time since an account was opened) and summary statistics about past portfolio characteristics and investment performance to the future characteristics and performance of each account.

We have three main results. First, account performance improves with account age, and stocks whose individual investors have older accounts tend to outperform the value-weighted Indian stock market. Second, several empirical proxies for investment mistakes are less prevalent among older accounts. Third, accounts that experience unusual underperformance associated with investment mistakes appear to respond by reducing such behavior in the future. The first two results suggest that investors learn from stock market participation, while the third suggests that investment experiences influence the rate of learning.

The organization of the paper is as follows. Section 2 describes our data, defines the empirical proxies we use for investment mistakes and style tilts, and presents some summary statistics. Section 3 relates account age to characteristics and investment performance. Section 4 shows that information about the investor base of each Indian stock can be used to predict the returns of that stock. Section 5 concludes.

2 Data and Summary Statistics

2.1 Electronic stock ownership records

Our data come from India's National Securities Depository Limited (NSDL), with the approval of the Securities and Exchange Board of India (SEBI), the apex capital markets regulator in India. NSDL was established in 1996 to promote dematerialization, that is, the transition of equity ownership from physical stock certificates to electronic records of ownership. It is the older of the two depositories in India, and has a significantly larger market share (in terms of total assets tracked, roughly 80%, and in terms of the number of accounts, roughly 60%) than the other depository, namely, Central Depository Services Limited (CDSL).

While securities in India can be held in both dematerialized and physical form, settlement of all market trades in listed securities in dematerialized form is compulsory, and statistics from the Bombay Stock Exchange (BSE) and the National Stock Exchange (NSE) highlight that virtually all stock transactions take place in dematerialized form. To facilitate the transition from the physical holding of securities, the stock exchanges do provide an additional trading window, which gives a one time facility for small investors to sell up to 500 physical shares, however the buyer of these shares has to dematerialize such shares before selling them again, thus ensuring their eventual dematerialization.

The sensitive nature of these data mean that there are certain limitations on the demographic information provided to us. While we are able to identify monthly stock holdings and transactions records at the account level in all equity securities on the Indian markets, we have sparse demographic information on the account holders. The information we do have includes the state in which the investor is located, whether the investor is located in an urban, rural, or semi-urban part of the state, and the type of investor. We use investor type to classify accounts as beneficial owners, domestic financial institutions, domestic non-financial institutions, foreign institutions, foreign nationals, government, and individual accounts.[4]

[4]We classify any account which holds greater than 5% of an stock with market capitalization above 500 million Rs (approximately $10 million) as a beneficial owner account if that account is a trust or "body

This paper studies only the last category of individual accounts.

A single investor can hold multiple accounts on NSDL; however, a requirement for account opening is that the investor provides a Permanent Account Number (PAN) with each account. The PAN is a unique identifier issued to all taxpayers by the Income Tax Department of India. NSDL provided us with a mapping from PANs to accounts, so in our empirical work, we aggregate all individual accounts associated with a single PAN. PAN aggregation reduces the total number of individual accounts in our database from about 13.7 million to 11.6 million. It is worth noting here that PAN aggregation may not always correspond to household aggregation if a household has several PAN numbers, for example, if children or spouses have separate PANs.

Table 1 summarizes the coverage of the NSDL dataset, beginning in 2002 when we have reliable data on the characteristics of newly issued accounts. The first two columns report the total number of securities (unique International Securities Identification Numbers or ISIN) and the total number of Indian equities reported in each year. Securities coverage grows considerably over time from just over 8,000 in 2002 to almost 23,000 in 2012, but the number of Indian equities covered is much more stable. Starting at 2,136 in 2002, the number of equities reaches a peak of 3,500 in 2010 before declining back to 2,464 in 2012.

The third column shows the market capitalization of the BSE at the end of each year. The dramatic variation in the series reflects both an Indian boom in the mid-2000s, and the impact of the global financial crisis in 2008.

The fourth column of Table 1 shows the fraction of Indian equity market capitalization that is held in NSDL accounts. The NSDL share grows from about 45% at the beginning of our sample period to about 70% at the end. The fifth column reports the fraction of NSDL market capitalization that is held in individual accounts. The individual share starts at about 20% in 2002, but declines to just below 10% in 2012, likely reflecting a secular increase in intermediated investment over our sample period.

corporate" account, or would otherwise be classified as an individual account. This separates accounts with significant control rights from standard investment accounts. Otherwise our account classifications are many-to-one mappings based on the given investor types.

Figure 1 illustrates the expansion of equity ownership in India by plotting the number of individual accounts active at each point in time. From the beginning to the end of our sample period, this number grew from 2.6 million to 6.6 million, that is, by 156%. Equity ownership expanded throughout the decade, but the rate of growth is correlated with the return on the aggregate Indian market (illustrated by the dashed line in the figure). Growth was particularly rapid in 2003-04 and 2007, and much slower in the period since the onset of the global financial crisis.

2.2 Characteristics of individual accounts

Table 2 describes some basic characteristics of the individual accounts in our dataset. Because this dataset is an unbalanced panel, with accounts entering and exiting over time, we summarize it in two ways. The first set of three columns reports time-series moments of cross-sectional means. The first column is the time-series mean of the cross-sectional means, which gives equal weight to each month regardless of the number of accounts active in that month. The second and third columns are the time-series maximum and minimum of the cross-sectional mean, showing the extreme extent of time-variation in cross-sectional average account behavior. The second set of three columns reports cross-sectional moments of time-series means calculated for each account over its active life, giving equal weight to each account regardless of the number of months in which it is active. Since the cross-sectional dimension of the dataset is much larger than the time-series dimension, we report the 10th percentile, median, and 90th percentile of the cross-sectional distribution.

For this table and all subsequent analysis, the data used represents a stratified random sample of our full dataset, an approach we also use (and describe more fully) in the regression analysis of the next section.

Account size and location

We begin by reporting account sizes both in rupees (using Indian conventions for comma placement), and in US dollars, both corrected for inflation to a January 2012 basis. The cross-sectional average account size varies across months from under $10,000 in 2004 to about \$70,000 in June 2008, with a time-series of mean of \$17,670. The median account size is however much smaller at \$1,674, and even the 90th percentile account size is only \$13,842, reflecting positive skewness in the distribution of account sizes. These differences imply that the weighting scheme used in summary statistics and regressions will have an important influence on the results. Given our focus on household finance questions, as opposed to the determination of Indian asset prices, we equally weight accounts in most of our empirical analysis as advocated by Campbell (2006).

Next we document the fraction of our accounts that are urban, semi-urban, or rural. About 55% of individual accounts are associated with urban account addresses, 31% with rural addresses, and 13% with semi-urban addresses. The relative weighting of these account types, particularly the urban and rural shares, does change somewhat over time.

Account performance

The fourth row of Table 2 looks at monthly account returns, calculated from beginning-of-month stock positions and monthly returns on Indian stocks.[5] These returns are those that an account will experience if it does not trade during a given month; in the language of Calvet et al. (2009a), it is a "passive return". It captures the properties of stocks held, but will not be a perfectly accurate measure of return for an account that trades within a month.

The table shows that on average, individual accounts have slightly underperformed the Indian market (proxied by a value-weighted index that we have calculated ourselves). The time-series mean of cross-sectional average underperformance is 5 basis points per month, and the cross-sectional median of time-series average underperformance is 26 basis points per month. There is considerable variation over time in the cross-sectional average, with individual accounts underperforming in their worst months by as much as 6.1% or overperforming in their best months by as much as 7.6%. This variation is consistent with the literature on institutional and individual performance in US data, and can largely be explained by style preferences of individual investors. There is also dramatic variation across investors

[5] The Data Appendix provides details on our procedures for calculating Indian stock returns.

in their time-series average performance, with the 10th percentile account underperforming by 2.64% per month and the 90th percentile account overperforming by 1.23% per month.

Underdiversification

The next set of three rows examines account-level statistics that proxy for the investment mistakes described in the introduction. The idiosyncratic share of portfolio variance is calculated from estimates of each stock's beta and idiosyncratic risk, using a market model with the value-weighted universe of Indian stocks as the market portfolio. This procedure closely follows Calvet et al. (2007), except that we use a local market index rather than a global index. In order to reduce noise in estimated stock-level betas, however, we do not use past stock-level betas but instead use fitted values from a panel regression whose explanatory variables include stock-level realized betas (in monthly data over the past two years), the realized betas of stocks in the same size, value, and momentum quintiles, industry dummies, and a dummy for stocks that are less than two years from their initial listing. To reduce noise in estimated idiosyncratic risk, we estimate idiosyncratic variance from a GARCH(1,1) model.[6]

The average idiosyncratic share is just over 40% in both the time-series and cross-sectional moments. This number is somewhat lower than the median idiosyncratic share of 55% reported by Calvet et al. (2007), the difference probably resulting from our use of an Indian rather than a global market index. Once again there is considerable variation over time (from 21% to 51%) and across accounts (from 19% at the 10th percentile to 63% at the 90th percentile).

Turnover

Turnover is estimated by averaging sales turnover (the fraction of the value of last month's holdings, at last month's prices, that was sold in the current month) and purchase turnover (the fraction of the value of this month's holdings, using this month's prices, that was purchased in the current month). This measure of turnover is not particularly high on

[6]The GARCH model is first estimated for each stock, then is re-estimated with the GARCH coefficients constrained to equal the median such coefficient estimated across stocks. This approach deals with stocks for which the GARCH model does not converge or yields unstable out of sample estimates.

average for Indian individual accounts. The time-series mean of the cross-sectional mean is 3.7% per month (or about 45% per year), and the cross-sectional median turnover is only 1.2% (or 14% per year). Turnover this low should not create large differences between the passive return we calculate for accounts and the true return that takes account of intra-month trading.

Once again, however, there is important variation over time and particularly across accounts. The 10th percentile account has no turnover at all (holding the same stocks throughout its active life), while the 90th percentile account has a turnover of over 10% per month (120% per year).

Following Odean (1999), we have compared the returns on stocks sold by individual Indian investors to the returns on stocks bought by the same group of investors over the four months following the purchase or sale. In India, the former exceeds the latter by 3.23%, which makes it more difficult to argue that trading by individuals is not economically harmful. By comparison, the difference Odean finds in US discount brokerage data is a much smaller 1.36%. At a one year horizon following the purchase or sale, we find that stocks sold outperform stocks bought by 5.57% compared to 3.31% in Odean's data.

The disposition effect

We calculate the disposition effect using the log ratio of the proportion of gains realized (PGR) to the proportion of losses realized (PLR). This is a modification of the previous literature which often looks at the simple difference between PGR and PLR. PGR and PLR are measured within each month as follows. Gains and losses on each stock are determined relative to the cost basis of the position if the position was established after account registry with NSDL (i.e. if the cost basis is known). Otherwise, we use the median month-end price over the 12 months prior to NSDL registry as the reference point for determining gains and losses (we do this in roughly 30% of cases). Sales are counted only if a position is fully sold, although this convention makes little difference to the properties of the measure. When computing the measure, we winsorize PGR and PLR below at 0.01, and if either PGR or PLR are missing, we substitute these missing values with 0.01.[7] This avoids dropping large

[7] In our regression specifications, when we include the lagged disposition effect as a control, we include

numbers of small accounts that may have only gains, or only losses, at a particular time.

The disposition effect is important on average for Indian individual accounts. On average across months, the cross-sectional mean proportion of gains realized is 0.43 log points or 54% larger than the proportion of losses realized, while the median account has a PGR that is 0.28 log points or 32% larger than its PLR. While both time-series and cross-sectional variation in the disposition effect are substantial, it is worth noting that nearly 90% of accounts in the sample exhibit this effect.

Figure 2 compares the disposition effect in our Indian data with US results reported by Odean (1998). The figure plots the log mean ratio of PGR to PLR by calendar month, a series that can be compared with Odean's numbers. The Indian disposition effect is considerably stronger on average than the US effect. In both India and the US, the disposition effect is weaker towards the end of the tax year (calendar Q4 in the US, and calendar Q1 in India).

Style tilts

Table 2 also reports several measures of individual accounts' style tilts. We construct account-level betas with the Indian market by estimating stock-level betas as described earlier, and then value-weighting them within each account. The average beta is almost exactly one both in the time-series and cross-sectional moments. The cross-sectional mean betas vary over time by about 0.05 in each direction, and the cross-sectional spread from the 10th to the 90th percentile runs from 0.94 to 1.10.

In US data, individual investors overweight small stocks, which of course implies that institutional investors overweight large stocks (Falkenstein 1996, Gompers and Metrick 2001, Kovtunenko and Sosner 2004). We measure this tendency in our Indian dataset by calculating the value-weighted average market-capitalization percentile of stocks held in individual accounts, relative to the value-weighted average market-capitalization percentile of stocks

---

a dummy for the inclusion of such observations. This is similar to the procedure employed in Calvet et al. (2009b), who set PGR (PLR) equal to the cross-sectional mean for households that do not have gains (losses) in a given month. It is also worth noting that this substitution is never an issue when the disposition effect is on the left-hand side of regressions as we always employ account-months in which losses and gains are both present.

in the market index. We find only a modest individual-investor tilt towards small stocks: the time-series mean percentile of market cap held by individual investors is only 2% lower than the market index. It varies from about 1% to about 3% over time, but never switches sign. Across accounts, the 10th percentile account has a 10% small-cap tilt while the 90th percentile account has a 3% large-cap tilt.

Individual Indian investors have a very small tilt towards value stocks. Ranking stocks by their book-market ratio and calculating percentiles in the same manner that we did for market capitalization, we find that the time-series mean percentile of value held by individual investors is only 1.6% greater than the market index. This value tilt varies over time and does switch sign, reaching -7% in the month that is most tilted towards growth. There are also very large differences across accounts in their orientation towards growth or value, with a spread of almost 30% between the 10th and 90th percentiles of accounts.

Finally, individual investors have a strong contrarian, or anti-momentum tilt. Ranking stocks by momentum and calculating the momentum tilt using our standard methodology, we find that both the time-series mean and cross-sectional median momentum tilts are about -5%. This pattern is consistent with results reported for US data by Cohen, Gompers, and Vuolteenaho (2002), and with short-term effects (but not longer-term effects) of past returns on institutional equity purchases estimated by Campbell, Ramadorai, and Schwartz (2009).

Cross-sectional correlations of characteristics

Table 3 asks how the account characteristics described in Table 2 are correlated across accounts. We calculate cross-sectional correlations of account characteristics for each month, and then report the time-series mean of these correlations. To limit the influence of outliers, we winsorize account-level stock returns at the 1st and 99th percentiles, and winsorize account value below at 10,000 rupees (approximately $200).

There are a number of intriguing patterns in Table 3. Older accounts tend to be larger, and account age is negatively correlated with all three of our proxies for investment mistakes—an effect we explore in detail in the next section. Among the mistake proxies, there is a 0.38 correlation between turnover and the disposition effect; this is partly a mechanical effect since accounts that do not trade cannot exhibit a disposition effect. Turnover also has a weak 0.15 correlation with the idiosyncratic share of variance, implying that underdiversified accounts tend to trade more. All the mistake proxies are positively correlated with accounts' market betas and negatively correlated with their size tilts, implying that accounts holding high-beta and small-cap stocks tend to be less diversified, trade more, and have a stronger disposition effect. The log of account value correlates negatively with beta and value, and positively with size and momentum tilts. This implies that larger individual accounts look more like institutional accounts in that they prefer lower-beta stocks, growth stocks, large stocks, and recent strong performers. Finally, there is a strong negative correlation of -0.42 between the size tilt and the value tilt, implying that individuals who hold value stocks also tend to hold small stocks. This effect too is somewhat mechanical given the correlation of these characteristics in the Indian universe.

3 Account Performance, Experience, and Behavioral Biases

In this section we explore the dynamic relation between account characteristics and performance, that is, the monthly returns of the portfolio of stocks held at the beginning of each month. We calculate returns without subtracting transactions costs (which we do not measure).

3.1 Predicting account returns

Table 4 reports regression results for five different panel regressions predicting account returns. The regression in column 1 includes only dummy variables for quintiles of account age (the length of time since each account was opened). These dummies capture the effect of investment experience in Indian equities on account performance. New accounts are of course in the first quintile, and it takes about one, two, three, and five years, respectively, for accounts to reach the older quintiles.

Since older accounts were disproportionately opened earlier in our sample period, column

2 adds control variables for cohort-level characteristics, specifically, the initial cohort means of log account size, log number of equity positions, log state income (averaged over 2002-2011) for the states where accounts are held, state literacy rate (also averaged over 2002-2011), percentage of rural accounts, and percentage of urban accounts. Column 3 adds our three proxies for investment mistakes, smoothing the last two over the past year to reduce noisy monthly variation: the idiosyncratic variance share, average monthly turnover over the past year, and the average log ratio of PGR to PLR over the past year. Column 4 adds account size, market beta, style tilts, and current account location, and finally column 5 adds the average monthly outperformance over the past year to capture momentum effects and persistent investment skill.

These regressions are estimated on a stratified random sample, drawing 5,000 individual accounts from each Indian state with more than 5,000 accounts, and all accounts from states with fewer than 5,000 accounts. Regression weights account for this sampling strategy, and in most of our specifications, we include state fixed effects as well as dummies for accounts located in rural areas. About 2.9 million account months spanning January 2004 through September 2011 are used in each regression once account-months with all computable characteristics and controls are included. This constitutes a cross-sectional average of about 35% of all accounts in our sample (44% of accounts by value), or about 62% of all accounts in our sample opened after January 2002 (74% by value). Pre-2002 cohorts are excluded from these regressions because initial cohort-level characteristics are unavailable for cohorts which opened prior to the first month in our database (February 2002). The remaining account months are excluded as a result of undefined right hand side variables (typically affecting very recently opened accounts).

All regressions are estimated using a Fama-MacBeth methodology. A sequence of monthly cross-sectional regressions is estimated, and the table reports time-series average coefficients and uses the time-series variation of the monthly coefficients to calculate standard errors. These standard errors are adjusted upwards as in Fama and French (2002) where the coefficients are positively serially correlated. The table reports panel R^2, which increases from 0.4% where only account age dummies are used, to 1.3% with all variables included.[8]

Figure 3 illustrates our first main result, that performance increases with age. The increase is monotonic in age across all specifications. The difference in performance between the oldest and youngest accounts is 35 to 40 basis points per month in the first three specifications, and about 20 basis points per month in the last two specifications that include style tilts. This suggests that about half the raw age effect is attributable to the relative style tilts of older accounts, and that the remainder is unexplained by any other account characteristics that we measure.

The account characteristics that we control for predict account performance in Table 4 in an economically meaningful fashion. The two strongest effects are that accounts with value tilts overperformed in this sample period, and that accounts with high turnover underperformed. The latter result is particularly striking because we are not subtracting transactions costs from measured returns; the effect we are picking up is simply that high-turnover accounts picked worse performing stocks. A somewhat weaker effect is that accounts with a high idiosyncratic variance share overperformed, consistent with the idea that skilled investors hold concentrated portfolios (Cremers and Petajisto 2009). Our third measure of investment mistakes, the disposition effect, has a weak negative effect on performance, but this effect is not statistically significant.

Table 5 concentrates on the return difference between the portfolios held by the oldest and youngest accounts, and asks to what extent this difference is explained by compensated risks. In other words, it calculates raw excess returns and multi-factor alphas for a long-short portfolio that value-weights within each account and equal-weights across accounts, going long for the oldest accounts and short for the youngest ones. The first column of the table reports a raw excess return of 34 basis points per month, which is not statistically significant because of noise created by market movements. The second column shows that this corresponds to a statistically significant CAPM alpha of 50 basis points per month, and a negative market beta of -0.14, reflecting the fact that older accounts tend to hold somewhat

[8] Panel R-squared is equal to the average cross-sectional variance of fitted values (produced with full sample coefficients) scaled by the variance of the dependent variable.

lower-beta stocks even while delivering a higher return. The third and fourth columns show that the alpha increases to 54 basis points per month in a Fama-French-Carhart four-factor model including momentum, and 56 basis points per month in a six-factor model that includes factors for short-term reversals and illiquidity (proxied by a long-short portfolio constructed by sorting the universe of stocks on turnover).

3.2 Predicting account behavior

We now ask whether our three proxies for investment mistakes are predictable using information about accounts' past characteristics and performance. In Table 6 we regress the idiosyncratic variance share (columns 1A-C), turnover (columns 2A-C), and the log ratio of PGR to PLR (columns 3A-C) on these characteristics. The A columns include account age quintile dummies and cohort characteristics on the right-hand side, the B columns combine these with measures that capture the past returns associated with investment mistakes, and the C columns further combine these controls with lagged account characteristics.

Figure 4 plots the account age dummy coefficients estimated in the A columns of Table 6 for the idiosyncratic variance share, turnover, and disposition effect. In every specification the estimated coefficients decline with age, showing that our three proxies for investment mistakes are less prevalent among older accounts.[9] These declines are not only statistically significant, but also economically large. The magnitude of the coefficient on the oldest quintile of accounts corresponds to roughly 10%, 50%, and 25% of the unconditional mean of the idiosyncratic share of portfolio variance, turnover, and disposition bias respectively.

We now turn to the B columns of Table 6, which add measures of past performance to the set of explanatory variables. Lagged account outperformance may encourage investors to assess their investing skills more optimistically, leading them to pick larger idiosyncratic bets. Past performance does predict the subsequent idiosyncratic share of portfolio variance. However, some caution in interpretation is urged as this relationship may be partly mechanical; very large returns in the absence of rebalancing may lead to high variance portfolios.

[9] We do not report age dummy coefficients from the B or C columns because the presence of lagged dependent variables in these columns makes the age dummies hard to interpret.

For turnover, we add the average monthly outperformance over the past year due to trading, measured as the difference between the monthly returns actually experienced and those that would have been experienced if the portfolio had maintained its initial holdings from a year ago. This variable strongly predicts turnover, implying that trading profits strengthen the tendency to trade stocks frequently, a result consistent with those of Linnainmaa (2011), who employs information on a set of high-frequency traders from Finland.

Finally, we measure the outperformance of stocks held at a gain minus the outperformance of stocks held at a loss over the past year. This is a measure of the penalty associated with the disposition effect. When this measure is high, the account's winners have tended to outperform the account's losers—in other words, the account holdings have displayed momentum—and reinforcement learning might then lead the account holder to avoid disproportionately selling winners. Consistent with our expectations, this measure enters the regression with a negative sign, suggesting that painful experiences with disposition-effect trading teach investors to avoid this behavior.

The coefficients on lagged account characteristics, reported in the C columns of Table 6, show some interesting patterns. Unsurprisingly, there is positive serial correlation in our proxies for investment mistakes, so the lagged dependent variables are highly statistically significant. The idiosyncratic variance share is the most persistent variable, and the log ratio of PGR to PLR is the least persistent because intermittent trading makes this series very noisy. We also see that lagged turnover predicts a high idiosyncratic variance share, since trading more often moves accounts away from index weights than towards them, and predicts the log ratio of PGR to PLR, because accounts with low turnover have little tendency to realize either gains or losses.

Underdiversification also appears to be associated with small accounts, ownership of high-beta and value stocks, and location in a semi-urban or rural area. Turnover, on the other hand, is typically greater in large accounts. Finally the disposition effect is predicted by ownership of small stocks, value stocks, and stocks with negative momentum. This is not surprising since stocks that have performed poorly are likely to appear even more attractive to investors that prefer stocks with low valuations, and the disposition effect is defined by the tendency to hold losers and sell winners, just as prescribed by a negative momentum strategy.

In conclusion, Table 6 provides suggestive evidence of reinforcement learning among Indian equity investors. Our interpretation might be challenged if there is reverse causality, for example if skilled traders generate trading profits and continue to trade frequently in the future, or if certain investors specialize in holding mean-reverting stocks for which realizing gains and holding losses is a systematically profitable strategy. However, reverse causality should imply that high-turnover and disposition-effect accounts outperform other accounts, and Table 4 showed that this is not the case. The only proxy for investment mistakes that is associated with outperformance is underdiversification, and we find only a weak effect of past performance on the future tendency to hold a concentrated portfolio.

4 Investor Base and Stock Performance

In this section we change our focus from the performance of individual accounts to the performance of the stocks they hold, as predicted by the investor base of those stocks. This is somewhat analogous to the recent literature on the performance of mutual funds' stock picks, as opposed to the overall performance of the funds themselves (Wermers 2000, Cohen, Polk, and Silli 2010).

We begin by sorting stocks into quintiles of market capitalization, based on the average age of the individual accounts that hold them. Within each quintile, we value-weight the stocks to create portfolios that can be held at reasonable transactions costs and whose properties are not overly influenced by extremely small stocks. Table 7 reports summary statistics. The top panel of the table shows median characteristics for stocks within each quintile. Stocks favored by young accounts tend to be stocks of young companies with a higher market capitalization.[10] There is only a weak relation between the average account

[10] As a robustness check, we deleted initial public offerings (stocks less than six months old) from our stock and account level analysis. Our results were barely affected by this change, which provides reassurance that the underperformance of young accounts is not driven by the well-known phenomenon of IPO long-run underperformance.

age of the investor base and a stock's market beta. Stocks with the youngest and oldest account holders have lower book-market ratios than other stocks. Stocks remain in experience sorted portfolios, 54% of stocks in the quintile with newest investor accounts remain in the same quintile one year later, while about 82% of stocks in the quintile with the oldest investor accounts remain in that quintile one year later.

The bottom panel shows the average portfolio weights for the five stock quintiles within the oldest and youngest quintile of accounts. For both types of accounts, the portfolio weights decline with the average account age of the investor base, implying that stocks that appeal to young accounts also appeal to all individual investors. However, the decline is much steeper for young accounts than for old accounts. That is, old accounts look more similar to institutional investors, and have relatively higher portfolio allocations to stocks with larger institutional and insider ownership.

This observation raises the possibility that the superior investment performance of older individual accounts is driven in part by similarities between these accounts and institutional investors. Since institutional investors have gained market share over our sample period, stocks favored by such investors may rise in price just because they control more capital over time (Gompers and Metrick 2001). If older individual accounts are more like institutions, and hold similar stocks, this transitional effect may benefit long-established individual investors as well as institutions. We therefore attempt to control for this possible explanation of our results towards the end of this section.

In Table 5 in the previous section, we formed a long-short portfolio using the holdings of the oldest and youngest accounts. The implied weights of this portfolio are also reported in the bottom panel of Table 7, along with a long-short portfolio that goes long the quintile of stocks with the highest average account age and short the quintile with the lowest account age. This approach increases the spread in average account age between the long side and short side of the portfolio.

Figure 5 shows the relation between the outperformance of the oldest accounts, relative to the youngest ones—the object of study in the previous section—and the returns to a zero-cost portfolio formed by going long the decile of stocks with the highest average account age and short the decile with the lowest average account age. The common variation of these two series is obvious. In addition, the figure illustrates the high volatility of the long-short portfolio based on account age of the ownership base. Even though the returns are averaged over six months, there are periods with returns greater than 6% per month in absolute value.

Figure 6 illustrates that stocks favored by young accounts tend to underperform other stocks. More generally, stock performance increases with the account age of the ownership base, but the relation is not always monotonic and the increase in performance occurs mostly between the youngest-investor-base quintile and the median-investor-base quintile.

Stocks can similarly be sorted on the behavioral biases of their investor base, for example the turnover or disposition effect of the investor base. Figure 7 shows a weak tendency for stock performance to improve with reduced disposition effect and turnover of the investor base. Again the effect is not always monotonic, and in the case of turnover, it appears to be primarily driven by outperformance of the quintile of stocks with the lowest-turnover (buy-and-hold). Figure 8 plots the cumulative returns and six-factor alphas for zero-cost portfolios formed on all three investor-base sorts, along with the cumulative Indian equity premium.

Table 8 reports the risk-adjusted performance of investor-base stock-picking strategies. In panel A, a strategy that goes long the quintile of stocks with the highest average account age and short the quintile with the lowest average account age has a positive monthly excess return of 109 basis points, significant at the 10% but not the 5% level. Because this strategy overweights low-beta stocks, it has a negative market beta of about -0.4, and its CAPM alpha is a 5%-significant 151 basis points. Further factor adjustments using four or six factors slightly reduce the alpha, because the strategy loads positively on small stocks and momentum stocks. However the alpha remains at least 120 basis points and significant at the 5% level.

Panels B and C report results for long-short portfolios that are long (short) stocks whose investors have a particularly small (large) disposition effect (panel B) or particularly low (high) turnover (panel C). In all cases excess returns and alphas are positive, but they are not always statistically significant. A striking feature of these portfolios is that they have very large positive loadings on UMD, implying that low-disposition-effect and low-turnover individual investors earn high returns in part by avoiding contrarian investing practiced by their high-disposition-effect and high-turnover peers.

Finally, it might be the case that our results arise from noisily estimated factor loadings, which may be better approximated by stock characteristics. To account for this possibility we independently double sort stocks on investor base and the characteristics listed in column headers [5] to [9], and average the returns of long-short portfolios along the investor-base dimension within each characteristic-sorted portfolio. We then report the six-factor alphas of these "characteristic-neutralized" portfolio returns in these columns. One of these characteristics is the contemporaneous level of institutional ownership of the stock, in an attempt to control for the Gompers-Metrick effect mentioned earlier.[11]

These columns of Table 8 show that despite the attenuation in statistical significance arising from the smaller (and hence noisier) double-sorted portfolios used to construct returns, the economic magnitude of the outperformance of the investor-base strategy is barely diminished, and in some cases greater. Of course, at this stage we cannot rule out that our results may be partially a result of a more subtle version of the Gompers-Metrick effect in which the stocks favored by more experienced, low turnover, or low disposition-effect investors may rise in price just because they control more capital over time.

5 Conclusion

In this paper we have studied the investment strategies and performance of indiovidual investors in Indian equities over the period from 2002 to 2012. We find strong effects of account age, the number of years since a particular account begins holding Indian stocks and appears in our dataset. Older accounts outperform younger ones, in part by tilting profitably towards small stocks and stocks with positive momentum, but also over and above controls for these style tilts. We find similar patterns in individual stock returns when we sort stocks

[11] We also used the level of institutional ownership and the lagged change in institutional ownership as alternative characteristics, with very similar results. We do not present them here in the interests of brevity.

by the average account age of their investor base.

Our evidence also suggests that learning is important among Indian individual investors. Older accounts have a smaller tendency to underdiversify, lower turnover, and a smaller disposition effect. Moreover, accounts that have experienced low returns from their trading during the past year tend to reduce their turnover in the future, while poor returns associated with the disposition effect lower this effect in the future. These results suggest that Indian individual investors learn, not only from the experience of stock market participation itself, but also from the returns they experience from popular investment behaviors.

Data Appendix

We collect stock-level data on monthly total returns, market capitalization, and book value from three sources: Compustat Global, Datastream, and Prowess. Prowess further reports data sourced from both of India's major stock exchanges, the BSE and NSE. In addition, price returns can be inferred from the month-end holding values and quantities in the NSDL database. We link the datasets by ISIN.[12]

To verify reliability of total returns, we compare total returns from the (up to three) data sources, computing the absolute differences in returns series across sources. For each stock-month, we use returns from one of the datasets for which the absolute difference in returns with another dataset is smallest, where the exact source is selected in the following order of priority: Compustat Global, Prowess NSE, then Prowess BSE. If returns are available from only one source, or the difference(s) between the multiple sources all exceed 5% then we compare price returns from each source with price returns from NSDL, We then use total returns from the source for which price returns most closely match NSDL price returns, provided the discrepancy is less than 5%.

After selecting total returns, we drop extended zero-return periods which appear for non-traded securities. We also drop first (partial) month returns on IPOs and re-listings, which are in many cases very extreme and reported inconsistently. For the 25 highest and lowest remaining total monthly returns, we use internet sources such as Moneycontrol and Economic Times to confirm that the returns are indeed valid. The resulting data coverage is spotty for the very smallest equity issues, which could lead to survivorship issues. Therefore, we drop returns for all stock-months where the aggregate holdings of that stock across all account types in NSDL is less than 500 million Rs (approximately $10 million) at the end of the prior month.

We follow a similar verification routine for market capitalization and book value, confirming that the values used are within 5% of that reported by another source. Where market

[12] Around dematerialisation, securities' ISINs change, with some data linked to pre-dematerialisation ISINs and other data linked to post-dematerialisation ISINs. We use a matching routine and manual inspection to match multiple ISINs for the same security.

capitalization cannot be determined for a given month, we extrapolate it from the previous month using price returns. Where book value is unknown, we extrapolate it forward using the most recent observation over the past year.

References

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

Blume, Marshall, and Irwin Friend, 1975, The asset structure of individual portfolios and some implications for utility functions, *Journal of Finance* 30, 585–603.

Calvet, Laurent, John Y. Campbell, and Paolo Sodini, 2007, Down or out: Assessing the welfare costs of household investment mistakes, *Journal of Political Economy* 115, 707–747.

Calvet, Laurent, John Y. Campbell, and Paolo Sodini, 2009a, Fight or flight? Portfolio rebalancing by individual investors, *Quarterly Journal of Economics* 124, 301–348.

Calvet, Laurent, John Y. Campbell, and Paolo Sodini, 2009b, Measuring the financial sophistication of households, *American Economic Review Papers and Proceedings* 99, 393–398.

Campbell, John Y., 2006, Household finance, *Journal of Finance* 61, 1553–1604.

Campbell, John Y., Tarun Ramadorai, and Allie Schwartz, 2009, Caught on tape: Institutional trading, stock returns, and earnings announcements, *Journal of Financial Economics* 92, 66–91.

Campbell, John Y. and Luis M. Viceira, 2002, *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*, Oxford University Press.

Chiang, Yao-Min, David Hirshleifer, Yiming Qian, and Ann E. Sherman, 2011, Do investors learn from experience? Evidence from frequent IPO investors, *Review of Financial Studies* 24, 1560–1589.

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, 409–62.

Cohen, Randolph B., Christopher Polk, and Bernhard Silli, 2010, Best ideas, unpublished paper, Harvard Business School, London School of Economics, and Goldman Sachs Group, Inc.

Cremers, Martijn and Antti Petajisto, 2009, How active is your fund manager? A new measure that predicts performance, *Review of Financial Studies* 22, 3329–3365.

Falkenstein, Eric G., 1996, Preferences for stock characteristics as revealed by mutual fund portfolio holdings, *Journal of Finance* 51, 111–135.

Fama, Eugene F. and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fama, Eugene F. and Kenneth R. French, 2002, Testing tradeoff and pecking order predictions about dividends and debt, *Review of Financial Studies* 15, 1–33.

Gompers, Paul A. and Andrew Metrick, 2001, Institutional ownership and equity prices, *Quarterly Journal of Economics* 116, 229–260.

Greenwood, Robin, and Stefan Nagel, 2009, Inexperienced investors and bubbles. *Journal of Financial Economics* 93, 239–258.

Grossman, Sanford J. and Joseph E. Stiglitz, 1980, On the impossibility of informationally efficient markets, *American Economic Review* 70, 393–408.

Kaustia, Markku and Samuli Knüpfer, 2008, Do investors overweight personal experience? Evidence from IPO subscriptions, *Journal of Finance* 63, 2679–2702.

Kelly, Morgan, 1995, All their eggs in one basket: Portfolio diversification of U.S. households, *Journal of Economic Behavior and Organization* 27, 87–96.

Kovtunenko, Boris and Nathan Sosner, 2004, Avoidance of small stocks and institutional performance, unpublished paper, Blue Fir Capital and AQR Capital Management, available online at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1132770.

Linnainmaa, Juhani, 2011, Why do (some) households trade so much? *Review of Financial Studies* 24, 1630-1666.

Lusardi, Annamaria and Olivia Mitchell, 2007, Baby boomer retirement security: The roles of planning, financial literacy, and household wealth, *Journal of Monetary Economics* 54, 205–224.

Malmendier, Ulrike and Stefan Nagel, 2012, Learning from inflation experiences, unpublished paper, UC Berkeley and Stanford University.

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.

O'Hara, Maureen, 2003, Presidential address: Liquidity and price discovery, *Journal of Finance* 58, 1335–1354.

Shefrin, Hersh and Meir Statman, 1985, The disposition to sell winners too early and ride losers too long: Theory and evidence, *Journal of Finance* 40, 777–790.

Siegel, Jeremy J., 2007, *Stocks for the Long Run*, 4th ed., McGraw-Hill.

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: NSDL Database Summary Statistics

The percentages below are computed for each monthly cross-section, and the average of these monthly percentages within each year appear in the table. The number of unique securities and equities are determined by the average number of unique ISIN appearing in the NSDL database in each month in the given year. Individual accounts exclude beneficial owners. BSE market capitalization (from the World Federation of Exchanges), is from the end of each year (except 2012, where data is from October), and represents the market capitalization of all equities listed on the BSE, representing the vast majority of Indian equities.

	Number of Unique Securities	Number of Unique (Indian) Equities	Market Capitalization of BSE (Billions of US$)	% of Indian Equity Market Capitalization in NSDL Accounts	% of NSDL Equity Value in Individual Accounts
2002	8,245	2,136	$130.4	44.38%	19.60%
2003	11,054	2,316	$278.7	45.32%	19.36%
2004	12,264	2,511	$386.3	51.24%	17.59%
2005	13,487	2,785	$553.1	58.04%	15.86%
2006	15,279	2,980	$818.9	63.74%	15.04%
2007	17,091	3,153	$1,819.1	66.72%	12.87%
2008	17,511	3,324	$647.2	65.26%	11.94%
2009	17,458	3,413	$1,306.5	64.73%	11.29%
2010	19,458	3,500	$1,631.8	67.69%	10.84%
2011	22,663	3,140	$1,007.2	69.67%	10.16%
2012	22,696	2,464	$1,202.9	70.22%	9.90%

Table 2: Summary Statistics for Individuals' NSDL Accounts

Statistics are computed on the basis of all individuals' account months used in the regression models. Sampling weights are used to reflect the stratified manner in which the random sample was drawn. Time-series averages of the variables are computed only for accounts which appear in the data for at least 12 months.

	Time Variation in Cross-Sectional Means			Cross-Sectional Variation in Time-Series Means		
	Mean	**Min**	**Max**	**10th**	**50th**	**90th**
Account Value, Jan 2012 Rs	Rs 9,01,170	Rs 4,40,708	Rs 36,11,019	Rs 5,046	Rs 85,383	Rs 7,05,923
Account Value, Jan 2012 US$	$17,670	$8,641	$70,804	$99	$1,674	$13,842
Number of Equity Positions	8.09	6.08	10.44	1	4	18
Urban Accounts	55.53%	54.17%	58.95%	0	1	1
Semi-Urban Accounts	13.41%	12.40%	17.73%	0	0	1
Rural Accounts	31.06%	26.44%	32.41%	0	0	1
Monthly Account Stock Return Minus Market	-0.05%	-6.07%	7.65%	-2.64%	-0.26%	1.23%
Idiosyncratic Share of Portfolio Variance	0.41	0.21	0.51	0.19	0.40	0.63
Turnover	3.73%	1.59%	7.79%	0.00%	1.18%	10.53%
Disposition Effect - ln(PGR/PLR)	0.43	-0.13	1.45	-0.06	0.28	1.09
Stock Portfolio Beta	1.01	0.95	1.05	0.94	1.00	1.10
Size Percentile of Stocks Held Relative to Market Portfolio	-1.89	-2.98	-0.84	-10.37	-0.19	3.21
Book-Market Percentile of Stocks Held Relative to Market Portfolio	1.62	-7.11	14.05	-9.06	0.64	20.22
Momentum Percentile of Stocks Held Relative to Market Portfolio	-4.96	-12.40	4.02	-20.13	-6.36	3.58

Table 3: Cross-Sectional Correlations of Account Level Variables

Statistics are computed on the basis of all account months used in the regression models. Sampling weights are used to reflect the stratified manner in which the random sample was drawn. Cross-sectional correlations are computed for each month, and the average cross-sectional correlation is reported below. Account stock returns are winsorized at the 1st and 99th percentiles, and log account value is winsorized below at approximately 10,000 Rs (approximately $200).

		[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
Account Age	[1]	1.00									
Log Account Value	[2]	**0.24**	1.00								
Account Stock Return Over the Past Year	[3]	0.03	0.09	1.00							
Idiosyncratic Share of Portfolio Variance	[4]	-0.06	**-0.33**	**0.12**	1.00						
Turnover Over the Past Year	[5]	**-0.16**	**-0.13**	0.04	**0.15**	1.00					
Disposition Effect Over the Past Year	[6]	**-0.11**	-0.06	0.03	0.01	**0.38**	1.00				
Stock Portfolio Beta	[7]	-0.08	**-0.10**	0.04	0.10	**0.22**	**0.11**	1.00			
Size Percentile of Stocks Held	[8]	0.01	0.09	-0.03	**-0.28**	**-0.19**	**-0.14**	**-0.37**	1.00		
Book-Market Percentile of Stocks Held	[9]	0.00	-0.07	-0.07	**0.11**	0.08	0.08	**0.12**	**-0.42**	1.00	
Momentum Percentile of Stocks Held	[10]	0.07	**0.18**	**0.45**	-0.02	-0.01	**-0.10**	0.01	**0.10**	**-0.19**	1.00
Urban Account	[11]	0.02	0.06	0.00	-0.03	0.00	-0.01	0.00	0.02	-0.02	0.02
Semi-Urban Account	[12]	0.00	-0.01	0.00	0.00	0.00	0.01	0.02	-0.02	0.02	0.00
Rural Account	[13]	-0.02	-0.06	0.00	0.03	0.00	0.00	-0.02	-0.01	0.00	-0.02

Table 4: Account Equity Return Regressions

Results are constructed from a stratified random sample: 5,000 individual accounts are drawn at random from each Indian state, with all individual accounts drawn from states with less than 5,000 individual accounts. Regression weights account for this sampling strategy. About 2.9 million account months spanning January 2004 through September 2011 are used in each regression below. These observations account for a cross-sectional average of about 35% of all accounts (44% of accounts by value), or about 62% of all accounts opened after January 2002 (74% by value). Pre-2002 cohorts are excluded from these regressions as the cohort level characteristics are unavailable for cohorts which opened prior to the first month in our database (February 2002). The remaining account months are excluded as a result of undefined right hand side variables (typically affecting very recently opened accounts). Regressions are conducted by a Fama MacBeth procedure, with standard errors in ( ) adjusted upwards as in Fama and French (2002) where the coefficients are positively serially correlated. Coefficients that are significant at a five percent level are in bold type, and coefficients that are significant at a ten percent level are in italics. All coefficients (except dummies and lagged average monthly outperformance) are scaled by the average cross-sectional standard deviation of the corresponding independent variable, and all coefficients and standard errors are further multiplied by 100 for readability. Panel R-squared is equal to the average cross-sectional covariance of fitted values (using the full sample coefficients) scaled by the variance of the dependent variable. The small stock, value, and momentum tilts are defined as (standardized) percentiles of negative market capitalization, book-market, and momentum respectively. The cohort characteristic controls indicated below include mean log account values, log number of equity positions, log state income averaged over the period 2002-2011, state literacy rate, % rural, and % urban, where the mean is taken as of the account opening month for all accounts opened in the same month that are present in the given cross-section/current month. Specifications [3] through [5] include an ancillary dummy equal to one where the lagged disposition bias measure is unavailable and set equal to zero as a result.

Dependent Variable: Monthly Equity Portfolio Return (Unconditional Mean: 1.49%)

		[1]	[2]	[3]	[4]	[5]
Lagged Account Returns and Behavior	Average monthly outperformance of equity portfolio in past year					0.043
						(0.048)
	Idiosyncratic share of portfolio variance based on last month's holdings			0.249	*0.191*	*0.183*
				(0.154)	*(0.101)*	*(0.100)*
	Average monthly account turnover over the past year			-0.072	**-0.093**	**-0.094**
				(0.093)	*(0.046)*	*(0.045)*
	Average ln(PGR/PLR) over the past year			-0.035	-0.002	-0.008
				(0.036)	*(0.017)*	*(0.019)*
Account Size	Log(account value)				-0.012	-0.010
					(0.093)	*(0.090)*
Account Composition / Tilts	Portfolio stock market beta				-0.134	-0.149
					(0.199)	*(0.192)*
	Small stock tilt				-0.064	-0.059
					(0.174)	*(0.173)*
	Value tilt				**0.427**	**0.413**
					(0.127)	*(0.124)*
	Momentum tilt				0.295	0.239
					(0.200)	*(0.156)*
Account Location	Semi-urban accounts				0.026	0.019
					(0.033)	*(0.035)*
	Rural accounts				-0.002	0.000
					(0.042)	*(0.042)*
	State dummies				Y	Y
Account Age Quintile Dummies		Y	Y	Y	Y	Y
p-value: Oldest Age Quantiles Dummy is Zero		0.044	0.012	0.012	0.026	0.037
Cohort Characteristics			Y	Y	Y	Y
Panel R-Squared		0.004	0.003	0.005	0.013	0.013

Table 5: Performance Evaluation of the Difference in Returns on Old and Young Account Quintiles

For each month (in the period January 2004 through January 2012) and account age quintile, we compute the average portfolio weight across accounts for each stock with a market capitalization greater than 500 million Rs (approximately $10 million). Sampling weights are used. The difference in portfolio weights between the top and bottom age quintiles are used as weights for the zero-cost portfolio analyzed below. This portfoliois formed monthly. Portfolio returns are adjusted using unconditional CAPM, four, or six factor models, where the factor returns (except Illiq) are constructed in an analogous way to the factor returns from Ken French's website. The yield on three-month Indian Treasury bills is used as the risk free rate. The illiquidity factor is constructed from a independent double sort on size and turnover over the past 12 months, Illiq=0.5 x (Small, Low Turnover-Small, High Turnover)+0.5 x (Large, Low Turnover-Large, High Turnover). All standard errors are computed using a Newey West adjustment for serial correlation (with three lags).

		No Factors/ Raw Return	CAPM	Four Factor	Six Factor
		[1]	[2]	[3]	[4]
Monthly Alpha		0.34%	**0.50%**	**0.54%**	**0.56%**
		(0.21%)	*(0.20%)*	*(0.18%)*	*(0.18%)*
Factor Loadings					
	Market Beta		**-0.14**	**-0.08**	**-0.07**
			(0.03)	*(0.02)*	*(0.02)*
	SMB			**0.06**	**0.06**
				(0.02)	*(0.02)*
	HML			**-0.13**	**-0.16**
				(0.03)	*(0.03)*
	UMD			**0.13**	**0.14**
				(0.03)	*(0.03)*
	Short Term Reversals				-0.05
					(0.04)
	Illiq (Based on Turnover)				0.04
					(0.05)

Table 6: Account Behavior Regressions

All regressions below use an unbalanced panel spanning January 2004 through September 2011, covering accounts opened after January 2002. Accounts opened earlier are excluded from these regressions as the cohort-level characteristics are unavailable for accounts which opened prior to the first month in our database (February 2002). The data is from a stratified random sample of accounts (see notes to Table 4 for details). In specifications [1] and [2], an average of about 70% of accounts opened after January 2002 are included in the regressions, with the remainder excluded primarily due to one or more undefined right hand side variables. In specification [3] only about 35% of account-months by value are included as disposition bias is undefined for the many accounts which do not hold both positions at a gain and positions at a loss. Regressions are conducted by a Fama MacBeth procedure, with standard errors in ( ) adjusted upwards as in Fama and French (2002) where the coefficients are positively serially correlated. Coefficients that are significant at a five percent level are in bold type, and coefficients that are significant at a ten percent level are in italics. All coefficients (except dummies and the lagged dependent variable for each regression) are scaled by the average cross-sectional standard deviation of the corresponding independent variable, and all coefficients and standard errors are further multiplied by 100 for readability. Panel R-squared is equal to the average cross-sectional covariance of fitted values (using the full sample coefficients) scaled by the variance of the dependent variable. The cohort characteristic controls indicated below include mean log account values, log number of equity positions, log state income averaged over the period 2002-2011, state literacy rate, % rural, and % urban, where the mean is taken as of the account opening month for all accounts opened in the same month that are present in the given cross-section/current month. Specifications [1] and [2] include an ancillary dummy equal to one where the lagged disposition bias measure is unavailable as a control and set equal to zero as a result. All independent variables based on portfolio returns are winsorized at the 1st and 99th percentiles.

Dependent Variable		Idiosyncratic Share of Portfolio Variance			Monthly Turnover			Disposition Effect: ln(PGR/PLR)		
		[1A]	[1B]	[1C]	[2A]	[2B]	[2C]	[3A]	[3B]	[3C]
Observations (Account-Months)			2,837,081			2,249,818			592,370	
Unconditional Mean of Dependent Variable X 100			42.568			3.690			48.147	
Feedback-Based Experience	Average monthly outperformance of equity portfolio in past year		**1.537**	0.011			0.011			0.211
			(0.704)	*(0.013)*			*(0.083)*			*(0.561)*
	Average monthly outperformance over the past year due to trading					0.220	**0.110**			
						(0.173)	*(0.029)*			
	Outperformance of stocks held at a gain minus outperformance of stocks held at a loss over the								0.461	-1.166
									(1.082)	*(0.746)*
Lagged Account Behavior	Idiosyncratic share of portfolio variance based on last month's holdings			**0.964**			0.075			**-7.093**
				(0.003)			*(0.062)*			*(0.743)*
	Average monthly account turnover over the past year			**0.207**			**0.756**			**15.400**
				(0.032)			*(0.039)*			*(1.664)*
	Average ln(PGR/PLR) over the past year			0.016			-0.065			**0.283**
				(0.017)			*(0.122)*			*(0.024)*
Account Size	Log(account value)			**-0.188**			**0.185**			*1.288*
				(0.026)			*(0.031)*			*(0.759)*
Account Composition / Tilts	Portfolio stock market beta			**0.038**			0.022			-0.053
				(0.013)			*(0.040)*			*(0.785)*
	Small stock tilt			-0.005			-0.023			**-3.595**
				(0.010)			*(0.038)*			*(0.754)*
	Value tilt			**0.064**			0.011			0.770
				(0.007)			*(0.023)*			*(0.534)*
	Momentum tilt			0.004			0.046			**-2.386**
				(0.015)			*(0.107)*			*(1.101)*
Account Location	Semi-urban accounts			**0.085**			*0.125*			-0.668
				(0.029)			*(0.072)*			*(1.699)*
	Rural accounts			**0.054**			0.018			0.507
				(0.020)			*(0.059)*			*(1.174)*
	State dummies			Y			Y			Y
Account Age Quintile Dummies		Y	Y	Y	Y	Y	Y	Y	Y	Y
Cohort Characteristics		Y	Y	Y	Y	Y	Y	Y	Y	Y
Panel R-Squared		0.027	0.032	0.940	0.011	0.011	0.288	0.015	0.015	0.044

Table 7: Characteristics and Allocations of Stocks Sorted by Mean Account Age

The median for each characteristic is computed for each quintile of stocks (sorted by average age of investor's accounts) in each month, and the statistics reported below are the time-series medians of the cross-sectional medians. The allocations in the lower panel are computed for each month, and the average of these monthly allocations over time appear in the table. The account-age quintiles are constructed so that each quintile represents approximately 20% of the Indian stock market's capitalization after stocks with capitalization below 500 million Rs (approximately $10 million) are excluded.

Characteristics:

	Age of Company (Years)	Market Capitalization (Jan 2012 US$, millions)	Market Beta	Book-Market Ratio
Stocks with Youngest Average Account Age of Investors	16.0	$150.0	1.06	0.64
2	19.0	$122.0	1.08	0.72
3	19.0	$96.6	1.08	0.72
4	21.5	$80.3	1.07	0.73
Stocks with Oldest Average Account Age of Investors	27.0	$72.0	1.02	0.67

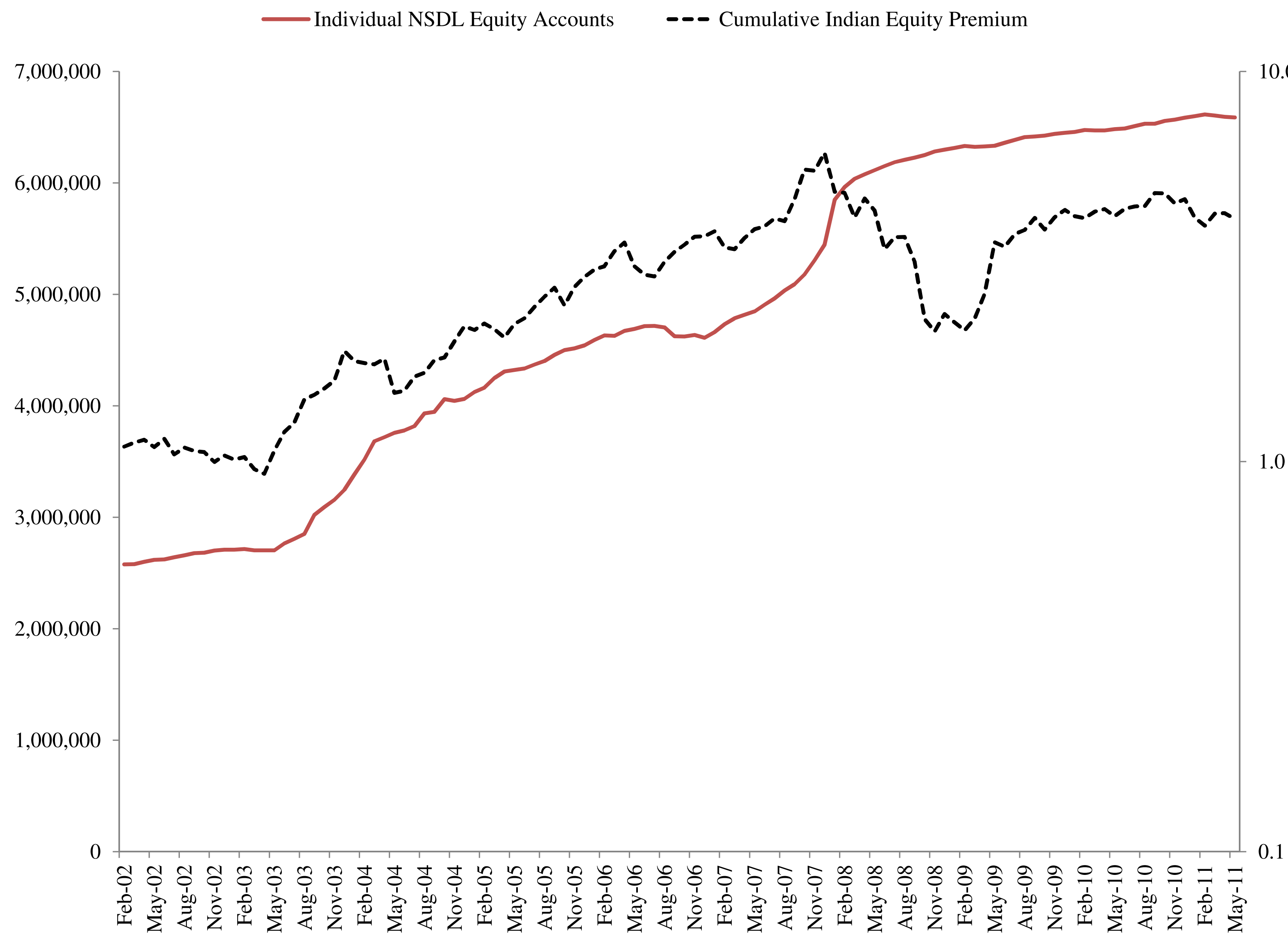
Stock Allocations:	Average Weights Within		Portfolio Weights in	
	Oldest Quintile of Individual Accounts	Newest Quintile of Individual Accounts	Account Level Analysis (Table 5)	Stock Level Analysis (Table 8)
Stocks with Youngest Average Account Age of Investors	27.51%	48.26%	-20.75%	-100.00%
2	21.93%	19.01%	2.92%	0.00%
3	18.79%	15.25%	3.54%	0.00%
4	18.78%	11.00%	7.78%	0.00%
Stocks with Oldest Average Account Age of Investors	12.99%	6.48%	6.52%	100.00%

Table 8: Performance Evaluation - Zero Cost Portfolios Formed on the Basis of Sophistication of Investors in the Stock

For each stock with a market capitalization of at least 500 million Rs (approximately $10 million) and month in the period January 2004 through January 2012, we compute the average age of investors' accounts, and the average disposition bias (measured by ln(PGR/PLR) and turnover of accounts over the past year. All portfolios are value weighted, and formed monthly. The strategies in [1] through [4] buy stocks in the most sophisticated quintile (accounts are old, and exhibit low disposition bias and turnover) and sell stocks in the least sophisticated quintile, where each quintile is constructed to represent about 20% of the total market capitalization. These returns are adjusted using unconditional CAPM, four, or six factor models, where the factor returns (except Illiq) are constructed in an analogous way to the factor returns from Ken French's website. The yield on three-month Indian Treasury bills is used as the risk free rate. The illiquidity factor is constructed from a independent double sort on size and turnover over the past 12 months, Illiq=0.5 x (Small, Low Turnover-Small, High Turnover)+0.5 x (Large, Low Turnover-Large, High Turnover). In [5] through [9], the portfolios are formed through independent 5 X 5 double sorts on the measure of average investor sophistication (with 20% of the market capitalization in each quintile) and one of size, book-market, momentum, etc (with 20% of the stocks by count in each quintile). Specification [9] sorts on institutional ownership percentage to test if the high returns of experienced investors can be attributed to institutional buying patterns (Gompers and Metrick 2001). These sorts are used to form 5 zero-cost portfolios; one that goes long high sophistication stocks and short low sophistication stocks for each quintile of the other sorting characteristic (e.g. size, book, market, etc.). The alphas below are then produced by regressing the average return of these 5 zero-cost portfolios on the six factors used in specification [4]. All standard errors are computed using a Newey West adjustment for serial correlation (with three lags).

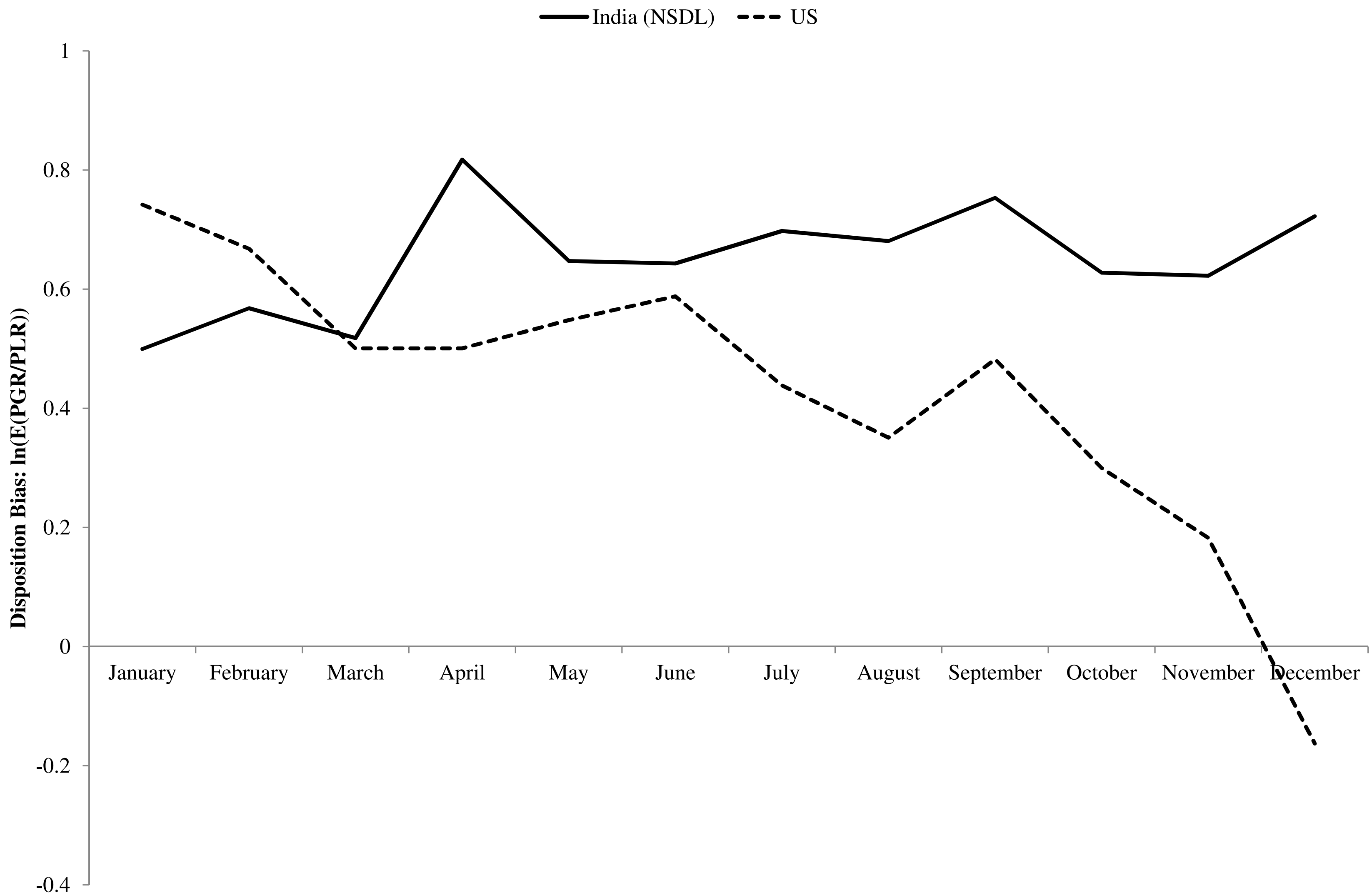
					Six-Factor Model plus Independent Double Sorts on the Sophistication Proxy and				
	Raw Return	CAPM	Four Factor	Six Factor	Size	Book-Market	Momentum	Liquidity (Turnover)	Institutional Ownership
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Panel A: Buy Stocks with the Greatest Average Account Age, and Vice Versa									
Monthly Alpha	*1.09%*	**1.51%**	**1.20%**	**1.26%**	*1.03%*	**1.20%**	**1.50%**	*1.14%*	0.91%
	(0.59%)	*(0.51%)*	*(0.57%)*	*(0.55%)*	*(0.59%)*	*(0.60%)*	*(0.58%)*	*(0.60%)*	*(0.57%)*
Factor Loadings									
Market Beta		**-0.41**	**-0.32**	**-0.30**					
		(0.09)	*(0.06)*	*(0.07)*					
SMB			**0.15**	**0.16**					
			(0.06)	*(0.06)*					
HML			-0.08	-0.13					
			(0.14)	*(0.14)*					
UMD			**0.30**	**0.31**					
			(0.10)	*(0.11)*					
Short Term Reversals				-0.09					
				(0.11)					
Illiq (Based on Turnover)				0.07					
				(0.16)					
Panel B: Buy Stocks with the Least Average Account Turnover, and Vice Versa									
Monthly Alpha	0.54%	*1.08%*	0.69%	0.76%	0.61%	0.77%	*1.11%*	0.74%	0.78%
	(0.84%)	*(0.58%)*	*(0.58%)*	*(0.63%)*	*(0.81%)*	*(0.72%)*	*(0.61%)*	*(0.74%)*	*(0.75%)*
Factor Loadings									
Market Beta		**-0.52**	**-0.32**	**-0.30**					
		(0.18)	*(0.05)*	*(0.07)*					
SMB			-0.04	-0.03					
			(0.09)	*(0.10)*					
HML			-0.22	-0.27					
			(0.20)	*(0.17)*					
UMD			**0.66**	**0.66**					
			(0.14)	*(0.15)*					
Short Term Reversals				-0.10					
				(0.15)					
Illiq (Based on Turnover)				0.07					
				(0.22)					
Panel C: Buy Stocks with the Least Average Account Disposition Bias, and Vice Versa									
Monthly Alpha	0.62%	**1.15%**	0.82%	**1.12%**	1.70%	**2.01%**	**1.52%**	0.83%	1.14%
	(0.77%)	*(0.57%)*	*(0.52%)*	*(0.54%)*	*(1.07%)*	*(0.70%)*	*(0.75%)*	*(0.72%)*	*(0.78%)*
Factor Loadings									
Market Beta		**-0.50**	**-0.33**	**-0.31**					
		(0.14)	*(0.05)*	*(0.05)*					
SMB			-0.05	-0.04					
			(0.06)	*(0.07)*					
HML			-0.19	**-0.29**					
			(0.14)	*(0.12)*					
UMD			**0.58**	**0.60**					
			(0.08)	*(0.09)*					
Short Term Reversals				-0.19					
				(0.11)					
Illiq (Based on Turnover)				-0.01					
				(0.17)					

Figure 1: Number of Individual Equity Accounts (Millions) and Cumulative Indian Equity Premium



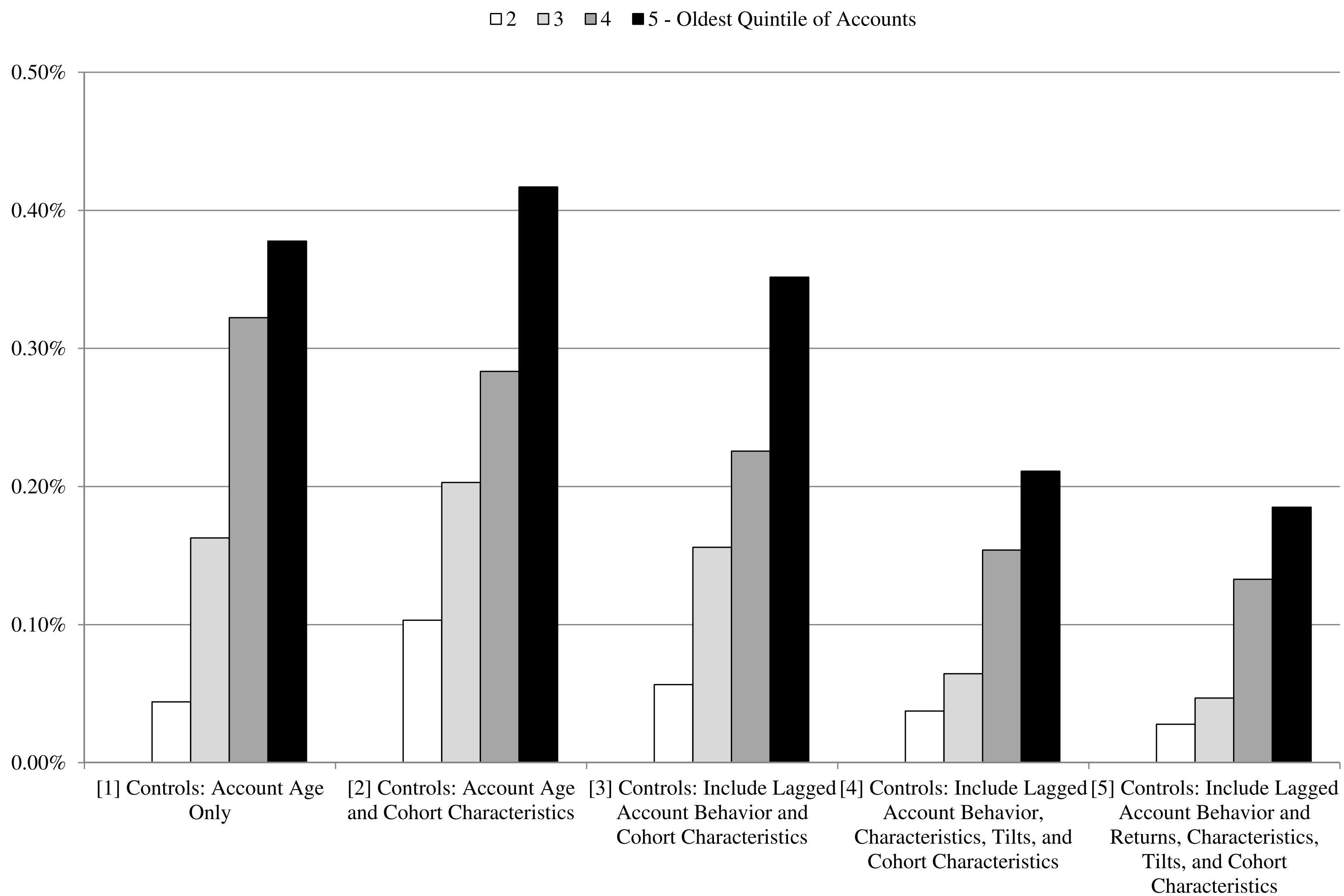
Equity accounts are aggregated by Permanent Account Number (PAN) which uniquely identify Indian investors. The Indian equity premium is computed using the yield on three-month Indian Treasury bills, and total returns and market capitalization of all Indian stocks for which we have such information.

Figure 2: Disposition Bias of Individual NSDL Accounts vs US Discount Brokerage Accounts (Odean 98, 99)



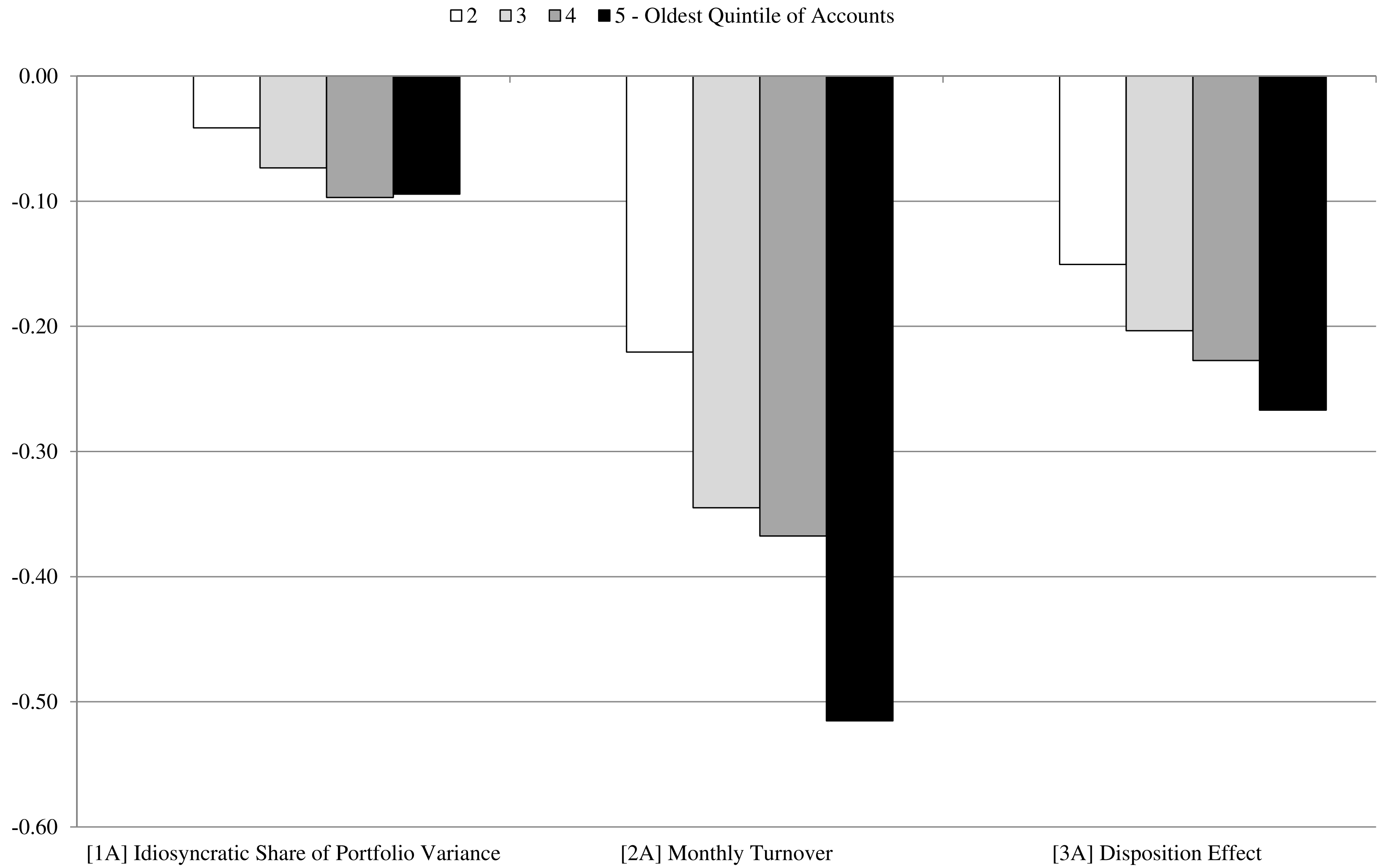
PGR/PLR is computed by month aggregating across accounts and years as in Odean (1998), which is the source for the US brokerage based statistics plotted. We take the log of both measures to make the units of the measure more comparable to our cross-sectional analysis. Levels are not quite comparable as this plot is a measure of ln(E(PGR/PLR)), not E(ln(PGR/PLR)), and the Odean measure applies greater weight to accounts with more transactions.

Figure 3: Account Age Quintile Dummies from Account Equity Return Regressions



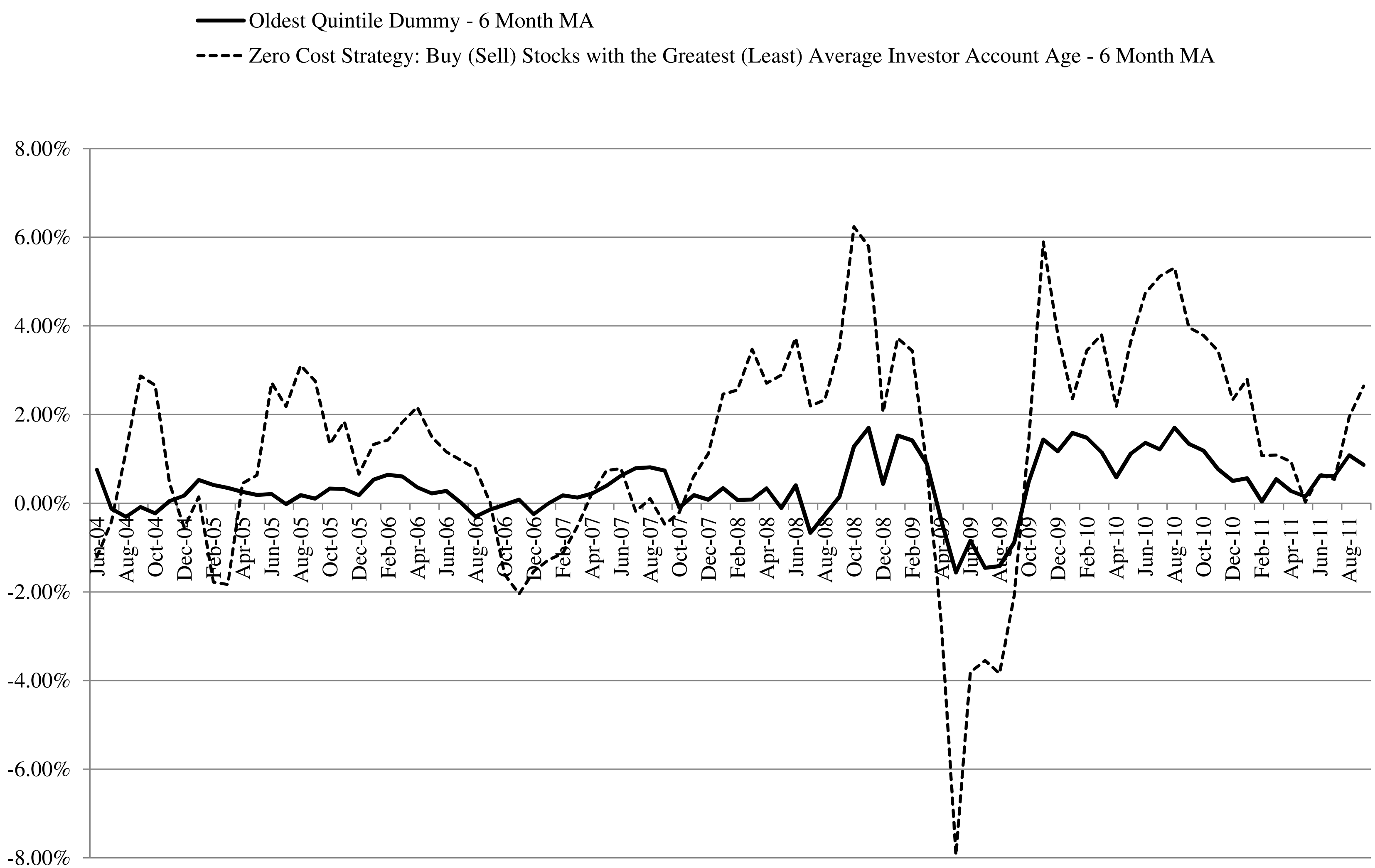
Bars represent the estimated dummy coefficients, i.e. additional monthly return, on accounts in each age quintile of accounts compared to the youngest accounts (for which such coefficients equal zero by construction). See Table 4 for details on the regressions to which these dummies correspond.

Figure 4: Account Age Quintile Dummies from Account Behavior Regressions



The plotted series represent the account age quintiles from regression specifications [1A], [2A], and [3A] of Table 5. The dummies from the idiosyncratic share of portfolio variance and monthly turnover regressions are scaled by the unconditional mean of the dependent variable to simplify the interpretation of the economic magnitude. Disposition effect dummies can be interpreted as log changes in the disposition effect predicted by account age quintile. These specifications include cohort characteristics as controls. For scaling purposes, account age dummies are divided by 10 in the disposition effect regression. The t-statistics on the oldest account quintile are all lower than -3.7.

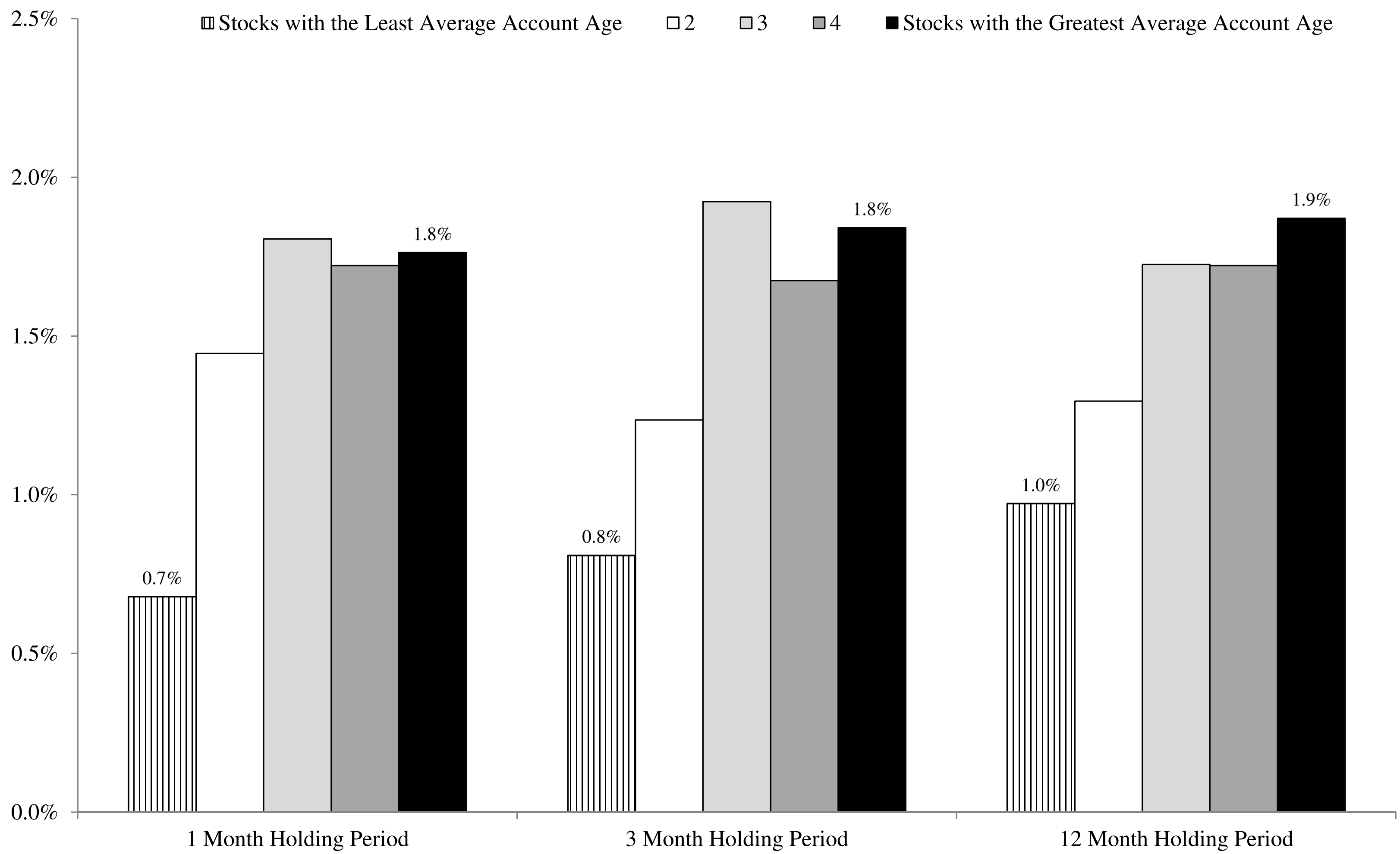
Figure 5: Additional Return to Older Accounts from Account Equity Regressions and a Stock-Level Strategy Based on Mean Age of Accounts Holding the Stock, 6 Month Moving Averages



The coefficients on the oldest account quintile dummy are extracted from regressions of specification [1] in Table 4. To construct the zero-cost portfolio represented in the plot, we compute the average age of accounts that hold the stock for all stocks with market capitalization in excess of 500 million Rs (approximately $10 million). The zero-cost portfolio, which is formed monthly, buys stocks in the top average account age quintile (where each quintile represents 20% of the Indian equity market by capitalization) as of the end of the previous month and sells stocks in the bottom such quintile (value weighting the stocks in the long and short portfolios).

Figure 6: Returns on Stock Portfolios Formed on the Basis of Average Account Age of Investors in the Stock

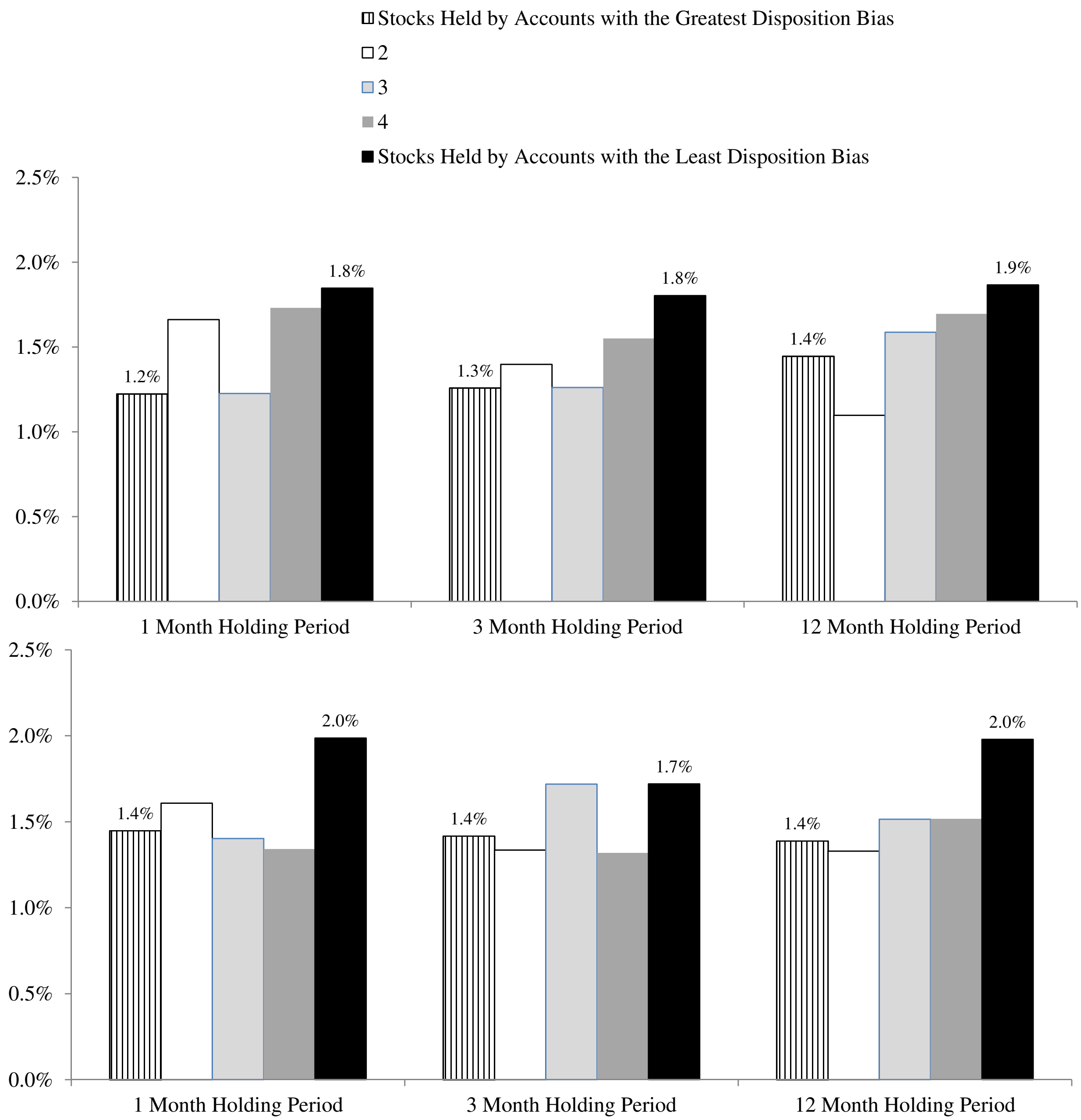
Value Weighted Portfolios, Average Monthly Returns Jan 2004-Jan 2012



The plotted series are raw monthly returns on value weighted portfolios of stocks formed on the basis of the mean age of accounts that hold the stock at the end of the previous month. Portfolios are held for one, three, or six months, with the returns on the three or six month holding period portfolios being the average return on portfolios formed over the past three or six months. Each portfolio represents stocks accounting for approximately 20% of the entire Indian stock market's capitalization after all stocks with a market capitalization below 500 million Rs (approximately $10 million) are removed.

Figure 7: Average Monthly Returns on Stock Portfolios Formed on the Basis of Past Disposition Bias (Top) and Turnover (Bottom) of Investors in the Stock

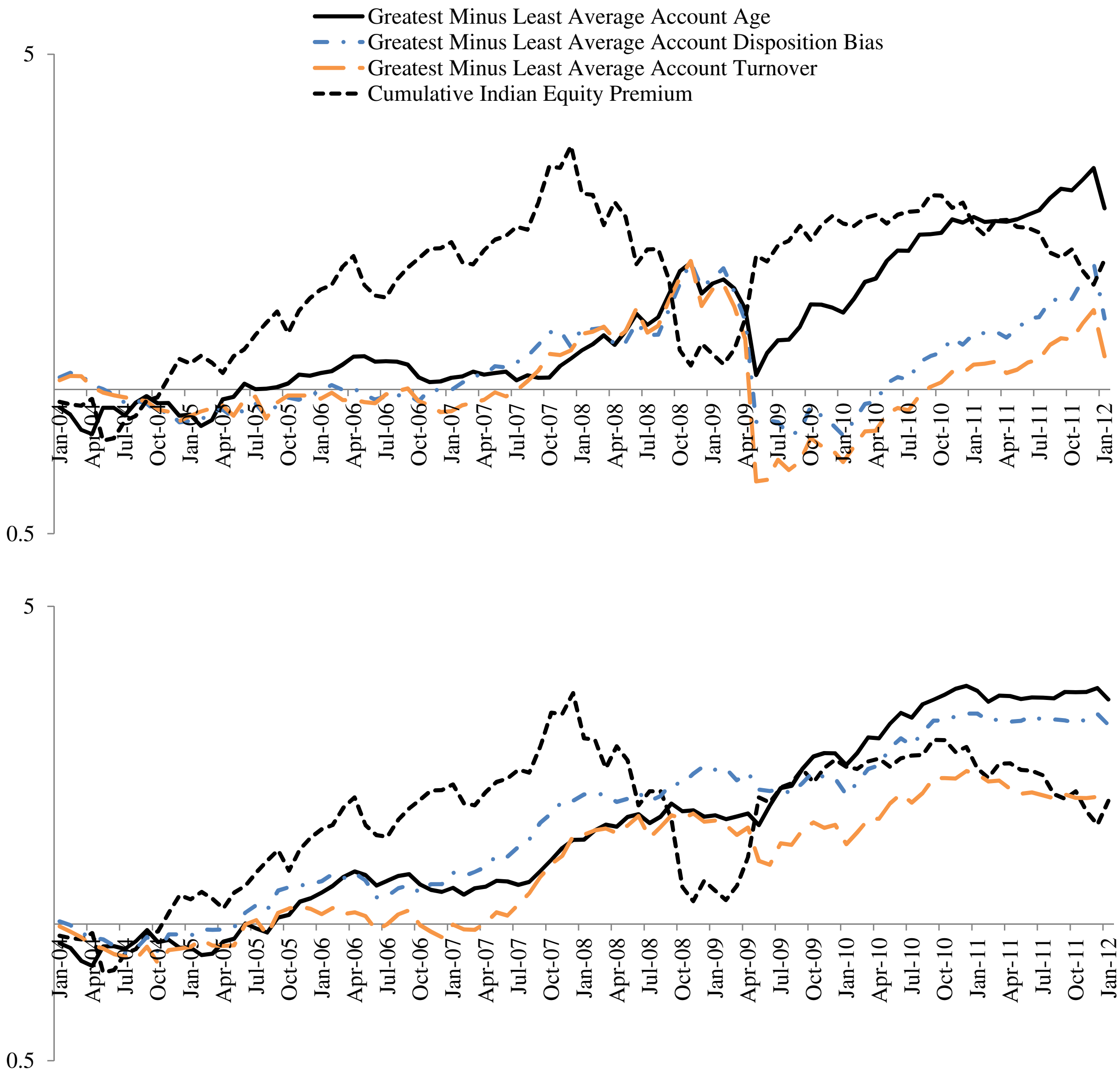
Value Weighted Portfolios, Average Monthly Returns Jan 2004-Jan 2012



Construction follows that of Figure 6, with portfolios here formed on the basis of mean disposition bias and turnover of accounts that hold the stock as of the end of the previous month (and for which such measures are available). Both disposition bias, ln(PGR/PLR), and turnover are measured over the past year of the account.

Figure 8: Cumulative Returns on Zero-Cost Portfolios Formed on the Basis of the Sophistication of Investors in the Stock
Jan 2004-Jan 2012

(Top Plot: Raw Zero-Cost Portfolio Returns, Bottom Plot: Six-Factor Alphas)



The zero-cost portfolios shown buy the top age quintile portfolios (seen in Figure 6) or bottom disposition bias or turnover portfolios (seen in Figure 7), and sell the bottom age quintile portfolios or top disposition bias or turnover portfolios. The top plot compares the raw returns on these strategies with the Indian equity premium. The bottom plot instead first adjusts strategy returns for exposure to six risk factors: market returns, size (SMB), value (HML), momentum (UMD), illiquidity, and short-term reversals. (Factor loadings are seen in Table 8.) All but the illiquidity factor are constructed by the same procedure as is used to construct the risk factors found on Ken French's website. The illiquidity factor is constructed from a independent double sort on size and turnover over the past 12 months, Illiq=0.5 x (Small, Low Turnover-Small, High Turnover)+0.5 x (Large, Low Turnover-Large, High Turnover).