Getting Better:
Learning to Invest in an Emerging Stock Market*

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

March 29, 2013

Abstract

This paper reports evidence that individual investors in Indian equities hold better performing portfolios as they become more experienced in the equity market. Experienced investors tilt their portfolios profitably towards value stocks and stocks with low turnover, but these tilts do not fully explain their good performance. Experienced investors also tend to have lower turnover and disposition bias. These behaviors, as well as underdiversification, diminish when investors experience poor returns resulting from them, consistent with models of reinforcement learning. Indian stocks held by experienced, well diversified, low-turnover and low-disposition-bias investors deliver higher average returns even controlling for a standard set of stock-level characteristics.

*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, Gavin Boyle, Stefano Della Vigna, Luigi Guiso, Michael Haliassos, Matti Keloharju, Ralph Koijen, Nagendar Parakh, Prabhakar Patil, Gagan Rai, Renuka Sane, Manoj Sathe, Ajay Shah, U.K. Sinha, Jayesh Sule, and seminar participants at NSDL, SEBI, NIPFP, NSE, the Einaudi Institute of Economics and Finance, EDHEC, the Oxford-Man Institute of Quantitative Finance, the IGIDR Emerging Markets Finance Conference, and the Sloan-Russell Sage Working Group on Behavioral Economics and Consumer Finance for numerous useful comments and discussions, and Vimal Balasubramaniam 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.

It's a little better all the time. (It can't get no worse.)

Lennon and McCartney, "Getting Better," 1967.

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, the equity premium, and many details of the household's financial situation, including 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 engage in certain widely-studied investment behaviors. Three such investment behaviors 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 counterparties in a market that is not strong-form efficient, and thus rewards investors who possess private information (Grossman and Stiglitz 1980, O'Hara 2003).

Households can control suboptimal investment behaviors in several ways. They can hold mutual funds as a way to gain equity exposure without trading stocks directly. This, however, may result in trade-offs between households' tendencies to engage in these behaviors, the level of fees charged by intermediaries, and the possibility that mutual fund managers may themselves be susceptible to these behaviors. Households can also learn from observing overall patterns in the market, or from their own investment experience (Nicolosi et al. 2009, Seru et al. 2010, Malmendier and Nagel 2011, 2012). 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 inexperienced investors, learning may be faster and easier to detect than in better established equity markets. Second, as discussed more fully below, 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. 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 history of listed Indian stocks over a decade. The relatively long time dimension of our panel allows us to measure investors' performance using their

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

realized returns, a method that is vulnerable to common shocks when applied to a short panel. Moreover, our data are 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, and occupation (Calvet et al. 2007, 2009a), their IQ test scores (Grinblatt and Keloharju 2011), 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 behavior and investment performance to the future behavior and performance of each account.

We have four main results. First, investment performance improves with account age. Second, older accounts have several profitable tilts in their portfolio weights, particularly towards value stocks and stocks with low turnover. However, these style tilts leave much of the outperformance of older accounts unexplained. Third, two of the three potentially harmful investment behaviors that we focus on, namely high turnover and the disposition effect, are less prevalent among older accounts. Fourth, all three investment behaviors diminish in response to painful experiences, including account underperformance, large losses in a single month, and poor returns from past trading and sales of gains. Putting these results together, investors appear to learn from stock market participation, at a rate that is influenced by their investment experiences.

1.1 Related Literature

The behavior of individual investors in equity markets has been of interest to financial economists studying market efficiency ever since the efficient markets hypothesis was first formulated. Shleifer (2000) succinctly summarizes the importance of this line of inquiry for the study of market efficiency, outlining that theoretical defenses of the efficient markets hypothesis rest on three pillars, the first of which is rational decision making and securities valuation by individuals, the second, the absence of correlated deviations from rationality even if some investors deviate from rational decision making, and the third, limits to arbitrage.

Understanding the behavior of individual investors is also important for the field of household finance (see Campbell 2006, for example). There has been much work on theoretically optimal investment in risky assets, and deviations from such idealized behavior by households have important implications for the evolution of the wealth distribution in the economy.

While the theoretical motivation for the study of individual investors has been clear for some time, empirical work in this area has been hampered by the difficulty of obtaining detailed data on individual investors' portfolios as well as by the large computational burden imposed by the study of such large datasets. These constraints have gradually been surmounted, and this field of study has increasingly become one of the most active areas of empirical research in financial economics.

Early work in the area (Cohn et al. 1975, Schlarbaum et al. 1978, Badrinath and Lewellen 1991) utilized relatively small samples of trader accounts from retail or discount brokerages to shed light on the stocks held by individual investors, the returns they earned, and the practice of tax-loss selling. The first set of empirical studies with a primary focus on questions related to rationality and market efficiency followed in the late 1990s, also using data sourced from discount brokerages, identifying that individual investors exhibit the disposition effect (Odean 1998), and trade excessively in the sense that their transactions costs outweigh any stock-picking ability they may possess (Odean 1999, Barber and Odean 2000). These tendencies were found to vary with the demographic characteristics and trading technologies of investors such as gender, marital status, and access to online trading (Barber and Odean 2001, 2002).

A characteristic of this early literature, and continuing to the present day, is the focus on *trading* rather than *investment* decisions of individual investors. While many questions in household finance are about the performance and risk properties of the entire risky asset portfolio of individual households, much of the literature has concentrated on performance evaluation of individual investors' purchases and sales at different post-trade horizons (see, for example, Coval et al. 2005, Barber et al. 2008, Seru et al. 2010), and on contrasting individual returns with those achieved by domestic and foreign institutional investors (Grinblatt and Keloharju 2000, Kaniel et al. 2008). A related focus has been on characterizing the trading strategies of individual investors through the lens of various behavioral biases such as the disposition effect, overconfidence, or inattention (see, for example, Barber and Odean 2008 and references above), and demonstrating the types of stocks (large, hard-to-value) in which these biases are most likely to manifest themselves (Ranguelova 2001, Kumar 2009).

This focus on trades rather than on investment arises quite naturally from the limitations of the data used to study investor behavior. In the US, discount brokerage accounts from a single service provider may not be truly representative of the entire portfolio of an individual investor, a problem made significantly worse when investors also have untracked mutual fund or 401(k) investments.[2] And some international datasets, such as the Taiwanese stock exchange data used by Barber et al. (2008), track all individual investor transactions but have little detail on holdings.

Our use of Indian data on direct equity holdings and trades helps us to partially surmount this obstacle. We have a relatively high-quality proxy for total household investment in risky assets, because equity mutual fund ownership by individual investors in India is very much smaller than direct equity ownership. As explained in the next section, we estimate that Indian households' equity mutual fund holdings are between 8% and 16% of their direct equity holdings over our sample period.

There are some other countries, such as Sweden and Finland, in which both direct equity ownership and mutual fund holdings are tracked. In principle this allows for a fuller characterization of household investment, but most previous studies using data from these countries have pursued different objectives than our focus on learning to invest. For example, Grinblatt et al. (2011) show that IQ affects stock market participation using data from the Finnish registry which provides detailed information on direct equity portfolios

[2] Calvet et al. (2007), show that mutual fund investments are an important source of diversification for Swedish investors.

combined with an indicator for whether the household invested in mutual funds in the year 2000. Grinblatt et al. (2012) highlight the impacts of IQ on mutual fund choice by Finnish investors using detailed data on mutual fund choices alongside less detailed information on direct equity investment. Calvet et. al (2007, 2009) use comprehensive data on Swedish investors' total wealth to shed light on stock-market participation and portfolio rebalancing, but the annual frequency of their data makes it difficult for them to evaluate higher-frequency phenomena such as momentum investing and turnover

Several papers, including those referenced in the previous section, share our focus on learning by individual investors, but emphasize different facets of this important issue. Feng and Seasholes (2005) use data on over 1500 individual accounts from China over the 1999 to 2000 period, and find that both experience (measured by the number of positions taken) and sophistication (measured by variables that include the idiosyncratic variance share) attenuate the disposition effect. Our analysis differs from theirs in our use of a more comprehensive set of portfolio characteristics, including the idiosyncratic variance share, and our exploration of feedback effects on future investing behavior. Linnainmaa (2010) estimates a structural model of learning and trading by investors in Finland, focusing on high-frequency traders, who make at least one round-trip trade in a given day. He finds, intriguingly, that traders appear to experiment with high-frequency trading to better understand their levels of skill, and cease trading if they experience poor returns. Our estimated feedback effects on underdiversification suggest that households also experiment with the composition of their equity portfolios, choosing to underdiversify more aggressively if they beat the market. This finding of experimentation is also consistent with Seru et. al. (2010), who carefully study the trading behavior of Finnish investors, focusing on the disposition effect. Seru et al. find that investors stop trading ("exit") after inferring that their ability is poor, and that trading experience weakens the disposition effect.[3] Our work is distinguished from this literature by our focus on investments rather than trades; to provide an instructive example, "exit" in

[3]Related work on the positive effect of trader experimentation and trader experience on returns and bias attenuation includes Dhar and Zhu (2006), Mahani and Bernhardt (2007), and Nicolosi et al. (2009). Korniotis and Kumar (2011), in contrast, find that the adverse effects of aging dominate the positive effects of experience.

our setting is the relatively uncommon exit of an investor from all equity positions, whereas Seru et al. use this term to refer to a period of time during which no trading occurs.[4]

Other authors have demonstrated the impacts of learning, including reinforcement learning, in other settings, such as trend following by mutual fund managers during the technology boom (Greenwood and Nagel 2009), individual investment in IPOs (Kaustia and Knüpfer 2008, Chiang et al. 2011) and household choice of credit cards (Agarwal et al., 2006, 2008). Agarwal et al. (2008) find that households learn how best to reduce fees on their credit card bills, and estimate that knowledge depreciates by roughly 10% per month, i.e., they find evidence that households learn and subsequently forget. While our current specifications do not explore this possibility, this is an important avenue that we intend to pursue in future work.

Finally, while we explore the role of personal feedback and investment experience in households' learning about investment, we do not currently consider the important topic of how social interaction or local networks affect learning (Hong et al., 2004, Ivkovic and Weisbenner, 2005, 2007).

The organization of the remainder 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 investment performance and behaviors. Section 4 shows that past performance predicts account behavior, while Section 5 shows that the behavior of the investor base predicts the returns on Indian stocks. Section 6 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

[4] While the frequency of exits is relatively low in our data, we estimate two alternative specifications to account for any potential biases caused by exits that are driven either by skill or luck.

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 equity 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. 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. Statistics from the Bombay Stock Exchange (BSE) and the National Stock Exchange (NSE) highlight that virtually all stock transactions take place in dematerialized form.

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.[5] 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

[5]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 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 detailed investor types we observe.

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. 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 12,200 in 2004 to almost 23,000 in 2012, as does the number of unique Indian equities covered. Starting at 4,510 in 2004, the number of equities reaches a peak of 7,721 in 2012. When we match these data to price, returns, and corporate finance information from various datasets, we are able to match between 95% and 98% of the market capitalization of these equities, and roughly the same fraction of the individual investor ownership share each year.

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 50% 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 18% in 2004, but declines to just below 10% in 2012, reflecting changes in NSDL coverage of institutions, as well as an increase in institutional investment over our sample period.

The sixth column shows the mutual fund share of total equities, which accounts for a little over 3.5% of total assets in the NSDL data in 2004, growing to a maximum of 4.72% in 2006, and declining to 3.97% by 2012. While comparing the fifth and sixth columns of

Table 1 demonstrates the magnitude of direct household equity ownership relative to mutual funds, this simple comparison would lead to an overestimate of mutual fund ownership by households. SEBI data in 2010 show that roughly 60% of mutual funds in India are held by corporations.[6] Assuming that this share has been static over our sample period, and that corporations and individuals hold roughly the same fraction of equity and bond mutual funds, this leads us to estimate that mutual fund holdings were between 8% and 16% of household direct equity holdings over the sample period. We note also that a 2009 SEBI survey of Indian equity-owning households found that about 65% of such households did not own any bonds or mutual funds.

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 2.7 million to roughly 6.1 million, that is, by 125%. 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 2004 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

[6] See SEBI website, http://www.sebi.gov.in/mf/unithold.html.

calculated for each account over its active life, giving equal weight to each account which is active for at least twelve months. 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, number of stocks held, and location

In the first panel of Table 2, 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 $4,000 in 2004 to about $68,000 in June 2008, with a time-series mean of $24,760. The median account size is however much smaller at $1,330, and even the 90th percentile account size is only $10,494, reflecting positive skewness in the distribution of account sizes. This positive skewness also explains the time-series variability of cross-sectional average account size, which is strongly influenced by the entry and exit of very large accounts. The large difference between mean and median account sizes implies 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).

The number of stocks held in each account is also positively skewed. The average number of stocks held across all accounts and time periods is almost 7, but the median account holds only 3.4 stocks on average over its life. The 10th percentile account holds 1 stock, while the 90th percentile account holds 14.2 stocks.

The next row shows that around 56% of individual accounts are associated with urban account addresses, 32% with rural addresses, and 12% with semi-urban addresses. These relative shares do change somewhat over time.[7]

[7]See the Data Appendix for a description of the method used to classify accounts into location-based

Account performance

The second panel of Table 2 looks at monthly account returns, calculated from beginning-of-month stock positions and monthly returns on Indian stocks.[8] 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). There is considerable variation over time in the cross-sectional average, with individual accounts underperforming in their worst months by as much as 7.8% or overperforming in their best months by as much as 9.7%. This variation is consistent with the literature on institutional and individual performance in US data (e.g. Grinblatt and Titman 1993, Kovtunenko and Sosner 2004, Kaniel et al. 2008), and can be explained in part by style preferences of individual investors. There is also dramatic variation across investors in their time-series average performance, with the 10th percentile account underperforming by 1.89% per month and the 90th percentile account overperforming by 1.56% 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, using a procedure very similar to that employed in Calvet et al. (2007). 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

categories.

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

years from their initial listing. To reduce noise in estimated idiosyncratic risk, we estimate idiosyncratic variance from a GARCH(1,1) model.[9]

The average idiosyncratic share is 45% in both the time-series and cross-sectional moments, which is slightly 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 23% to 54%) and across accounts (from 23% at the 10th percentile to 67% at the 90th percentile). However, the idiosyncratic variance share is not skewed to the same degree as the number of stocks held (reported in the top panel of the table), reflecting the convex declining relation between the number of stocks held in a portfolio and the portfolio's idiosyncratic risk.

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 average for Indian individual accounts. The time-series mean of the cross-sectional mean is 5.6% per month (or about 67% per year), and the cross-sectional median turnover is only 2.3% (or 28% 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 15.8% per month (190% 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

[9]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.

four months following the purchase or sale. In India, the former exceeds the latter by 2.78%, 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 5.25% 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 where the account executes a sale 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.

The disposition effect is important for Indian individual accounts. On average across months, the cross-sectional mean proportion of gains realized is 1.24 log points or 245% larger than the proportion of losses realized, while the median account has a PGR that is 1.37 log points or 293% larger than its PLR. While both time-series and cross-sectional variation in the disposition effect are substantial, it is worth noting that over 90% of accounts in the sample with 12 or more months with sales 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 very slightly greater than one at 1.02 in both the time-series and cross-sectional moments. The cross-sectional mean betas have modest variation over time from 0.94 to 1.08, and the cross-sectional variation in the time-series average beta is also small.

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 in the market index. We find a modest individual-investor tilt towards small stocks: the time-series mean percentile of market cap held by individual investors is 4.6% lower than the market index. This tilt varies modestly over time, but never switches sign. The small-cap tilt is skewed across accounts: the 10th percentile account has an 18% small-cap tilt while the 90th percentile account has a 3% large-cap tilt.

Individual Indian investors have a very small tilt on average 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 3.2% greater than the market index. This value tilt varies over time and does switch sign, reaching almost -6% 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 over 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 et al. (2002), and with short-term effects (but not longer-term effects) of past returns on institutional equity purchases estimated by Campbell et al. (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 investment behavior proxies – an effect we explore in detail in the next section. Among the proxies, turnover also has a 0.34 correlation with the idiosyncratic share of variance, implying that underdiversified accounts tend to trade more. All the investment behavior 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 -0.46 between the size tilt and the value tilt, implying that individuals who hold value stocks also tend to hold small stocks. This effect is somewhat mechanical given the correlation of these characteristics in the Indian universe.

3 Account Age Effects on Performance and Behavior

3.1 Regression specifications

In this section we explore the relation between the age of an account—our measure of overall investor experience and sophistication—and the account's performance and behavioral biases. In order to do this, we work with two alternative regression specifications. Defining an outcome (account return or behavior) for investor i at time t as Y_{it}, and the cross-sectional average of Y_{it} at time t as Y_t, we first estimate

$$Y_{it} - Y_t = \beta(A_{it} - A_t) + s_i + \varepsilon_{it,} \tag{1}$$

where A_{it} is a measure of the age of account i at time t, A_t is the cross-sectional average age measure for all accounts at time t, and s_i is an investor fixed effect that captures the inherent sophistication of investor i. We include the investor fixed effect to address the concern that more sophisticated investors may enter the market earlier and exit the market later than unsophisticated investors, which would make older accounts disproportionately sophisticated and would bias the estimation of a pure age effect. Equation (1) is our baseline specification.

A potential weakness in this approach is that the disposition effect – the tendency of investors to sell gains rather than losses – could lead to the disproportionate exit of investors who have earned high returns, presumably largely due to luck (Calvet et al. 2009a). As a consequence, older accounts may disproportionately be held by investors who had poor returns when their accounts were newer. In the presence of investor fixed effects, this biases upwards the estimated effect of account age on portfolio returns. To deal with this potential source of bias, we also estimate an alternative specification:

$$Y_{i,t} = \delta_t + \beta A_{it} + \theta C_i + \varepsilon_{it}, \tag{2}$$

where δ_t represents an unobserved time fixed effect. The vector C_i contains measured attributes of investor i which proxy for sophistication. The C_i include initial account value, initial number of stocks held, investor location type (urban or rural), and the income and literacy levels of the Indian state in which the investor resides at the time that the account was opened. In addition we include cohort-level means of these characteristics to capture the idea that accounts opened at a time when most other accounts are sophisticated are more likely to be sophisticated themselves.

In specification (2), account exits driven by lucky returns have no effect on the estimated age effect, but early entry and late exit by sophisticated, skilled investors does bias upward the age effect to the extent that the variables in C_i do not fully capture investor sophistication. For these reasons, we estimate both specifications to check the robustness of our results to these two potential sources of bias. As we continue this research we plan to estimate an auxiliary model of exit in order to estimate the possible size of exit-related bias.

Other explanatory variables can be added to these regressions. One natural choice is account size, which we know from Table 3 is correlated with both account age and investment behaviors. We note however that account size is mechanically correlated with past returns. In the presence of an investor fixed effect, as in specification (1), this can lead to a spurious negative effect of size on returns along with a spuriously positive fixed effect for accounts that experience high early returns. Accordingly we exclude account size from our regressions predicting returns, although we have confirmed that the inclusion of size in specification (2) has little impact on the reported results. Investment behaviors can also be added as regressors in both specifications (1) and (2).

In both regression specifications we consider several possible forms for the account age effect. First and most simply, we consider linear age effects: $A_{it} = Age_{it}$. Since there is no particular reason why an investor's expected returns or behavior should be a linear function of account age, we also model account age effects as a piecewise linear form of account age. The curvature of the piecewise linear age effects suggests age effects of the form $A_{it} = Age_{it}^{0.5}$, which we adopt as our benchmark for the non-linear functional form of account age effects.

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. Figure 3 shows the distribution of NSDL accounts from various states. The size of the bubbles in the plot are proportional to the population of each state. The Y-axis shows the number of people in each state per NSDL account, and the X-axis plots the per-capita income of the state in 2011. For example, in Bihar, a poor state with a per-capita annual income of roughly \$350 per annum, 1 in 1400 people invest in the stock market and are captured in NSDL data, whereas the small, relatively wealthy state of

Delhi, with per-capita annual income of roughly \$2600, has 1 in 33 people participating and captured in NSDL. Given that the NSDL share of total equity capitalization is around 70% in 2012, these fractions are relatively accurate representations of total participation (without accounting for pure indirect equity ownership) by individual households in the stock market.

Our return regressions are estimated using 4 million account months of data spanning January 2004 through January 2012, and our regressions of account behaviors use somewhat fewer observations, as these measures cannot be defined for as many account months. We estimate panel regressions applying equal weight to each cross-section, and within each cross-section, we use weights to account for the sampling strategy. Standard errors are computed by bootstrapping months of data, to account for any possible contemporaneous correlation of the residuals.

3.2 How performance improves with age

Table 4 reports four variants of our basic regression approach and documents the relationship of our behavior measures to returns at the account level. The first three columns predict account returns relative to the cross-sectional average of all account returns (specification 1), while the next three columns (specification 2) allow for a time effect. The effect of age is estimated either to be linear (columns [a] and [c]) or square-root (columns [b]), where the coefficients reported give the expected performance of a one-year-old account relative to a brand new account. The linear age effect is estimated to be about 14 basis points per month in specification 1, and 11 basis points per month in specification 2, and both effects are statistically significant at the 5% level. The square-root age effect is greater for a one-year-old account (39 basis points in specification 1 and 30 basis points in specification 2), but of course it dies off more rapidly and the coefficient is only statistically significant at the 10% level for specification 2.

Columns [c] of Table 4 show that the age effect in returns barely changes once we control for lagged investor behaviors. The superior performance of older accounts cannot be purely attributed to any propensity of older accounts to better diversify, trade less, or show less disposition bias.

Figure 4 illustrates the choice between a linear and a square-root functional form, with results for account returns shown in the top left panel of the figure. The solid line shows the estimates from a more general piecewise linear function of age, while the dashed lines illustrate three parametric models, linear, square-root, and cube-root. The piecewise linear function is upward-sloping but somewhat jagged, and evidence for concavity is quite weak.

An important question is how more experienced investors achieve higher average returns. In Table 5 we attempt to answer this question by forming a zero-cost portfolio that goes long stocks held by a representative experienced investor (a stratified-sample-weighted average of the portfolio weights of accounts in the oldest quintile), and goes short stocks held by a representative novice investor (a stratified-sample-weighted average of the portfolio weights of accounts in the youngest quintile). Figure 5 illustrates the cumulative excess returns to the long and short legs of this portfolio relative to the Indian short rate, along with the overall excess return of the Indian equity market, over the period January 2004-January 2012. By the end of this period the cumulative excess return on the experienced-investor portfolio was 122%, while the cumulative excess return on the Indian market index was 87%, and the cumulative excess return on the novice-investor portfolio was only 29%.

In the first column of Table 5, we regress the portfolio weights in the zero-cost portfolio onto a vector of stock characteristics, to see what characteristics are preferred or avoided by experienced investors relative to novice investors. In the second column, we decompose the returns on the zero-cost portfolio into unconditional and timing effects related to either stock characteristic tilts or a residual that we call "selectivity" following Wermers (2000). The top half of this column reports the unconditional contribution of each stock characteristic tilt to returns, reporting standard errors that take into account the sampling error in returns to characteristics as well as uncertainty in the characteristic tilt itself. The lower part of this column reports the overall performance contribution of all unconditional characteristic tilts, together with the contributions of unconditional stock selectivity, and stock characteristic and other stock timing effects. The third and fourth columns of the table repeat this exercise adding a variable for large, attention-grabbing initial public offerings, to capture the idea that such events might be important contributors to the performance of novice investors.

Table 5 shows that relative to novice investors, experienced Indian investors tilt their portfolios towards low-beta stocks; this has a minimal effect on return while reducing risk. Experienced investors also have little systematic preference for momentum. However, they do have a number of other important characteristic tilts. They favor small stocks, value stocks, stocks with low turnover, stocks without large beneficial ownership, stocks held by institutions, and older stocks. All of these tilts except for the size tilt, which contributes negligibly, are return-enhancing. In particular, more experienced investors enjoy higher returns from their tilts towards value stocks and low turnover stocks.

Taken together, the stock characteristics explain 22 basis points per month out of a total excess return of 39 basis points. The remainder is not explained by characteristic timing, which makes an insignificant negative contribution of -2 basis points. The remaining 19 basis points of performance are split between non-characteristic related stock selection (7 basis points) and stock timing effects (12 basis points). Results are generally similar when we add in a dummy for large IPOs, though the apparent preference of older accounts for older stocks appears to be entirely due to their avoidance of large IPOs.

The characteristic tilts documented in Table 5 suggest that performance evaluation of experienced investors relative to novice investors may need to correct for exposures to systematic risk factors. Table 6 compares raw excess returns to CAPM and multi-factor alphas for the long-short portfolio constructed in Table 5. The first column of the table reports a raw excess return of 39 basis points per month, which is statistically significant only at the 10% level because of noise created by market movements. The second column shows that this corresponds to a CAPM alpha of 54 basis points per month—significant at the 5% level—and a negative market beta of -0.15, reflecting the fact that older accounts tend to hold somewhat lower-beta stocks even while delivering a higher return. The third and fourth columns show that the alpha increases to 64 basis points per month in a Fama-French-Carhart four-factor model including momentum, and 93 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). Interestingly, the long-short portfolio has negative loadings on an Indian version of the Fama-French HML factor and our illiquidity factor, despite the preference of experienced investors for stocks with high book-market ratios and lower turnover documented in Table 5.

3.3 How behavior changes with age

We now ask whether our three proxies for investment behaviors change with the age of the account. Table 7 predicts the idiosyncratic variance share, turnover, and disposition bias measured by the log ratio of PGR to PLR, again using our two specifications (1) and (2) and allowing for either a linear or square-root age effect. While a positive linear age effect fit performance better, turnover is better captured by a negative square-root function of age. This is shown by the incremental R^2 statistics reported in the table, which measure the contribution of the age variable to the overall fit of the regression, and are markedly higher for the square-root specification. The piecewise linear regressions shown in the lower left panel of Figure 4 is also clearly declining and convex.

The age effects documented in Table 4 are not only statistically significant, but large in economic magnitude. To see this, the vertical axes on the plots in Figure 4 are scaled to have a range equal to twice the cross-sectional standard deviation in returns or behavior. Over the course of five years, monthly turnover declines by 11 percentage points and disposition bias declines by 56 log percentage points, both of which are on par with or greater than the cross-sectional standard deviation. In contrast, the portfolio share of idiosyncratic variance changes little with age. This may not be surprising when considering the results of Ivkovic et al. (2008), who suggest that underdiversification may also represent extreme sophistication – they find that individual trader performance improves as the number of stock holdings decrease, holding other determinants of performance constant. In addition, Table 5 showed that experienced Indian investors have a preference for small value stocks, which have unusually high idiosyncratic volatility.

4 Investment Experience and Behavior

Since behavior changes dramatically with account age, it is plausible that it may also be affected not only by the fact of investing, but also by the experiences that investors have in the market. We explore this possibility in Table 8, which uses fixed-effect regressions (1) to predict our three proxies for investment mistakes. All regressions include square-root age effects and account size controls as in the previous section.

Panel A of Table 8 predicts the idiosyncratic share of portfolio variance. The predictor variables are two summaries of past investment success: the cumulative outperformance of the account relative to the market, and the worst monthly return experienced by each account. Cumulative account outperformance may lead investors to assess their investing skills more optimistically, encouraging them to make larger idiosyncratic bets. Large negative returns may remind investors of the risks of stock market investing in general, and undiversified investing in particular. Both variables enter strongly, with positive and negative signs respectively. However, this result must be interpreted with some caution because the effect of cumulative outperformance may result in part from inertia. If an account has a diversified component and an undiversified bet, the weight of the undiversified bet increases with its return if the account is not rebalanced, and this will mechanically increase the idiosyncratic share of variance.

In panel B of Table 8 we predict turnover from the cumulative increase in returns due to trades, a measure of an account's past trading success. For each month, the return to trades is calculated as the difference between actual returns in the current month and the returns that would have been experienced if the account had stopped trading three months earlier. This return to trades is then cumulated over the life of the account. This variable strongly predicts turnover, implying that trading profits strengthen the tendency to trade stocks frequently. This result is consistent with the findings of Linnainmaa (2011), who employs information on a set of high-frequency traders from Finland. The two variables from panel A also enter the turnover regression significantly.

It should be noted that the effect of recent trading profits on turnover may result in part from the disposition effect. If recent trading is profitable, then an account has tended to purchase winners which are more likely to be sold if the investor has disposition bias. Such sales, and subsequent purchases of replacement stocks, increase turnover.

Finally, in panel C we predict disposition bias using the returns to past sales of winners and losers. We calculate excess returns relative to the market index on stocks that each account sold, during the three month period following each sale, and compare the excess returns to losers sold relative to winners sold, weighting by the value of each sale and finally cumulating this measure over the life of the account. The idea of this measure is that if an account holds mean-reverting stocks, disposition bias tends to be profitable because winners sold underperform losers sold after the sale date, encouraging further disposition bias. If an account holds stocks that display short-term momentum, however, disposition bias tends to be unprofitable and may be discouraged by experience. This variable enters the regression with the expected sign, but is not statistically significant.

Figure 6 illustrates the relative importance of account age and investment experience in predicting each of our three investment behaviors. For all accounts that opened in December 2003, the figure shows the predicted behaviors from January 2004 through the end of the sample, using the all predictor variables except account value from the specification in column [2] of Table 8. The figure illustrates the median and the 10th and 90th percentiles of predicted behaviors. In both the disposition effect and turnover plots, the dominant influence of the age effect is clearly visible in the figure, but the spread in predicted behaviors across accounts is meaningful in the case of all investment behavior proxies. Declines in predicted behaviors occur rapidly at the beginning of the period, because of a strong early age effect and a market downturn in the spring of 2004. There is also a marked decline in the fall of 2008, again resulting from poor stock returns.

The empirical results of this section provide 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. The presence of account level fixed effects in our specifications should significantly reduce concerns on this score, as the investor's average skill at trading should be absorbed by these account level effects. In addition, our regressions in Table 4 showed that turnover and disposition bias are associated with lower account returns, not higher returns as reverse causality would require.

5 Stock Returns and the Investor Base

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 et al. 2010).

Table 9 uses Fama-MacBeth regressions to predict the returns of Indian stocks with at least 10 individual investors in our sample of individual accounts. Column 1 shows that the average age of the accounts that hold a stock predicts the return to that stock, consistent with the account-level results reported in Table 4. Column 2 adds information on the behavior of the investor base: the average share of idiosyncratic variance in the portfolios of the stock's investors, the turnover of these portfolios, and the disposition bias of the stock's investors. A high turnover investor base in particular predicts lower returns. The age effect, though somewhat diminished, remains significant.

Column 3 adds a standard set of stock characteristics to the regression. The book-market ratio and momentum enter positively, and stock turnover enters negatively, consistent with evidence from developed markets. The effect of account age in the investor base is now much weaker, but stocks with undiversified investors have lower average returns (significant at the 5% level), and stocks with disposition-biased investors have lower average returns. The effect of a high-turnover investor base remains negative, but it is smaller in magnitude because it is correlated with turnover in the stock itself.

The institutional ownership of stocks is included in Table 9 to addresses one possible concern about our finding of a positive age effect. 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. However, this story is contradicted by the fact that in Table 9, the coefficient on institutional ownership is negative rather than positive.

6 Conclusion

In this paper we have studied the investment strategies and performance of individual investors in Indian equities over the period from 2004 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 value stocks and stocks of longer-established companies, but also by picking stocks that perform well after controlling for their characteristics. Older accounts also have lower turnover and a smaller disposition effect.

Our evidence also suggests that learning is important among Indian individual investors. Accounts that have experienced low returns relative to the market, and low returns in a single month, increase their diversification and reduce their turnover and disposition bias. Moreover, accounts that have experienced low returns from their trading decisions tend to reduce their turnover in the future, while poor returns associated with the disposition effect have an imprecisely estimated negative effect on future disposition bias. These results suggest that Indian individual investors learn, not only from the experience of stock market participation itself, but also from the returns generated by their investment behaviors.

If investment behaviors are related to investor financial sophistication, and if sophisticated investors are able to pick stocks with high expected returns, then the characteristics of a stock's investor base can be used to predict the stock's returns. We present evidence that this is the case, even controlling for the stock's own characteristics.

There are several interesting questions we have not yet explored, but plan to examine in the next version of this paper. We can ask whether the effect of experience on behavior is permanent, as implicitly assumed by our specification that predicts behavior using cumulative past returns, or whether the effect of experience decays over time as suggested by Agarwal et al. (2006, 2008). Second, we can explore whether the effect of experience on behavior varies with age, as might be the case if investors update priors about their skill or about the merits of selling winning positions, and gradually become more confident in their beliefs. Finally, we can ask whether all three of the behaviors studied in this paper can be aggregated into a single index of financial sophistication, as suggested by Calvet et al. (2009b).

Data Appendix

Classification of Investor Account Geography (Urban/Rural/Semi-Urban)

We provided NSDL with a mapping of PIN codes (Indian equivalent of ZIP codes) to an indicator of whether the PIN is a rural, urban, or semi-urban geography. To make this determination, PIN codes were matched to state and district in an urbanization classification scheme provided by Indicus. In cases where urbanization at the district level is ambiguous, we use use postal data, noting that the distribution of number of large postal branches and small sub-branches in a PIN is markedly different in urban and rural geographies.

Stock Data

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.[10]

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 reported inconsistently. For the 25 highest and lowest remaining total monthly returns,

[10] 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.

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, in computing account returns we 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 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

Agarwal, Sumit, Souphala Chomsisengphet, Chunlin Liu, and Nicholas S. Souleles, 2006, Do consumers choose the right credit contracts?, unpublished paper, available online at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=843826.

Agarwal, Sumit, John C. Driscoll, Xavier Gabaix, and David Laibson, 2008, Learning in the credit card market, NBER Working Paper 13822.

Badrinath, S., and Wilbur Lewellen, 1991, Evidence on tax-motivated securities trading behavior, *Journal of Finance* 46, 369–382.

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.

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

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

Barber, Brad M., and Terrance Odean, 2008, All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors, *Review of Financial Studies*, 21, 785–818.

Barber, Brad M., Yi-Tsung Lee, Yu-Jane Liu, and Terrance Odean, 2009, Just how much do individual investors lose by trading?, *Review of Financial Studies* 22, 609–632.

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.

Cohn, Richard A., Wilbur G. Lewellen, Ronald C. Lease and Gary G. Schlarbaum, 1975, Individual Investor Risk Aversion and Investment Portfolio Composition, *Journal of Finance* 30, 605–620.

Coval, Joshua, David Hirshleifer, and Tyler Shumway, 2005, Can individual investors beat the market?, Unpublished paper, available online at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=364000

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.

De, Sankar, Naveen R. Gondhi, and Bhimasankaram Pochiraju, 2010, Does sign matter more than size? An investigation into the source of investor overconfidence, unpublished paper, Indian School of Business, Hyderabad, available online at

http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1657926.

Dhar, Ravi, and Ning Zhu, 2006, Up Close and Personal: An Individual Level Analysis of the Disposition Effect, *Management Science*, 52, 726–740.

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.

Feng, Lei and Mark S. Seasholes, 2005, Do investor sophistication and trading experience eliminate behavioral biases in financial markets?, *Review of Finance* 9, 305–351.

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.

Grinblatt, Mark and Matti Keloharju, 2000, The investment behavior and performance of various investor types: a study of Finland's unique data set, *Journal of Financial Economics* 55, 43–67.

Grinblatt, Mark, Matti Keloharju, and Juhani Linnainmaa, 2011, IQ and stock market participation, *Journal of Finance* 66, 2121–2164.

Grinblatt, Mark, Ikaheimo, Seppo, Keloharju, Matti and Knüpfer, Samuli, 2012, IQ and Mutual Fund Choice, Unpublished paper, available online at

http://ssrn.com/abstract=2021957

Grinblatt, M., and S. Titman, 1993, Performance Measurement without Benchmarks: An Examination of Mutual Fund Returns, *Journal of Business* 66, 47–68.

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

Hong, H., J. D. Kubik, and J. C. Stein, 2004, Social Interaction and Stock-Market Participation. *Journal of Finance* 59:137–63.

Ivkovic, Z., C. Sialm, and S. J. Weisbenner, 2008, Portfolio Concentration and the Performance of Individual Investors. *Journal of Financial and Quantitative Analysis* 43:613–55.

Ivkovic, Z., and S. J. Weisbenner, 2005, Local Does as Local Is: Information Content of the Geography of Individual Investors Common Stock Investments. *Journal of Finance* 60:267–306.

Ivkovic, Z. Z., and S. Weisbenner, 2007, Information Diffusion Effects in Individual Investors' Common Stock Purchases: Covet Thy Neighbors' Investment Choices. *Review of Financial Studies* 20:1327–57.

Kaniel, Ron, Gideon Saar, and Sheridan Titman, 2008, Individual investor trading and stock returns, *Journal of Finance,* 63, 273–310.

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.

Kumar, Alok, and George Korniotis, 2011, Do Older Investors Make Better Investment Decisions? *Review of Economics and Statistics*, 93, 244–265.

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.

Kumar, Alok, 2009, Hard-to-value stocks, behavioral biases, and informed trading, *Journal of Financial and Quantitative Analysis,* 44, 1375–1401.

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.

Mahani, Reza and Dan Bernhardt, 2007, Financial speculators' underperformance: Learning, self-selection, and endogenous liquidity, *Journal of Finance* 62, 1313–1340.

Malmendier, Ulrike and Stefan Nagel, 2011, Depression babies: Do macroeconomic experiences influence risktaking?, *Quarterly Journal of Economics* 126, 373–416.

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

Mukherjee, Saptarshi and Sankar De, 2012, Are investors ever rational?, unpublished paper, Indian School of Business, available online at

http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2156047.

Nicolosi, Gina, Liang Peng, and Ning Zhu, 2009, Do individual investors learn from their trading experience?, *Journal of Financial Markets* 12, 317–366.

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.

Pastor, Lubos and Pietro Veronesi, 2009, Learning in financial markets, *Annual Review of Financial Economics* 1, 361–381.

Ranguelova, Elena, 2001, Disposition effect and firm size: New evidence on individual investor trading activity. Unpublished paper, available online at

http://ssrn.com/abstract=293618.

Schlarbaum, Gary, Wilbur Lewellen, and Ronald Lease, 1978, Realized returns on common stock investments: The experience of individual investors, *Journal of Business* 51, 299–325.

Seru, Amit, Tyler Shumway, and Noah Stoffman, 2010, Learning by trading, *Review of Financial Studies* 23, 705–739.

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	% of NSDL Equity Value in Mutual Funds
2004	12,264	4,510	$386.3	51.24%	17.59%	3.51%
2005	13,487	4,818	$553.1	58.04%	15.86%	3.73%
2006	15,279	5,126	$818.9	63.74%	15.04%	4.72%
2007	17,091	5,479	$1,819.1	66.72%	12.87%	4.55%
2008	17,511	5,949	$647.2	65.26%	11.94%	4.46%
2009	17,458	6,367	$1,306.5	64.73%	11.29%	4.56%
2010	19,458	6,846	$1,631.8	67.69%	10.84%	4.35%
2011	22,663	7,448	$1,007.2	69.67%	10.16%	4.00%
2012	22,696	7,721	$1,202.9	70.22%	9.90%	3.97%

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 the given data appear 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 12,62,757	Rs 2,00,160	Rs 34,66,283	Rs 7,403	Rs 67,833	Rs 5,35,213
Account Value, Jan 2012 US$	$24,760	$3,925	$67,966	$145	$1,330	$10,494
Number of Equity Positions	6.81	4.73	8.38	1.00	3.43	14.18
Urban Accounts	55.66%	54.54%	57.31%	0	1	1
Semi-Urban Accounts	12.25%	11.76%	12.77%	0	0	1
Rural Accounts	32.09%	30.39%	33.15%	0	0	1
Monthly Account Stock Return Minus Market	-0.06%	-7.77%	9.71%	-1.89%	-0.19%	1.56%
Idiosyncratic Share of Portfolio Variance	0.45	0.23	0.54	0.23	0.45	0.67
Monthly Turnover	5.59%	1.93%	12.36%	0.00%	2.26%	15.76%
Disposition Effect - ln(PGR/PLR)	1.24	-1.26	2.42	0.26	1.37	2.42
Stock Portfolio Beta	1.02	0.94	1.08	0.95	1.02	1.12
Size Percentile of Stocks Held Relative to Market Portfolio	-4.60	-6.10	-3.08	-18.24	-1.58	3.27
Book-Market Percentile of Stocks Held Relative to Market Portfolio	3.22	-5.96	17.54	-9.71	1.89	22.02
Momentum Percentile of Stocks Held Relative to Market Portfolio	-5.43	-17.18	6.05	-21.35	-5.59	4.60

Table 3: Cross-Sectional Correlations of Account Level Variables

Statistics are computed on the basis of individuals' account months used in the regression models for which all variables are defined. 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]	**0.32**	1.00								
Account Stock Return Over the Past Year	[3]	0.03	0.08	1.00							
Idiosyncratic Share of Portfolio Variance	[4]	**-0.17**	**-0.46**	0.03	1.00						
Turnover Over the Past Year	[5]	**-0.40**	**-0.31**	0.02	**0.34**	1.00					
Disposition Effect Over the Past Year	[6]	**-0.12**	**-0.15**	0.01	0.03	0.00	1.00				
Stock Portfolio Beta	[7]	**-0.10**	**-0.14**	-0.01	**0.16**	**0.17**	0.04	1.00			
Size Percentile of Stocks Held	[8]	0.02	**0.14**	0.02	**-0.30**	**-0.13**	-0.05	**-0.33**	1.00		
Book-Market Percentile of Stocks Held	[9]	-0.01	**-0.11**	-0.09	**0.15**	0.06	0.05	**0.12**	**-0.46**	1.00	
Momentum Percentile of Stocks Held	[10]	0.05	**0.18**	**0.27**	-0.06	0.02	**-0.15**	0.00	**0.15**	**-0.26**	1.00
Urban Account	[11]	0.05	0.07	0.01	-0.03	-0.02	-0.02	-0.01	0.01	-0.01	0.02
Semi-Urban Account	[12]	0.00	-0.02	0.01	0.01	0.00	0.01	0.02	-0.02	0.02	0.00
Rural Account	[13]	-0.05	-0.06	-0.01	0.03	0.02	0.02	-0.01	0.00	0.00	-0.02

Table 4: Account Age Effects in Equity Portfolio Returns

Results are constructed from a stratified random sample: 5,000 individual accounts are drawn at random from each Indian state/territory, with all individual accounts drawn from states with less than 5,000 individual accounts. A bit over 4 million account months spanning January 2004 through January 2012 are used in the regressions. Pre-2002 cohorts are excluded from all regressions as the cohort level characteristics C_i are unavailable for cohorts which opened prior to the first month in our database (February 2002). Specification 1 is $(R_{it}-R_t)=\beta(A_{it}-A_t)+\kappa(B_{it}-B_t)+s_i+\varepsilon_{it}$, and specification [2] is $R_{it}=\delta_t+\beta A_{it}+\kappa B_{it}+\theta C_i+\varepsilon_{it}$, where R_{it} represents the returns of investor i in month t, A represents the account age (in columns [a] and [c]) or its square root (in column [b]), B are the lagged account behaviors appearing in columns [c], and δ and s are time and individual fixed effects. Lagged behaviors are averages over the past 12 months, with values winsorized at the 1st and 99th percentile of accounts for which at least five observations of the given behavior are available over the past year. Where missing, the cross-sectionally de-meaned (specification 1) or cross-sectionally then individually de-meaned values (specification 2) of lagged behaviors are imputed as zeros. The account characteristics C_i include initial log account value, initial log number of equity positions, rural/urban account address dummies, and the time-series average income and recent literacy rate of the state in which the account is from, as well as the cohort level means of each of these account-level characteristics. Panel regressions are run using weights that account for sampling probability and further apply equal weight to each cross-section (month). Standard errors in ( ) are computed from bootstraps of monthly data. 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. Incremental R-squared is the ratio of the variance of the fitted age effects to the variance of the dependent variable.

Dependent Variable: Account Monthly Return in Excess of Risk-Free Rate (bp) (Mean: 110.2bp)

		Specification 1:			Specification 2:		
		[a]	[b]	[c]	[a]	[b]	[c]
	Account Age	**14.34**		*13.00*	**11.02**		**11.32**
		(7.11)		*(7.27)*	*(4.77)*		*(5.06)*
	Account Age$^{1/2}$		38.73			*29.71*	
			(24.62)			*(16.37)*	
Investor Behavior	Lagged Idio. Share of Portfolio Var.			54.73			102.02
				(82.98)			*(83.62)*
	Lagged Portfolio Turnover			**-159.34**			-110.75
				(73.69)			*(75.21)*
	Lagged Disposition Bias			-2.45			-0.45
				(2.02)			*(1.97)*
Incremental R^2		0.00040	0.00031	0.00033	0.00023	0.00018	0.00025

Table 5: Decomposition of the Difference in Returns on Old and New Accounts

For the period January 2004 through January 2012, a zero-cost portfolio is formed which buys the each stock in proportion to its average weight in the oldest quintile of accounts and sells each stock in proportion to its average weight in the newest quintile of accounts. Stocks with market capitalization below 500 million Rs (approximately $10 million) are excluded during formation of the portfolio, leaving 2,677 stocks j in the sample. Columns [1] and [3] report the time-series average of coefficients, φ_{bar}, from the Fama MacBeth regression $W_{jt}=\varphi_t X_{jt}+\varepsilon_{jt}$ of portfolio weights W on the set X of cross-sectionally de-meaned stock characteristics below. Normalized rank transforms are used to measure market capitalization, book-market, prior returns (momentum), turnover, and beneficial and institutional ownership shares. In columns [2] and [4], we decompose the returns in the zero-cost portfolio. Total returns on the zero-cost portfolio are first broken into timing effects $\{\Sigma_j W_{jt}R_{jt}-\Sigma_j W_{bar,j}R_{bar,j}\}$ and selection effects $\{\Sigma_j W_{bar,j}R_{bar,j}\}$. To decompose timing and selection effects, we run Fama MacBeth regressions of returns on stock characteristics $\{R_{jt}=\psi_t X_{jt}+\eta_{jt}\}$. Selection effects are decomposed into "stock characteristic selection" $\{\Sigma_j(\varphi_{bar}X_{bar,j})'(\psi_{bar}X_{bar,j})\}$ and "additional stock selection" $\{\Sigma_j\varepsilon_{bar,j}\eta_{bar,j}\}$ effects. We further decompose the "stock characteristic selection" effect into components attributed to marginal returns associated with each stock characteristic c $\{\Sigma_j\psi_{bar,c}x_{bar,c,j}(\varphi_{bar}X_{bar,j})\}$. Timing effects are decomposed into "stock characteristic timing" $\{\Sigma_j[(\varphi_t X_{jt})'(\psi_t X_{jt})-(\varphi_{bar}X_{bar,j})'(\psi_{bar}X_{bar,j})]\}$ and "additional stock timing" $\{\Sigma_j(\varepsilon_{jt}\eta_{jt}-\varepsilon_{bar,j}\eta_{bar,j})\}$, where the t-subscriped coefficients are from the cross-sectional regressions run in Fama MacBeth estimation. Standard errors given in ( ) are computed by bootstrap, with standard errors in the top half of columns [2] and [4] accounting for the uncertainty in coefficients in columns [1] and [3]. Coefficients statistically significant at the five and ten percent level are indicated by bold and italicized type respectively.

Dependent Variable:	1000 x Portfolio Weight (Old minus New) [1]	Contribution to Difference in Returns (bp/mo) [2]	1000 x Portfolio Weight (Old minus New) [3]	Contribution to Difference in Returns (bp/mo) [4]
Market beta	-0.603	0.79	-0.656	0.78
	(0.420)	*(1.99)*	*(0.420)*	*(2.04)*
Market capitalization	**-0.555**	-0.14	-0.230	-2.60
	(0.274)	*(2.22)*	*(0.223)*	*(2.33)*
Book-market	**0.337**	**3.92**	**0.238**	**3.24**
	(0.100)	*(1.65)*	*(0.113)*	*(1.57)*
Momentum (t-2:t-12 returns)	0.077	**2.24**	0.046	*1.62*
	(0.170)	*(1.00)*	*(0.158)*	*(0.92)*
Stock turnover	**-0.972**	**10.29**	**-0.988**	**8.19**
	(0.186)	*(2.30)*	*(0.186)*	*(1.99)*
Beneficial ownership	**-0.673**	1.33	**-0.639**	1.17
	(0.248)	*(3.79)*	*(0.218)*	*(3.66)*
Institutional ownership	**0.769**	2.01	**0.770**	2.05
	(0.228)	*(3.45)*	*(0.226)*	*(3.47)*
Ln(1+stock age)	**0.493**	1.77	-0.114	2.24
	(0.121)	*(4.53)*	*(0.080)*	*(3.16)*
Large IPOs (market cap if age<1 year)			**-12.690**	-0.53
			(3.108)	*(1.95)*
Stock characteristic selection		**22.20**		**16.17**
		(6.78)		*(6.70)*
Additional stock selection		6.78		6.57
		(13.04)		*(14.11)*
Stock characteristic timing		-1.82		-1.80
		(11.72)		*(16.50)*
Additional stock timing		11.52		17.74
		(18.37)		*(19.72)*
Total difference in old and new account returns		38.67		38.67
		(26.94)		*(26.94)*

Table 6: Performance Evaluation of the Difference in Returns on Old and Young Account Quintiles

The zero-cost portfolio evaluated below is the difference in the representative portfolio held by the oldest quintile of accounts and the representative portfolio held by the youngest quintile of accounts, and is the same as that used in Table 5. 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 [1]	CAPM [2]	Four Factor [3]	Six Factor [4]
Monthly Alpha		*0.39%*	**0.54%**	**0.64%**	**0.93%**
		(0.21%)	*(0.22%)*	*(0.19%)*	*(0.27%)*
Factor Loadings					
	Market Beta		**-0.15**	*-0.11*	*-0.11*
			(0.04)	*(0.06)*	*(0.06)*
	SMB			0.01	0.03
				(0.03)	*(0.03)*
	HML			-0.08	*-0.14*
				(0.08)	*(0.08)*
	UMD			0.05	0.06
				(0.05)	*(0.05)*
	Short Term Reversals				*-0.12*
					(0.07)
	Illiq (Based on Turnover)				-0.09
					(0.09)

Table 7: Account Age Effects in Individuals' Equity Investing Behavior

Results are constructed from the subsample of data used in Table 4 where the equity investing behaviors are measurable. About 3.3 million account-months are used in the idiosyncratic variance share and turnover regressions, and about 400 thousand in the disposition bias regressions (disposition bias is only defined for account months in which there are both gains and losses, and trading occurs). Specification [1] is $(Y_{it}-Y_t)=\beta(A_{it}-A_t)+\lambda(V_{it}-V_t)+s_i+\varepsilon_{it}$, and specification [2] is $Y_{it}=\delta_t+\beta A_{it}+\lambda V_{it}+\theta C_i+\varepsilon_{it}$, where Y_{it} represents the indicated behavior of investor i in month t and V is log account value from the end of the previous month (winsorized below at 10,000Rs or about $200). See Table 4 for definitions of other terms. Panel regressions are run using weights that account for sampling probability and further apply equal weight to each cross-section (month). Standard errors in ( ) are computed from bootstraps of monthly data. 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. Incremental R-squared is the ratio of the variance of the fitted age effects to the variance of the dependent variable.

Dependent Variable:	Idiosyncratic Share of Portfolio Variance (%)		Monthly Turnover (%)		Disposition Bias - ln(PGR/PLR) x 100	
Mean:	44.72%		5.06%		124.18	
Specification:	[1]	[2]	[1]	[2]	[1]	[2]
Effect						
Age	**0.57**	-0.08	**-1.31**	**-1.39**	**-9.33**	**-6.05**
	(0.08)	*(0.05)*	*(0.09)*	*(0.09)*	*(1.07)*	*(0.68)*
Log(Account Value)	**-5.58**	**-6.18**	**0.80**	**0.60**	-1.48	**-4.42**
	(0.09)	*(0.08)*	*(0.07)*	*(0.04)*	*(1.52)*	*(1.27)*
Incremental R^2	0.0027	0.0001	0.0250	0.0282	0.0051	0.0021
[B] Age Effect=$Age^{1/2}$						
Age	*0.36*	**-1.09**	**-5.59**	**-5.21**	**-29.94**	**-19.05**
	(0.22)	*(0.18)*	*(0.27)*	*(0.24)*	*(2.97)*	*(2.17)*
Log(Account Value)	**-5.40**	**-6.12**	**1.08**	**0.69**	-0.76	**-4.14**
	(0.10)	*(0.08)*	*(0.07)*	*(0.04)*	*(1.52)*	*(1.26)*
Incremental R^2	0.0000	0.0000	0.0349	0.0346	0.0056	0.0023

Table 8: Response of Individual Investor Behavior to Feedback

Results are constructed from the random sample used in Table 4. The specification tested is $(Y_{it}-Y_t)=\beta(A_{it}-A_t)+\lambda(V_{it}-V_t)+\eta(F_{it}-F_t)+s_i+\varepsilon_{it}$. The s_i are account fixed effects and the terms Y, A, V, and F are cross-sectionally de-meaned account behavior, square root of account age, log account value (winsorized below at 10,000Rs or about $200), and the feedback measures used below. The increase in returns due to trades for a given month is computed as the difference between actual returns in the current month and the returns that would have obtained if no trades had been made in the past three months. The cumulative value of this measure is used below. The cumulative increase in returns due to selling off gains versus losses is computed by comparing the three-month returns following past sales with market returns over that period, with each gain and loss weighted in proportion to the value of the sale relative to the investor's stock portfolio and the outperformance of gains counting negatively in the measure. All feedback measures are defined such that positive coefficients indicate that the feedback reinforces in the given behavior. Panel regressions use weights that account for sampling probability and further apply equal weight to each cross-section (month). Standard errors in ( ) are computed from bootstraps of monthly data. 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.

[A] Dependent Variable: Idiosyncratic Share of

		[1]	[2]
Feedback Measures	Cumulative outperformance relative to the market	**4.15**	**3.92**
		(0.32)	*(0.32)*
	Size of worst monthly stock portfolio return experienced		**-13.23**
			(1.55)
Age Effects: (Account Age)$^{1/2}$		Y	Y
Log(Account Value)		Y	Y

[B] Dependent Variable: Monthly Turnover

		[1]	[2]
Feedback Measures	Cumulative increase in returns due to trades	**3.81**	**3.20**
		(0.39)	*(0.37)*
	Cumulative outperformance relative to the market		**0.75**
			(0.11)
	Size of worst monthly stock portfolio return experienced		**-14.41**
			(1.20)
Age Effects: (Account Age)$^{1/2}$		Y	Y
Log(Account Value)		Y	Y

[C] Dependent Variable: Disposition Bias - ln(PGR/PLR) x 100 (Mean=35.52)

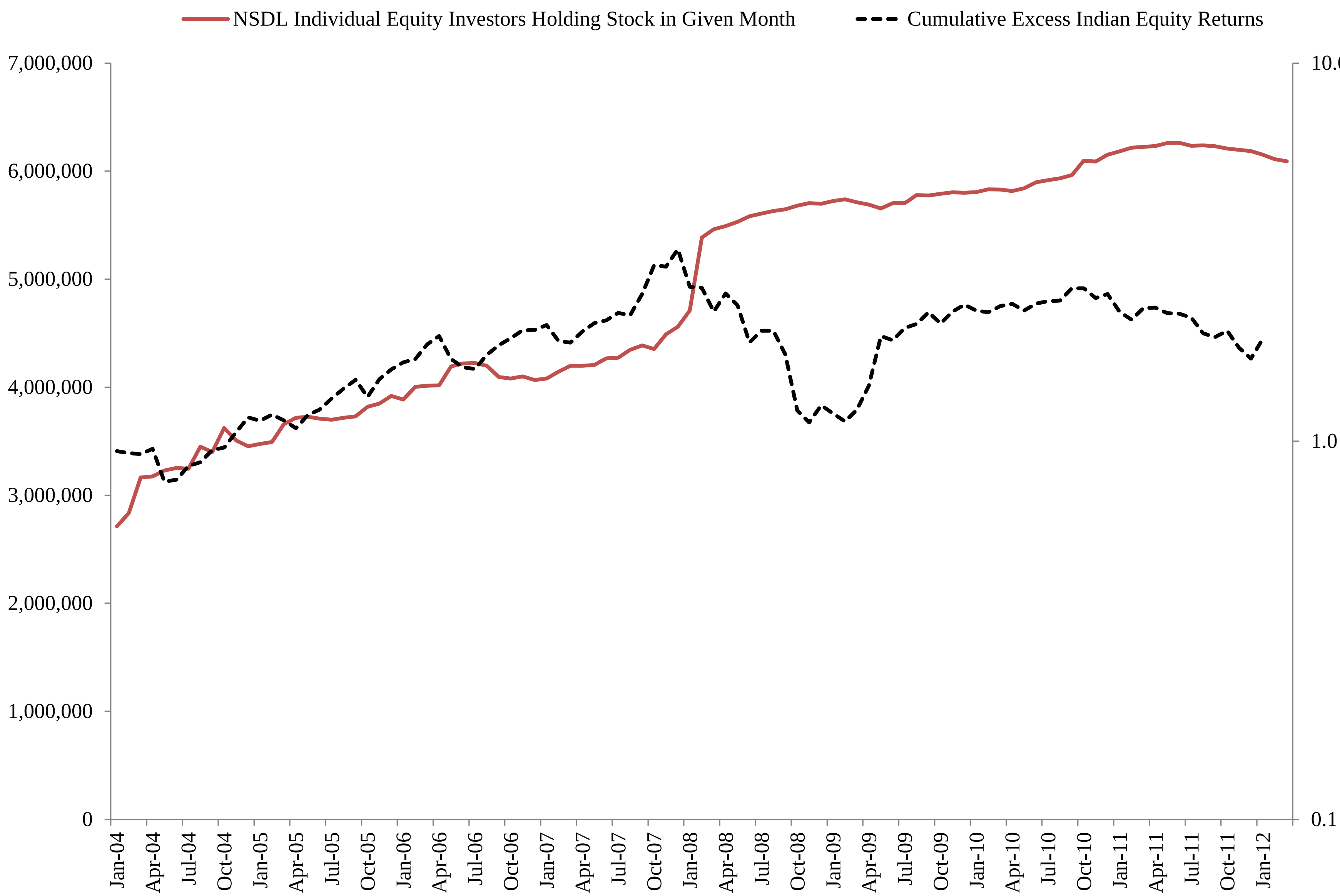
		[1]	[2]
Feedback Measures	Cumulative increase in returns due to selling off gains versus losses	6.46	8.45
		(6.48)	*(6.56)*
	Cumulative outperformance relative to the market		**14.21**
			(3.44)
	Size of worst monthly stock portfolio return experienced		-17.63
			(17.57)
Age Effects: (Account Age)$^{1/2}$		Y	Y
Log(Account Value)		Y	Y

Table 9: Predicting Indian Stock Returns Using Characteristics of Investors

The dependent variable is monthly stock returns from January 2004 through September 2011 for each of 3,614 stocks with at least 10 individual investors from our sample individual accounts. Stockholder account age is the average account age of investors in the stock in the given month. For behavioral characteristics of stockholders, we similarly use the average behavior of across individual investors, where the behavior from a given individual investor is taken as the cumulative average of a cross-sectionally de-meaned measure of the behavior (idiosyncratic share of portfolio variance, monthly turnover, or ln(PGR/PLR)). Average investor account age and behavior measures, as well as market capitalization, book-market, momentum, turnover, and beneficial and institutional ownership share measures are converted to normalized rank form. The regressions below are carried out by the Fama MacBeth procedure, and a serial correlation adjustment (Newey West, 3 monthly lags) is applied. All coefficients are multiplied by 100 for readability, and statistical significant at the five and ten percent level are indicated by bold and italicized type respectively.

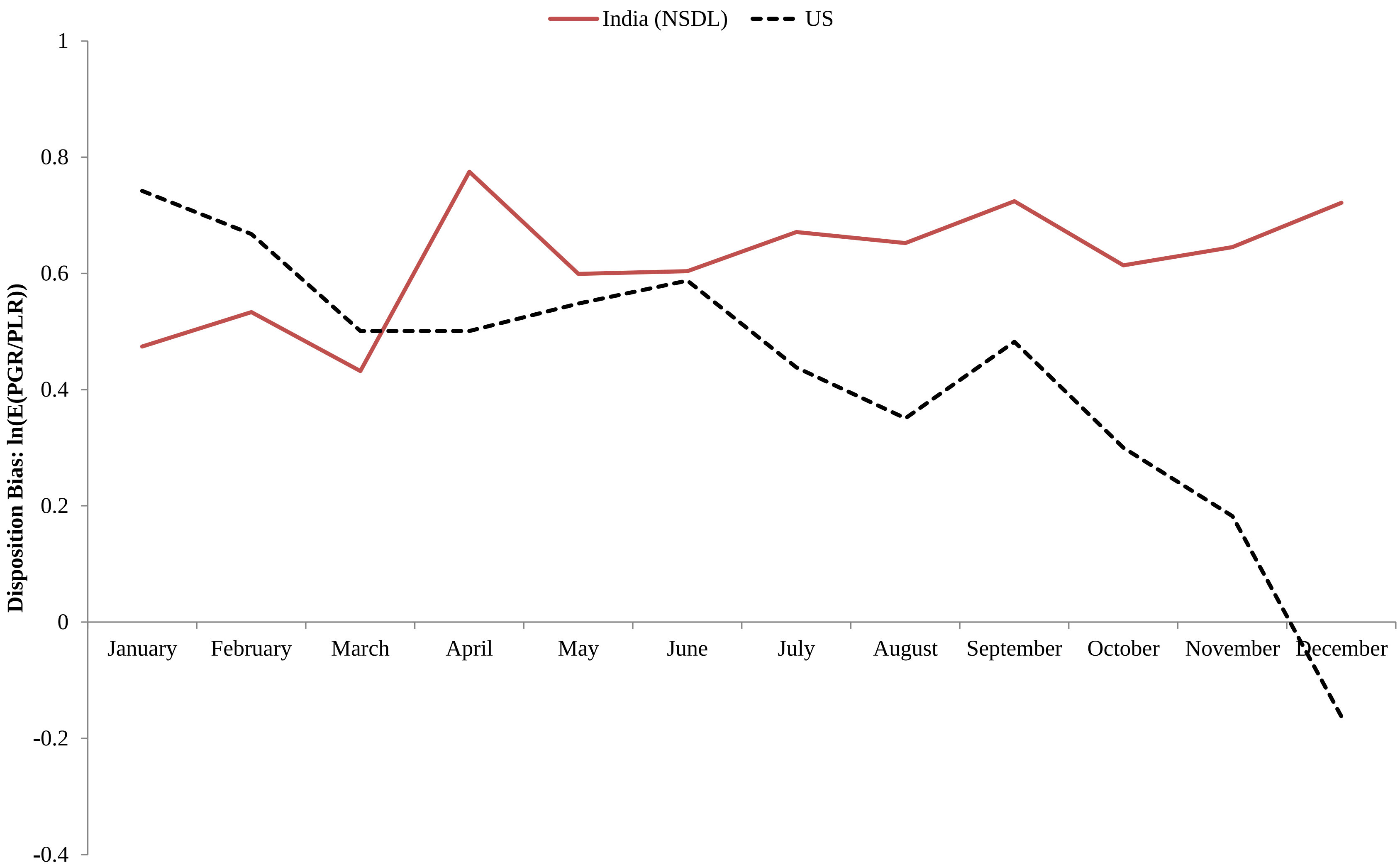
		[1]	[2]	[3]
Investor Characteristics	Account Age	**1.85**	**1.21**	0.13
		(0.55)	*(0.57)*	*(0.26)*
	Idio. Share of Portfolio Var.		0.93	**-0.63**
			(0.82)	*(0.29)*
	Portfolio Turnover		**-1.75**	**-0.89**
			(0.51)	*(0.32)*
	Disposition Bias		-0.11	-0.26
			(0.48)	*(0.32)*
Stock Characteristics	Market beta			0.22
				(1.22)
	Market capitalization			-1.51
				(1.56)
	Book-market			**3.84**
				(0.64)
	Momentum			**3.20**
				(0.63)
	Stock turnover			**-1.52**
				(0.39)
	Beneficial ownership			-0.67
				(0.42)
	Institutional ownership			-0.75
				(0.63)
	Ln(1+stock age)			0.06
				(0.11)

Figure 1: Individual Equity Investors and Cumulative Excess Indian Equity Returns



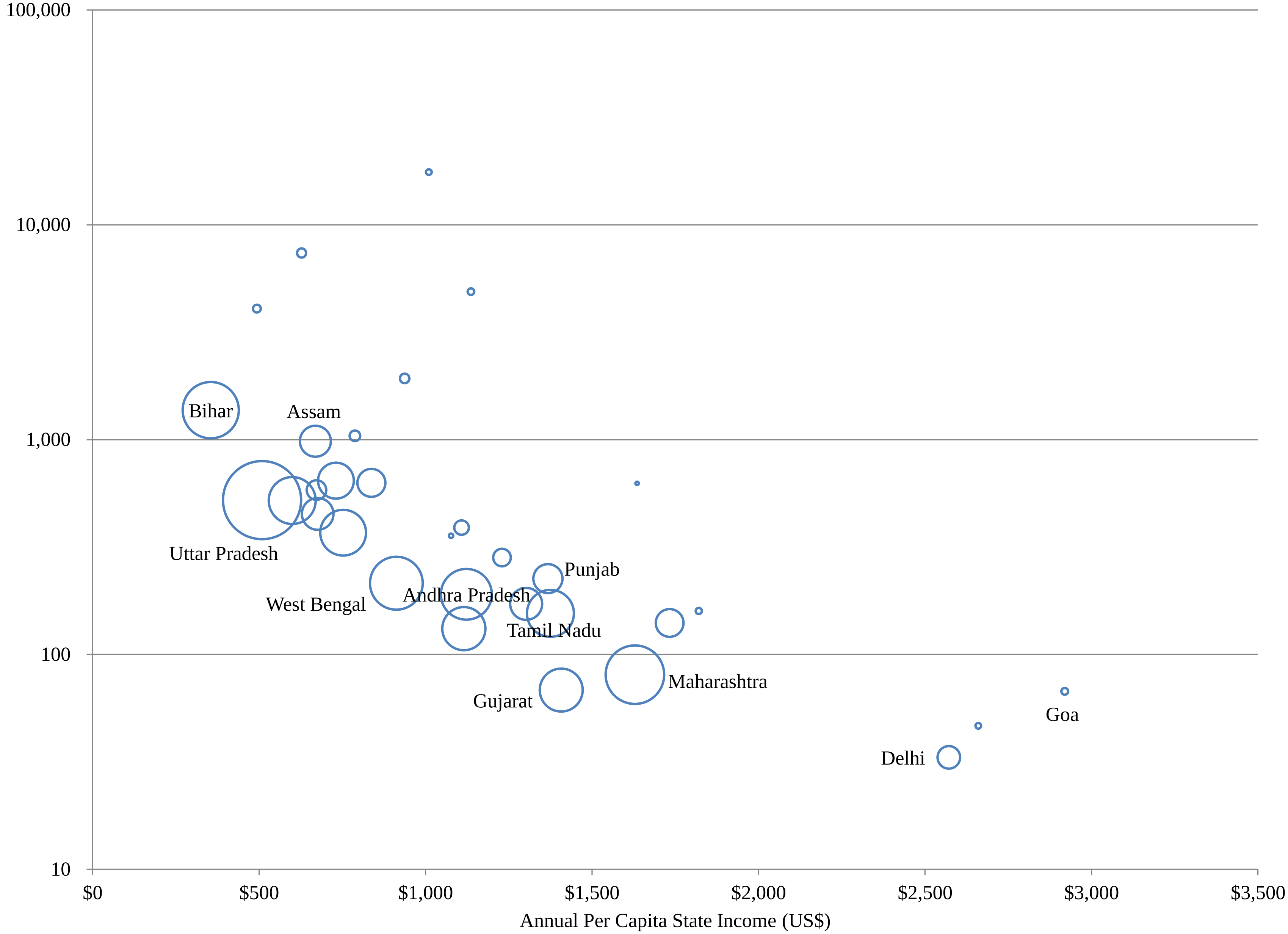
Equity investors are defined by the aggregation of accounts by Permanent Account Number (PAN) which uniquely identify individuals. Excess Indian equity returns are computed using the yield on three-month Indian Treasury bills, and total returns and market capitalization of all Indian stocks for which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

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



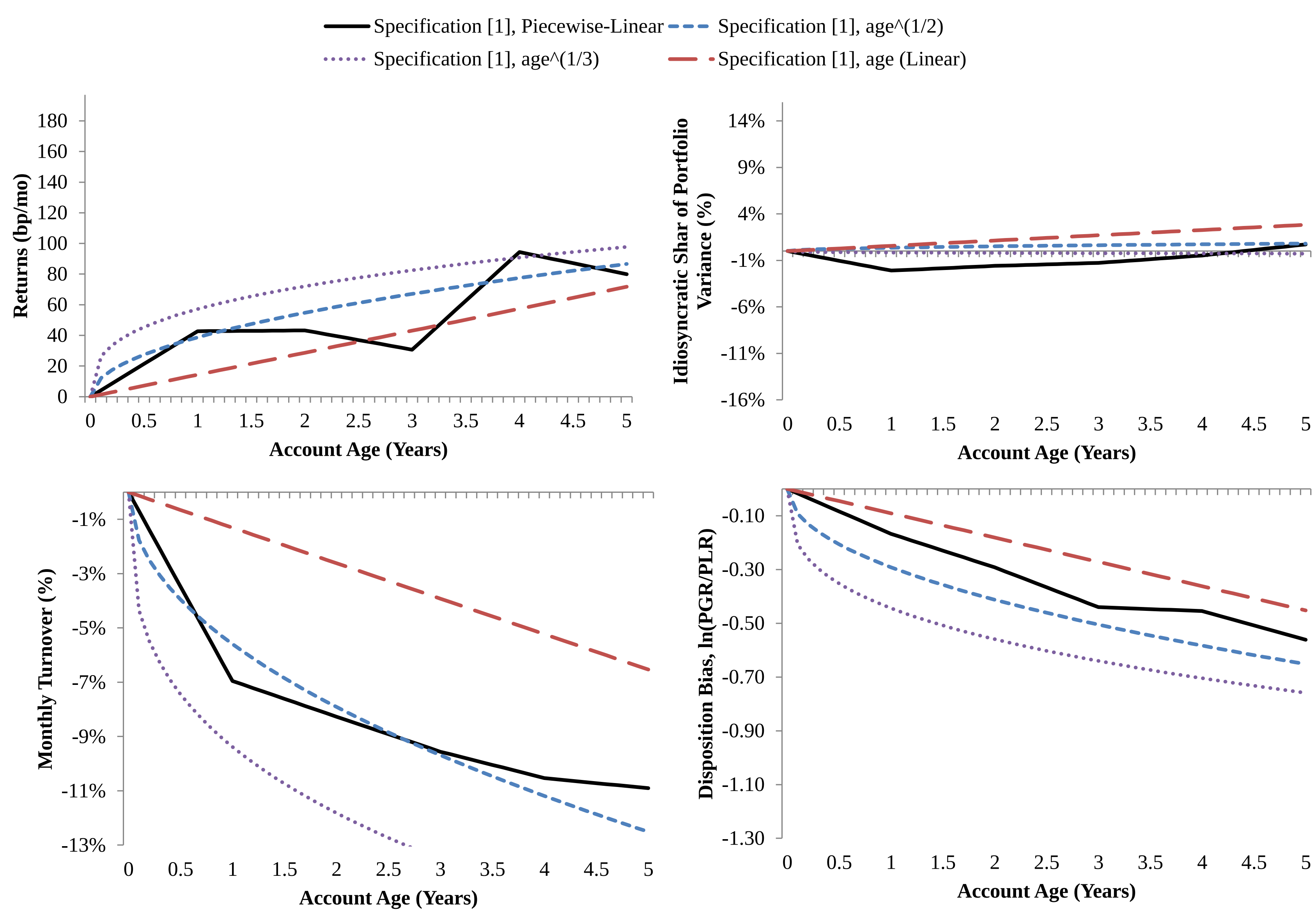
PGR/PLR is computed for each month, aggregating across accounts and years as in Odean (1998), which is the source for the US brokerage based statistics plotted. We take the log of both measures to make the units somewhat more comparable to our cross-sectional analysis. Levels are not quite comparable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

Figure 3: State Population per Individual NSDL Account by Per Capita State Income (2011)



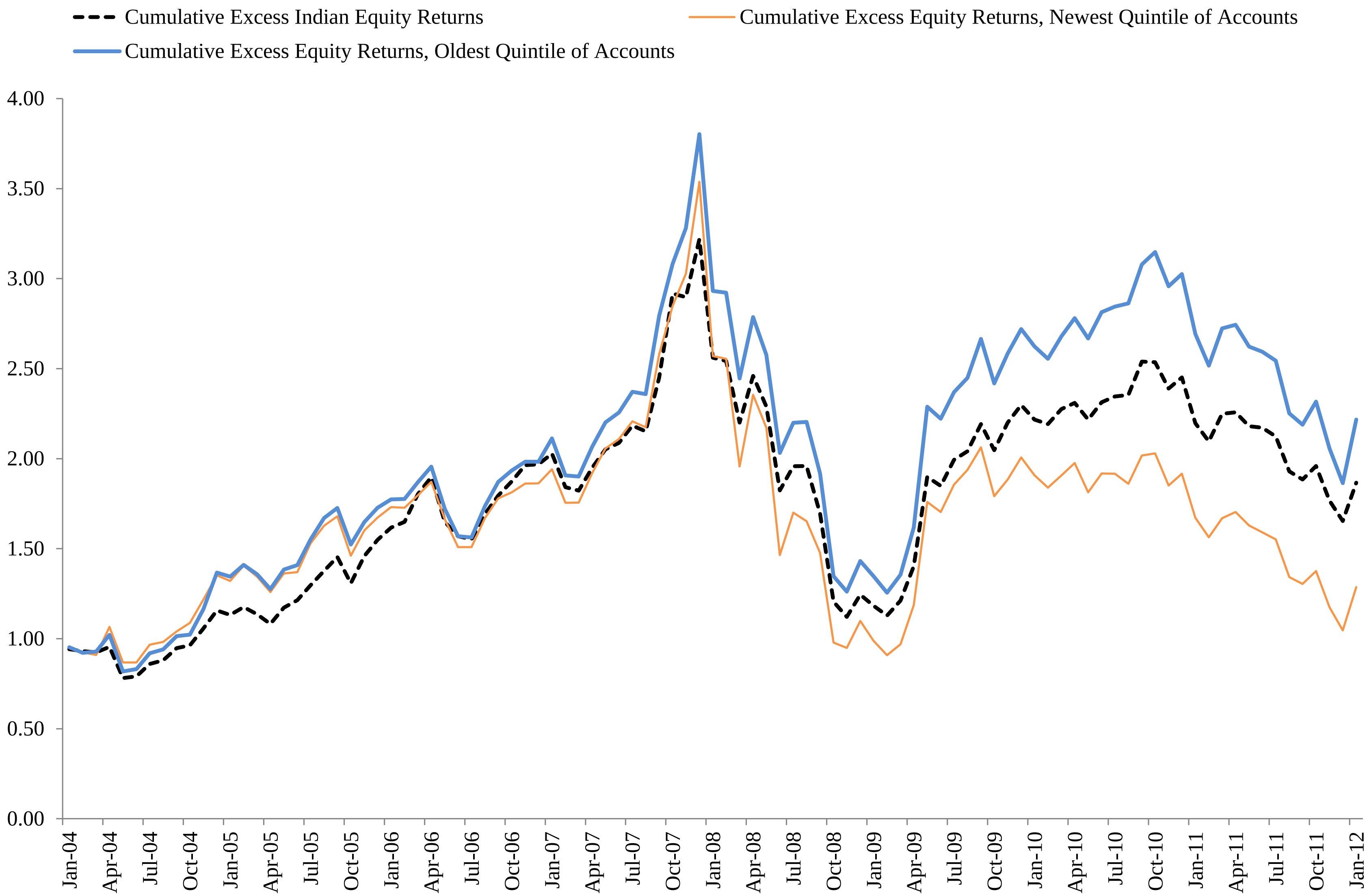
The size of each bubble is proportional to the state population in 2011 (Indian Census). State per capita income data is as of March 2011 from the Reserve Bank of India.

Figure 4: Account Age Effects in Equity Portfolio Returns and Investing Behavior



The blue and red dashed curves represent the age effects estimated in Tables 4 and 7 under Specification [1] (individual fixed effects). The purple dashed curve represents estimated age effects when the impact of age on the given behavior is a function of age^(1/3). Similarly, the solid black line is produced by specifying age effects as a piecewise-linear function with breakpoints every year (out to five years). The vertical axis in each plot is scaled to a range equal to twice the cross-sectional standard deviation of the measured behavior.

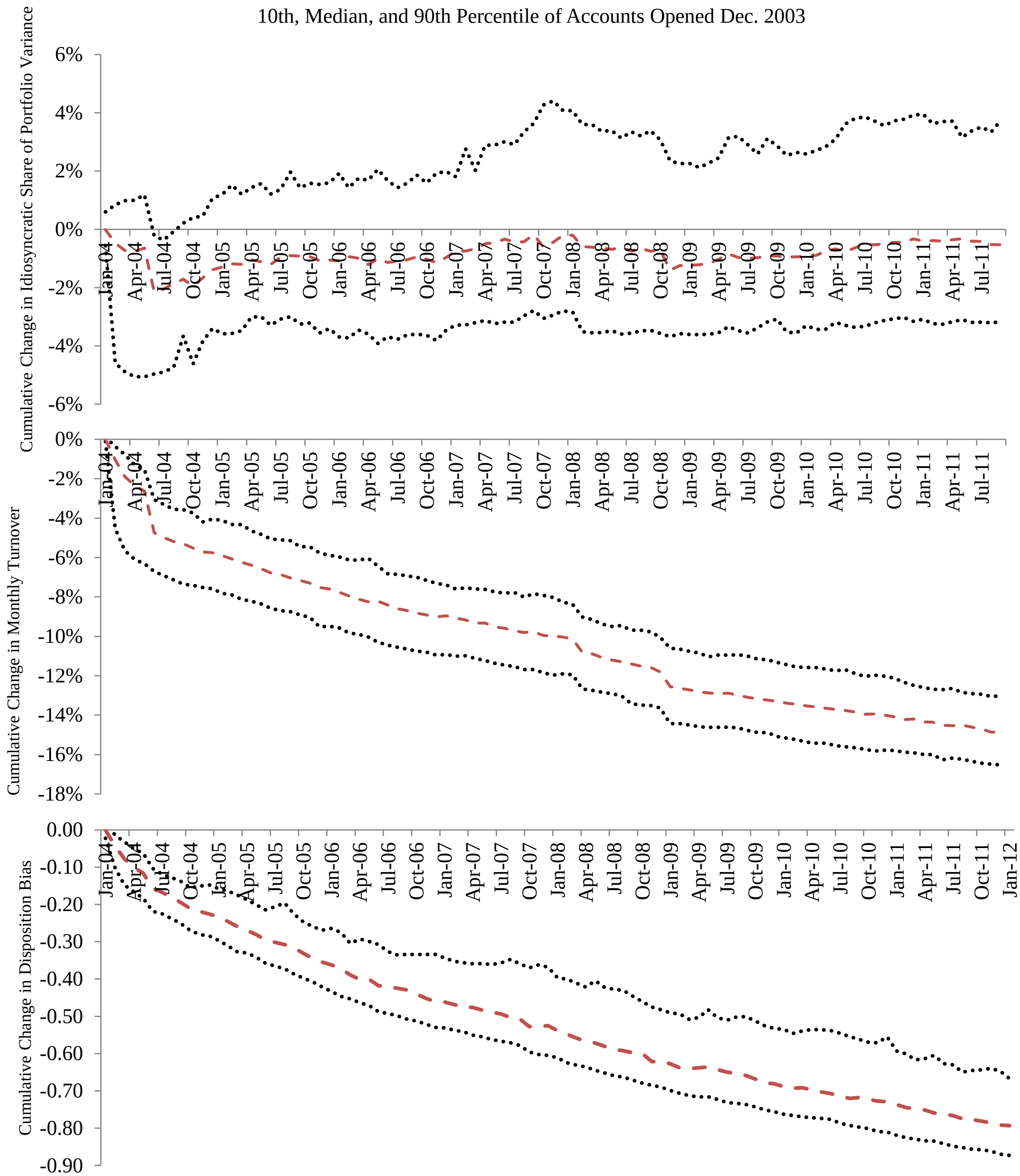
Figure 5: Cumulative Indian Excess Equity Return Received by Oldest and Newest Quintile of Accounts



Each month, individual investors are sorted into quintiles based on their account age, and a representative portfolio of holdings of oldest and newest quintile accounts is formed by equally weighting across the actual portfolio weights of accounts in that age quintile. The plot above subtracts the yield on three-month Indian Treasury bills from returns on the portfolios formed from the oldest and newest account quintiles as well as the market portfolio.

Figure 6: Simulated Cumulative Change in Investor Behaviors from Age Effects and Feedback

10th, Median, and 90th Percentile of Accounts Opened Dec. 2003



These figures are produced using age and feedback coefficients in specifications [b] of Table 8 combined with the actual age and feedback received by individual investor accounts opened in December 2003. This feedack consists of cumulative market outperformance and worst monthly return experienced, return improvement due to trading (turnover plot), and return improvement due to selling gains versus losses (disposition bias plot).