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

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

December 17, 2013

Abstract

Using a large representative sample of Indian retail equity investors, many of them new to the stock market, we show that both feedback from investment returns and years of investment experience have a significant effect on investor behavior, style tilts, and performance. We identify two channels of feedback: overall performance relative to the market, and feedback from the impact on performance of specific behavior and style tilts. Consistent with models of reinforcement learning, feedback has strong predictive ability for future behavior and style tilts. We show that experienced investors have lower portfolio turnover, exhibit a smaller disposition effect, and invest more heavily in value stocks than novice investors, although these behaviors do not fully explain their better performance. We also find that Indian stocks held by experienced investors, or investors whose strategies resemble those of experienced investors (with low turnover and a value tilt), deliver abnormal returns even controlling for standard stock-level characteristics.

*We gratefully acknowledge NSDL and SEBI for providing us with access to the data, the Sloan Foundation for financial support, and the use of the University of Oxford Advanced Research Computing (ARC) facility. 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, EDHEC, the Oxford-Man Institute, the IGIDR Emerging Markets Finance Conference, and the Sloan-Russell Sage Working Group on Behavioral Economics and Consumer Finance for many useful comments, and Vimal Balasubramaniam and Sushant Vale for excellent and dedicated research assistance. The opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the Board of Governors or other employees of the Federal Reserve System.

†Campbell: Department of Economics, Littauer Center, Harvard University, Cambridge MA 02138, USA, and NBER. 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. tarun.ramadorai@sbs.ox.ac.uk. Ranish: Federal Reserve Board, Washington DC 20006, USA. ben.ranish@frb.gov.

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, Gomes and Michaelides 2008).

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 investment behaviors that appear to be quite prevalent. 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. This has several important implications. Most obviously, 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. Finally, underdiversification of directly held Indian stock

[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 market where fundamentals determine risk prices 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.

portfolios creates a rich cross-section of investment experiences that we can use to identify investors' responses to their experiences. This source of identification is considerably stronger than the variation in experience across cohorts exploited by Malmendier and Nagel (2011, 2012).

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 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 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, Betermier et al. 2013), 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 the behavior and investment performance of each account to account age (the length of time since an account was opened) and summary statistics about past account behavior and performance.

Both age since account opening and feedback from past investment performance have large and statistically significant effects on a range of investing behaviors, and ultimately, on the returns experienced by Indian retail investors. We measure this feedback both using past account performance relative to the market, and in a more targeted fashion for each behavior and style tilt that we study. For example, each month our investor receives new feedback regarding the desirability of portfolio turnover from comparing the total returns garnered by their trades over the past three months to the buy-and-hold return they would have earned instead.

We find that both turnover and the disposition effect decline substantially with (account) age. Over the first eight years of investing experience, controlling for feedback, we find that turnover falls by 84% and the disposition effect by 52% relative to their means. However there is little effect of age on the idiosyncratic variance of retail investors' portfolios. We find that behavior-specific feedback has a strong impact on idiosyncratic variance share and turnover, and a weaker impact on the disposition effect. Overall account performance appears to encourage aggressive investment behavior: less diversification, heavier trading, and a greater disposition effect.

Investors' style investing behavior is also affected by both overall time spent in the market and feedback. We find a strong tendency for more experienced investors to accumulate value stocks, a more modest tendency for investors to accumulate greater quantities of small stocks as they become more experienced, and a small and statistically insignificant effect on the accumulation of momentum stocks as investors age.

Feedback also has important effects on the style tilts of Indian retail investors. "Style chasing" behavior is a feature of a number of different theoretical models, most notably that of Barberis and Shleifer (2003), although sharp evidence on the effects of directly experienced feedback on investing behavior is scarce. In the short term, we find that investors decumulate outperforming styles in a manner consistent with the disposition effect. However, we find a less precisely estimated but far longer-lived tendency to accumulate or "chase" styles in which the investor has experienced positive returns. We also find that investors who achieve high overall returns tend to accumulate large, growth, and high momentum stocks.

Account age and feedback ultimately affect retail investor performance both directly, and indirectly through their impacts on behavior and style tilts. Experienced investors appear to have substantially higher returns than novice investors, although this result is imprecisely estimated. Investors who behave like experienced investors—those with low turnover and a low disposition effect—perform better with and without controls for standard risk factors. Experienced Indian investors tend to tilt their portfolios towards the type of stocks which have higher returns: small stocks, stocks with low turnover, and stocks held by institutions. Experienced Indian investors are also more likely to avoid large, attention-grabbing initial public offerings.

The effects of investor experience can also be detected in the cross-section of Indian stock returns. Controlling for stock characteristics and factor loadings, Indian stocks held by older investors, those with low portfolio turnover, and those with portfolios tilted towards value stocks have significant returns unexplained by standard factors and characteristics.

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 presents our methodology for estimating age and feedback effects. Section 4 applies this methodology to estimate age and feedback effects on behavior, while Section 5 applies it to age and feedback effects on style tilts. Section 5 draws implications for individuals' account performance given their behavior and the performance of stocks given behavior of their investor base. Section 6 concludes.

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 has been hampered by the difficulty of obtaining detailed data on individual investors' portfolios and the computational burden imposed by the study of such large datasets. These constraints have gradually been surmounted, and this field is now 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) used 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 stocks, 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 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, and a recent study by Betermier et al. (2013) examines the value tilt of Swedish investors, finding a tendency for this tilt to increase over the life cycle. However the annual frequency of the Swedish 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

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

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 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.

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. In a similar spirit our empirical specification allows us to compare the short- and long-run effects of investment performance

[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.

feedback on household investment decisions.

2 Data and Summary Statistics

2.1 Electronic stock ownership records

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

While 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.[4] This paper studies only the category of individual accounts.

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

Table 1 summarizes the coverage of the NSDL dataset. 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 12,350 in 2004 to almost 23,000 in 2011, as does the number of unique Indian equities covered. Starting at 4,533 in 2004, the number of equities reaches a peak of 7,735 in 2012. When we match these data to price, returns, and corporate finance information from various datasets, we are able to match between 96% 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 just above 50% at the beginning

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

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.[5] 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.

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

2.2 Characteristics of individual accounts

Table 2 describes some basic characteristics of Indian individual accounts. Because this dataset is an unbalanced panel, with accounts entering and exiting over time, we summarize it in two ways. The first set of three columns reports time-series moments of cross-sectional means. The first column is the time-series mean of the cross-sectional means, which gives equal weight to each month regardless of the number of accounts active in that month. The second and third columns are the time-series maximum and minimum of the cross-sectional mean, showing the extreme extent of time-variation in cross-sectional average account behavior.

The second set of three columns reports cross-sectional moments of time-series means calculated for each account over its active life, giving equal weight to each account for which the given characteristic can be measured in 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 (described more fully in the next section) of individual accounts opened after January 2002. For accounts which opened earlier, which represent about 14.4% of all individual accounts, we do not observe the full investing history, do not know when the account first invested in stocks, and do not observe the initial account characteristics. In our appendix, we show that our results are qualitatively unchanged when we perform our analyses with all individual accounts after making assumptions required to make use of the additional older accounts.

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 $5,000 in 2004 to about $66,000 in June 2008, with a time-series mean of $24,771. The median account size is however much smaller at $1,327, and even the 90th percentile account size is only \$10,815, 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 14.3 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.[6]

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. 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.[7]

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

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

[7]The online appendix, Campbell, Ramadorai, and Ranish (2013), provides details on our procedures for calculating Indian stock returns. The appendix also shows that our results are robust to consideration of "active" returns from intra-month trading.

underperforming in their worst months by as much as 4.8% or overperforming in their best months by as much as 10.2%. 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 1.75% per month and the 90th percentile account outperforming by 1.52% 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 years from their initial listing. To reduce noise in estimated idiosyncratic risk, we estimate idiosyncratic variance from a GARCH(1,1) model.[8]

The average idiosyncratic share is about 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 25% to 55%) and across accounts (from 24% at the 10th percentile to 68% 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

[8]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.

between the number of stocks held in a portfolio and the portfolio's idiosyncratic risk.

Turnover

Turnover is estimated by averaging sales turnover and purchase turnover. Sales turnover equals the value of last month's holdings (at last month's prices) that were sold in the current month divided by the geometric average of the value of last month's holdings and the current month's holdings. This value is winsorized at 100%. Purchase turnover equals the value of the current month's holdings (at current prices) that were bought in the current month, divided by the same denominator and winsorized in the same manner. This measure of turnover is not particularly high on average for Indian individual accounts. The time-series mean of the cross-sectional mean is 5.7% per month (or about 68% per year), and the cross-sectional median turnover is only 2.6% (or 31% 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 16.3% per month (196% per year).

Following Odean (1999), we have compared the returns on stocks sold by individual Indian investors to the returns on stocks bought by the same group of investors over the four months following the purchase or sale. In India, the former exceeds the latter by 2.79%, 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 5.22% 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. By calculating a log ratio we eliminate any mechanical relation between the level of turnover and our measure of the disposition effect. To avoid extreme values of the ratio we winsorize PGR and PLR below at 0.01, and find that our results are robust to reasonable variation in the winsorization threshold.

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 35% of cases). Sales are counted only if a position is fully sold, although this convention makes little difference to the properties of the measure.

The disposition effect is important for Indian individual accounts. On average across months, the cross-sectional mean proportion of gains realized is 1.23 log points or 242% larger than the proportion of losses realized, while the median account has a PGR that is 1.35 log points or 286% 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.

In the online appendix to this paper, Campbell, Ramadorai, and Ranish (2013), we compare the disposition effect in our Indian data with US results reported by Odean (1998). Specifically, we plot 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, and 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 1.03 in both the time-series and cross-sectional moments. The cross-sectional mean betas have modest variation over time from 0.95 to 1.09, 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 3.5% greater than the market index. This value tilt varies over time and does switch sign, reaching around -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 -6%. 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

In the online appendix we ask 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 these correlations. 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 0.33 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 -0.47 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 The Effects of Age and Feedback: Methodology

In this section we describe our approach to understanding how Indian investors learn. We consider two potential sources of learning. The first is the amount of time spent by an investor in the market, which we proxy by the time elapsed since the investor opened an account and held stock. The second is the set of specific experiences that each investor has had in the market. Our econometric specifications identify the first of these sources as an age effect in the panel of investors, and the second using the fact that variation in portfolio holdings across investors and time generates variation in their "experienced returns." We use these specifications to explain and forecast the evolution of investment behaviors, investor style tilts, and overall investor performance.

Our empirical specifications are constructed to capture variation in behaviors, style tilts, and performance arising from age and feedback. However, to ensure that we correctly identify these effects, we need to control for the effect of broader temporal fluctuations in the Indian market, as well as the possibility that individuals differ in their levels of inherent ability or sophistication.

To begin, consider a specification which provides for estimation of all these sources of potential variation in behavior, style tilts, and performance, represented generically as an outcome Y_{it} below:

$$Y_{it} = s_i + \delta_t + \beta A_{it} + \gamma X_{it} + \varepsilon_{it,} \tag{1}$$

where s_i is an investor fixed effect, δ_t represents an unobserved time fixed effect, A_{it} is a measure of the age of account i at time t, and X_{it} is a predictor variable such as the feedback experienced by investor i at time t. In the case where Y_{it} measures investment performance, we might think of s_i as capturing the inherent sophistication or investment ability of investor i.

We can re-write equation (1) in cross-sectionally demeaned form as:

$$Y_{it} - Y_t = (s_i - s_t) + \beta(A_{it} - A_t) + \gamma(X_{it} - X_t) + \varepsilon_{it,} \tag{2}$$

where s_t is the cross-sectional average fixed effect of investors in the market at time t. As investors enter and exit the market, s_t varies over time.

The fatal drawback of equations (1) and (2) is that they are not identified on account of perfect collinearity. This is the usual problem with any specification containing a linear transformation of unrestricted age effects, unrestricted cohort or individual effects, and unrestricted time effects (Ameriks and Zeldes 2004, Guiso and Sodini 2013).

To estimate the objects of interest (β and γ), we therefore apply restrictions on s_t in equation (2). The simplest such restriction that we employ is that $s_t = 0$, which in the case of a performance regression implies that the average inherent sophistication of investors in the market does not change with time. Applying this restriction, we arrive at our baseline specification:

$$Y_{it} - Y_t = s_i + \beta(A_{it} - A_t) + \gamma(X_{it} - X_t) + \varepsilon_{it}. \tag{3}$$

Our baseline specification is vulnerable to several econometric difficulties. First, it is possible that the average inherent sophistication of Indian investors has been changing over time as market participation expands. To address this possibility, in the online appendix (Campbell, Ramadorai, and Ranish 2013), we model these changes using the cross-sectional average of a set of investor attributes, i.e., by estimating:

$$Y_{it} - Y_t = (s_i - \alpha C_t) + \beta(A_{it} - A_t) + \gamma(X_{it} - X_t) + \varepsilon_{it} \tag{4}$$

where C_t includes the cross-sectional average of investor initial account value and investor initial number of equity positions, as well as the income and literacy rates of the states in which investors are located, and the share of the investor population residing in rural and urban areas. Put differently, specification (4) simply attempts to fit cross-sectional average sophistication with the set C_t of cross-sectional average investor attributes. We find that the majority of our results are unaffected by the introduction of these variables C_t, despite the fact that their introduction weakens our identification of age effects.

Second, it is well known that panel estimation with fixed effects can deliver biased estimates when explanatory variables are not strictly exogenous. Intuitively, if the time dimension of the panel is short, and if high values of Y_i early in the sample predict high future values of X_i, then relative to its sample mean Y_i must be low later in the sample, and will spuriously appear to be negatively predicted by X_i. This is a particular problem if we use account size as an explanatory variable to predict returns, since account size is mechanically driven by past returns. Similar issues may arise when we use investment behaviors or style tilts as explanatory variables, if their prevalence is behaviorally influenced by past returns.

Even the use of account age as an explanatory variable may suffer from this problem if the disposition effect – the tendency of investors to sell gains rather than losses – leads to disproportionate exit of investors who have been lucky (Calvet et al. 2009a). In this case, experienced investors may disproportionately be investors who had poor returns when they were novices. In the presence of investor fixed effects, this can produce an upward bias in the estimated effect of account age on portfolio returns.

Fortunately this problem is less serious in our application than in many panel estimation exercises, because our panel has a relatively long time dimension. Furthermore, in the online appendix we respond to the problem by estimating an alternative specification:

$$Y_{it} - Y_t = \theta(C_i - C_t) + \beta(A_{it} - A_t) + \gamma(X_{it} - X_t) + \varepsilon_{it} \tag{5}$$

This specification restricts the individual fixed effects used in (3), modeling them using the same set of investor attributes C described above. By eliminating the use of sample mean Y_i to estimate fixed effects, the specification protects against the bias discussed above. The online appendix shows that our inferences about the impact of variables (such as feedback variables) for which we might be concerned about bias arising from violations of strict exogeneity are unaffected in this new specification.

With regard to the specific issue of luck-driven account exit, in the online appendix we model the relationship of account exit to past returns and use this to simulate the survival bias in account age effects using our primary specification. We find that the account exit rate is too modest, and too weakly related to past returns for our inferences to be affected significantly.

Our baseline regressions of investment behaviors also include lagged behavior, Y_{it-1}, as a regressor and our figures and tables come from specifications which incorporate these regressors. The inclusion of lagged outcome variables makes the model one in which there is partial adjustment to an age-dependent target, although in the current version of the paper, we do not calculate the effect of age or other right-hand-side variables on the target, but simply report the coefficients from the estimated specifications directly, meaning that their effects are interpretable as conditional on past levels of the outcome variables.

When estimating our specifications, we focus primarily on a flexible piecewise linear form for the account age effect. In addition, we consider linear age effects, $A_{it} = Age_{it}$, in our account return regressions. This represents a tradeoff of generality for an improvement in statistical power.

We consider two sources of feedback in our empirical estimation. The first, which we term "account performance feedback" is the historical *total* outperformance of the investor relative to the market. The coefficients on account performance feedback capture the effects on the outcome variables of interest (behavior and style tilts) of the investor performing relatively well or relatively poorly over a period of time. The second is "behavior-specific" or "style-specific" feedback. We measure this source of feedback using historical experienced returns attributable *specifically* to the past behavior or style tilt of the investor which we seek to explain. For example, when forecasting investor turnover, the turnover-specific feedback is measured as the increase in returns due to trading activity. This 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.

Our 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. The internet appendix shows that, as expected, state participation rates are highly correlated with per-capita state income. Our return regressions are estimated using about 4.2 million account months of data spanning January 2004 through January 2012, and our regressions of account behaviors and style tilts 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. This estimation methodology is in the spirit of the well-known Fama-Macbeth regression method (since it gives each time period equal weights, and assumes errors are cross-sectionally correlated within each period but uncorrelated across periods), although it differs in its inclusion of account fixed effects.

4 Age and Feedback Effects on Behavior

4.1 How behavior changes with age

We first ask whether our three proxies for investment behaviors change with the age of the account. In our specifications, we predict the idiosyncratic variance share, turnover, and disposition bias measured by the log ratio of PGR to PLR using specification (3), allowing for a piecewise linear age effect. Figure 2 shows our preferred presentation of these age effects; this presentation carries through our analysis of style tilts and performance. In the plot, the full length of each bar represents the point estimate from the regression. The solid portion of the bar lies outside the 95% confidence interval, allowing the reader to visually focus on the progression of the statistically significant coefficients across the age spectrum. We scale behaviors by the time-series average of their cross-sectional means as reported in Table 2.

The age effects documented in Figure 2 are large in economic magnitude. Over the course of five years, monthly turnover declines by a statistically significant 38% of the time-series average cross-sectional mean, with this number becoming an even larger 52% for an eight-year old account relative to a novice account. The disposition bias declines by an even larger 48% for a five-year account relative to a novice, although the internet appendix shows that a considerable portion of this effect may be attributable to the fact that early cohorts appear more sophisticated along the dimension of the disposition bias. 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 in some cases result from stock-specific information possessed by sophisticated investors – they find that individual trader performance improves as the number of stock holdings decrease, holding other determinants of performance constant.

4.2 How behavior changes with feedback

Since two of our behavior measures vary strongly with account age, it is plausible that behavior may also be affected not only by the fact of investing, but also by the experiences that investors have in the market. Figure 3 shows the impacts of feedback on the three behaviors that we consider (in rows), with the two columns showing, respectively, the impacts of past total account performance relative to the market, and past behavior-specific feedback.

In the first row, we consider the impact of the outperformance of the account relative to the market on the future idiosyncratic share of portfolio variance. The figure shows that account outperformance leads investors to make larger idiosyncratic bets, especially over the first quarter following an increase in performance. The estimated coefficient implies that the idiosyncratic variance share becomes 9% higher than the time-series average cross-sectional mean idiosyncratic share for a 100% increase in outperformance relative to the market in the previous month. This may be because past outperformance encourages investors to assess their investing skills more optimistically, in turn leading to them increasing their idiosyncratic bets.[9]

The second row of Figure 3 shows the results from predicting turnover. Again, outperformance of the account relative to the market appears to have a significant effect on turnover, with a rise of over 60% relative to mean turnover following an increase in returns of 100% relative to the market. The effect is also longer-lived than the impact on the idiosyncratic variance share, although it continues to show a relatively rapid decline following the first quarter after the elevated account outperformance.

In the same row, the second column shows the impact of the 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

[9]Note that the presence of the lagged idiosyncratic variance share in the regression controls for any mechanical impact (given less than complete rebalancing within the month) of the return to an undiversified account on the end-of-month idiosyncratic variance share of the account. That is, we measure account return during the month leading up to the measurement date for the lagged idiosyncratic variance share, not the month following that measurement date. In this way we guarantee that the effects we estimate are behavioral and do not result mechanically from imperfect rebalancing.

returns that would have been experienced if the account had stopped trading three months earlier. 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.

It should be noted that the effects of recent account performance and 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. However, the stronger response of turnover to trading profits than to account performance suggests that the disposition effect is not the only factor driving turnover.

Finally, in the third row, we predict disposition bias. The figure in the first column of this row shows that high account outperformance relative to the market substantially increases the short-term tendency to sell winners rather than losers. The second column uses a more specific measure of feedback, in which 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. 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.[10] This variable appears to predict the future disposition bias with the expected sign, but is not statistically significant.

Figure 4 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 all the predictor variables except account value from our specifications. The figure illustrates the median and the 10th and 90th percentiles of predicted behaviors. In

[10] Consistent with this view, Ranguelova (2001) finds that disposition bias is attenuated among investors who hold small US stocks with greater momentum in their returns.

both the disposition effect and turnover plots, the dominant influence of the age effect is clearly visible in the figure. The spread in predicted behaviors across accounts is meaningful for both idiosyncratic variance share and turnover.

This section provides evidence that there may be 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, we will show during our analysis of the impacts of behavior on performance that both turnover and the disposition bias are associated with lower account returns, not higher returns as reverse causality would require. We now turn to the impacts of age and feedback on another important dimension of investment behavior, namely, investors' portfolio tilts towards particular styles.

5 Age and Feedback Effects on Style Tilts

5.1 How style tilts change with age

We focus on measures of style demand and supply. Style demand is defined as the cross-sectionally demeaned percentile of the portfolio of stocks bought by the investor multiplied by the purchase turnover of that investor. Thus, demand for value can be high when an investor buys a sizable amount of stocks with modest value tilt or a modest amount of stocks with a sizable value tilt. Style supply is defined similarly, but for the investor's sales. We are especially interested in net style demand, which is the difference between style demand and style supply.

The three plots of Figure 5 show how these measures of style investing behavior vary with age for three different style dimensions, namely size, value, and momentum. The top panel shows that when purchasing stocks, more experienced investors tend to more strongly favor value stocks and small stocks than novice investors. However, these more experienced investors appear to slightly tilt away from momentum in their purchases relative to novice investors.

The second panel in the figure shows that when selling stocks, the main tendency of more experienced investors relative to novices is to sell growth stocks. An eight-year old investor has monthly value supply about 0.6 turnover-weighted-percentiles lower than a novice investor. This corresponds to sales with a value tilt about 10.5 percentiles lower than a new investors in a month with average turnover (0.6 divided by 5.7% equals 10.5). This tilt away from value in stock sales is the main driver of the effect observed in the third panel, which shows that the strongest of the age effects on style tilts is the substantial net demand by more experienced investors for value stocks. Our finding here is consistent with the results reported by Betermier et al. (2013) for older investors in Sweden, although it is important to keep in mind that Betermier et al. work with the age of underlying investors, not our measure of account age.

The third panel in Figure 5 also shows a tendency for investors to accumulate greater quantities of small stocks as they become more experienced. Finally, there is a modest U-shaped effect on the accumulation of momentum stocks as investors age – one cannot reject the hypothesis that eight-year-old accounts have momentum tilts that are comparable to those of novice investors, with a minimum rate of accumulation of winners at the five-year age mark. This may result from novices entering the market by purchasing well known stocks that have recently appreciated, then moving to a more neutral investing style, and then perhaps appreciating the evidence for momentum profits when they become extremely experienced.

5.2 How style tilts change with feedback

Figure 6 shows the impacts of the outperformance of the investor relative to the market and the style-specific feedback measures on investors' net style characteristic demands. The top panel of the figure shows the impacts of these feedback measures on net size demand, the second, on net value demand, and the third on net momentum demand.

The left column of the figure shows the impacts of account outperformance on net style demands. These impacts appear substantial and highly statistically significant, showing that outperformance predicts increasing accumulation of large, growth, and high momentum stocks. One possible interpretation is that these stocks tend to have similar characteristics to the best performing stocks among investors' current and recent holdings. The disposition effect implies that investors tend to sell their specific winners, but when their overall account performance has been good this tendency may be weaker, and they may also seek to replace these winners with other stocks that have similar characteristics at the date of sale. It is also likely that outperformance increases overconfidence, as suggested by Daniel, Hirshleifer, and Subrahmanyam (1998), Gervais and Odean (2001), Statman, Thorley, and Vorkink (2006), and Kruger (2013). However, while overconfidence might explain a preference for growth stocks it is not clear that it should generate tilts towards large-cap or momentum stocks.

The right-hand column of Figure 6 shows the impacts of style-specific feedback on the accumulation of stocks in those styles. This style-specific feedback is constructed by taking the total returns on the sub-portfolio of stocks held by the investor that are ranked above the cross-sectional average of all stocks in the same period on the given characteristic (i.e., size, value, and past-high-returns/momentum) minus the total returns on the sub-portfolio of stocks held by the investor ranked below average in the given characteristic. In cases in which the investor does not own stocks ranked above or below the average for a given characteristic, value-weighted market returns are substituted for the type of stocks that the investor does not own (e.g. growth, if an investor holds only value stocks).

The figures show that there are two impacts of style-specific feedback. The first is a more precisely estimated short-term effect, in which the investor decumulates the style that has outperformed. A figure in our appendix shows that this short-term effect is due to a spike in supply of the outperforming style, suggesting this is a manifestation of the disposition effect. However, for all three of the styles, the less precisely estimated but far longer-lived effect is a tendency to continue to accumulate styles in which the investor has experienced positive returns. This "style chasing" behavior is a feature of a number of different theoretical models, most notably that of Barberis and Shleifer (2003).

Figure 7 illustrates the relative importance of account age and investment experience in predicting each of the three style tilts. For all accounts that opened in December 2003, the figure shows the predicted changes to style tilts from January 2004 through the end of the sample, using all the predictor variables except account value from our specifications. The figure illustrates the median and the 10th and 90th percentiles of predicted changes in net style demands. In both the size and value plots, the dominant influence of the age effect is clearly visible in the figure, as older investors move into smaller, value stocks. The spread in predicted behaviors across accounts generated by differential feedback is meaningful in all three cases.

6 Implications for Performance

In the previous sections we have shown that both account age and investment experiences strongly affect behaviors and style tilts. How do these sources of learning feed into overall performance? In this section we analyze the impacts of age, behaviors, and style tilts on investor performance.

Performance is inherently difficult to measure, because account returns are subject both to considerable idiosyncratic volatility and to common shocks resulting from our measured style tilts and other systematic tilts. Accordingly we look at performance using three different approaches. First we measure performance directly at the account level; then we analyze the returns on portfolios of stocks held by novice and experienced investors; and finally we predict the returns on individual Indian stocks using the characteristics of the investor base as well as characteristics of the companies themselves.

6.1 Determinants of account performance

Figure 8 shows the impact of age on account performance, estimated from a piecewise linear model. While the reported age effects have substantial economic magnitudes (roughly 100 basis points a month higher for an eight year old investor relative to a novice), they are imprecisely estimated and only barely significant at the five percent level at seven years.

Table 3 reports more comprehensive results, showing that the age effect remains economically meaningful but statistically insignificant when it is restricted to be linear, and falls in magnitude by close to one-third when measures of investor behavior and style tilts are added as explanatory variables. Theses measures have various interesting effects. Underdiversified investors perform better, consistent with Ivkovic et al. (2008), and high-turnover investors perform worse, but these effects are statistically insignificant. Also, investors with a strong disposition bias tend to perform worse, unsurprisingly given the momentum effect in the Indian stock market. Of the style tilts, the strongest and most significant effect is that value investors perform better.

6.2 Aggregated novice- and experienced-investor portfolios

In Table 4 we decompose the returns to various zero-cost portfolios formed on the basis of stockholder account age. There is no equivalent of the fixed effects we have used in our study of individual accounts, so we do not attempt to distinguish stocks that are preferred by older accounts because of these accounts' constant characteristics from stocks that are increasingly preferred by accounts as they grow older.

We first form a zero-cost portfolio that goes long stocks held by a representative older individual account (a stratified-sample-weighted average of the portfolio weights of accounts in top quintile of accounts sorted by time since first stock investment), and goes short stocks held by a representative novice investor (i.e. formed from accounts in the bottom such quintile). The first column of Table 4 reports results for this portfolio. The second and third columns decompose it into long-short portfolios formed between the older and average (i.e. formed from accounts in the middle quintile) representative investor and average and novice representative investor.

Figure 9 illustrates the cumulative excess returns (relative to the Indian short rate) to the experienced-investor and novice-investor portfolios, 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 89%, while the cumulative excess return on the Indian market index was 79%, and the cumulative excess return on the novice-investor portfolio was only 10%.

In the first column of Table 4, we regress the portfolio weights in the older minus novice zero-cost portfolio onto a vector of stock characteristics, to see what characteristics are preferred or avoided by older investors relative to novice investors. The second and third columns of the table do the same for the older minus average and average minus novice portfolios respectively. In the bottom of the table, we show how the returns on the zero-cost portfolios can be attributed into unconditional and timing effects related to either stock characteristic tilts or a residual that we call "selectivity" following Wermers (2000).

The table shows that relative to novice investors, the most experienced Indian investors tilt their portfolios towards small stocks, value stocks, stocks with low turnover, stocks without large beneficial ownership, and stocks held by institutions. Experienced investors also avoid large, attention-grabbing initial public offerings. This is perhaps unsurprising considering that such IPOs are one of the main routes to initial investor participation in the Indian stock market.

In terms of their contribution to performance, the stock characteristics of older investors explain only 9 basis points out of a total outperformance relative to novice investors of 38 basis points per month. The remainder is not explained by characteristic timing, which makes an insignificant but negative contribution of 10 basis points. The performance differential is attributable mainly to stock timing effects (27 basis points) and non-characteristic related stock selection (13 basis points). Most of these differences are preserved when looking at the difference between average aged and novice accounts, implying that the initial mistakes made by inexperienced investors ("rookie mistakes") contribute to the performance differential between experienced and novice accounts.

In the first column of Table 5 we evaluate the older minus novice zero-cost portfolio in a different way, by regressing its return on six factors commonly used in the asset pricing literature: the market return, small minus big (SMB) return, value minus growth (HML) return, momentum (UMD) return, and factor portfolios capturing short-term reversals and illiquidity as measured by turnover. We find that the portfolio has a negative loading on HML, despite its slight tilt towards value characteristics, and has a significantly positive six-factor alpha. This result suggests that experienced investors add value not by taking compensated factor exposures, but by finding outperforming stocks whose factor exposures are generally poorly compensated. This suggests that the results of Coval et al. (2005) that following high-performing individual investors' trades generates high abnormal returns could also apply to the overall investments of individual investors.

The remaining columns of Table 5 repeat this exercise using zero-cost portfolios that go long stocks held by investors with high levels of behaviors or style tilts, and short stocks held by investors with unusually low levels of these behaviors or style tilts. It appears that underdiversified investors hold stocks with compensated factor exposures, but do not add value relative to the six-factor model, while high-turnover investors have similar factor exposures to underdiversified investors but have a significantly negative alpha relative to the model.

6.3 Stock returns and the investor base

Finally, 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 6 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—as well as the style tilts of the investor base. The age effect, though somewhat diminished, remains significant, and we find that an investor base with high turnover predicts lower returns.

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 underdiversified 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. Finally, we see that while large stocks have lower returns, stocks held by investors who favor large stocks—who may generally be larger, more sophisticated investors—tend to have higher returns.

The institutional ownership of stocks is included in Table 6 to address 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, in Table 6, the coefficient on institutional ownership is only weakly positive.

7 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 that feedback from investment performance and years of experience appear to be important drivers of investment behaviour and ultimately the returns experienced by individual investors. Both turnover and disposition bias decline with experience, but strengthen with good investment results, as does underdiversification. The tendency to invest in value stocks and small stocks increases with experience, but experiencing good investment performance relative to the overall market pushes investors towards growth stocks, large stocks, and high-momentum stocks. Overall, it appears that the Indian market rewards investment skill, which develops over time. Good investment performance early on, however, encourages what we might characterize as “bad” investment behaviour, because of the strong reinforcement learning effect that we identify.

There are several interesting questions we have not yet explored, but plan to examine in future research. First, we can ask whether feedback effects on behavior vary 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. Related to this, we can ask whether the response to feedback varies with investor characteristics that might proxy for sophistication. Second, we can explore nonlinearities in responses to feedback, for example asymmetries between positive and negative feedback, or different responses to extreme feedback. In a similar spirit, we can study how investors react to major personal shocks, such as losing a great deal of wealth in a company fraud. Finally, an extension of the Indian dataset with finer geographical resolution will make it possible to ask how social interaction or local networks affect learning (Hong et al., 2004, Ivkovic and Weisbenner, 2005, 2007).

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.

Ameriks, John, and Stephen Zeldes, 2004, How do household portfolio shares vary with age?, unpublished paper, Columbia University.

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.

Betermier, Sebastien, Laurent Calvet, and Paolo Sodini, 2013, Who are the value and growth investors?, unpublished paper, McGill University, HEC Paris, and Stockholm School of Economics.

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 Benjamin Ranish, 2013, Online appendix to Getting better: Learning to invest in an emerging stock market, available online at scholar.harvard.edu/campbell.

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.

Choi, James J., David Laibson, Brigitte C. Madrian, and Andrew Metrick, 2009, Reinforcement Learning and Savings Behavior, *Quarterly Journal of Economics* 124, 1639-1674.

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.

Cronqvist, Henrik, Stephan Siegel, and Frank Yu, 2013, Value versus Growth Investing: Why Do Different Investors Have Different Styles?, unpublished paper, available online at

http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2351123

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam, 1998, Investor psychology and security market under- and overreactions, *Journal of Finance* 53, 1839–1885.

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.

Froot, Kenneth and Melvyn Teo, 2008, Style Investing and Institutional Investors, *Journal of Financial and Quantitative Analysis* 43, 883-906.

Gervais, Simon and Terrance Odean, 2001, Learning to be overconfident, *Review of Financial Studies* 14, 1–27.

Gomes, Francisco, and Alexander Michaelides, 2008, Asset pricing with limited risk sharing and heterogeneous agents, *Review of Financial Studies* 21, 415–448.

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.

Guiso, Luigi, and Paolo Sodini, 2013, Household finance: An emerging field, Chapter 21 in George Constantinides, Milton Harris, and Rene Stulz eds., *Handbook of the Economics of Finance, Volume 2B*, North-Holland, 1397–1531.

Han, Bing and Alok Kumar, 2013, Speculative Retail Trading and Asset Prices, *Journal of Financial and Quantitative Analysis* 48, 377-404.

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., 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.

Kruger, Samuel, 2013, Disagreement and liquidity, unpublished paper, Harvard University.

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.

Statman, Meir, Steven Thorley, and Keith Vorkink, 2006, Investor overconfidence and trading volume, *Review of Financial Studies* 19, 1531–1565.

Strahilevitz, Michal Ann, Terrance Odean, and Brad M. Barber, 2011, Once Burned, Twice Shy: How Naïve Learning, Counterfactuals, and Regret Affect the Repurchase of Stocks Previously Sold, *Journal of Marketing Research* 48, S102-S120.

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 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	12,350	4,533	$386.3	51.02%	17.50%	3.51%
2005	13,613	4,844	$553.1	57.71%	15.78%	3.73%
2006	15,408	5,150	$818.9	63.12%	14.94%	4.72%
2007	17,262	5,516	$1,819.1	66.06%	12.81%	4.55%
2008	17,417	5,988	$647.2	64.60%	11.89%	4.46%
2009	17,592	6,398	$1,306.5	64.01%	11.22%	4.56%
2010	19,681	6,897	$1,631.8	66.82%	10.78%	4.35%
2011	22,794	7,493	$1,007.2	67.73%	10.12%	4.00%
2012	21,431	7,735	$1,263.3	68.55%	9.91%	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 12,63,311	Rs 2,50,767	Rs 33,73,772	Rs 7,288	Rs 67,684	Rs 5,51,585
Account Value, Jan 2012 US$	$24,771	$4,917	$66,152	$143	$1,327	$10,815
Number of Equity Positions	6.88	4.75	8.02	1.00	3.42	14.27
Urban Accounts	55.84%	54.60%	57.54%	0	1	1
Semi-Urban Accounts	12.31%	11.86%	12.79%	0	0	1
Rural Accounts	31.85%	30.05%	33.05%	0	0	1
Monthly Account Stock Return Minus Market	-0.03%	-4.80%	10.21%	-1.75%	-0.12%	1.52%
Idiosyncratic Share of Portfolio Variance	0.45	0.25	0.55	0.24	0.46	0.68
Monthly Turnover	5.72%	2.09%	12.28%	0.00%	2.57%	16.30%
Disposition Effect - ln(PGR/PLR)	1.23	-1.26	2.40	0.25	1.35	2.38
Stock Portfolio Beta	1.03	0.95	1.09	0.95	1.02	1.13
Size Percentile (Tilt) of Stocks Held Relative to Market Portfolio	-4.62	-6.17	-3.02	-18.28	-1.59	3.25
Book-Market Percentile (Tilt) of Stocks Held Relative to Market Portfolio	3.51	-6.35	18.66	-9.54	2.13	22.49
Momentum Percentile (Tilt) of Stocks Held Relative to Market Portfolio	-5.63	-17.16	7.38	-21.95	-5.75	4.43

Table 3: Account Age Effects in Equity Portfolio Returns

Results are constructed from a stratified random sample of individual accounts opened on or after February 2002: 5,000 accounts are drawn at random from each Indian state/territory, drawing all accounts from states with less than 5,000 accounts. A bit over 4.2 million account months spanning January 2004 through January 2012 are used in the regressions. The regression specification is $(R_{it}-R_t)=\beta(A_{it}-A_t)+\kappa(B_{it}-B_t)+s_i+\varepsilon_{it}$, where R_{it} represents the returns of investor i in month t, A represents the account age effect, B are the lagged account behaviors appearing in columns [c], and s are individual fixed effects. Lagged turnover and disposition bias are averages over the past 12 months, winsorized at the 1st and 99th percentile of accounts with at least 5 observations of the behavior in the 12 month period. Where missing, the cross-sectionally and then individually de-meaned values of lagged behaviors are imputed as zeros. 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 relative to the variance of monthly account excess returns.

Dependent Variable: Account Monthly Return in Excess of Risk-Free Rate (bp) (Mean: 96.7bp)

		[a]	[b]	[c]
Account Age Effect	Account Age (Linear)	*12.01*		8.25
		(7.10)		*(6.94)*
	Piecewise Linear		See Figure 4	
Investor Behavior	Lagged Idio. Share of Portfolio Var.			55.85
				(76.06)
	Lagged Portfolio Turnover			-97.64
				(68.57)
	Lagged Disposition Bias			**-3.69**
				(1.50)
	Size Tilt			177.90
				(181.70)
	Value Tilt			**554.30**
				(113.36)
	Momentum Tilt			-22.20
				(122.11)
Incremental R^2		0.00031	0.00039	0.00015

Table 4: Decomposition of the Difference in Returns on Old and New Accounts

For the period January 2004 through January 2012, portfolios are formed which buy each stock in proportion to its average weight amongst the oldest, average (middle), or newest quintile of accounts opened on/after February 2002. Zero-cost portfolios are formed using the differences in portfolio weights between the average and newest, oldest and average, and oldest and newest quintiles. Stocks with market capitalization below 500 million Rs (approximately $10 million) are excluded from all portfolios, leaving 2,677 stocks j in the sample. The top (portfolio tilts) part of each column reports the time-series average of coefficients, φ_{bar}, from Fama MacBeth regressions $W_{jt}=\varphi_t X_{jt}+\varepsilon_{jt}$ of portfolio weights W on the set X of cross-sectionally de-meaned stock characteristics listed in the table. Normalized rank transforms are used to measure market capitalization, book-market, momentum, turnover, and beneficial and institutional ownership shares. The bottom panel provides a decomposition of total returns, $\Sigma_j W_{jt}R_{jt}$, to these zero-cost portfolios. Returns are first broken into timing effects $\{\Sigma_j W_{jt}R_{jt}-\Sigma_j W_{bar,j}R_{bar,j}\}$ and selection effects $\{\Sigma_j W_{bar,j}R_{bar,j}\}$. Next, we run Fama MacBeth regressions of returns on stock characteristics $\{R_{jt}=\psi_t X_{jt}+\eta_{jt}\}$. Using these regressions, 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. 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 coefficients with t-subscripts are from the cross-sectional regressions run in Fama MacBeth estimation. Standard errors in ( ) are computed by bootstrap, and standard errors in the return decomposition account for uncertainty in both portfolio weight and return regressions. Statistically significant coefficients at the five and ten percent level are indicated by bold and italicized type respectively.

Zero-Cost Portfolio Represents:	Oldest minus Newest	Oldest minus Average	Average minus Newest
Portfolio Tilts (1000 x φ_{bar})	[1]	[2]	[3]
Market beta	-0.547	*-0.393*	-0.154
	(0.568)	*(0.224)*	*(0.428)*
Market capitalization	-0.318	**-0.448**	0.130
	(0.233)	*(0.185)*	*(0.272)*
Book-market	0.171	0.100	0.071
	(0.143)	*(0.095)*	*(0.179)*
Momentum (t-2:t-12 returns)	-0.003	0.113	-0.116
	(0.340)	*(0.255)*	*(0.175)*
Stock turnover	**-0.908**	-0.237	*-0.671*
	(0.262)	*(0.306)*	*(0.396)*
Beneficial ownership	*-0.604*	**-0.457**	-0.147
	(0.367)	*(0.192)*	*(0.246)*
Institutional ownership	**0.919**	**0.447**	0.472
	(0.356)	*(0.162)*	*(0.438)*
Ln(1+stock age)	0.010	*0.216*	**-0.207**
	(0.075)	*(0.118)*	*(0.104)*
Large IPOs (market cap if age<1 year)	**-13.358**	**-0.733**	**-12.625**
	(3.723)	*(0.358)*	*(3.625)*
Return Decomposition			
Stock characteristic selection	8.52	3.37	5.15
	(5.54)	*(2.34)*	*(3.54)*
Additional stock selection	12.90	4.72	8.19
	(14.55)	*(5.97)*	*(11.15)*
Stock characteristic timing	-9.63	1.16	-10.79
	(11.13)	*(5.73)*	*(7.17)*
Additional stock timing	26.60	-0.41	27.02
	(21.24)	*(7.35)*	*(21.53)*
Total difference in returns	38.40	8.83	29.56
	(28.34)	*(10.87)*	*(24.42)*

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

The zero-cost portfolios evaluated below are differences in representative portfolios held by [1] the oldest and newest accounts (the first column of Table 4), [2]-[4] the accounts most and least likely to engage in the given behaviors, and [5]-[7] the accounts with the strongest tilts towards and away from the given style characteristics of stocks. Account behaviors and characteristic tilts are determined by the cumulative average of cross-sectionally de-meaned behaviors or tilts. Portfolio returns are adjusted using a six factor model, 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 (Illiq) 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).

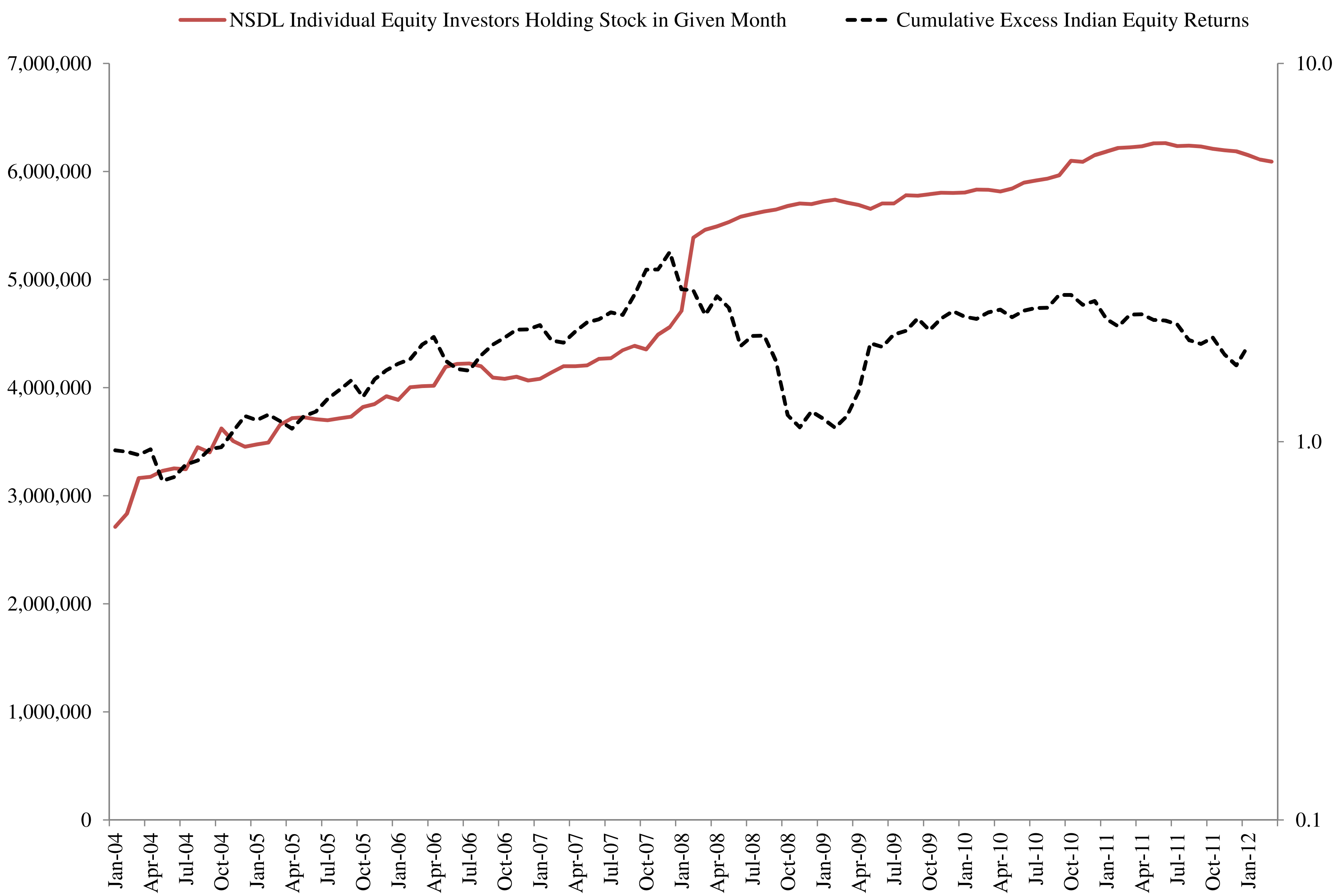
Investor Characteristic Sort Based On:	Account Age [1]	Idiosyncratic Share of Portfolio Variance [2]	Turnover [3]	Disposition Bias [4]	Size Tilt [5]	Value Tilt [6]	Momentum Tilt [7]
Raw Return	*0.38%*	0.30%	*-0.49%*	-0.05%	-0.03%	**0.83%**	-0.03%
	(0.23%)	*(0.19%)*	*(0.30%)*	*(0.11%)*	*(0.54%)*	*(0.38%)*	*(0.44%)*
Monthly Alpha	**1.11%**	-0.04%	**-0.79%**	-0.12%	**0.80%**	*0.82%*	-0.47%
	(0.33%)	*(0.33%)*	*(0.24%)*	*(0.12%)*	*(0.34%)*	*(0.43%)*	*(0.46%)*
Factor Loadings							
Market Beta	*-0.10*	**0.08**	*0.07*	0.00	**-0.12**	0.02	0.05
	(0.06)	*(0.04)*	*(0.04)*	*(0.01)*	*(0.04)*	*(0.04)*	*(0.05)*
SMB	*0.07*	**0.20**	**0.11**	**0.07**	**-0.51**	**0.29**	-0.10
	(0.04)	*(0.04)*	*(0.03)*	*(0.02)*	*(0.06)*	*(0.05)*	*(0.08)*
HML	**-0.20**	**0.19**	**0.19**	**0.08**	**-0.29**	0.08	-0.02
	(0.07)	*(0.07)*	*(0.06)*	*(0.03)*	*(0.08)*	*(0.08)*	*(0.09)*
UMD	0.03	**-0.23**	**-0.21**	**-0.08**	**0.24**	**-0.23**	**0.53**
	(0.05)	*(0.04)*	*(0.04)*	*(0.02)*	*(0.05)*	*(0.05)*	*(0.08)*
Short Term	-0.10	0.08	0.04	0.02	-0.01	0.01	-0.01
Reversals	*(0.07)*	*(0.05)*	*(0.06)*	*(0.02)*	*(0.06)*	*(0.06)*	*(0.10)*
Illiq (Based	-0.14	-0.13	-0.06	-0.06	0.00	-0.11	-0.03
on Turnover)	*(0.13)*	*(0.13)*	*(0.08)*	*(0.04)*	*(0.12)*	*(0.14)*	*(0.15)*

Table 6: 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 and style tilts (i.e. value weighted average style percentile of the investor's portfolio), we use the average behavior or style tilt across individual investors, where the measure from each investor is taken as the cumulative average of a cross-sectionally de-meaned measure of the behavior or style tilt. Average investor account age, behavior measures, style tilts, 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, with a Newey West serial correlation adjustment. All coefficients are multiplied by 100 for readability, and statistical significance at the five and ten percent level are indicated by bold and italicized type respectively.

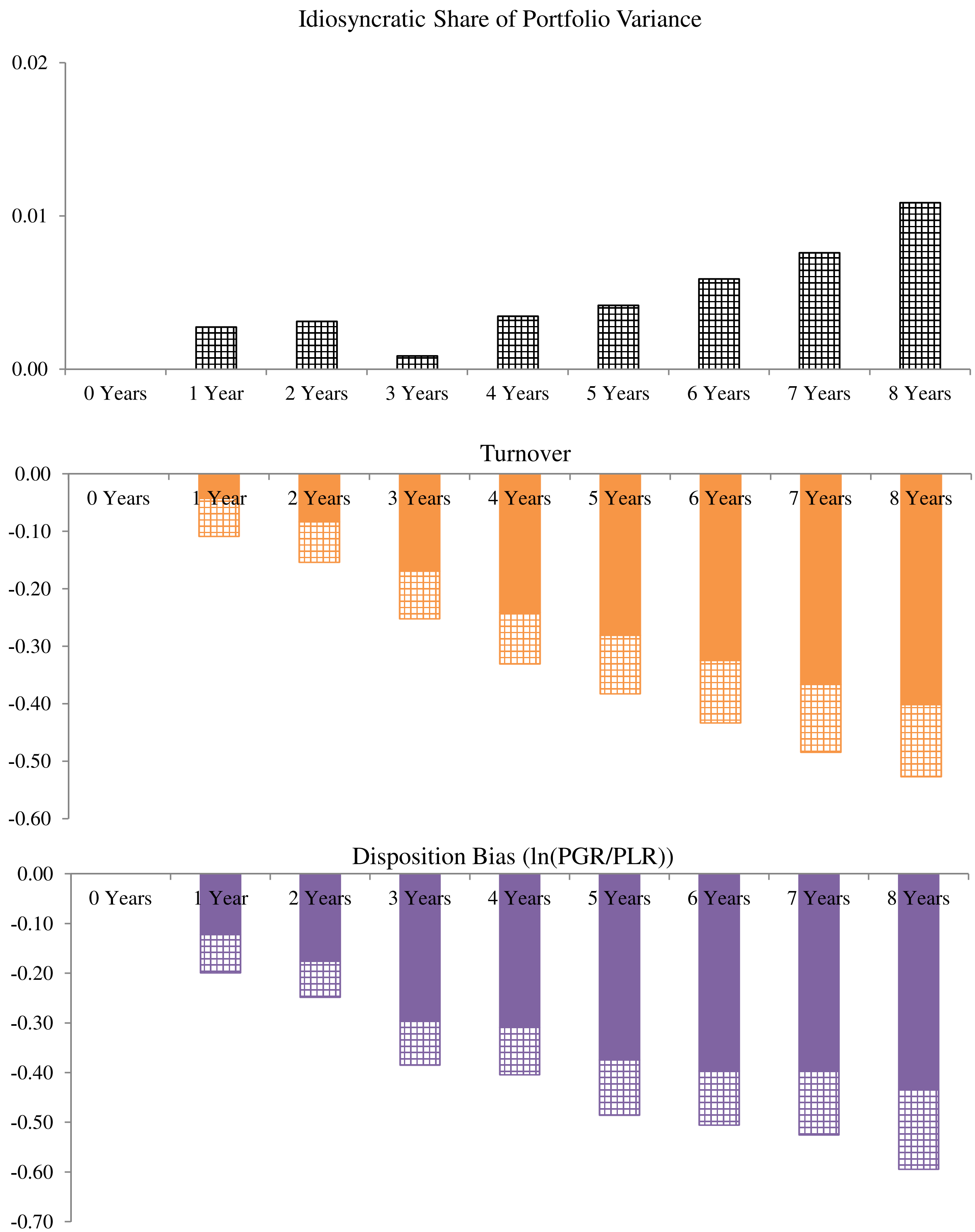
		[1]	[2]	[3]
Investor Characteristics	Account Age	**1.61**	**0.61**	0.14
		(0.58)	(0.29)	(0.23)
	Idio. Share of Portfolio Var.		-0.22	0.16
			(0.32)	(0.34)
	Portfolio Turnover		**-1.88**	**-0.64**
			(0.54)	(0.32)
	Disposition Bias		-0.16	-0.12
			(0.29)	(0.25)
	Size Tilt		-0.46	**1.32**
			(1.29)	(0.59)
	Value Tilt		**1.83**	-0.70
			(0.46)	(0.51)
	Momentum Tilt		*0.81*	0.19
			(0.47)	(0.32)
Stock Characteristics	Market beta			0.35
				(1.30)
	Market capitalization			**-3.40**
				(1.63)
	Book-market			**3.91**
				(0.67)
	Momentum			**3.16**
				(0.58)
	Stock turnover			**-1.73**
				(0.41)
	Beneficial ownership			**0.75**
				(0.33)
	Institutional ownership			0.28
				(0.37)
	Ln(1+stock age)			0.09
				(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.

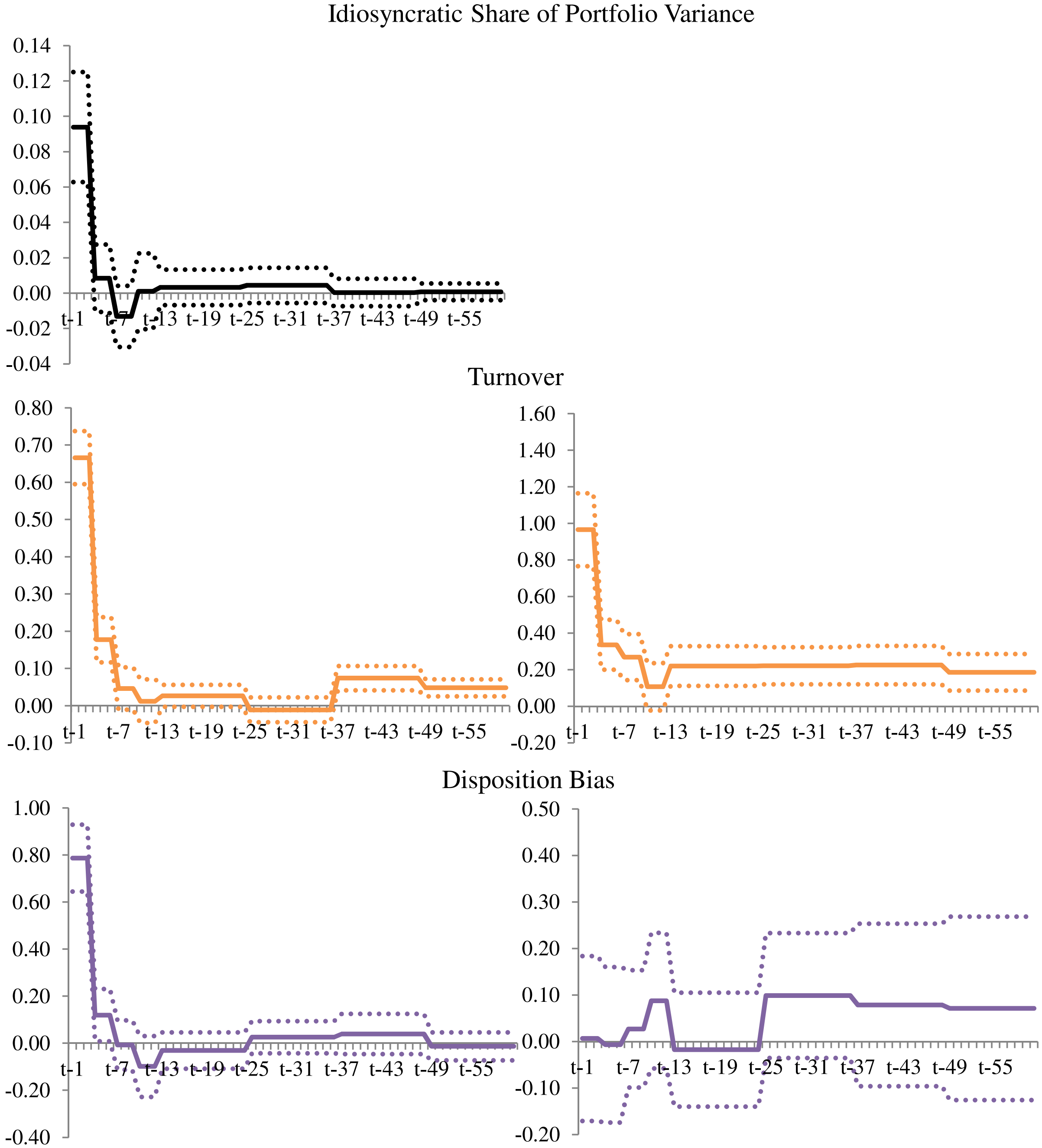
Figure 2: Account Age Effects for Investor Behaviors



The plots above are produced from account behavior regressions following Equation 3. The combined bars represent the point estimate given by the regression, while the solid part of the bar lies outside the 95% confidence interval. Account behaviors are scaled by the time-series average of their cross-sectional means as reported in Table 2.

Figure 3: Investing Behavior Response to Feedback at Various Horizons

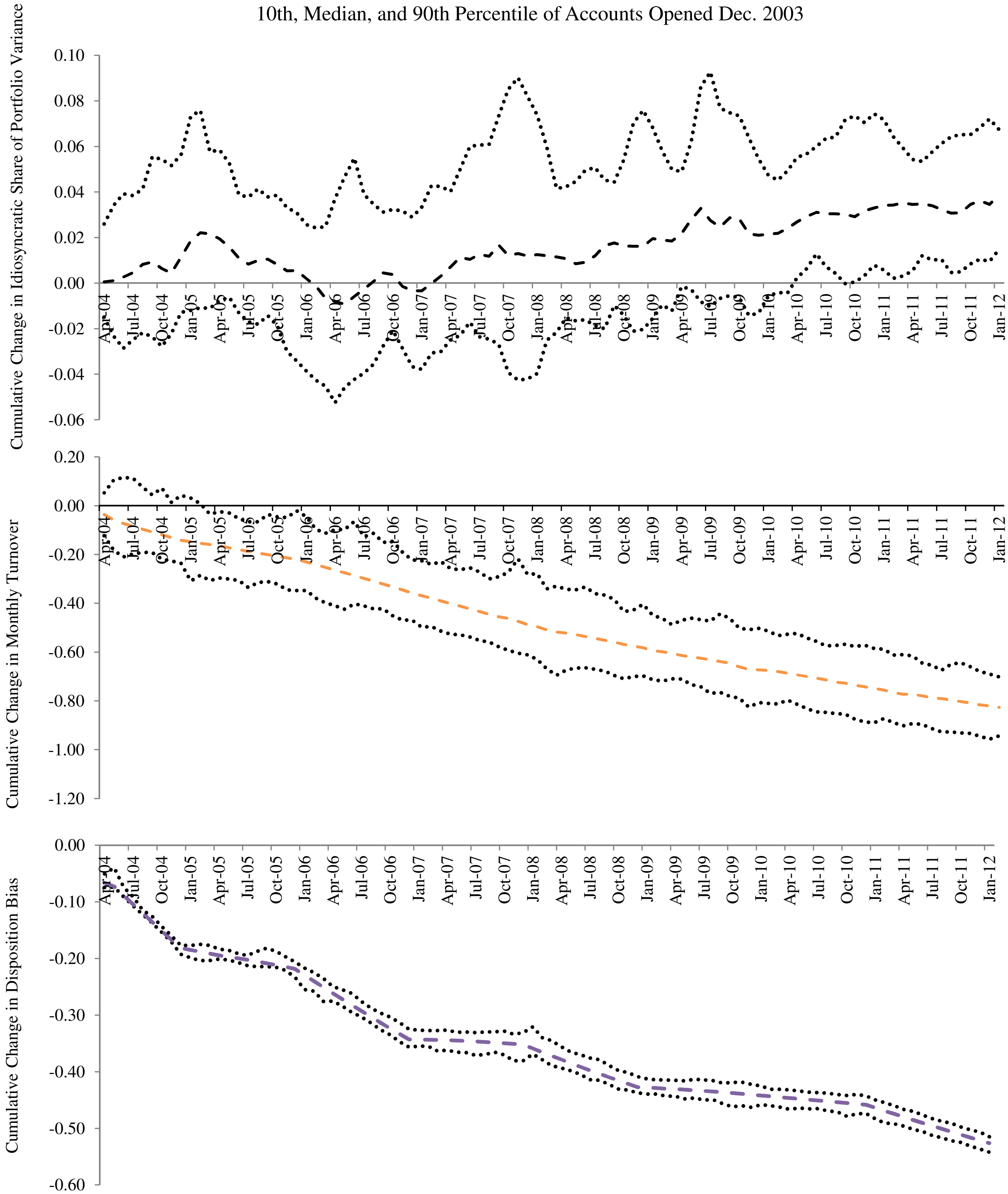
Account Performance Feedback (Left) and Behavior Feedback (Right)



Plots are produced from account behavior regressions following Equation 3.Dotted lines represent 95% confidence intervals. All behaviors are scaled by their mean, given in Table 2. All behavioral feedback measures are defined such that positive coefficients indicate that the feedback reinforces in the given behavior. Only one plot appears for idiosyncratic share of portfolio variance, as account outperformance is also used as a measure of behavioral feedback. Behavior feedback for turnover is taken as the increase in returns due to trading activity; 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. Behavior feedback for the disposition bias is taken as the increase in returns due to selling off gains versus losses; excess returns on stocks three-months following sales, with each sale weighted in proportion to the value of the sale (relative to the investor's portfolio) and with the outperformance of gains counting negatively in the measure.

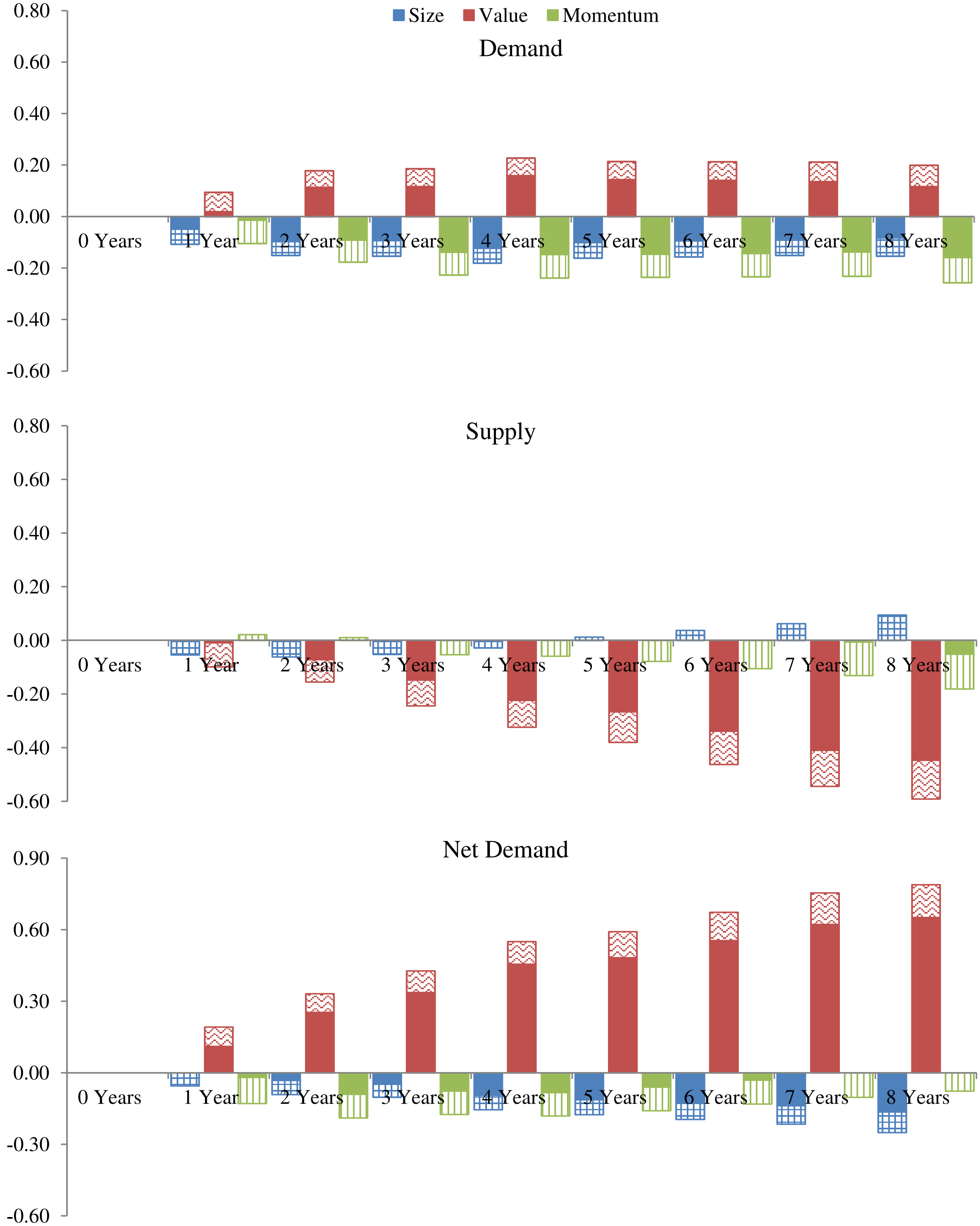
Figure 4: Simulated Cumulative Change in Investor Behaviors from Age Effects and Feedback

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



Plots above use the account behavior regressions following Equation 3.Specifically, age effects (see Figure 2), feedback coefficients (see Figure 3), and the coefficient on lagged behavior from this regression are combined with the actual age and feedback received by individual investor accounts opened in December 2003. The 10th, 50th, and 90th percentiles of the simulated distribution appear above. Cumulative changes in behavior are scaled by the time-series average of their cross-sectional means as reported in Table 2.

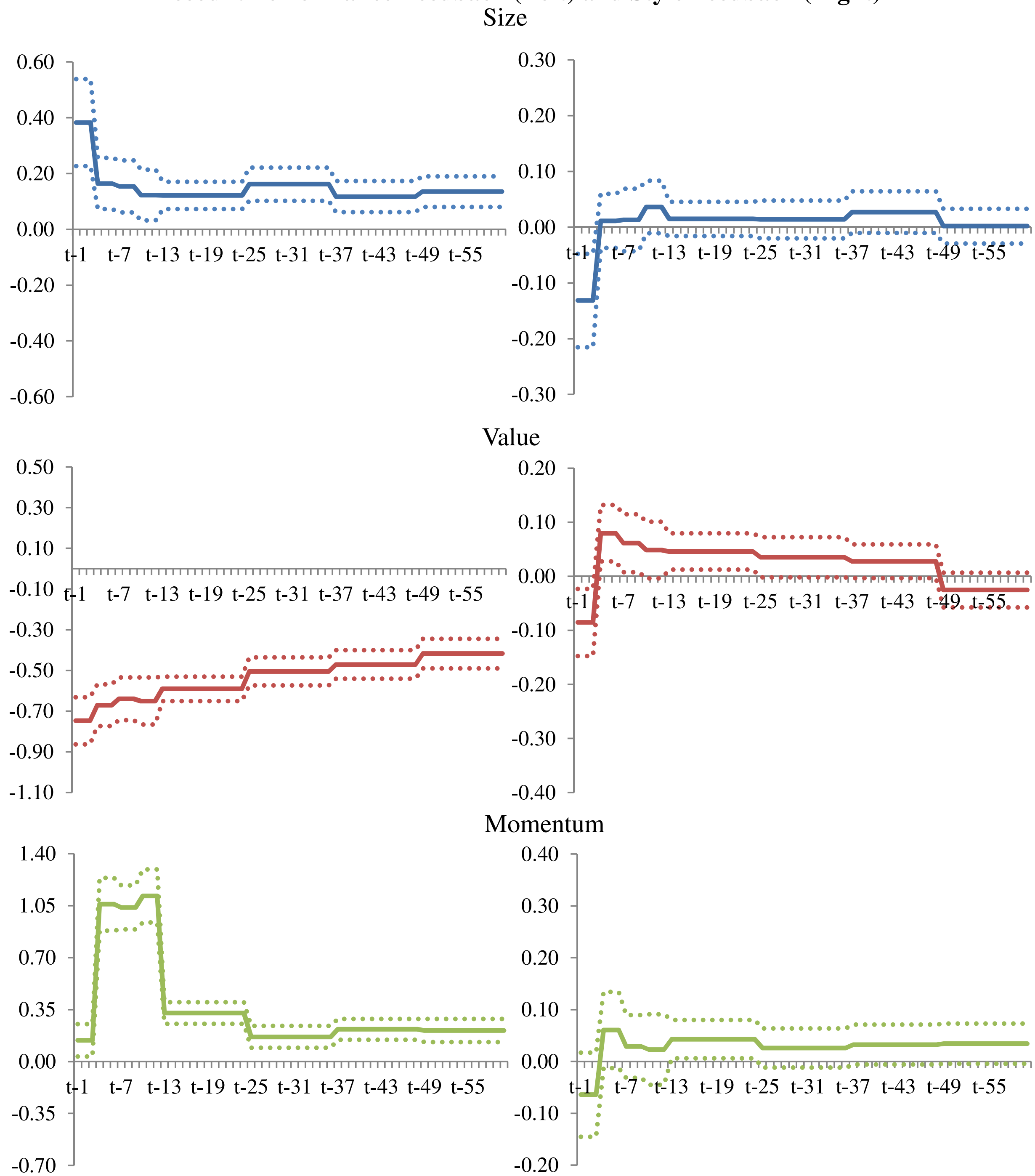
Figure 5: Account Age Effects for Style Characteristic Supply/Demand



The plots above are produced from style demand/supply regressions following Equation 3. The combined bars represent the point estimate given by the regression, while the solid part of the bar lies outside the 95% confidence interval.

Figure 6: Net Characteristic Demand Response to Feedback at Various Horizons

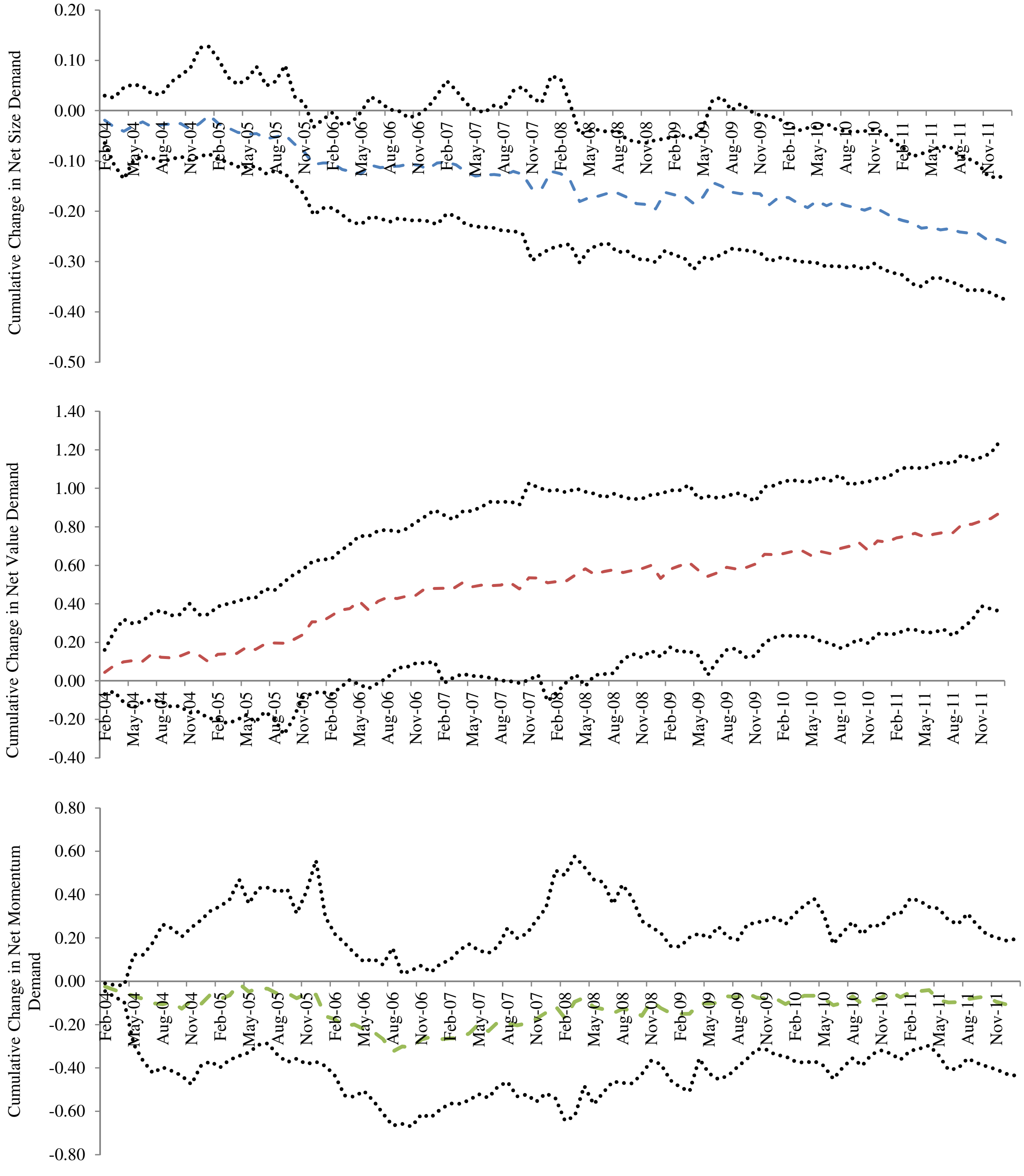
Account Performance Feedback (Left) and Style Feedback (Right)



Plots are produced from account net style characteristic demand regressions following Equation 3. Dotted lines represent 95% confidence intervals. Feedback is defined as the total returns of stocks held by the investor ranked above-average in the given characteristic (i.e. large/value/high-momentum) minus the total returns of stocks held by the investor ranked below-average in the given characteristic. Value-weighted market returns are substituted for returns for any type of stock the investor does not own (e.g. growth, if an investor holds only value stocks).

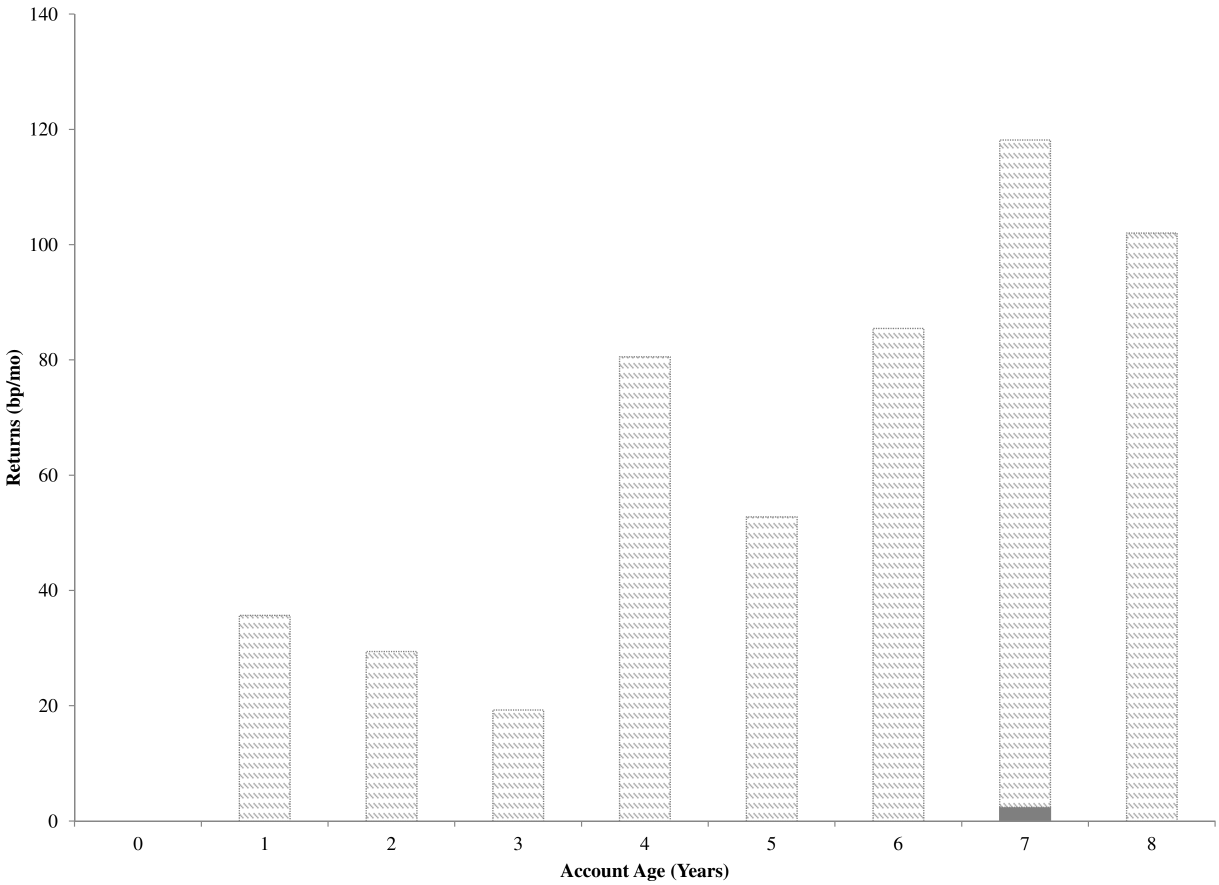
Figure 7: Simulated Cumulative Change in Net Characteristic Demand from Age Effects and Feedback

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



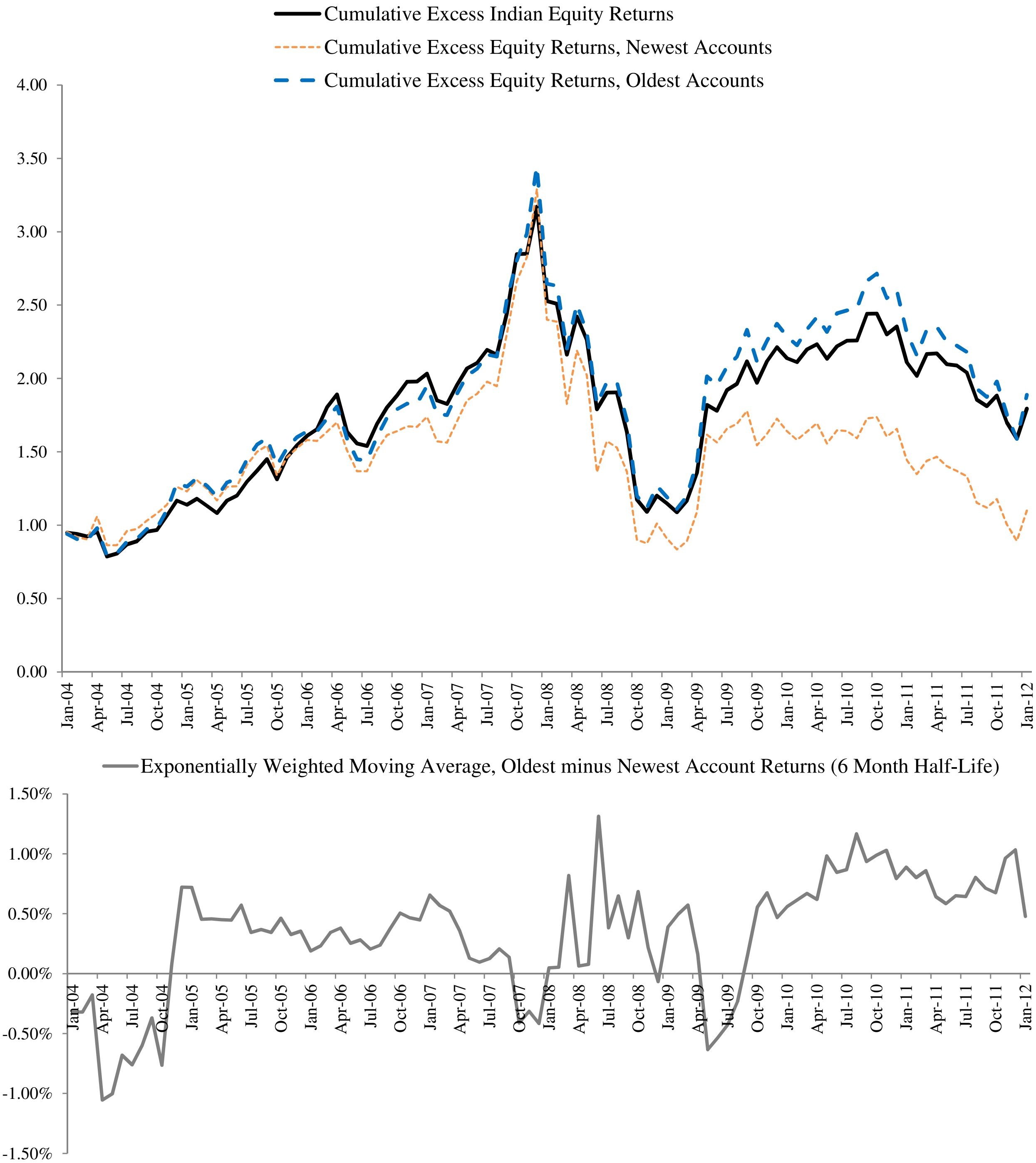
Plots are produced from account net style characteristic demand regressions following Equation 3. Specifically, age effects (see Figure 5) and feedback coefficients (see Figure 6) are combined with the actual age and feedback received by individual investor accounts opened in December 2003. The 10th, 50th, and 90th percentiles of the simulated distribution appear above.

Figure 8: Account Age Effects in Equity Portfolio Returns



The plots above are the piecewise linear age effects produced from regression Specification [1b] in Table 3. Breakpoints occur at years one, two, three, four, five, and seven. The combined solid and patterned bars represent the point estimate given by the regression, while the solid part of the bar (non-zero only at 7 years) lies outside the 95% confidence interval.

Figure 9: Cumulative Indian Excess Equity Return Received by Oldest and Newest Quintile of Accounts



Representative portfolios of account-age based groups are formed by equally weighting across the actual portfolio weights of accounts in each age group. The "oldest" and "newest" groups represent the oldest and newest quintile of accounts present in each month from within the set of accounts started on or after February 2002. The top plot produces excess returns by subtracting the yield on three-month Indian Treasury bills from returns on the indicated portfolios.

Getting Better: Learning to Invest in an Emerging Stock Market - Internet Appendix

December 19, 2013

1 Data Construction

1.1 Stock-Level 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.[1]

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, we use internet sources such as Moneycontrol and Economic Times to confirm that the returns are indeed valid. We also use internet sources to look up and confirm returns for stock-months where returns are missing and the stock comprises at least one percent of

[1] 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.

stock holdings for the representative individual investor for either the previous or current month.

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.

1.2 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.

2 Additional Exhibits

Table A1 provides the cross-sectional correlations of the account characteristics examined in Table 2.

Figure A1 provides a monthly measure of the disposition effect computed just as in Odean (1998), alongside Odean's measure based on US brokerage accounts. The Indian tax year ends in March, whereas the US tax year (relevant for Odean's investors) ends in December. The level of disposition effect is lower than the typical levels seen at the individual account level, as this aggregate monthly statistic effectively applies weight to accounts in proportion to the number of stock positions they hold, and investors with more stock positions exhibit a smaller disposition effect.

Figure A2 plots the population per NSDL account against per capita income at the state level. The size of the bubbles indicate state population. The largest contributor to the data is relatively populous and wealthy Maharashtra, which comprises over one-fifth of all accounts in the NSDL data.

3 Robustness Exercises

3.1 Addition of Accounts Opened Prior to February 2002

The data used throughout the paper excludes accounts opened prior to February 2002. For accounts which opened earlier, we do not observe the full investing history, do not know when the account first invested in stocks, and do not observe the initial account characteristics. Such accounts represent about 14.4% of all accounts present in our sample, though they represent a larger fraction of earlier (smaller) cross-sections and thus have potential for significant impact on our results.

To make use of this data in our basic analyses, we impute the first date of stock investment (from which account age is gauged) as equal to three months following the month the account opens. This is roughly equal to the mean time between account opening and stock investment that we observe for accounts opened on/after February 2002. We further assume that cross-sectionally then individually de-meaned feedback and account behaviors were zero for all accounts prior to February 2002 (i.e. all previous cross-sectional variation was explained by the individual account effect s_i).

Figures A3, A5, and A7 show that age effects in our account behavior, style, and return regressions, are qualitatively very similar. Figures A4 and A6 show that the same holds true for the response of investor behavior and style preferences to outperformance and behavior (or style)-specific performance. Table A2 shows that our account return regression allocates a similar share of increased returns to the pure age effect and investor behaviors and favored stock characteristics (which change with account age). Table A3 and Figure A8 show that the preferences over stock characteristics held by pre-2002 investors are similar in direction to those held by the oldest post-2002 investors, and returns on the two age groups of accounts are similar.

3.2 Sensitivity of Results to the Use of "Passive" Account Returns

We compute and use "passive" returns throughout the paper; the returns that the investor would have received if they did not trade during the given month. We can compute "active" returns which take account of trading, but assumptions are required since we do not know the exact timing or order of the purchases and sales which occur during the month. Here, we take two extreme assumptions about this timing, yielding a "high leverage" and "low leverage" measure of active returns.

First, we assume that as much investor capital as possible was tied-up during the month; purchases occurred at the beginning of the month and sales at the end. This will tend to bias net returns towards 0%. To compute this "low leverage" active return, we take the weighted average return on the portfolio of stocks j held at the beginning of the month and the portfolio of stocks bought during the month, where returns on stocks sold or bought during the month reflect partial-month returns. The resulting expression is given by Equation 1 below.

$$R_t^{active} = \frac{\sum_j (HoldingValue_{jt} + SalesValue_{jt})}{\sum_j (HoldingValue_{j,t-1} + PurchaseValue_{jt})} \tag{1}$$

Next, we assume that as little investor capital as possible was tied-up during the month; purchases occurred at the end of the month and sales at the beginning. This "high leverage" approach, given by Equation 2, will bias net returns away from 0%. Equation 2 is poorly behaved for account-months where starting and ending balances are very small relative to the purchase and sales values that occur during the month, so we drop account-months for which sales and purchases combined exceed ten times the account value at the beginning of the month (about 0.5% of all account months). There is really very little we can say with our data about the returns received by habitual day-traders.

$$R_t^{active*} = \frac{\sum_j (HoldingValue_{jt} + max(0, SalesValue_{jt} - PurchaseValue_{jt}))}{\sum_j (HoldingValue_{j,t-1} + max(0, PurchaseValue_{jt} - SalesValue_{jt}))} \tag{2}$$

Figure A9 shows that the age effects from our baseline account returns regression are similar whether passive returns or either form of active returns is used.

3.3 Controlling for Time-Variation in the Inherent Sophistication of the Average Individual Investor

Our baseline specification, Equation 3 below, implicitly assumes that the inherent sophistication of the average individual investor in the Indian market is constant, i.e. $s_t = 0$.

$$Y_{it} - Y_t = s_i + \beta(A_{it} - A_t) + \gamma(X_{it} - X_t) + \varepsilon_{it}. \tag{3}$$

It is possible that the average inherent sophistication of Indian investors has been changing over time as market participation expands. To address this possibility, we model these changes in s_t using the cross-sectional average of a set of investor characteristics C_t, resulting in Equation 4 below.

$$Y_{it} - Y_t = (s_i - \alpha C_t) + \beta(A_{it} - A_t) + \gamma(X_{it} - X_t) + \varepsilon_{it} \tag{4}$$

The investor characteristics in C include the (log) value and number of stock positions when the account was opened, the literacy rate and log income level of the state where the account was opened, and dummies indicating if the account was opened in a rural or urban area. Note that while the investor level set of these characteristics, C_i, may be a very noisy proxy for an individual's inherent sophistication, the cross-sectional mean of the characteristics C_t may yet provide a good proxy for time-variation in the average inherent sophistication of investors.[2]

Figure A10 shows plots of the fitted series $-\alpha C_t$, which represent changes in average inherent sophistication of investors over time. For most investor behavior series, the model suggests a modest decrease in inherent investor sophistication, which may be understandable in a market with a steadily increasing participation rate. However, the average inherent disposition effect grows dramatically over time. This is attributed to the fact that the average investor in later years opens their account with fewer stock positions, and such investors exhibit far greater disposition effect.

Figure A11 through A13 provide the age effects from regressions using Equation 4 alongside age effects from our baseline Equation 3. Consistent with the results in Figure A10, the age effects generally attenuate modestly, with the exception of the disposition effect, for which the age effect is absorbed by controls for average inherent investor sophistication. Table A4 shows that the relationship of account returns to account behaviors and stock characteristics does not change meaningfully.

Account age is a characteristic which only varies across-cohorts at a given point in time, but our feedback measures primarily vary *within* cohorts. Since only the cross-cohort variation can be potentially explained by changes in average inherent investor sophistication, our estimation of feedback effects is virtually unaffected.

3.4 Measuring Significance of Violations of Strict Exogeneity

Panel estimation with fixed effects can deliver biased estimates when explanatory variables are not strictly exogenous. Intuitively, if the time dimension of the panel is short, and if high values of Y_i early in the sample predict high future values of X_i, then relative to its sample mean Y_i must be low later in the sample, and will spuriously appear to be negatively predicted by X_i. This is a particular problem if we use account size as an explanatory variable to predict returns, since account size is mechanically driven by past returns. Similar issues may arise when we use investment behaviors or style tilts as explanatory variables, if their prevalence is behaviorally influenced by past returns.

As an alternative, we consider Equation 5 below, which restricts individual effects to the span of account characteristics C. These are the same account characteristics whose cross-sectional averages are used to model average inherent investor sophistication in Equation 4

[2] Of course, if the number of characteristics in C equals the time-dimension of our data, C will span s_t, but we lose identification.

from the last section. While Equation 5 loses the potentially important ability to control for account-specific propensities towards the behaviors and styles, by removing the individual fixed effect we no longer need to worry about violations of strict exogeneity.

$$Y_{it} - Y_t = \theta(C_i - C_t) + \beta(A_{it} - A_t) + \gamma(X_{it} - X_t) + \varepsilon_{it} \tag{5}$$

Figures A14 and A15 show that the response of investor behavior and net style demands to feedback are qualitatively similar when we use Equation 5. Table A5 shows that violations of strict exogeneity appear to lead to an overstatement of the returns to value investing (a strong portfolio value tilt is associated with relatively poor past returns), but otherwise conclusions are similar to our baseline model.

However, even account age is vulnerable to violations of strict exogeneity if the disposition effect – the tendency of investors to sell gains rather than losses – leads to disproportionate exit of investors who have been lucky (Calvet et al. 2009a). In this case, experienced investors may disproportionately be investors who had poor returns when they were novices.

To respond, we model the relationship of account exit to past returns and use this to simulate the survival bias in account age effects using our primary specification (Equation 3). Results of the simulation are given in Table A6. The account exit rate is too modest, and too weakly related to past returns for our inferences to be affected significantly. As some further evidence, Table A5 shows that age effects are similar under Equation 5.

Table A1: 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]	**0.33**	1.00								
Account Stock Return Over the Past Year	[3]	0.04	0.10	1.00							
Idiosyncratic Share of Portfolio Variance	[4]	**-0.17**	**-0.45**	0.01	1.00						
Turnover Over the Past Year	[5]	**-0.38**	**-0.31**	0.00	**0.33**	1.00					
Disposition Effect Over the Past Year	[6]	**-0.13**	**-0.15**	0.02	0.03	0.02	1.00				
Stock Portfolio Beta	[7]	**-0.10**	**-0.15**	-0.04	**0.18**	**0.19**	0.05	1.00			
Size Percentile of Stocks Held	[8]	0.02	**0.14**	0.06	**-0.30**	**-0.14**	-0.05	**-0.35**	1.00		
Book-Market Percentile of Stocks Held	[9]	-0.02	**-0.11**	**-0.11**	**0.15**	0.07	0.06	**0.15**	**-0.47**	1.00	
Momentum Percentile of Stocks Held	[10]	0.05	**0.19**	**0.29**	-0.06	0.00	**-0.15**	-0.04	**0.16**	**-0.27**	1.00
Urban Account	[11]	0.05	0.07	0.01	-0.03	-0.02	-0.02	0.00	0.00	-0.01	0.02
Semi-Urban Account	[12]	0.01	-0.02	0.00	0.00	0.00	0.01	0.01	-0.02	0.02	0.00
Rural Account	[13]	-0.06	-0.06	-0.01	0.02	0.02	0.01	-0.01	0.01	-0.01	-0.02

Table A2: Account Age Effects in Equity Portfolio Returns, Including Accounts Opened Prior to Feb. 2002

The regressions below are run identically to those in Table 3, except specifications [a'], [b'], and [c'] *include* accounts opened prior to February 2002. For these older accounts, account age (i.e. first month of stock investment) is imputed as three months following the account opening date.

Dependent Variable: Account Monthly Return in Excess of Risk-Free Rate (bp) (Mean: 96.9bp)

		[a]	[a']	[b]	[b']	[c]	[c']
Account Age Effect	Account Age (Linear)	*12.01*	9.09			8.25	6.68
		(7.10)	*(5.56)*			*(6.94)*	*(5.31)*
	Piecewise Linear			See Figure A7			
Investor Behavior	Lagged Idio. Share of Portfolio Var.					55.85	34.05
						(76.06)	*(73.49)*
	Lagged Portfolio Turnover					-97.64	-87.24
						(68.57)	*(64.13)*
	Lagged Disposition Effect					**-3.69**	**-4.06**
						(1.50)	*(1.43)*
	Size Tilt					177.90	118.54
						(181.70)	*(179.27)*
	Value Tilt					**554.30**	**493.57**
						(113.36)	*(108.29)*
	Momentum Tilt					-22.20	-5.75
						(122.11)	*(124.57)*
Incremental R^2		0.00031	0.00018	0.00029	0.00029	0.00015	0.00010

Table A3: Decomposition of the Difference in Returns on Old and New Accounts

In column [4], the analysis from Table 4 is replicated for a zero cost portfolio formed from the difference in portfolio weights between accounts opened prior to February 2002 and the oldest quintile of accounts opened on/after February 2002. The properities of portfolios formed from the difference in oldest and newest accounts opened on/after Februrary 2002 (a copy of Table 4 column [1]) is provided for comparison.

Zero-Cost Portfolio Represents: Portfolio Tilts (1000 x φ_{bar})	Oldest minus Newest [1]	Pre 2002 Accounts minus Oldest [4]
Market beta	-0.547	**-0.697**
	(0.568)	*(0.274)*
Market capitalization	-0.318	**-0.601**
	(0.233)	*(0.099)*
Book-market	0.171	**-0.735**
	(0.143)	*(0.200)*
Momentum (t-2:t-12 returns)	-0.003	-0.266
	(0.340)	*(0.167)*
Stock turnover	**-0.908**	1.067
	(0.262)	*(0.791)*
Beneficial ownership	*-0.604*	0.614
	(0.367)	*(0.519)*
Institutional ownership	**0.919**	**0.494**
	(0.356)	*(0.163)*
Ln(1+stock age)	0.010	**0.546**
	(0.075)	*(0.208)*
Large IPOs (market cap if age<1 year)	**-13.358**	0.447
	(3.723)	*(0.327)*
Return Decomposition		
Stock characteristic selection	8.52	-0.98
	(5.54)	*(8.55)*
Additional stock selection	12.90	-2.34
	(14.55)	*(3.64)*
Stock characteristic timing	-9.63	-0.31
	(11.13)	*(2.35)*
Additional stock timing	26.60	-3.39
	(21.24)	*(4.00)*
Total difference in returns	38.40	-7.02
	(28.34)	*(10.79)*

Table A4: Account Age Effects in Equity Portfolio Returns - with control for Cross-Sectional Average Innate Investor Sophistication

Columns [a'], [b'], and [c'] report coefficients from regression equation 4, where we add controls for cross-sectional average inherent investor sophistication (i.e. ability to invest profitably) to our baseline equation 3 estimates shown in Table 3. Cross-sectional average (innate) sophistication is proxied by the vector Ct, which includes the average state income, state literacy rate, initial account size and value, and rural/urban population of post-January 2002 individual equity investors present in the given month.

Dependent Variable: Account Monthly Return in Excess of Risk-Free Rate (bp) (Mean: 96.7bp)

		[a]	[a']	[b]	[b']	[c]	[c']
Account Age Effect	Account Age (Linear)	*12.01*	11.58			8.25	10.15
		(7.10)	*(10.39)*			*(6.94)*	*(10.79)*
	Piecewise Linear			See Figure A13			
Investor Behavior	Lagged Idio. Share of Portfolio Var.					55.85	-16.71
						(76.06)	*(74.38)*
	Lagged Portfolio Turnover					-97.64	-81.60
						(68.57)	*(67.65)*
	Lagged Disposition Effect					**-3.69**	**-3.69**
						(1.50)	*(1.49)*
	Size Tilt					177.90	198.08
						(181.70)	*(180.35)*
	Value Tilt					**554.30**	**540.00**
						(113.36)	*(112.75)*
	Momentum Tilt					-22.20	1.54
						(122.11)	*(10.74)*
Controls for (Cross-Sectional) Average Innate Investor Sophistication		No	Yes	No	Yes	No	Yes
Incremental R^2		0.00031	0.00029	0.00039	0.00043	0.00015	0.00022

Table A5: Account Age Effects in Equity Portfolio Returns - Robustness Check for Exogeneity Bias

Specifications [a'] and [c'] below are based on Equation 5 in the text: $(R_{it}-R_t)=\theta(C_i-C_t)+\beta(A_{it}-A_t)+\gamma(X_{it}-X_t)+\varepsilon_{it}$ (i.e. a set of 12 investor characteristics C_i described in the text proxy for the individual i's inherent sophistication). All other aspects of the regression and data used are identical across the two regression specifications.

Dependent Variable: Account Monthly Return in Excess of Risk-Free Rate (bp) (Mean: 96.7bp)

		From Table 3		Restricted Individual Effects	
		[a]	[c]	[a']	[c']
Account Age Effect	Account Age (Linear)	*12.01*	8.25	**12.34**	**9.83**
		(7.10)	*(6.94)*	*(4.73)*	*(4.51)*
	Piecewise Linear				
Investor Behavior	Lagged Idio. Share of Portfolio Var.		55.85		53.04
			(76.06)		*(60.33)*
	Lagged Portfolio Turnover		-97.64		*-117.48*
			(68.57)		*(63.86)*
	Lagged Disposition Effect		**-3.69**		-0.99
			(1.50)		*(1.48)*
	Size Tilt		177.90		227.04
			(181.70)		*(178.50)*
	Value Tilt		**554.30**		**298.79**
			(113.36)		*(91.73)*
	Momentum Tilt		-22.20		105.67
			(122.11)		*(126.57)*
Incremental R^2		0.00031	0.00015	0.00020	0.00012

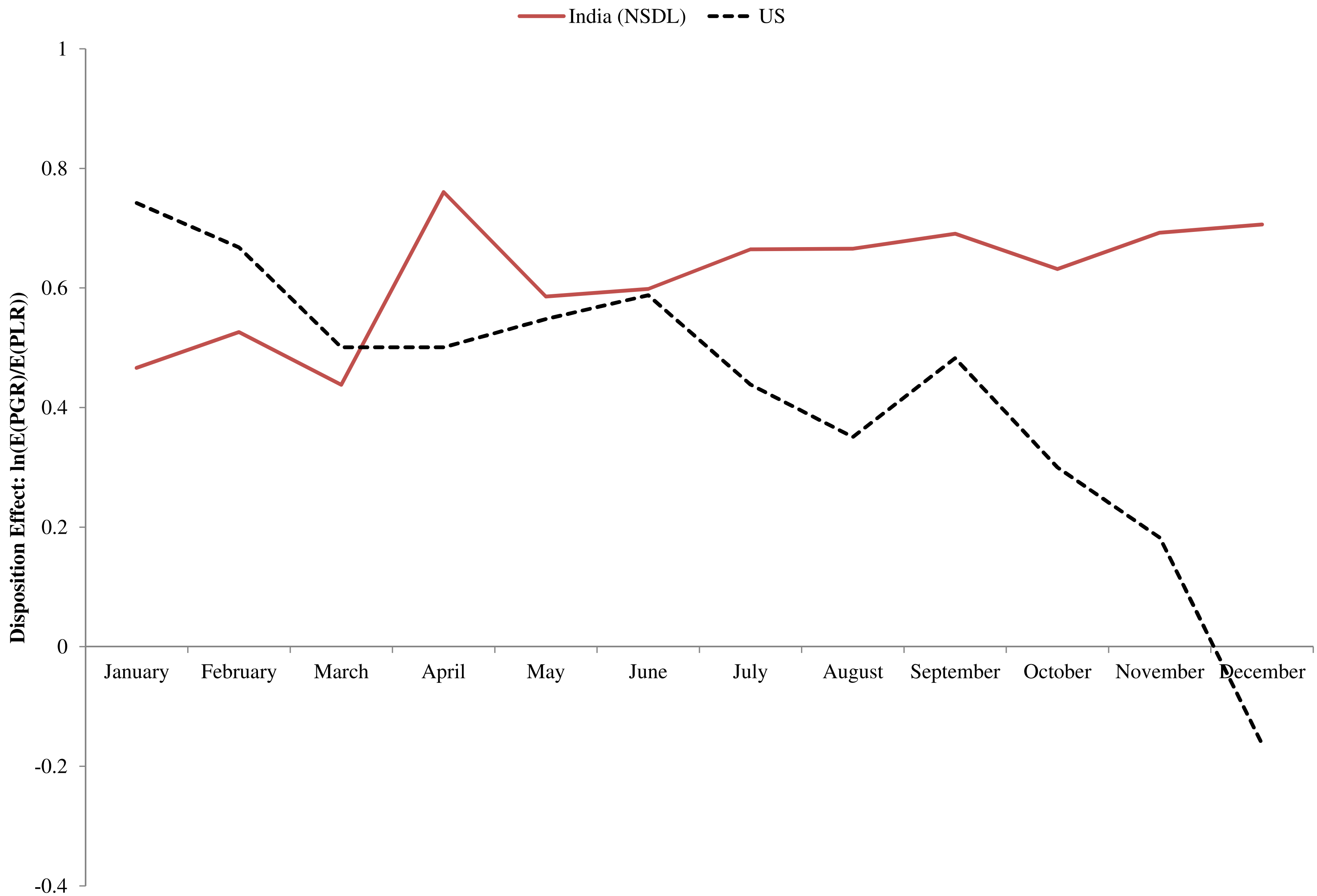
Table A6: Bias from Performance (Luck) Driven Exits in Return Regressions with Investor Fixed Effects

We use data from January 2004 through July 2011 to estimate a logit model where account exit (liquidation of all stockholdings with no subsequent stock purchases) is predicted by lagged account average outperformance versus the mean individual investor. Next, simulated investors (from five cohorts, one year apart) are assigned to idiosyncratic return volatility percentiles, with returns bootstrapped from the corresponding percentile of accounts in the sample. There are no age or investor sophistication effects in the simulation, which does not affect results. Exits are simulated for these investors using the logit model and our regressions of returns on account age with investor fixed effects is run. In variants [2], we add monthly fixed effects to the estimated logit model and simulation. In variant [3], we increase the sensitivity of exits to returns by five standard errors above that used in the baseline, while holding constant the unconditional exit rate. The simulation is run repeatedly to obtain the standard errors in ( ) on the estimated biases.

Estimated bias in Return Regression Specification [1a] - monthly returns (bp) per year of account age (estimated at 12.01bp/mo in Table 3)

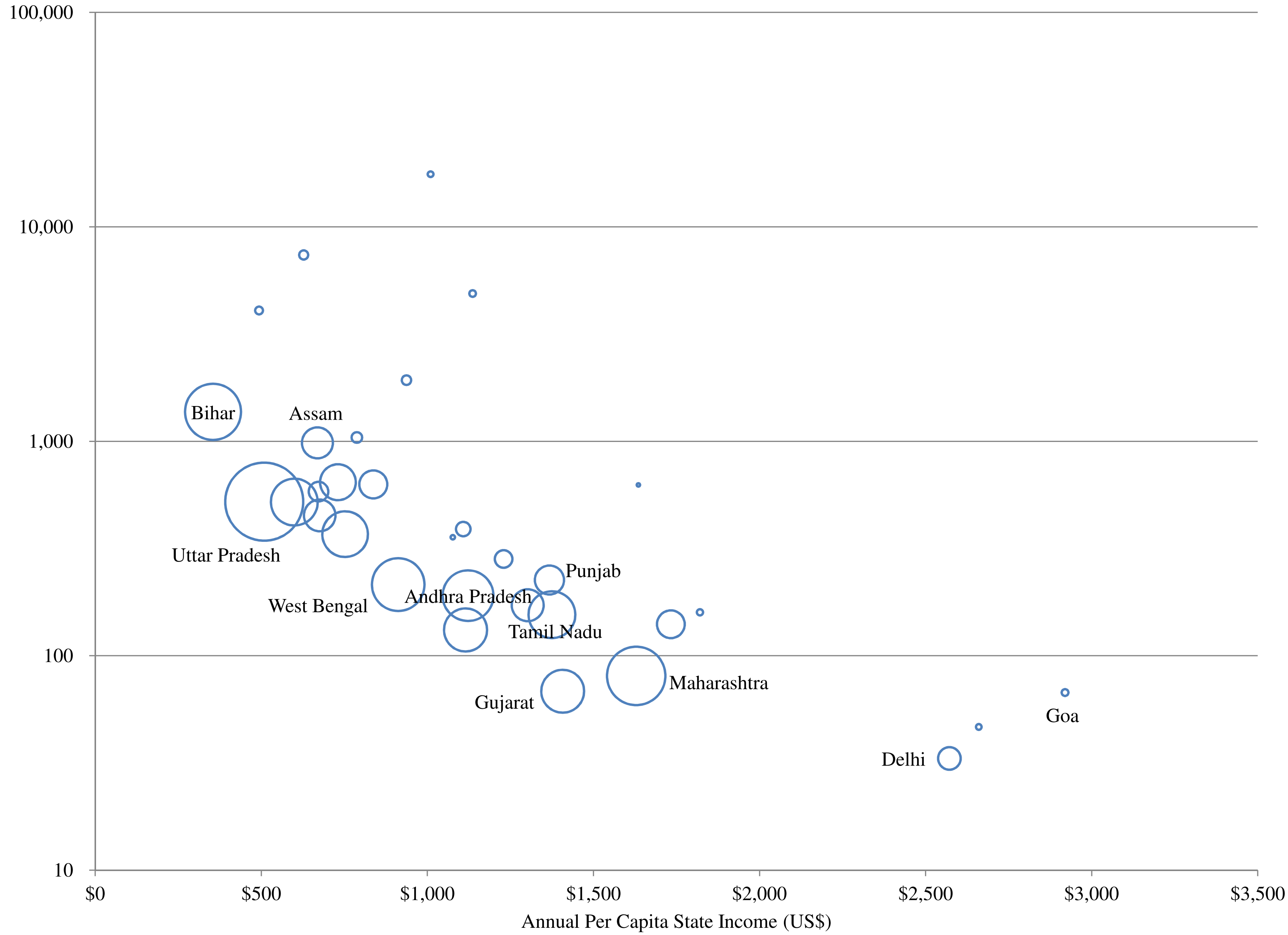
Baseline	[1]	0.74
		(0.03)
Logit model and simulations include monthly fixed effects	[2]	0.66
		(0.03)
Sensitivity of exits to performance equals logit model estimate plus five standard errors	[3]	1.30
		(0.06)

Figure A1: Disposition Effect for Individual NSDL Accounts vs US Discount Brokerage Accounts (Odean 98, 99)



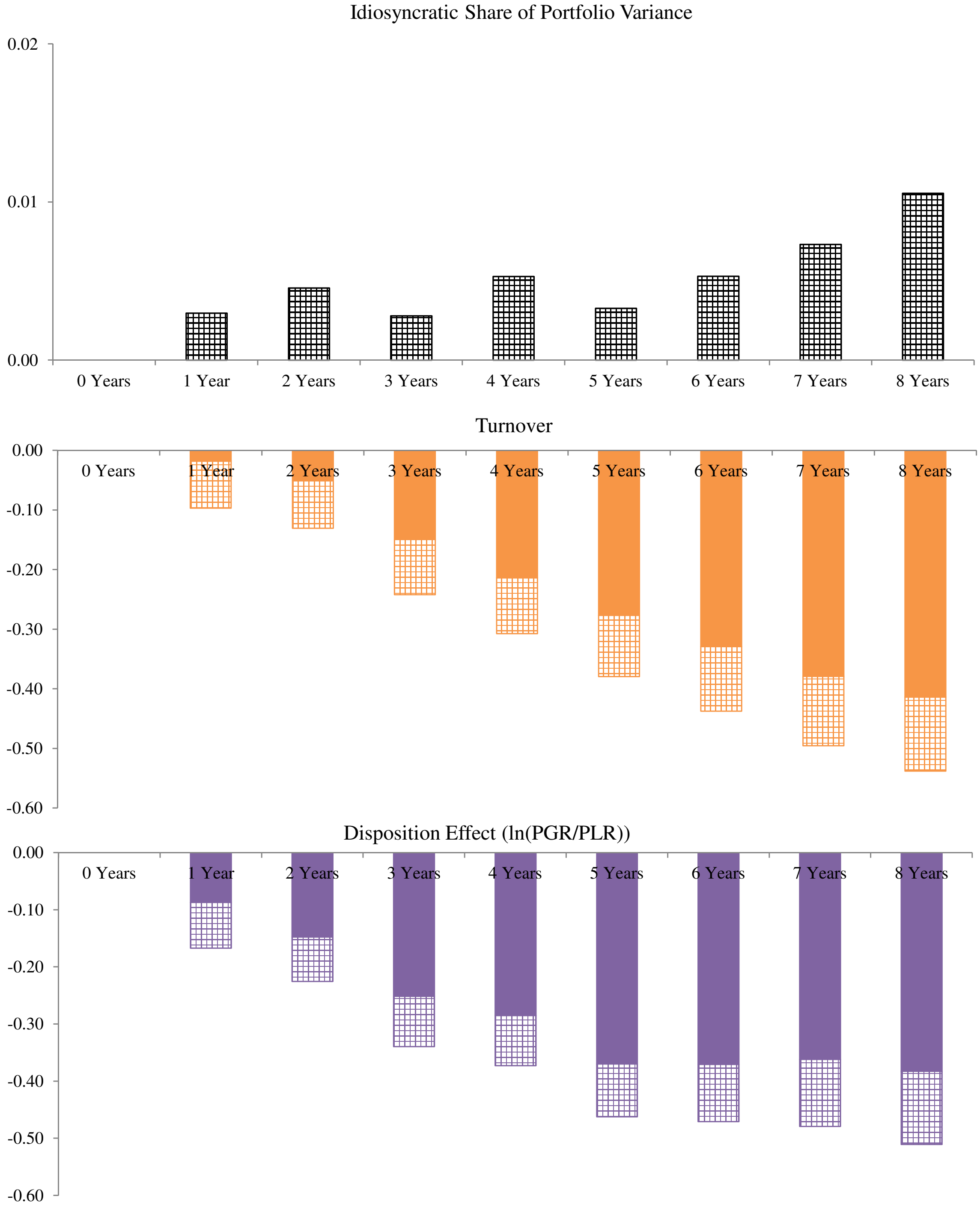
PGR/PLR is computed for each month, aggregating across accounts and years as in Odean (1998), which is the source for the US brokerage based statistics plotted. We take the log of both measures to make the units somewhat more comparable to our cross-sectional analysis. However, aggregating PGR and PLR across accounts (and implicitly weighting accounts with more trades more heavily) means that levels here are still not comparable to our account level disposition measure used elsewhere.

Figure A2: 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.

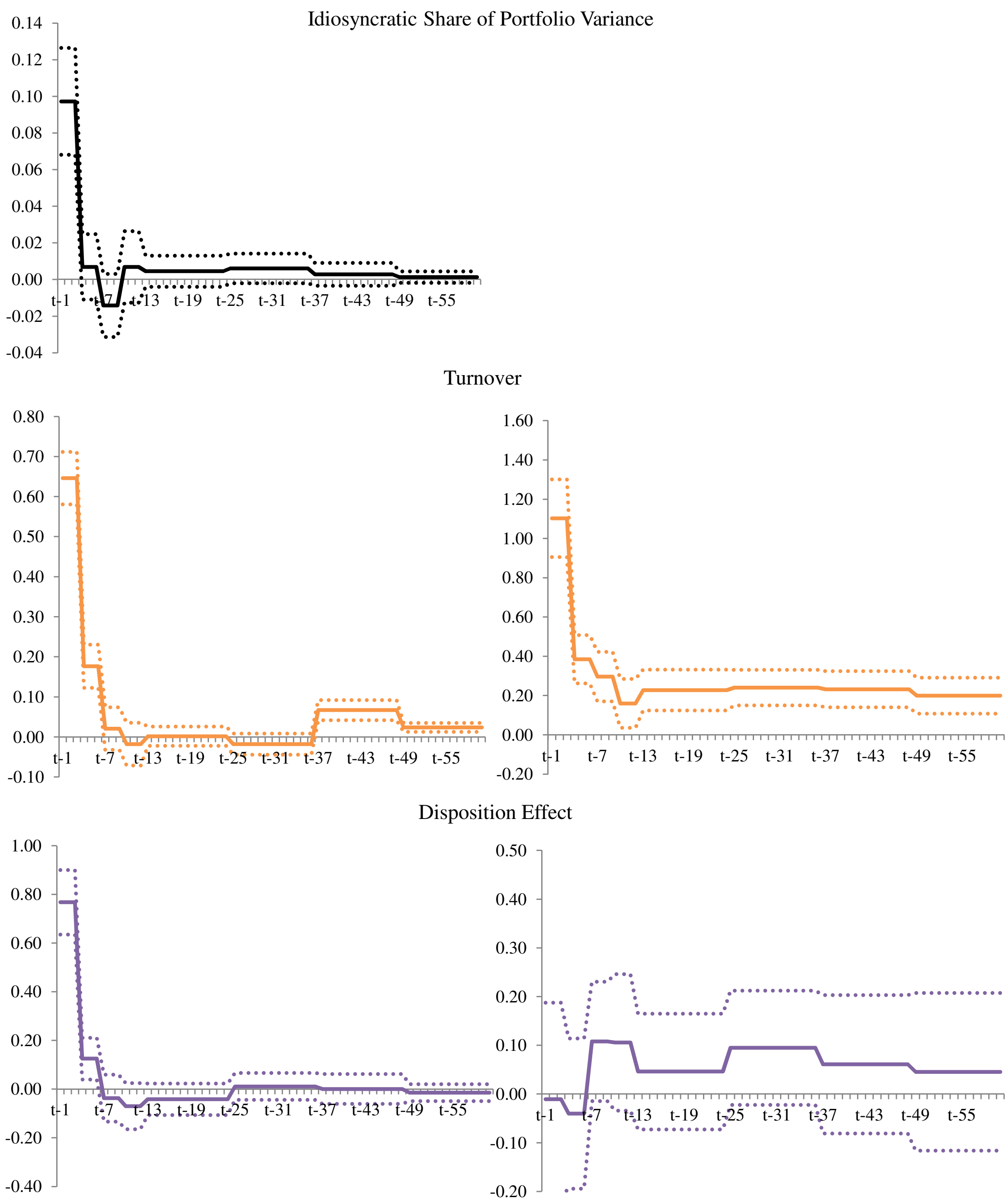
Figure A3: Account Age Effects for Investor Behaviors, Including Accounts Opened Prior to Feb. 2002



The plots above are produced in the same manner as in Figure 2, but include accounts opened prior to February 2002 in the scaling and regression model used.

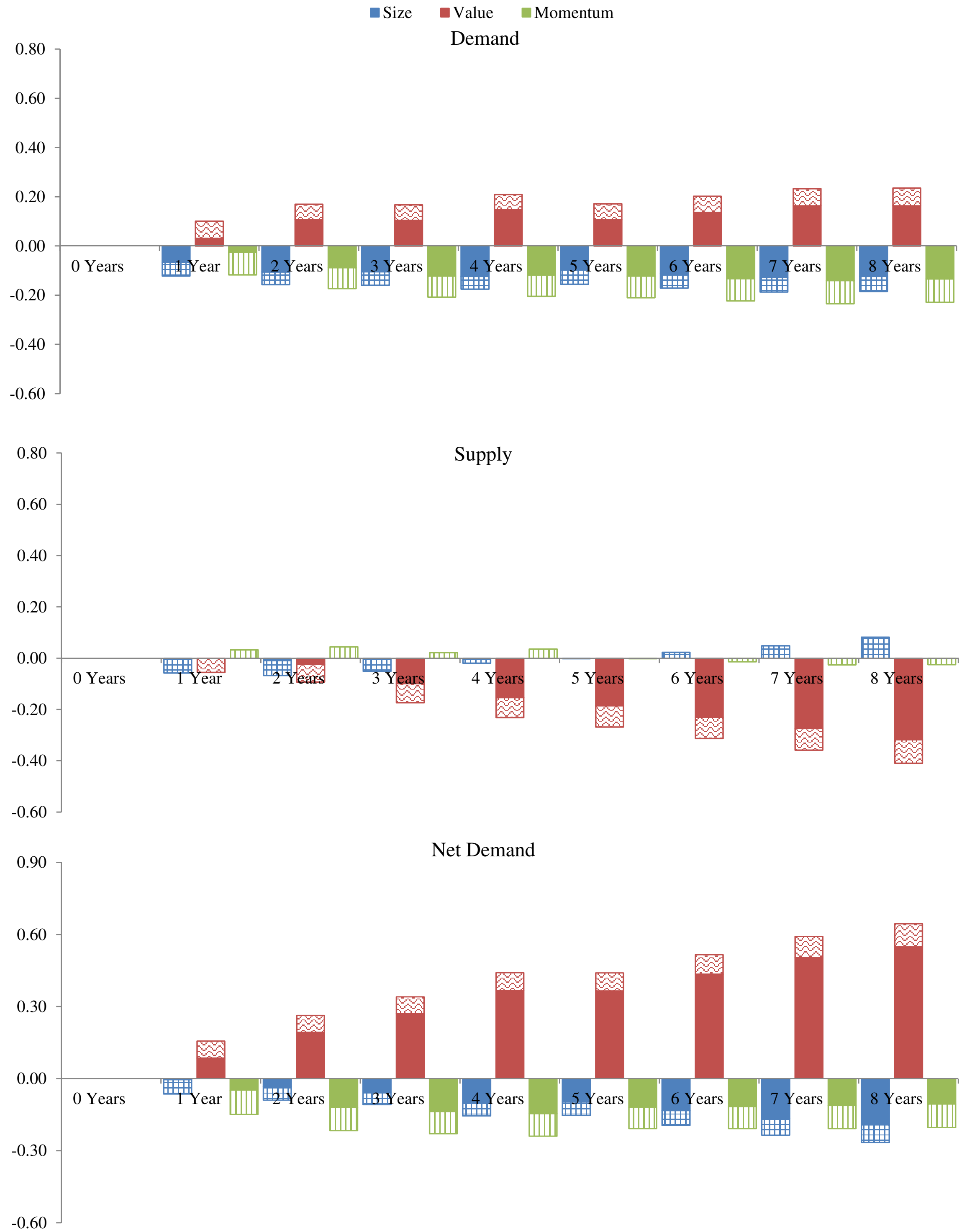
Figure A4: Investing Behavior Response to Feedback at Various Horizons, Including Accounts Opened Prior to Feb. 2002

Account Performance Feedback (Left) and Behavior Feedback (Right)



These plots are generated in the same was as Figure 3, but use accounts opened prior to February 2002 in the regression model and scaling applied.

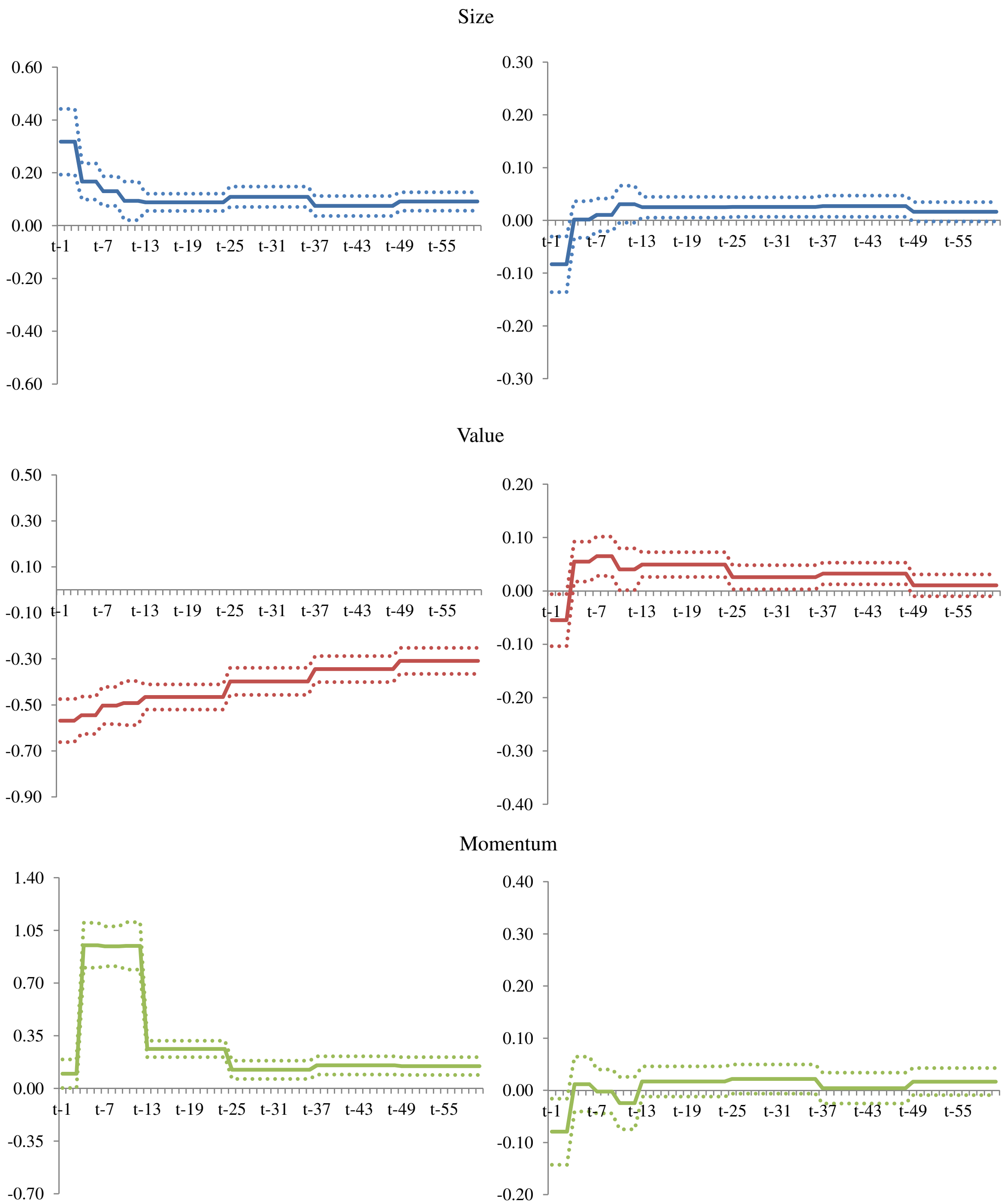
Figure A5: Account Age Effects for Style Characteristic Supply/Demand, Including Accounts Opened Prior to Feb. 2002



These plots are generated in the same was as Figure 5, but use accounts opened prior to February 2002 in the regression model.

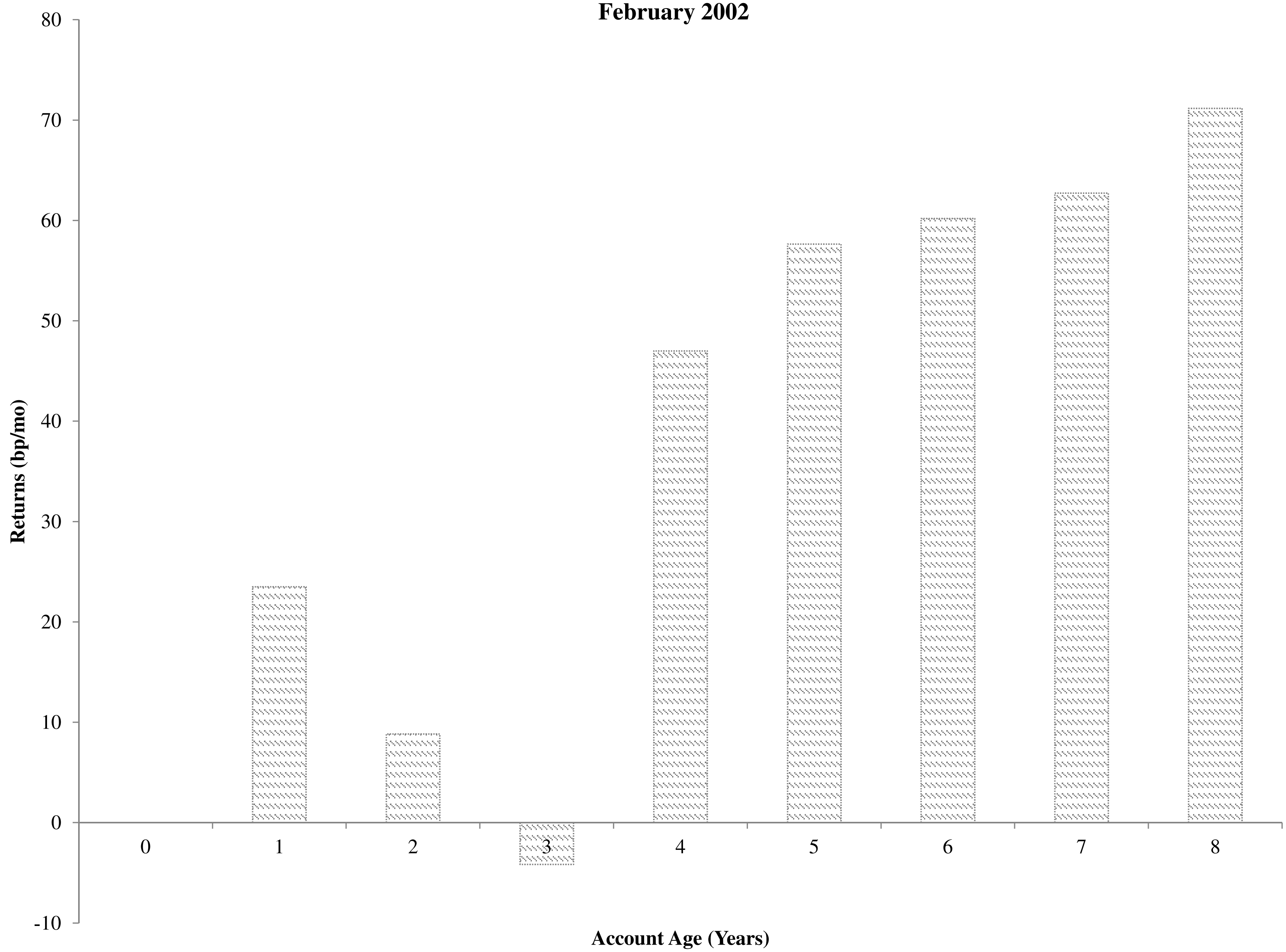
Figure A6: Net Characteristic Demand Response to Feedback at Various Horizons, Including Accounts Opened Prior to 2002

Account Performance Feedback (Left) and Style Feedback (Right)



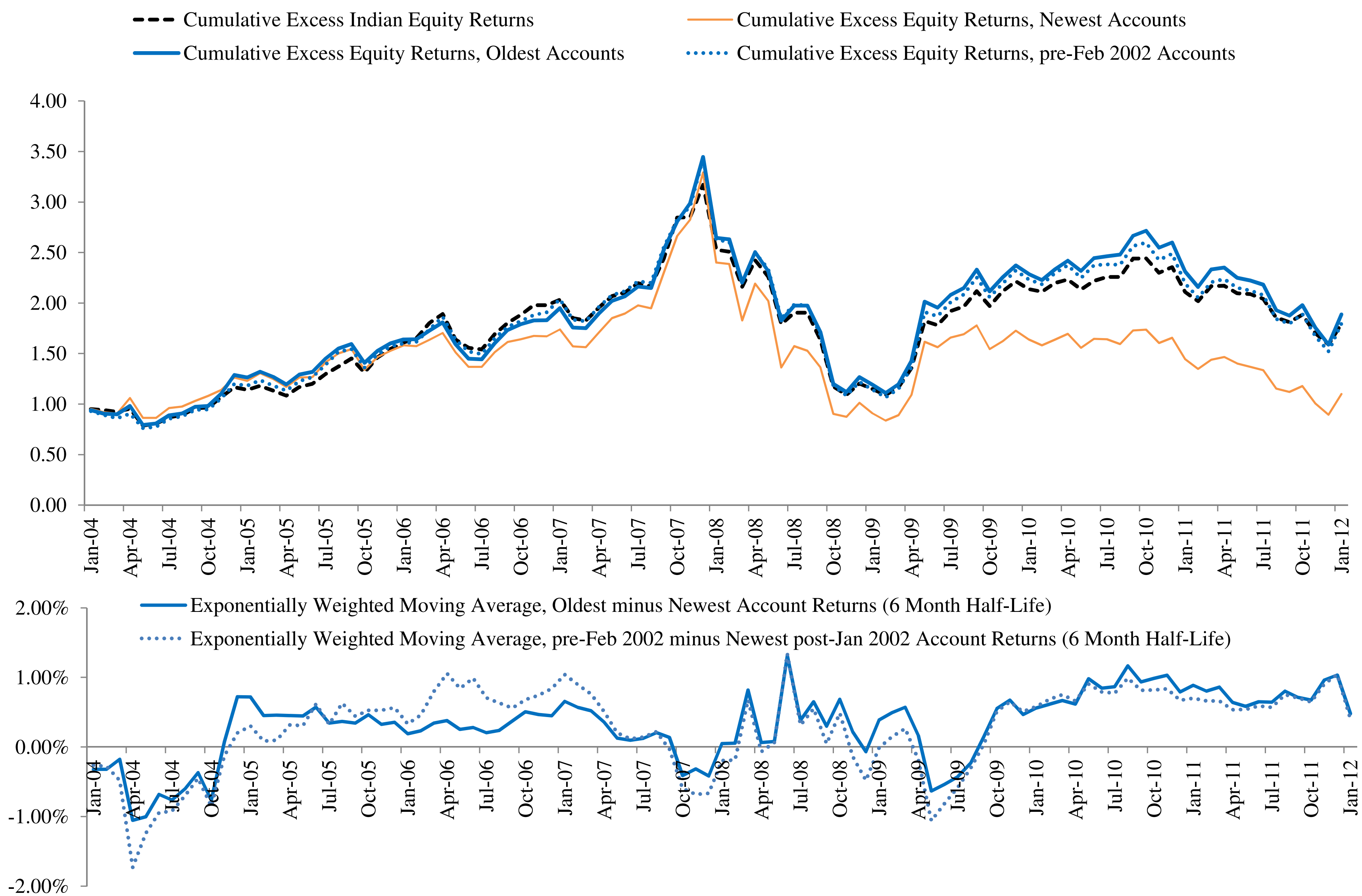
These plots are generated in the same was as Figure 6, but use accounts opened prior to February 2002 in the regression model.

Figure A7: Account Age Effects in Equity Portfolio Returns, Including Accounts Opened Prior to February 2002



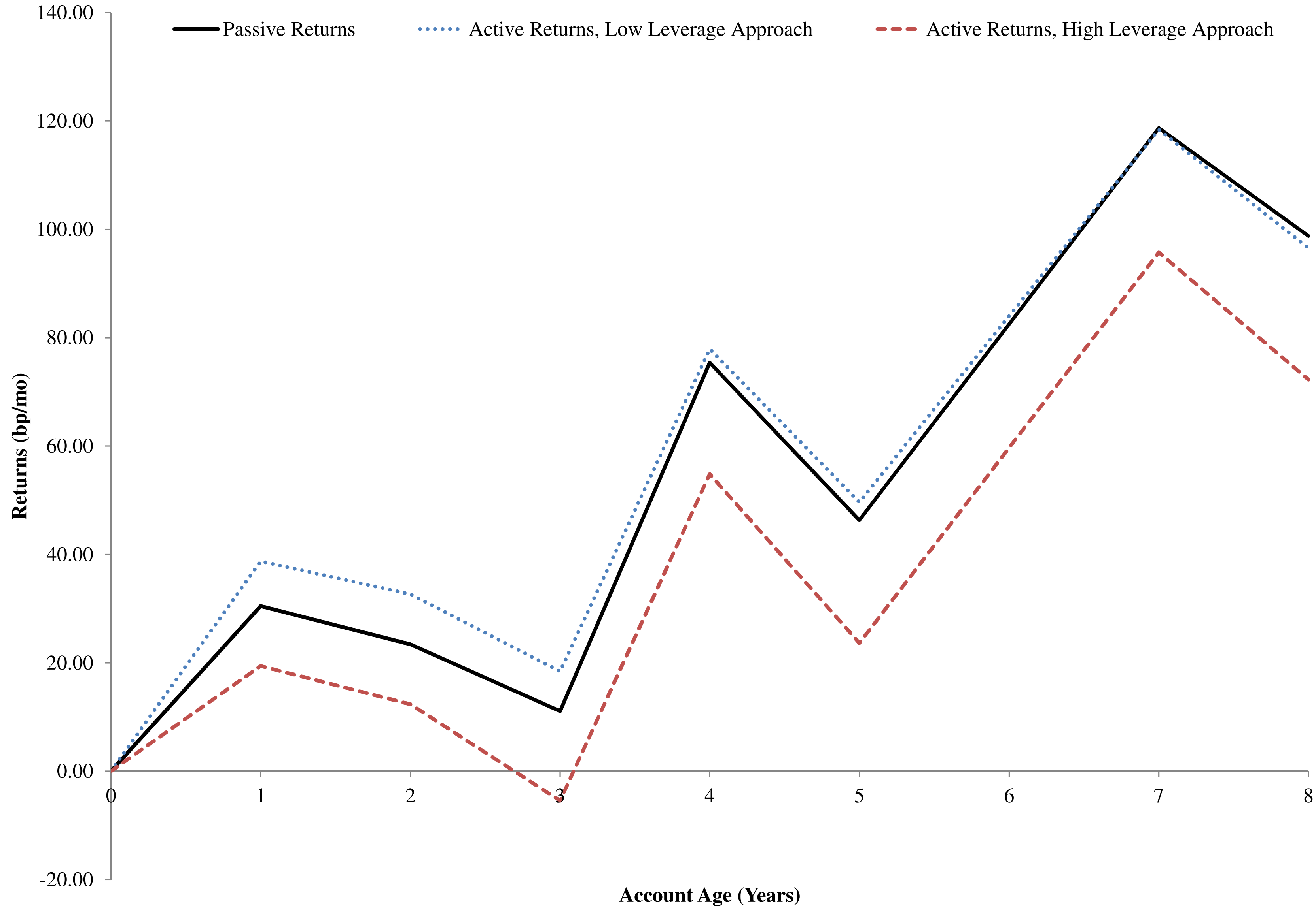
The plots above is generated identically to Figure 8, but the regression model used includes accounts opened prior to February 2002.

Figure A8: Cumulative Indian Excess Equity Return Received by Oldest and Newest Quintile of Accounts



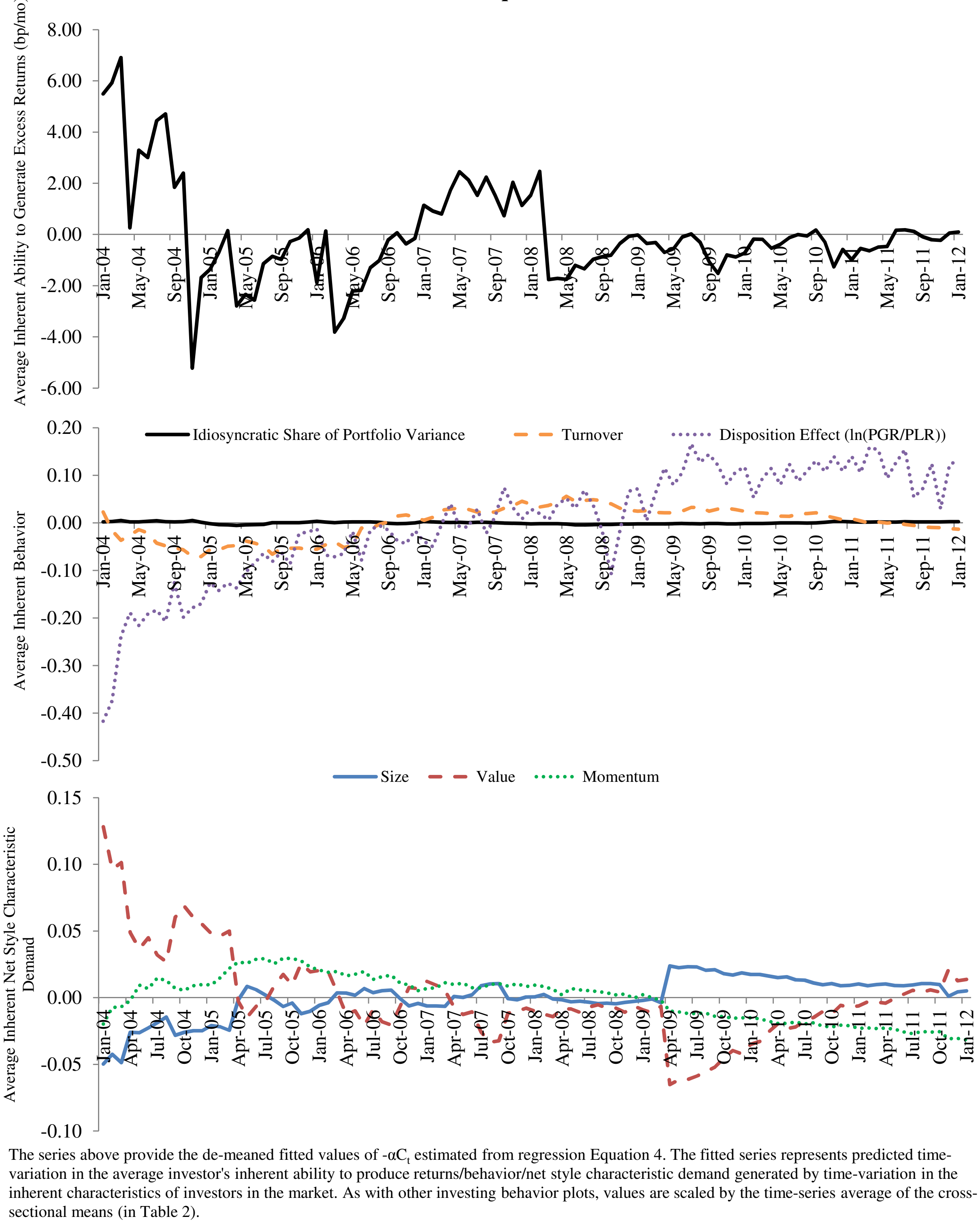
Representative portfolios of account-age based groups are formed by equally weighting across the actual portfolio weights of accounts in each age group. The "oldest" and "newest" groups represent the oldest and newest quintile of accounts present in each month from within the set of accounts started on or after February 2002. The top plot produces excess returns by subtracting the yield on three-month Indian Treasury bills from returns on the indicated portfolios.

Figure A9: Account Age Effects for Passive (Baseline) and Active Equity Portfolio Returns



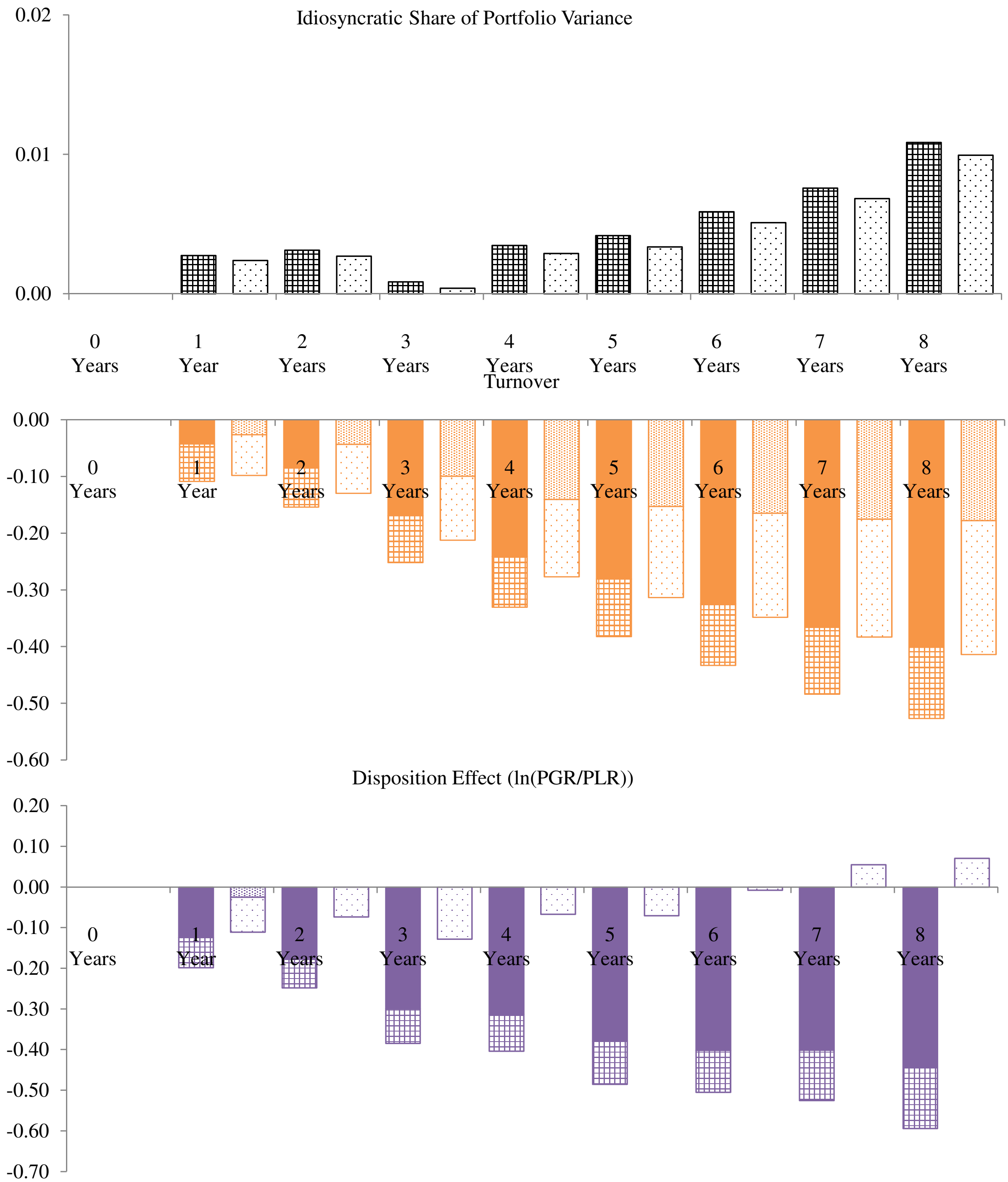
The regressions generating these curves use both passive and active account excess returns as the dependent variable, The passive return curve is a bit different from that given by the bars in Table 8 since regressions used here are based on the (slightly different, more restrictive, set of account-months) for which active returns and passive returns are defined. See text for a discussion of how active returns are computed.

Figure A10: Estimated Cross-Sectional Average Inherent Sophistication from Equation 4



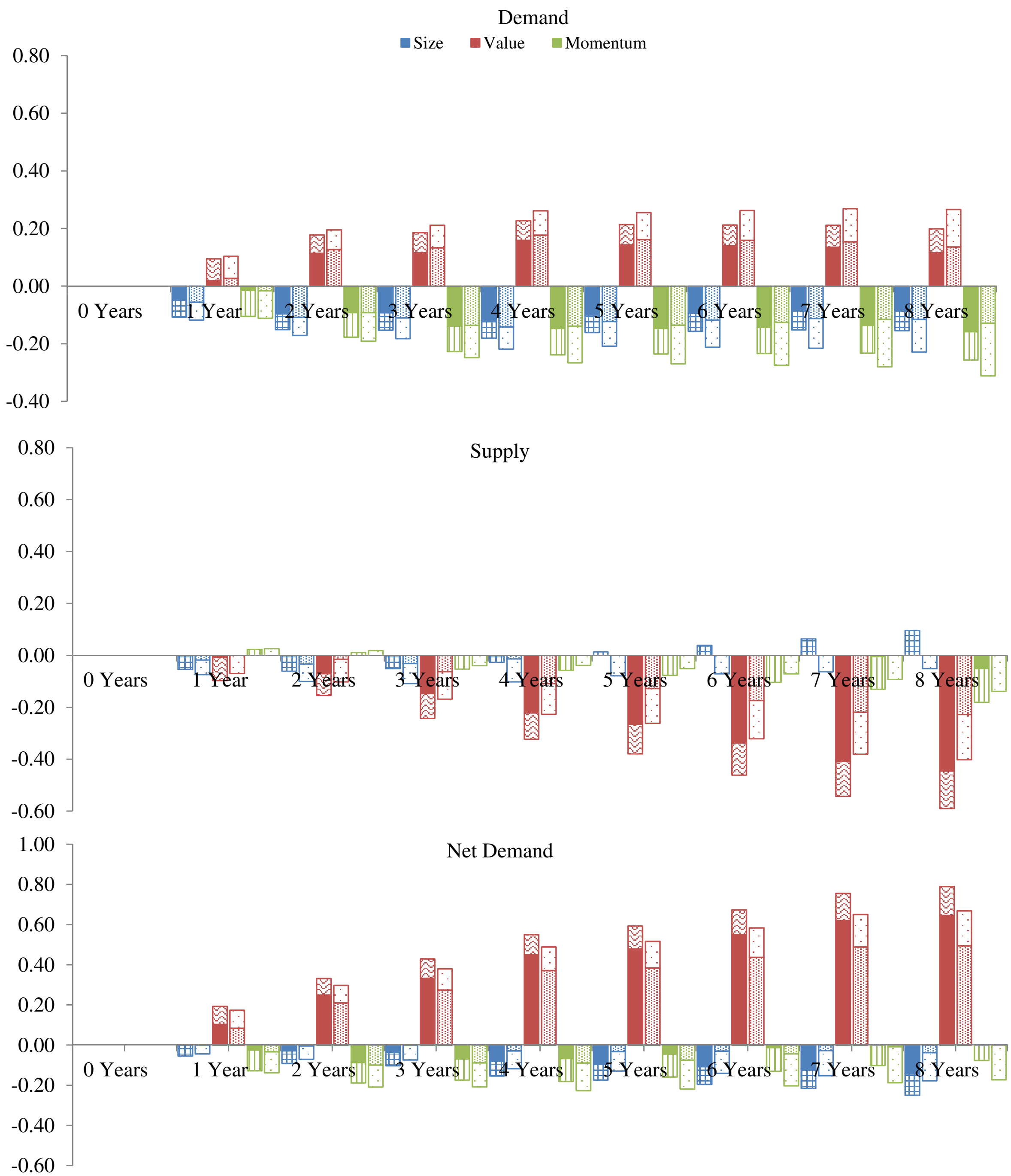
The series above provide the de-meaned fitted values of $-\alpha C_t$ estimated from regression Equation 4. The fitted series represents predicted time-variation in the average investor's inherent ability to produce returns/behavior/net style characteristic demand generated by time-variation in the inherent characteristics of investors in the market. As with other investing behavior plots, values are scaled by the time-series average of the cross-sectional means (in Table 2).

Figure A11: Account Age Effects for Investor Behaviors without (solid and cross-hatched) and with (dotted) control for Cross-Sectional Average Inherent Investor Sophistication



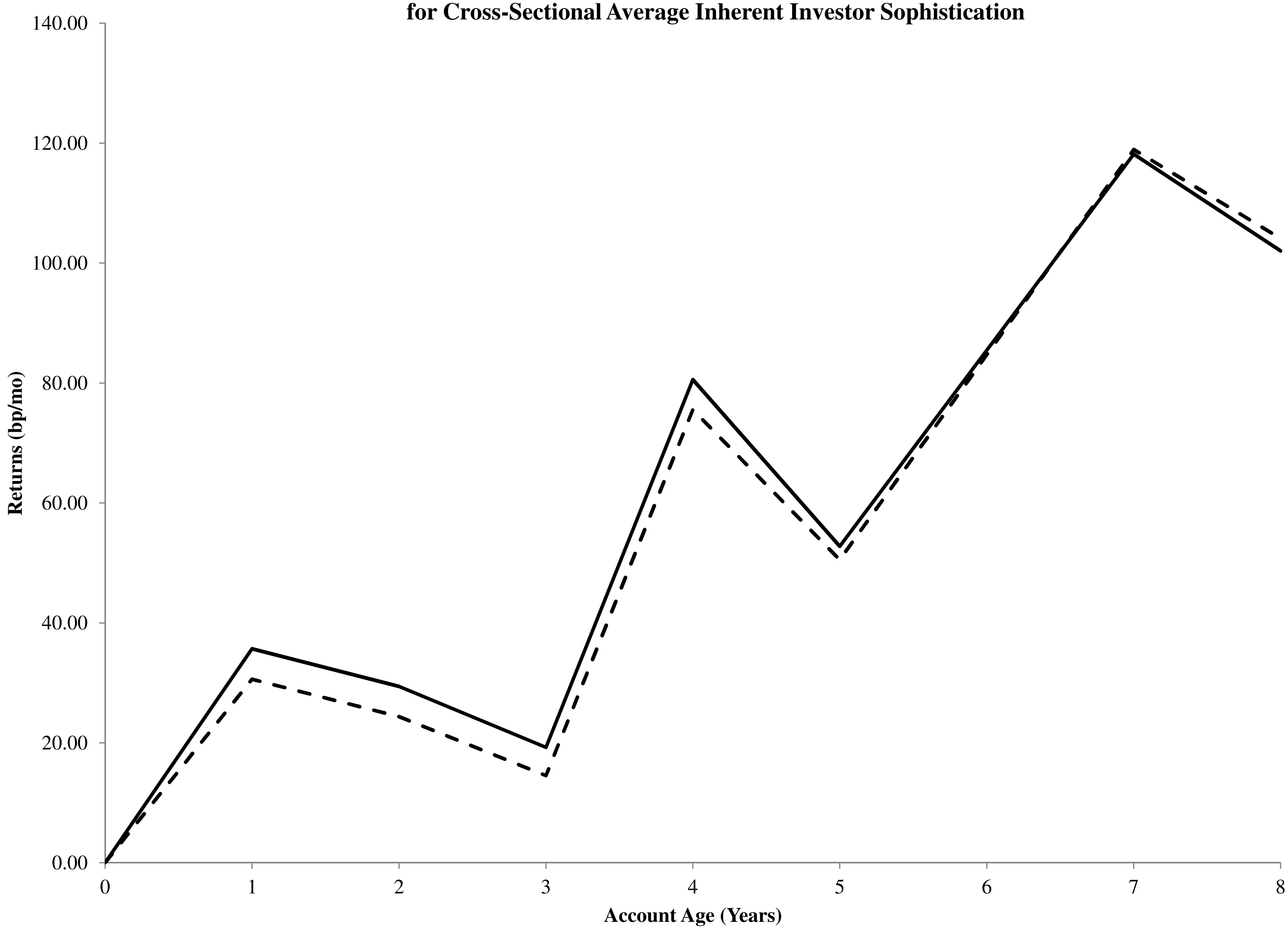
The plots above compare the age effects shown in Figure 2 (solid and cross-hatched bars) with those constructed when the regression specification is augmented with controls for the inherent sophistication of investors present in each month. See text for a discussion of these controls C_t.

Figure A12: Account Age Effects for Style Characteristic Supply/Demand
with (dotted) and without (solid and cross-hatched) control for Cross-Sectional Average Inherent Investor Sophistication



The plots above compare the age effects shown in Figure 5 (solid and cross-hatched bars) with those constructed when the regression specification is augmented with a set of controls for potential changes in the inherent sophistication of investors present in each month.

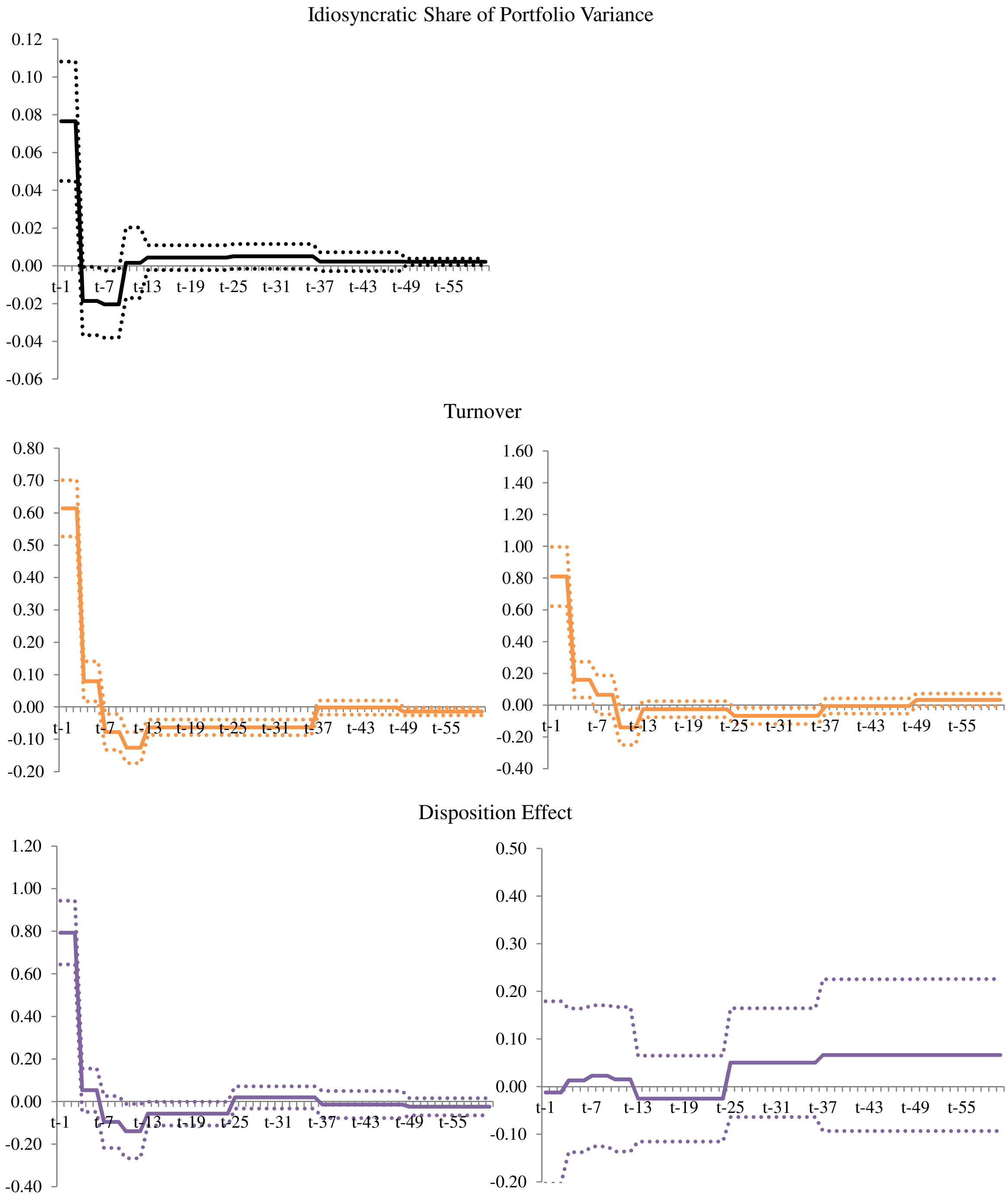
Figure A13: Account Age Effects in Equity Portfolio Returns - with (dashed) and without (solid) control for Cross-Sectional Average Inherent Investor Sophistication



The plotted curves represent account age effects produced when specifying the age effect as a piece-wise linear function with breakpoints at years one, two, three, four, five, and seven (Specifications [b] and [b'] in Table A4).

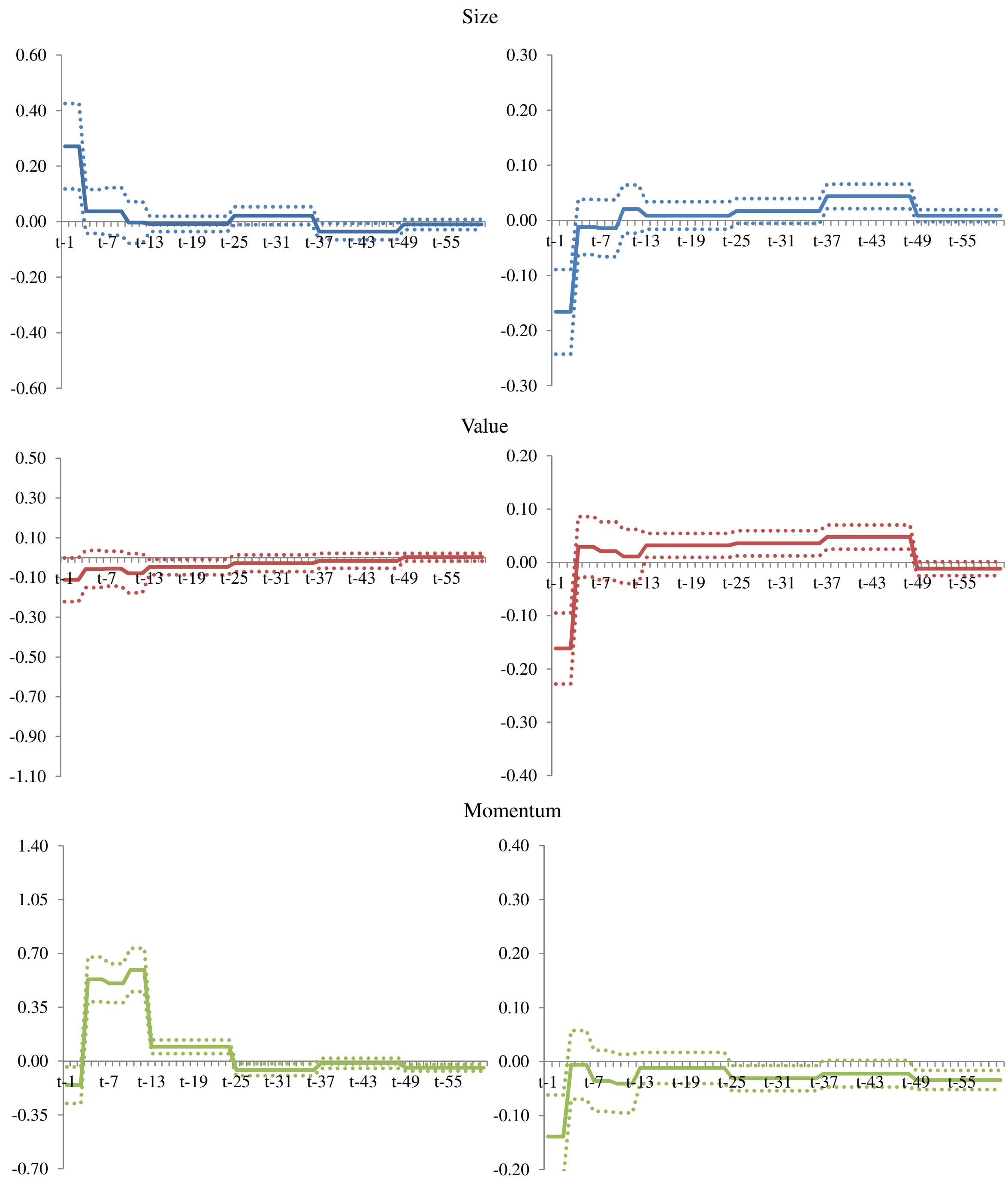
Figure A14: Investing Behavior Response to Feedback at Various Horizons from Equation 5 (Restricted Individual Effects)

Account Performance Feedback (Left) and Behavior Feedback (Right)



Plots are produced from account behavior regressions following Equation 5. Otherwise, the data and methods used to construct these plots are identical to Figure 3.

Figure A15: Net Characteristic Demand Response to Feedback at Various Horizons
Account Performance Feedback (Left) and Style Feedback (Right)



Plots are produced from account behavior regressions following Equation 5. Otherwise, the data and methods used to construct this plot are identical to Figure 6.