Internet Appendix to Getting Better or Feeling Better? How Equity Investors Respond to Investment Experience

John Y. Campbell, Tarun Ramadorai, and Benjamin Ranish[1]

This version: March 2014

[1]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.

Contents

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 reports data from both of India's major stock exchanges, the Bombay and National Stock Exchanges (BSE and NSE). In addition, monthly price returns can be inferred from the month-end holding values and quantities in the NSDL database. We link the datasets by ISIN.[2]

To verify reliability of total returns, we compare total returns from the available 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 returns match another dataset most closely, where the source from amongst those datasets 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 computing 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 appear 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 stock holdings for the representative individual investor for either the previous or current month.

The resulting data coverage is spotty for the smallest equity issues. Use of the returns we have on very small stocks could raise concerns that we are measuring returns for a non-representative set of small stocks. Therefore, in computing account returns, we use only stock-months where the aggregate holdings of that stock across all account types in NSDL is greater than 500 million Rs (approximately $10 million) at the end of the prior month. While this results in the loss of quite a few stock-months, the lost stock-months account for an average of only 2.3% of aggregate individual stock holdings, and about 6.5% of stock holdings for the representative individual account.[3]

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

[2] 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.

[3] Larger individual accounts have lower average portfolio weights in these excluded very small stocks.

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 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 Exercises, Explanations, and Extensions of Results

2.1 Indirect Individual Equity Ownership in India

Table A1 provides our estimate of the indirect share of individual stock ownership in India. We assume that indirect individual ownership occurs through mutual funds, unit trusts (state-sponsored mutual funds), and unit-linked insurance plans (insurance-investment plan hybrids popular in India).

We use comprehensive data from the Association of Mutual Funds of India to estimate the value of holdings in mutual funds and unit trusts. Funds classified as "growth" (called "equity" in some years) and "equity linked savings schemes" are assumed to be fully invested in stock, and funds classified as "balanced" are assumed to be invested half in stocks. Of these categories, "growth"/"equity" is by far the largest. We assume that individuals own a similar fraction of equity mutual funds and non-equity mutual funds, and obtain this fraction (which averages around 40%) from SEBI reports.[4]

We obtain the aggregate value of unit-linked insurance plan premiums from annual reports of the Insurance Regulatory and Development Authority. We assume that 50% of these premiums are invested in equity.

The value of equities held directly by individuals is extrapolated as 5/3 of that held by individual accounts registered in NSDL, based on NSDL having an approximately 60% share

[4] We use data from 2003 for 2004 through 2009, as we are unable to locate the figure for these intermediate dates.

of all such accounts.

2.2 Growth of Individual NSDL Accounts

Figure A1 plots the number of individual investors with NSDL accounts holding stock in each month. The number of investors increases rapidly along with stock prices from 2004 through late 2007. Following the market's decline, the growth in number of investors has been much lower.

2.3 Disposition Effect by Calendar Month - India versus US

Figure A2 provides a monthly measure of the disposition effect computed just as in Odean (1998), alongside Odean's measure based on US brokerage accounts. The start of the tax year in each country, January for the US and April for India, is signified by a square data point. 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.

2.4 Cross-Sectional Correlations of Account Level Characteristics

Table A2 provides the cross-sectional correlations of account age, plus account level characteristics examined in Table 2. The reported correlations are computed for each month (using sampling weights), with the average of these pure cross-sectional correlations reported in the table.

Consistent with our regressions, account age is significantly negatively correlated with each of the investment behaviors. Larger accounts also appear to behave better. Across accounts, disposition effect is barely correlated with idiosyncratic share and turnover, perhaps due to less precise measurement, but turnover and idiosyncratic share are highly correlated with each other. Rural accounts are more poorly behaved and have unhelpful portfolio tilts, but correlations are small. Finally, there are quite a few significant correlations amongst the style tilts of portfolios due to the fact that these styles are correlated within the population of underlying stocks.

2.5 Population Per NSDL Investor and Per Capita Income by State

We compute the population per individual NSDL investor with use of state population data from the 2011 Indian Census. We obtain data on per capita state income (in March 2011) from the Reserve Bank of India. These are produced as a bubble plot in Figure A3, where the area of each bubble represents the 2011 population of the Indian state. The largest share of NSDL data comes from relatively populous and wealthy Maharashtra, which comprises over one-fifth of all individual accounts.

2.6 Relative Significance of Account Performance and Specific Feedback Effects on the Evolution of Investor Behavior and Net Style Demand

In Figure A4, we use the regressions in Table 3 and actual experiences of the December 2003 cohort to separately simulate the evolution of the part of investor behavior that is due to account performance feedback, and the part that is due to behavior-specific feedback. We account for the longer-run impact of feedback through lagged behavior. The figure plots the 10th and 90th percentile of each. For turnover, behavior-specific feedback has a larger impact on the spread in subsequent behavior, while the magnitudes are comparable for disposition effect.

In Figure A5, we repeat this exercise for net style demands. In each case, account performance has a significantly greater impact on subsequent behavior.

2.7 Feedback Effects by Account Age and Size

To investigate whether larger and more experienced accounts respond differently to feedback, we estimate a version of regression equation (3) where we interact feedback terms with both account age and (log, inflation adjusted) initial account value. In Figures A6 and A7, we plot feedback effects on investor behaviors and account value respectively, for new accounts and eight-year-old accounts with the median initial account value. Grey bars represent differences in the two series, with dark grey portions representing the part of the difference which lies outside a 95% confidence interval.

There is only very tentative evidence that newer accounts respond more strongly to style-specific feedback; there is insufficient statistical power to say any more with confidence. Figures A8 and A9 are constructed similarly, but compare feedback effects on average-aged accounts at the 10th and 90th percentile of initial account value. Again, there is insufficient statistical power to conclude much.

2.8 Age and Feedback Effects for Style Supply and Demand

Figure A10 plots account age effects on style demand and supply separately. Investors both appear to increase purchases of small and value stocks over time as well as reducing sales of such stocks (conditional on the style tilt of their portfolio). Novice investors have a tendency to buy momentum stocks which rapidly fades. However, as accounts get older, the propensity to sell momentum also fades, perhaps related to weakening of the disposition effect. The combination of these two trends results in the U-shaped net momentum demand seen in Figure 4.

Figure A11 compares the effect of account performance feedback separately on style demand and supply. The increased net demand for large, growth, and momentum stocks following high returns is driven primarily by abnormally low sales of such stocks following good performance.

Figure A12 compares the effect of style-specific feedback separately on style demand and supply. Style supply spikes immediately following high style returns, presumably due to re-balancing and disposition effect related sales of winners. After this initial surge of sales, style demand remains high for a year or so. In the longer run following positive style returns, net demand is weakly positive primarily from low sales of the style that outperformed.

2.9 Effect of a Style Return Shock on Aggregate Individual Investors' Net Style Demand

In Figure A13, we show how the net style demands of individual investors as a whole respond to hypothesized style returns of +10% to all stocks ranked above average in the small/value/high-momentum style characteristics, and returns of -10% to all stocks ranked below average in those characteristics. At the average portfolio weights of the aggregate (i.e. portfolio-value-weighted, not representative) individual investor, such style returns also generate market outperformance of +0.45%, -0.21%, and -0.34% respectively; the aggregate individual investor's portfolio has a slight small, growth, low-momentum tilt.[5] The left hand side plots of Figure A13 combines these style returns and account outperformance with the estimated coefficients on style and account performance feedback. The right hand side plots cumulate the net style demand as was done in Figures 5 and 6.

For small and momentum, the indirect impact of account performance feedback offsets the style-specific feedback; high small and momentum stock performance for the aggregate individual accounts generate account performance feedback reducing demand for small and momentum stocks respectively. As a result, the cumulative response of the aggregate investor to small and momentum return shocks is not statistically significant. However, when value outperforms, the aggregate individual investor underperforms the market, which further bolsters net demand for value. In the few years following the style return shock, individual investors adjust portfolios towards value by an amount equivalent in impact to a shift around

[5] In contrast, the representative (i.e. non wealth-weighted) individual investor has a slight value tilt.

0.8% of the individual investor portfolio from growth to value stocks. About one-quarter of this affect is accounted for by the account performance feedback.

2.10 Regression Table for Account Age Effects on Account Returns

Table A3 provides our regressions of individual investor stock returns on account age effects and lagged investor behavior. In columns [1] and [2], account age is the only control. While age effects are only marginally significant, they could potentially be quite large. Our point estimate based on a linear account age effect is that investor returns increase by 12bp per month, per additional year of experience.

In column [3], we add controls for (recent) lagged investor behaviors and style tilts. Coefficients suggest accounts with low disposition effect and value tilts have particularly good returns, but the magnitudes of these coefficients are partly due to small sample time-series bias (i.e. "Stambaugh" bias). The reduction in the linear age effect in column [3] suggests improvements in the measured investor behaviors and style biases account for about one-third of the total account age effects in returns.

2.11 Moving-Average Difference in Returns on Old and New Accounts

Figure A14 takes the difference in returns on the oldest and newest quintile of individual accounts (cumulative plots of each are in the bottom panel of Figure 8), and plots it as an exponentially-weighted moving average. Only 2004, late 2007, and mid-2009 are periods of underperformance for the more experienced accounts. If anything, it appears that relative returns of experienced accounts have generally been growing over time, as might be expected given the growing spread in account age between the oldest and newest quintile since 2004.

2.12 Decomposition of the Difference in Returns on Old and New Accounts

The top part (portfolio tilts) of each column in Table 5 reports the time-series average of coefficients, $\overline{\phi}$, from Fama MacBeth regressions $W_{jt} = \phi_t X_{jt} + \varepsilon_{jt}$ of portfolio weights W on the set X of cross-sectionally de-meaned stock characteristics listed in the table.

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 ($\Sigma_j W_{jt} R_{jt} - \Sigma_j \overline{W_j R_j}$) and selection effects ($\Sigma_j \overline{W_j R_j}$). 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" ($\Sigma_j(\overline{\phi X_j})'(\overline{\psi X_j})$) and "additional stock selection" effects ($\Sigma_j \overline{\varepsilon_j \eta_j}$). Timing effects are decomposed into "stock characteristic timing" ($\Sigma_j[(\phi_t X_{jt})'(\psi_t X_{jt}) - (\overline{\phi X_j})'(\overline{\psi X_j})]$) and "additional stock timing" ($\Sigma_j(\varepsilon_{jt}\eta_{jt} - \overline{\varepsilon_j \eta_j})$), where the coefficients with t-subscripts are from the cross-sectional regressions run in Fama MacBeth estimation.

3 Robustness Exercises

3.1 Inclusion 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 meaningful 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 determined) as 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.

Figures A15 through A20 show that age and feedback effects are little affected by the inclusion of accounts opened prior to February 2002. For direct comparability, we still scale behaviors by their means from Table 2, which is based solely on accounts opened after January 2002 as with all other analyses in the main text. Table A4 provides an additional column to the age-based account return decomposition (Table 5), showing that several of the stock characteristics (e.g. low beta, small, higher institutional ownership) favored by older accounts within the set of post-2002 accounts are exaggerated further when looking at accounts opened even earlier. Returns on the oldest post-2002 accounts and the pre-2002 accounts are similar, but no higher. As a result, introducing these "oldest" accounts does lead to some reduction in the account age effects seen in Figure A20.

3.2 Use of "Passive" versus "Active" Account Returns

We compute and use "passive" returns throughout the rest of the paper. Passive returns reflect what the investor would have received if they did not trade during the given month.

Here, we compute "active" returns which take account of trading, but assumptions are required since we do not know the exact intra-month timing of purchases and sales.[6] Timing assumptions matter as they affect the average amount of wealth invested in equities over the month, which is the denominator of returns calculations.

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.

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

Next, we alternatively 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 (blows up) 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). We are unable to precisely estimate the returns experienced by these extremely active traders in our sample. However, they constitute a very small fraction of the total set of accounts, and consequently are likely to have a very small impact on our inferences.

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

Figure A21 compares the account age effects on account returns from Figure 8 alongside the account age effects similarly generated using both approaches to computing active returns above. In order to use the same set of observations as in Figure 8, we set cross-sectionally then individually de-meaned returns equal to zero in the few investor-months where active returns are undefined.

Since the excess equity returns are significantly positive on average in our sample, "high leverage" active returns tend to be greater than passive returns as the denominator is smaller. Since newer accounts trade less, this switch to "high leverage" active returns does lower the

[6]Since we do observe average sales and purchase prices for each stock in each account during a month, it might be possible to narrow down the timing a bit with the use of daily price ranges.

account age effect in returns, though it remains economically large.

3.3 Controlling for Time Variation in the Inherent Attributes 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 conceivable that the average inherent sophistication of Indian investors has been declining as market participation expands. If so, perhaps accounts appear to be "getting better" primarily because they are being compared to progressively "worse" newer investors. 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.[7]

Figure A22 shows plots of the fitted series αC_t, which represent the average inherent returns, behavior, and net style demands of investors present in the market over time. In general, the model suggests there has been a modest worsening in inherent investor behaviors and style preferences over time. 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.

Figures A23 through A25 provide age effects from regressions using equation (4) alongside age effects from our baseline equation (3). Consistent with the results in Figure A22, the age effects generally attenuate modestly with the exception of the disposition effect, for which the age effect is primarily explained by controls for average inherent investor sophistication.

[7] 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.

Account age varies only across (and not within) cohorts at a given point in time, whereas our feedback measures vary primarily within cohorts. Since only cross-cohort variation can be potentially explained by changes in average inherent properties of investors, our estimation of feedback effects is virtually unaffected and therefore not shown under equation (4).

3.4 Estimating Impact 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. As a result, Y_i will spuriously appear to be negatively predicted by X_i. For example, 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) from the last section. By removing the individual fixed effect, equation (5) addresses concerns about strict exogeneity, but loses the ability to control for account-specific propensities towards the behaviors and styles which are not picked up by C.

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

Figures A26 shows that the response of investor behavior to feedback is qualitatively similar when we use equation (5). Figures A27 and A28 show qualitatively similar patterns hold for the impact of feedback on net style demand, though responses are generally less positive, and the response of net momentum demand to momentum style returns is slightly negative.

However, there are good reasons to believe that the changes resulting from removal of individual fixed effects are not primarily attributed to time-series bias. For example, time-series bias should actually cut against the response of turnover to turnover-specific feedback; past high returns from trading are related to past high returns and therefore low present returns. It is plausible that the weaker result from equation (5) is instead due to the fact that investors typically lose by trading, and so investors who have low propensities to trade receive better than average signals of approximately zero. As another example, the impact of style feedback on net style demand will be understated in the absence of an investor fixed effect serving to disentangle the roles of inherent style preferences and lagged style returns on portfolio style tilts. Specifically, the investor who receives high style returns (and has a higher style tilt as a consequence), will have less favorable *inherent* preferences for that style than the average investor with the same style tilt.[8] The same argument applies to performance feedback, where high account performance tends to result in a tilt towards large, growth, momentum stocks (whereas other investors with those tilts may have stronger inherent preferences for those styles).

Unlike feedback, account age is deterministic. Even so, account age is vulnerable to violations of strict exogeneity if investor exit is influenced by the disposition effect and related to past "luck." In our data, a one standard deviation increase in average past monthly returns increases the probability of exit from around 0.68% to 0.72%. As a result, experienced investors may disproportionately be investors who had poor returns when they were novices.

To respond, we model the relationship of account exit, investor behaviors, and style demands to average past returns.[9] These models are estimated both with and without monthly fixed effects. Next, we use these model estimates along with draws from the distribution of estimated residuals and time shocks to exits and behaviors to jointly simulate account returns, exit, investor behavior, and net style demand.[10] In the simulation, there are no age or investor sophistication effects, so estimates of age effects on the simulated data using equation (3) reflect survivorship bias.

We report estimates of the simulated bias in the first three rows in Table A5. The fourth row provides estimates of a linear age effect for account returns, behavior, and net style demand as a basis for comparison. Survival bias in the age effect on account returns is quite small as investors do not exit our data frequently, and when they do, it is usually not related to past returns. Survival bias in age effects on investor behaviors and style demands are even smaller, as these behaviors are only partly related to past returns, and thus very tenuously related to luck-driven exits.

[8] Another way to state this is that equation (5) under-appreciates the tendency of style demands to depend negatively on style tilts when there are in fact un-modelled individual effects.

[9] We use a logit model for investor exit, and linear least squares models for the relationship of behavior and net style demands to past returns. For disposition effect, we model a separate definition of exit; where exit means an end to trading.

[10] We run 100 simulations with 20,000 investors in each of five cohorts over 100 months. Each simulated investor draws residuals from a randomly selected investor in our data sample.

4 Tables and Figures

Table A1: Indirect Share of Individual Stock Ownership

Data are as of the end of March, and amounts are stated in billions of US $.

	Equities Held Indirectly through		Equities Held Directly	% of Equities Directly Held
	Mutual Funds and Unit Trusts	Unit-Linked Insurance Plans		
2004	$2.6		$42.8	94.30%
2005	$3.9	$0.9	$63.4	93.00%
2006	$9.5	$1.8	$111.3	90.77%
2007	$12.1	$4.9	$126.6	88.16%
2008	$18.5	$8.7	$171.2	86.27%
2009	$9.2	$9.0	$75.9	80.63%
2010	$18.2	$12.9	$171.5	84.64%
2011	$10.7	$12.2	$186.1	89.02%
Average				88.35%

Table A2: 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. 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			
Small Tilt (i.e. Style of Holdings)	[8]	-0.02	**-0.14**	-0.06	**0.30**	**0.14**	0.05	**0.35**	1.00		
Value Tilt	[9]	-0.02	**-0.11**	**-0.11**	**0.15**	0.07	0.06	**0.15**	**0.47**	1.00	
Momentum Tilt	[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 A3: Account Age Effects on Individual Investor Returns

The regression specification follows Equation (3) in the paper. Lagged turnover and disposition bias are averages over the past year, winsorized at the 1st and 99th percentile of accounts with at least 5 observations of the behavior in the past year. Where missing, (cross-sectionally and then individually) de-meaned values of lagged behaviors are imputed as zeros. Incremental R^2 is the ratio of the variance of the fitted age effects relative to the variance of monthly account excess returns. 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 and ten percent level are in bold and italicized type respectively.

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

		[1]	[2]	[3]
Account Age Effect	Account Age (Linear)	*12.01*		8.24
		(7.22)		*(7.05)*
	Piecewise Linear		See Figure 8	
Investor Behavior and Style Tilts	Lagged Idio. Share of Portfolio Var.			55.88
				(75.56)
	Lagged Portfolio Turnover			-97.75
				(67.25)
	Lagged Disposition Effect			**-3.69**
				(1.49)
	Small Tilt			178.05
				(182.13)
	Value Tilt			**554.29**
				(113.76)
	Momentum Tilt			-22.29
				(122.25)
Incremental R^2		0.00031	0.00039	0.00015

Table A4: Decomposition of the Difference in Returns on Old and New Accounts

In column [4], the analysis from Table 5 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 after January 2002 (a copy of Table 4 column [1]) is provided for comparison.

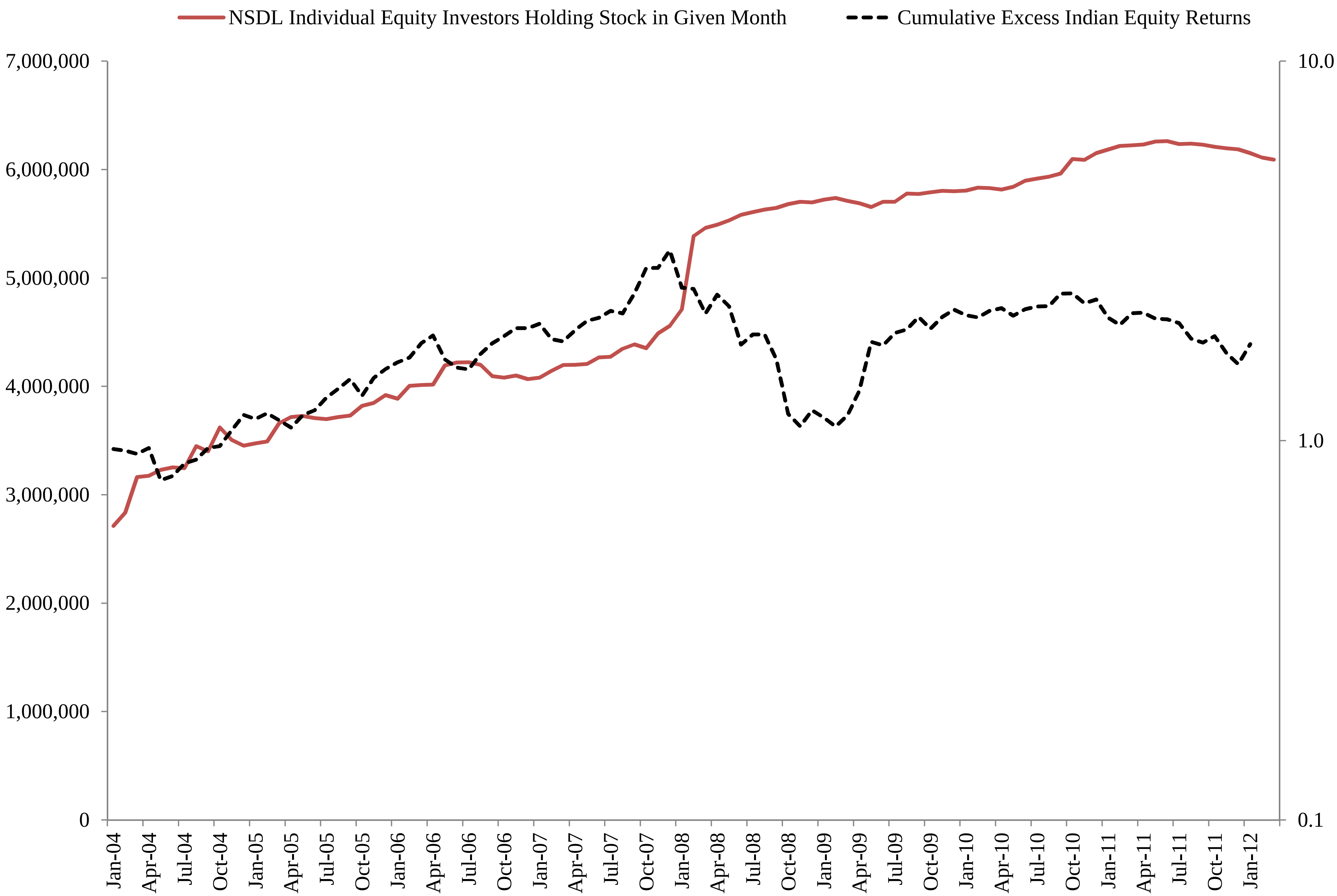
Zero-Cost Portfolio Represents:	Oldest minus Newest	Pre 2002 Accounts minus Oldest
Portfolio Tilts (1000 x φ_{bar})	[1]	[4]
Market beta	-0.547	**-0.697**
	(0.568)	*(0.274)*
Market capitalization	-0.318	**-0.601**
	(0.233)	*(0.099)*
Book-market	0.171	**-0.735**
	(0.143)	*(0.200)*
Momentum (t-2:t-12 returns)	-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 A5: Survival Bias in Account Age Effects

See Section 3.4 of the appendix for a discussion of the simulation. Linear age effect coefficients which are statistically significant at the five and ten percent level are indicated by bold and italicized type respectively.

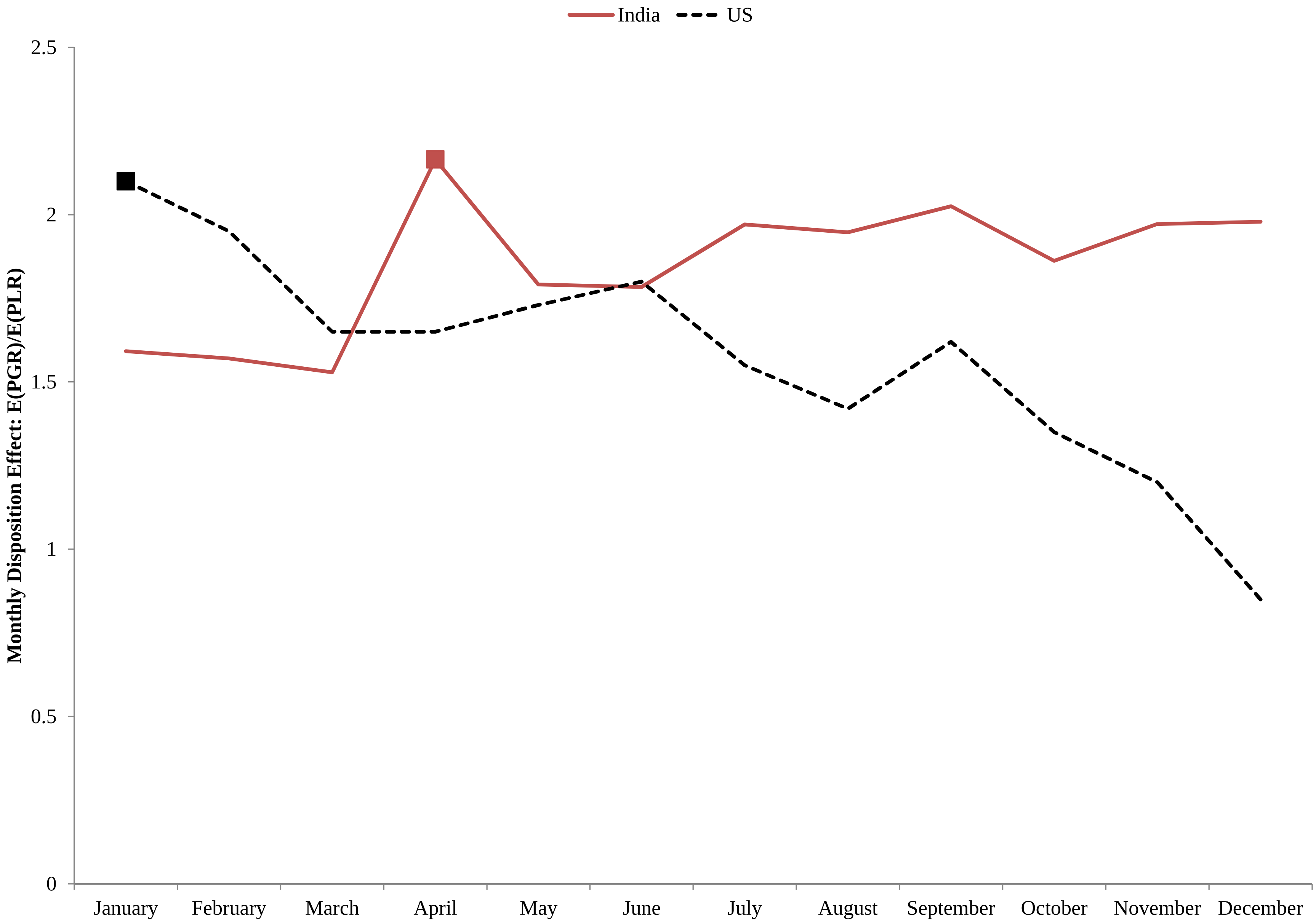
		Account Behavior			Net Style Demand		
	Account Returns	Id. Share of Port. Var.	Turnover	Disposition Effect	Small	Value	Momentum
Baseline simulation	0.72	0.0001	0.0001	0.000	0.0002%	0.0004%	-0.0004%
	(0.02)	*(0.0000)*	*(0.0001)*	*(0.000)*	*(0.0001%)*	*(0.0002%)*	*(0.0002%)*
Simulations include monthly fixed effects	0.58	-0.0004	-0.0002	0.000	-0.0011%	0.0007%	0.0006%
	(0.02)	*(0.0000)*	*(0.0001)*	*(0.000)*	*(0.0001%)*	*(0.0002%)*	*(0.0002%)*
Sensitivity of exits to lagged returns increased by five standard errors	1.34	0.0002	0.0000	0.000	-0.0001%	0.0001%	-0.0010%
	(0.02)	*(0.0000)*	*(0.0001)*	*(0.000)*	*(0.0001%)*	*(0.0002%)*	*(0.0002%)*
Point estimate for linear account age effect	*12.01*	0.0007	**-0.0705**	**-0.0724**	**0.0546%**	**0.2254%**	-0.0069%
	(7.22)	*(0.0009)*	*(0.0079)*	*(0.0093)*	*(0.0089%)*	*(0.0209%)*	*(0.0090%)*

Figure A1: Individual Equity Investors in NSDL 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 A2: Disposition Effect for India (Individual NSDL Accounts 2004-Jan 2012) vs US (Discount Brokerage Accounts 1987-1993 from Odean 98, 99)



PGR/PLR is computed for each calendar month, aggregating across accounts and years as in Odean (1998), which is the source for the US brokerage based statistics plotted.

Figure A3: Population per Investor and Per Capita Income by State, as of 2011

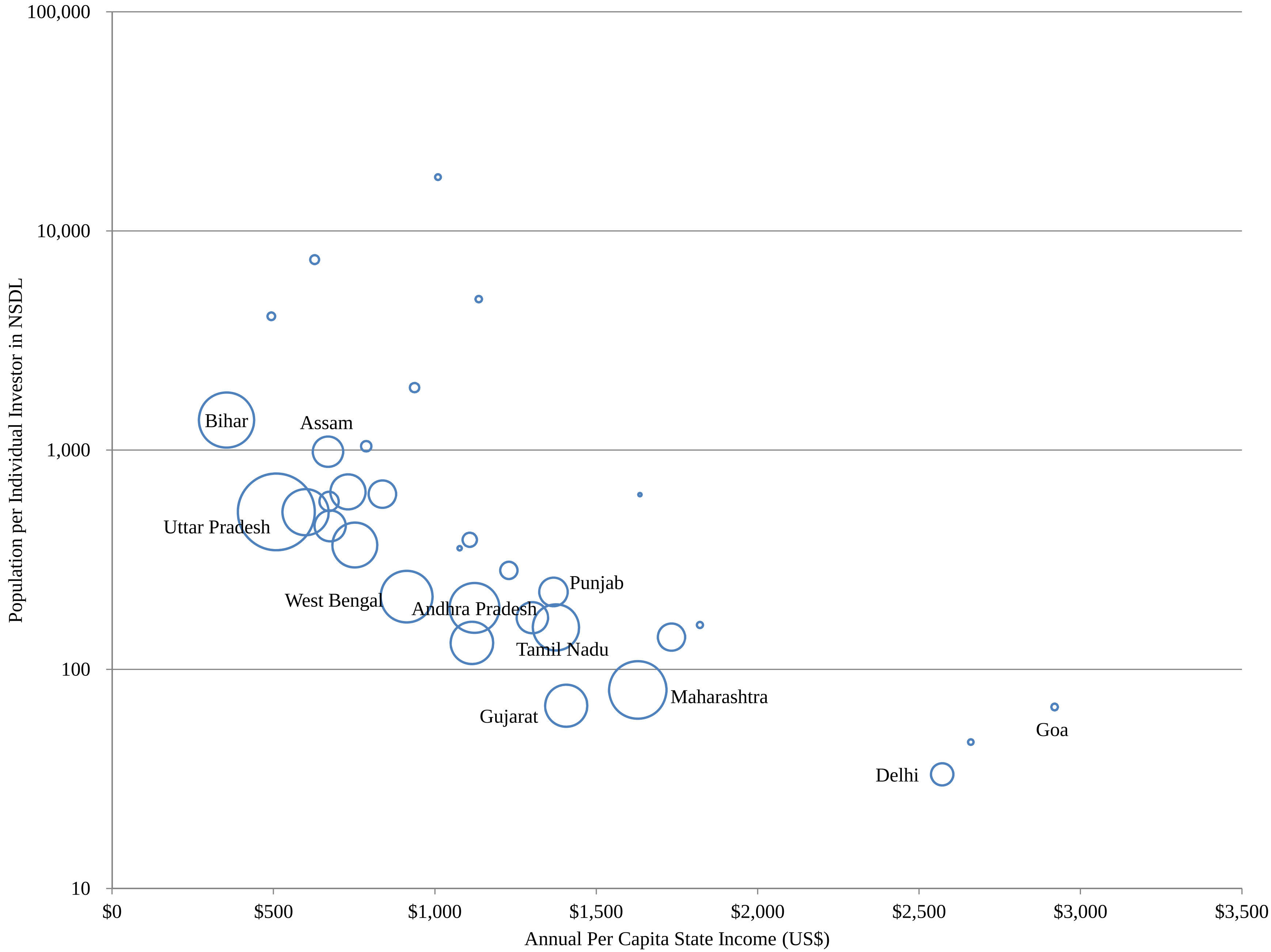
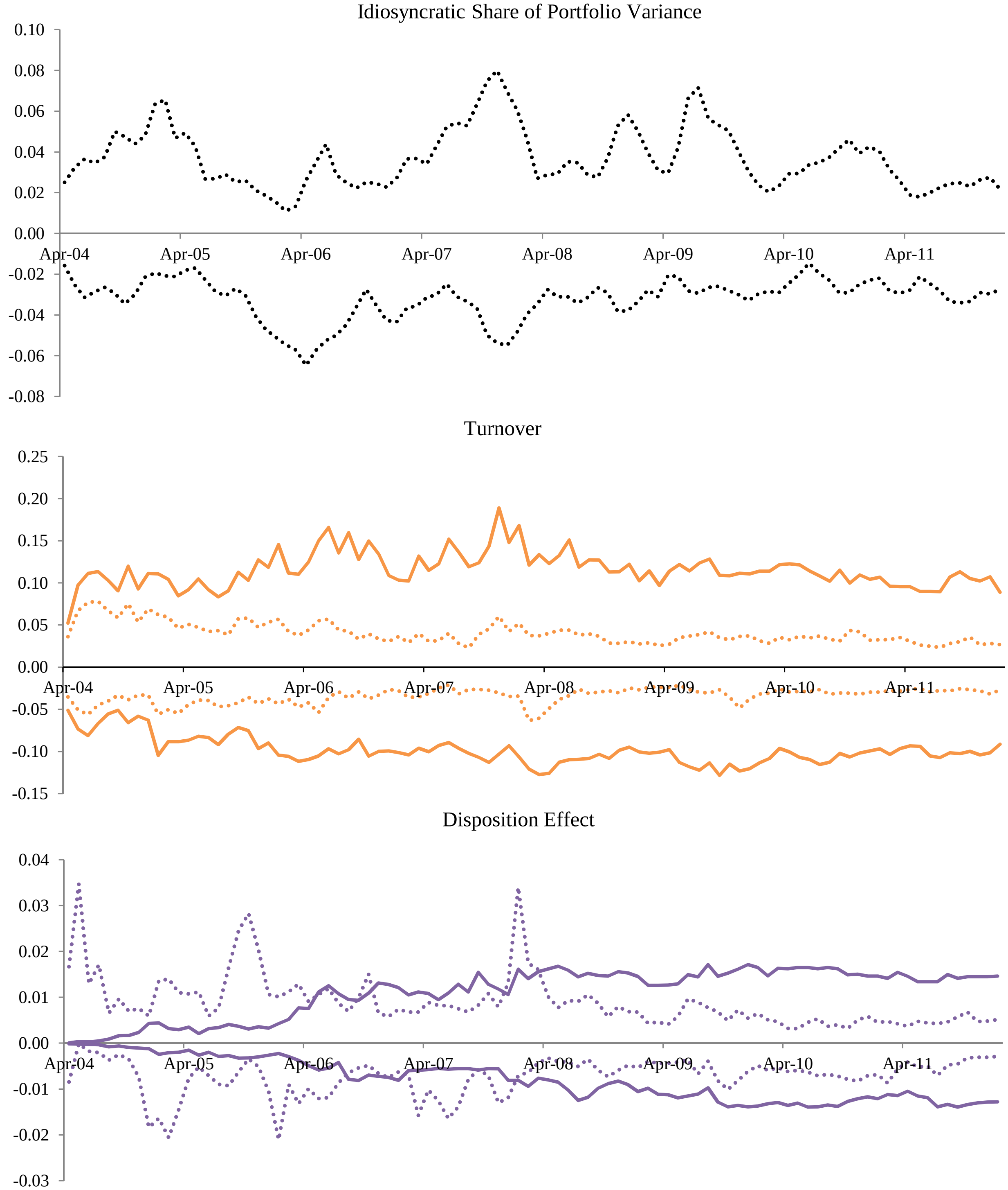


Figure A4: Evolution in Investor Behaviors from Performance (dotted lines) and Behavior-Specific (solid lines) Feedback

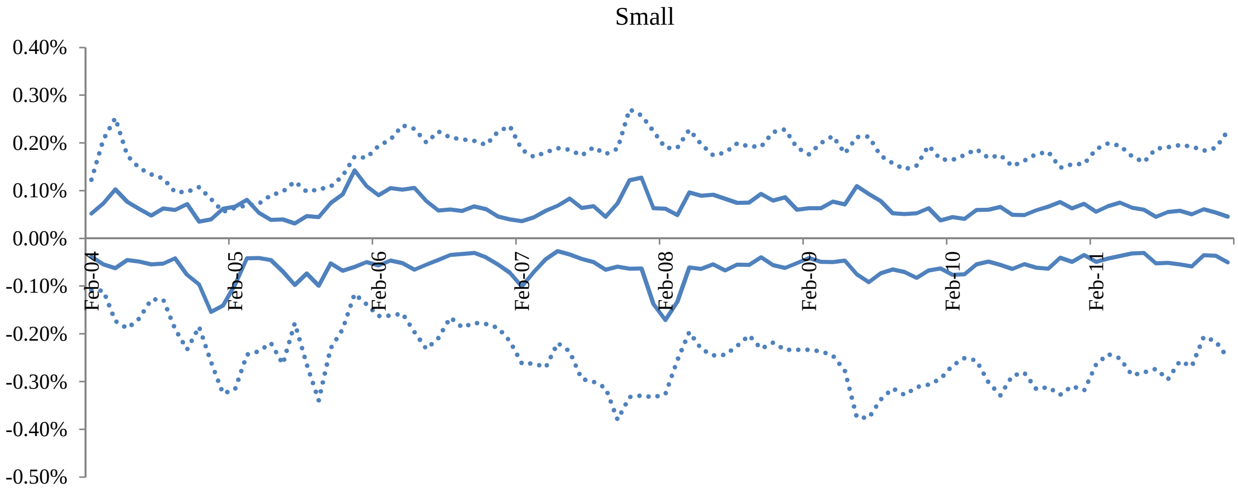
10th and 90th Percentile of Accounts Opened December 2003

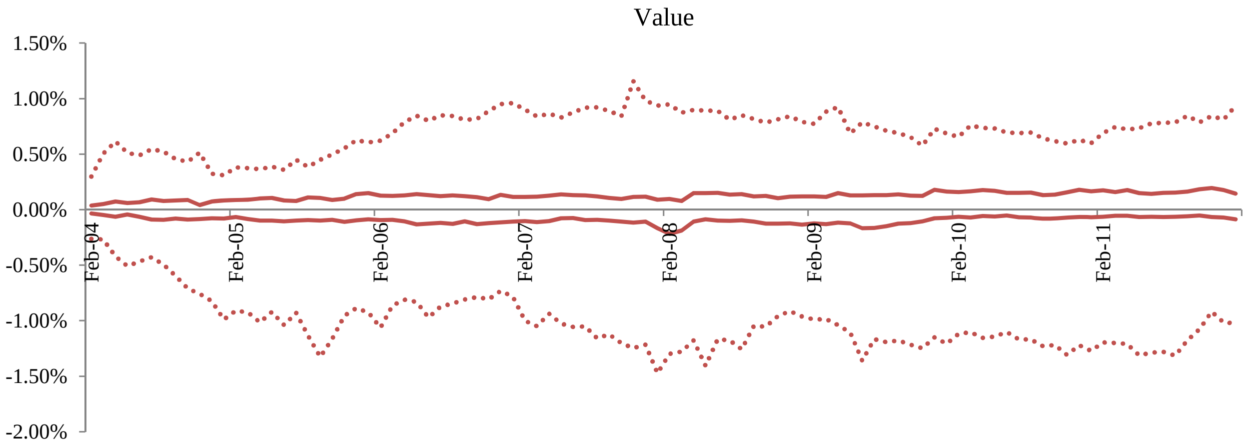


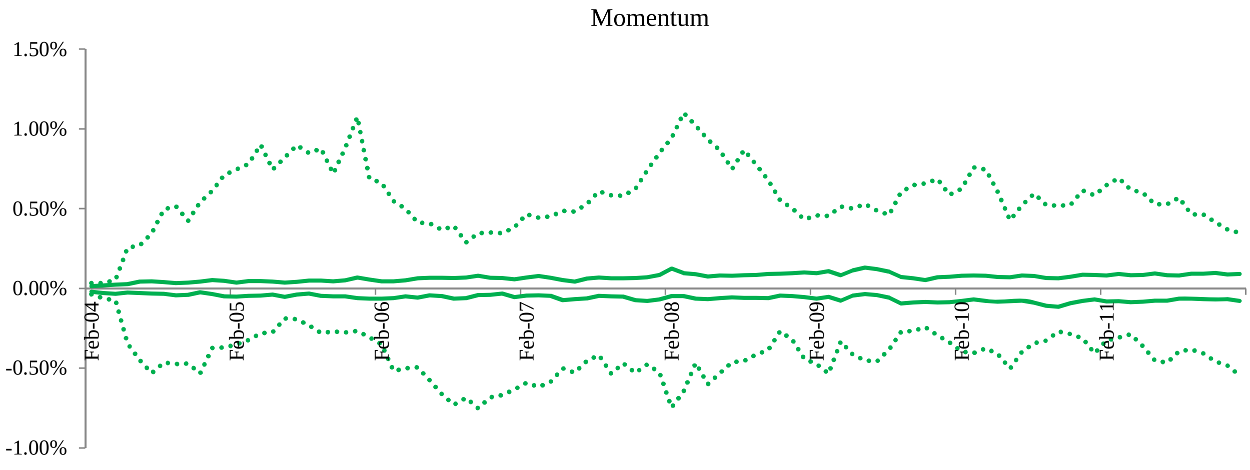
The plots separately fit account performance feedback and behavior-specific feedback coefficients (as well as the coefficient on lagged behavior) from investor behavior regressions in Table 3 with the actual feedback received by individual investor accounts opened in December 2003. The 10th and 90th percentiles of the simulated distribution appear above.

Figure A5: Evolution in Net Style Demand from Account Performance (dotted lines) and Style-Specific (solid lines) Feedback

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



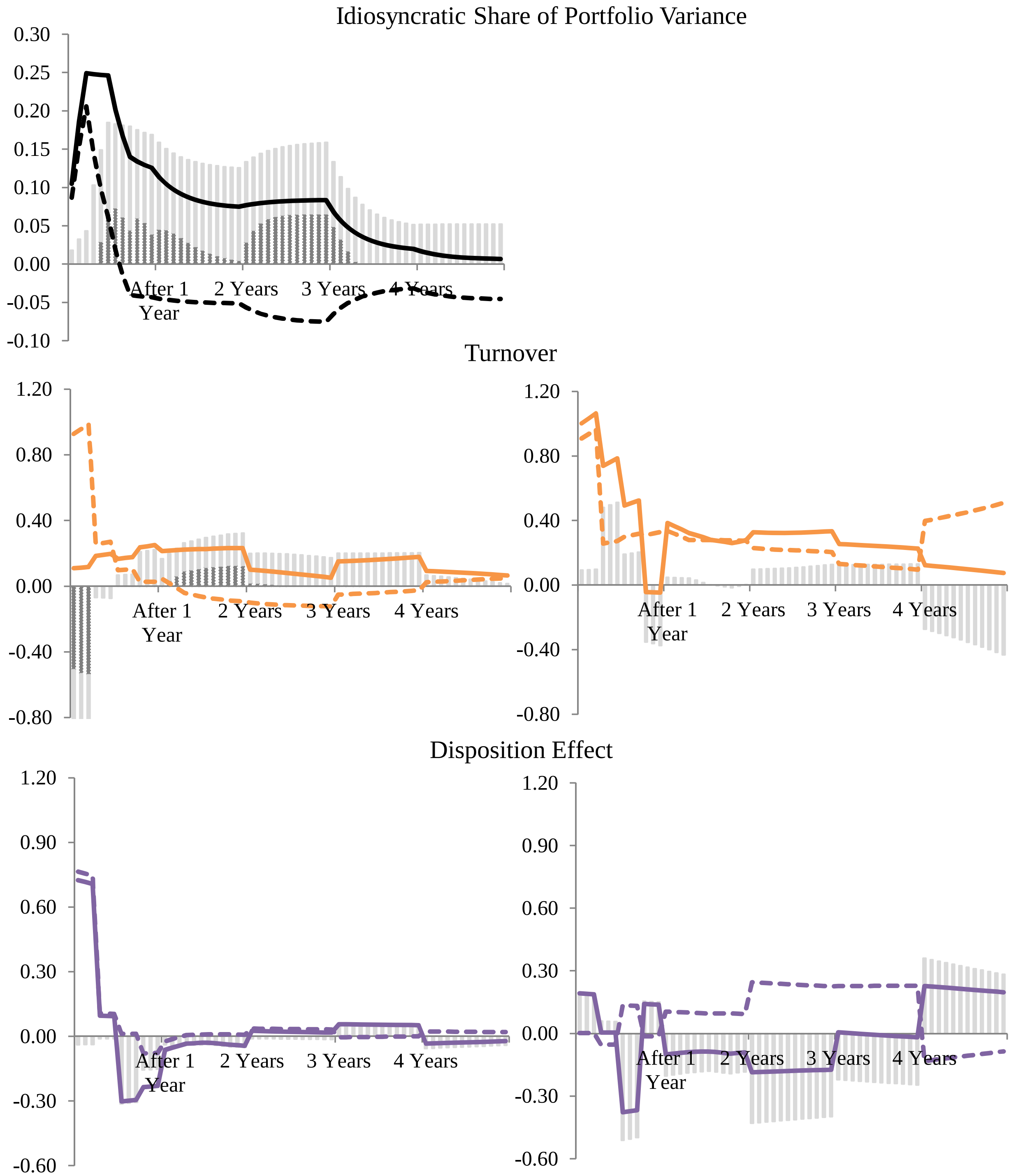




The plots separately fit account and style-specific feedback effects from investor net style demand regressions in Table 4 with the actual age and feedback received by individual investor accounts opened in December 2003. The 10th and 90th percentiles of the simulated distribution appear above.

Figure A6: Feedback Effects on Investing Behavior of Experienced (Solid) and Novice (Dashed) Investors

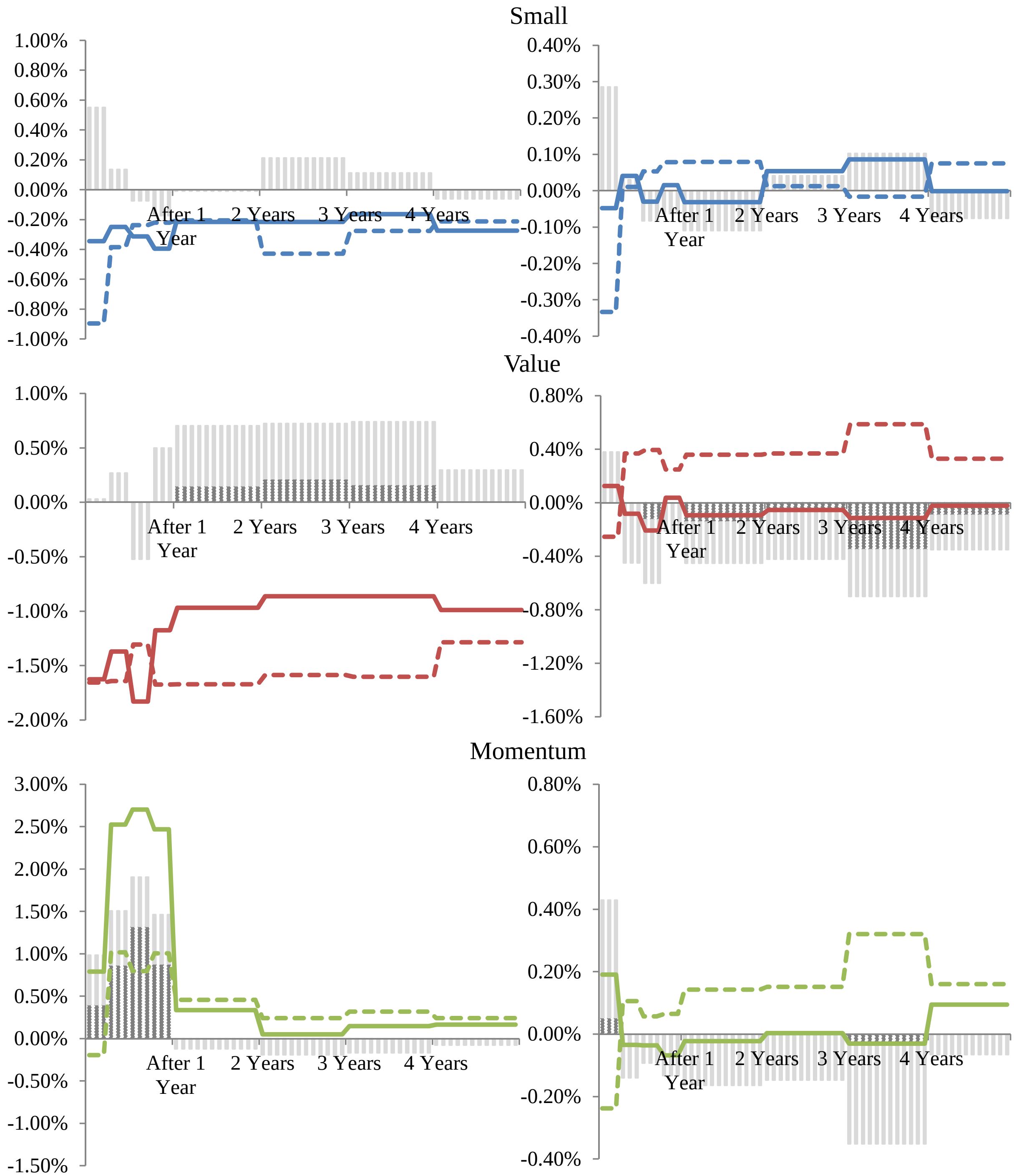
Account Performance Feedback (Left) and Behavior-Specific Feedback (Right)



Plots are produced similarly to Figure 2, but use coefficients from regressions which interact performance and behavior-specific feedback with both (inflation-adjusted) log initial account value and account age. The plotted fitted feedback responses use median initial account value, and account age of either zero or eight years. The bars in each plot represent the difference in the two series, with the dark part of each bar representing the part of the difference that lies outside a 95% confidence interval.

Figure A7: Feedback Effects on Net Style Demand of Experienced (Solid) and Novice (Dashed) Investors

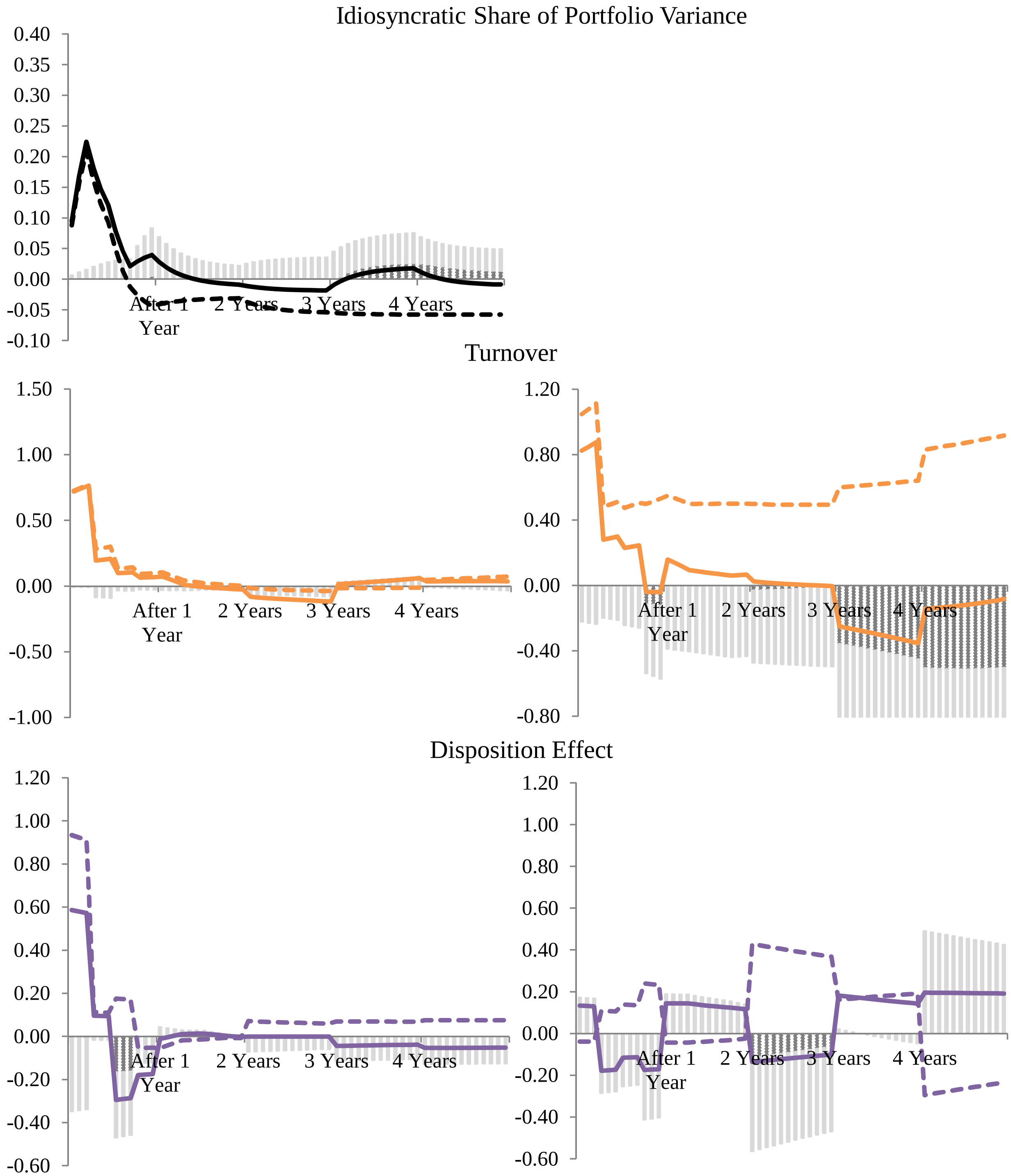
Account Performance Feedback (Left) and Behavior-Specific Feedback (Right)



Plots are produced similarly to the left hand side plots of Figures 5 and 6, but use coefficients from regressions which interact performance and style-specific feedback with both (inflation-adjusted) log initial account value and account age. The plotted fitted feedback responses use median initial account value, and account age of either zero or eight years. The bars in each plot represent the difference in the two series, with the dark part of each bar representing the part of the difference that lies outside a 95% confidence interval.

Figure A8: Feedback Effects on Investing Behavior of Large (Solid) and Small (Dashed) Investors

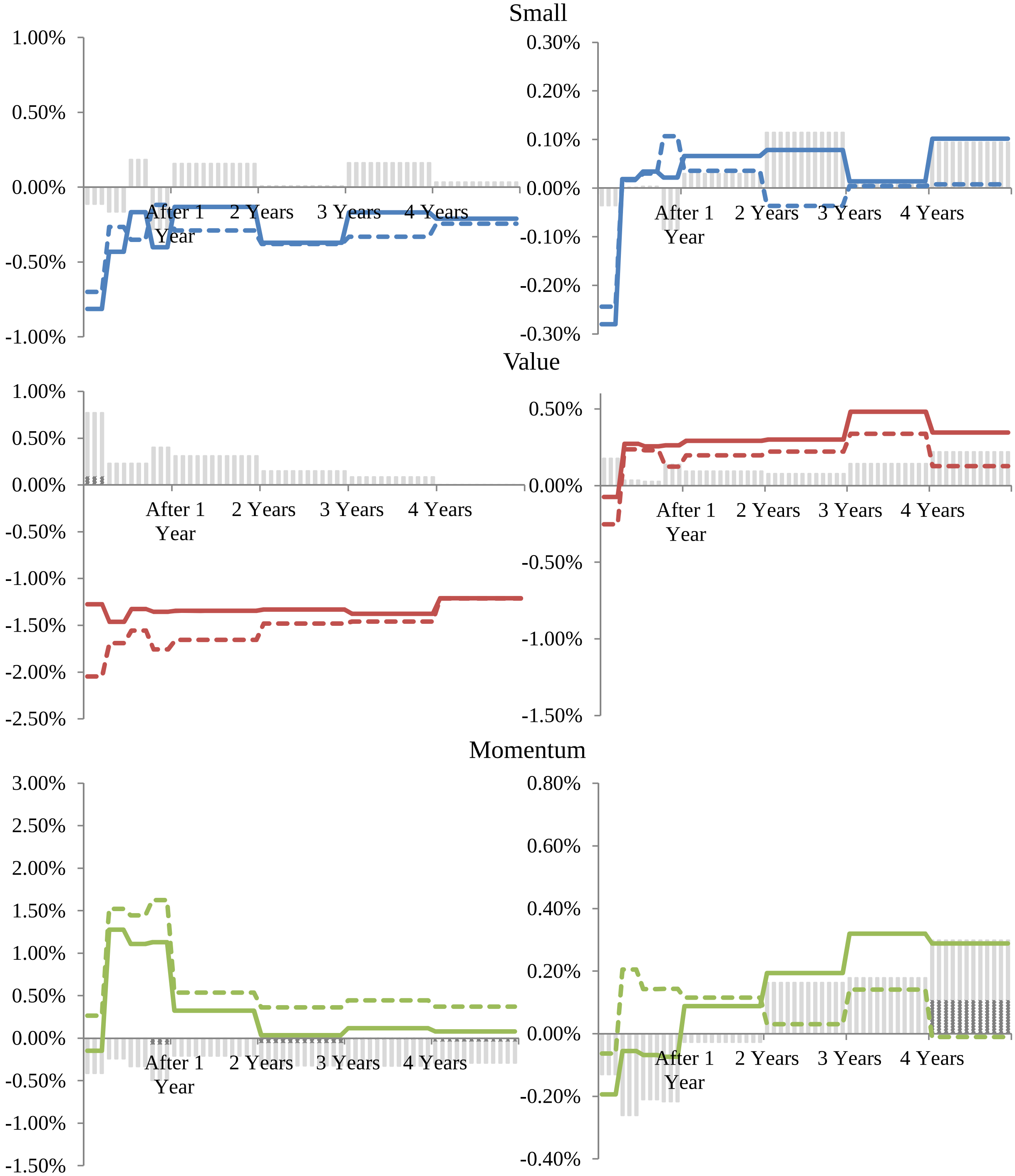
Account Performance Feedback (Left) and Behavior-Specific Feedback (Right)



Plots are analogous to Figure A6, but with the plotted fitted feedback responses use median account age, and account value set to either the 10th or 90th percentile of the distribution of (inflation-adjusted) log initial account value.

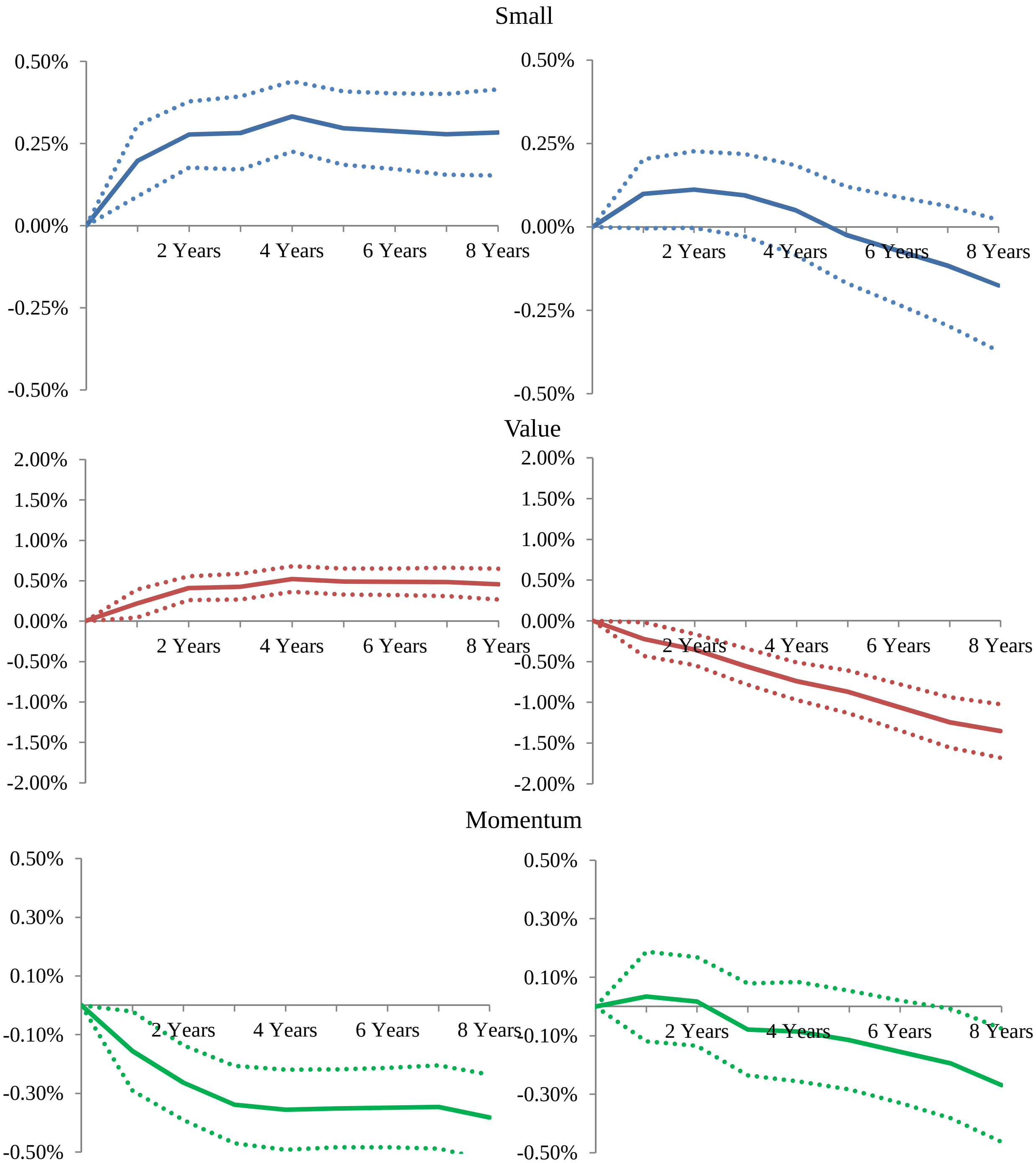
Figure A9: Feedback Effects on Net Style Demand of Large (Solid) and Small (Dashed) Investors

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



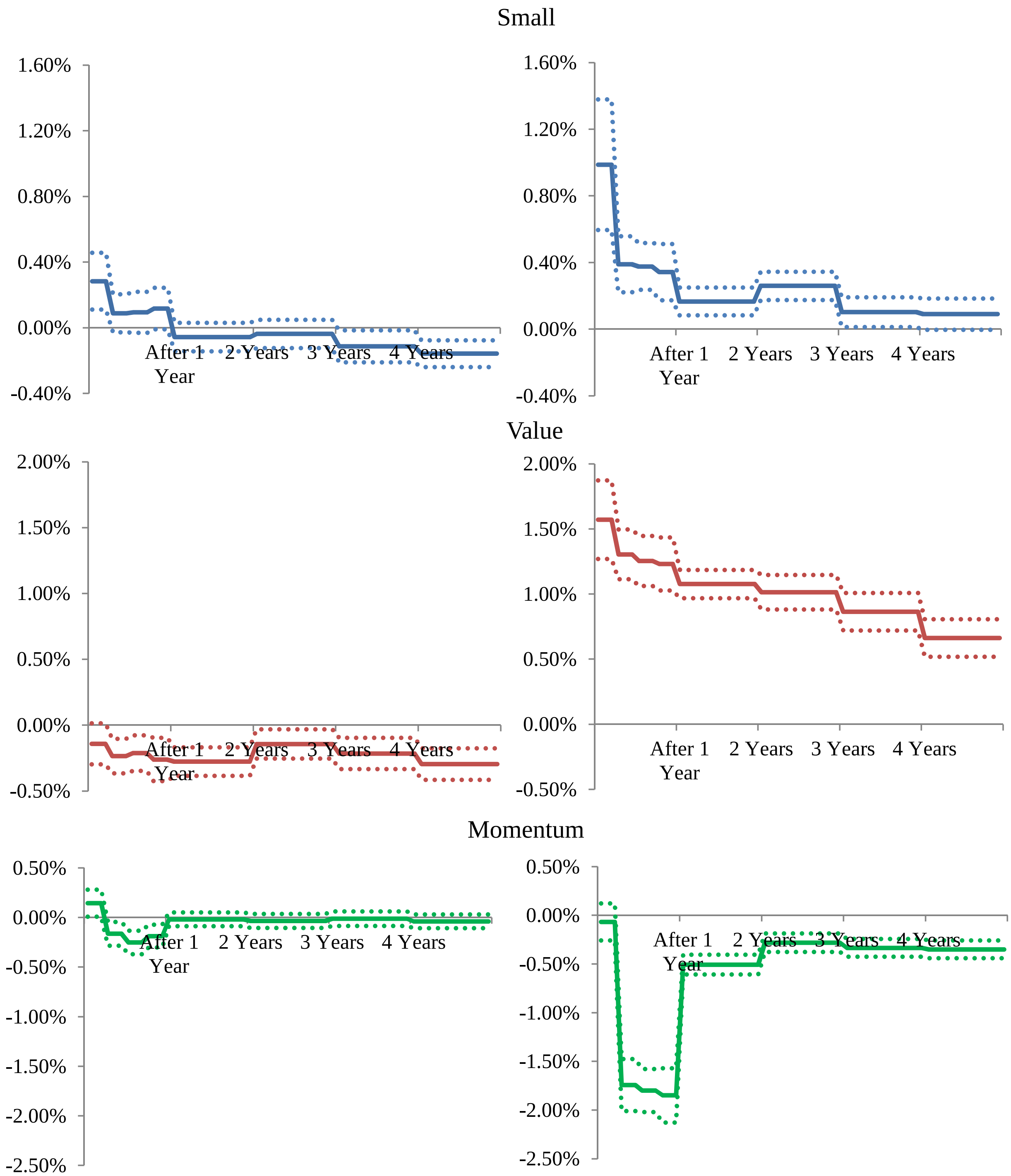
Plots are analogous to Figure A7, but with the plotted fitted feedback responses use median account age, and account value set to either the 10th or 90th percentile of the distribution of (inflation-adjusted) log initial account value.

Figure A10: Account Age Effects on Style Demand (Left) and Supply (Right)



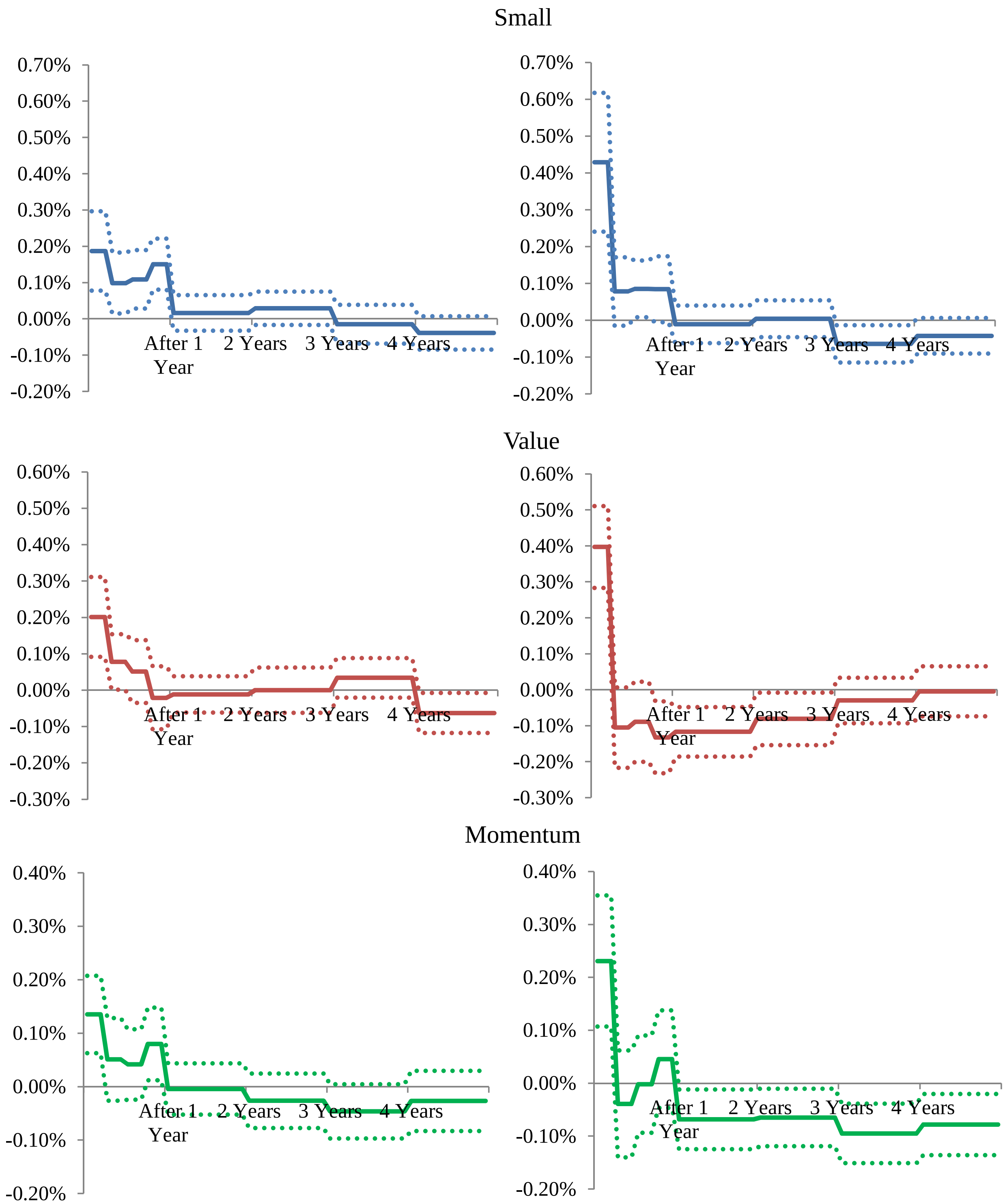
Plots analogous to Figure 4, but based on regressions of style demand and supply, instead of net style demand (which equals style demand minus style supply).

Figure A11: Account Performance Feedback Effect on Style Demand (Left) and Style Supply (Right)



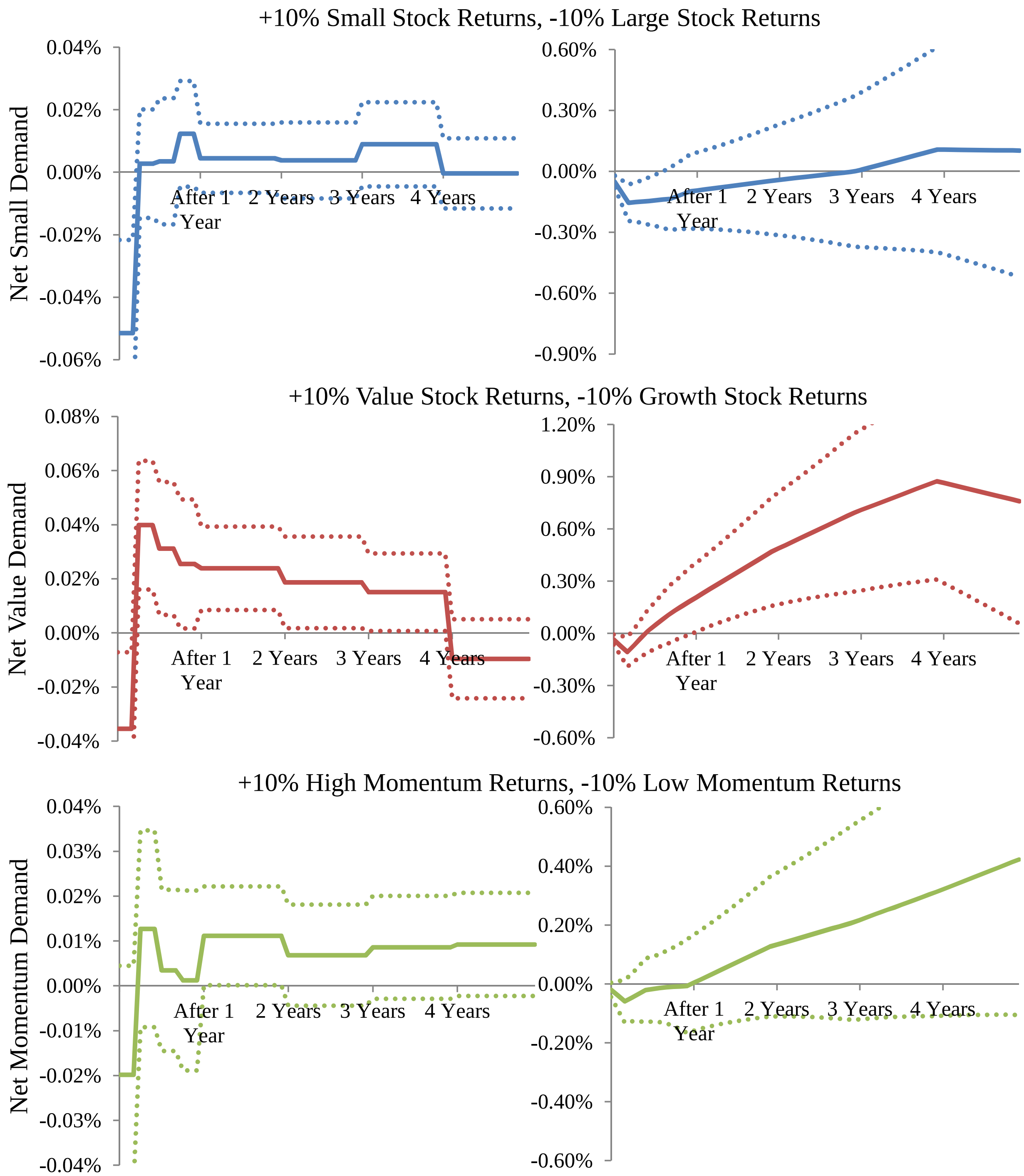
Plots are analogous to those on the left hand side of Figure 5, but produced from investor style demand and supply regressions.

Figure A12: Style Feedback Effect on Style Demand (Left) and Style Supply (Right)



Plots are analogous to those on the left hand side of Figure 6, but produced from investor style demand and supply regressions.

Figure A13: Effect of Style Returns on the Net Style Demands of Individual Investors in Aggregate



Style return shocks are defined as +10% returns to all stocks ranked above average in the given style, and -10% returns to all stocks ranked below average. Responses are based on a combination of style feedback of 20%, and account performance feedback based on the average market outperformance of the aggregate individual investor given the style return shock. Dotted lines represent 95% confidence intervals.

Figure A14: Returns on Oldest minus Newest Quintile of Accounts (Exponentially Weighted Moving Average with Six Month Half-Life)

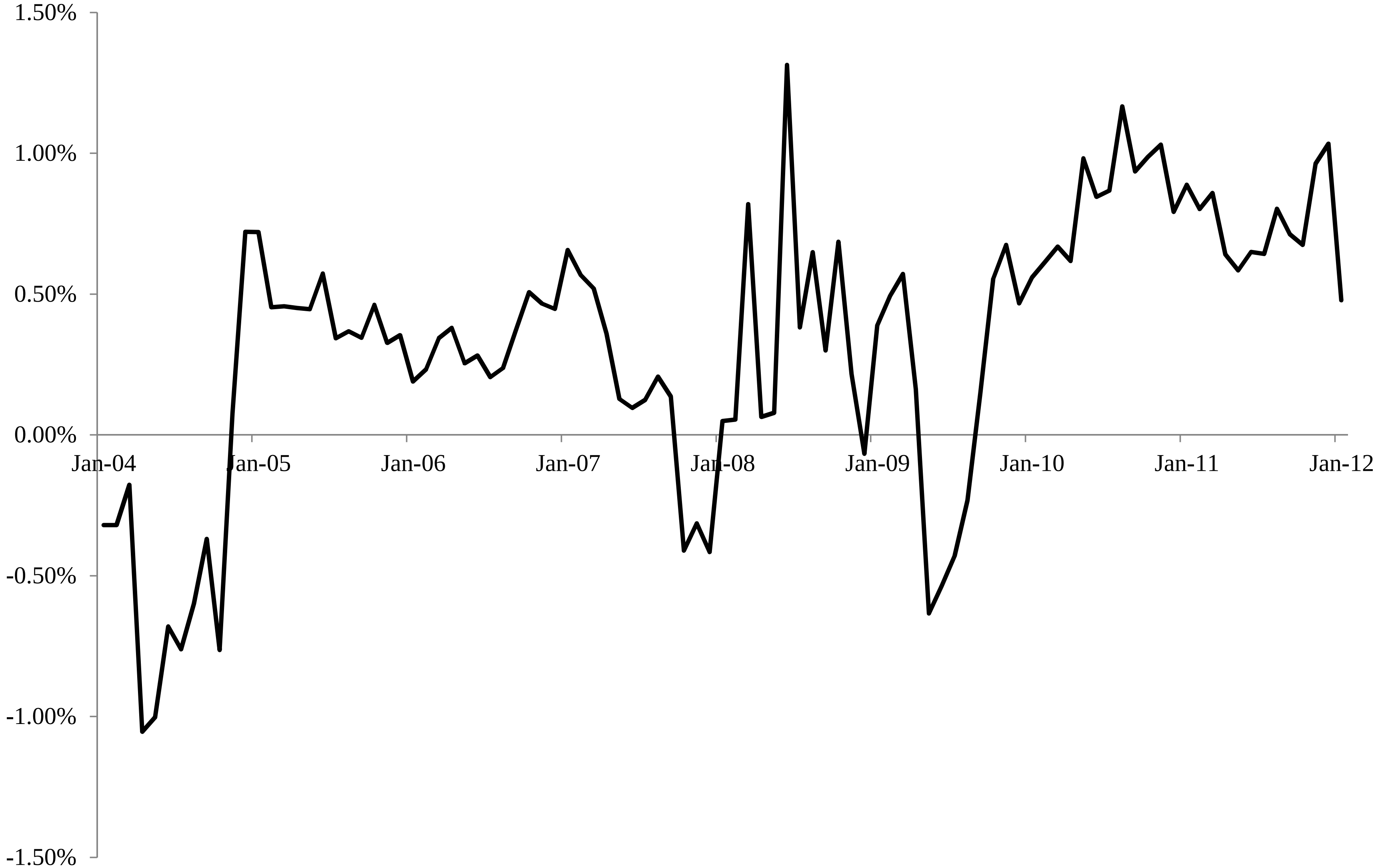
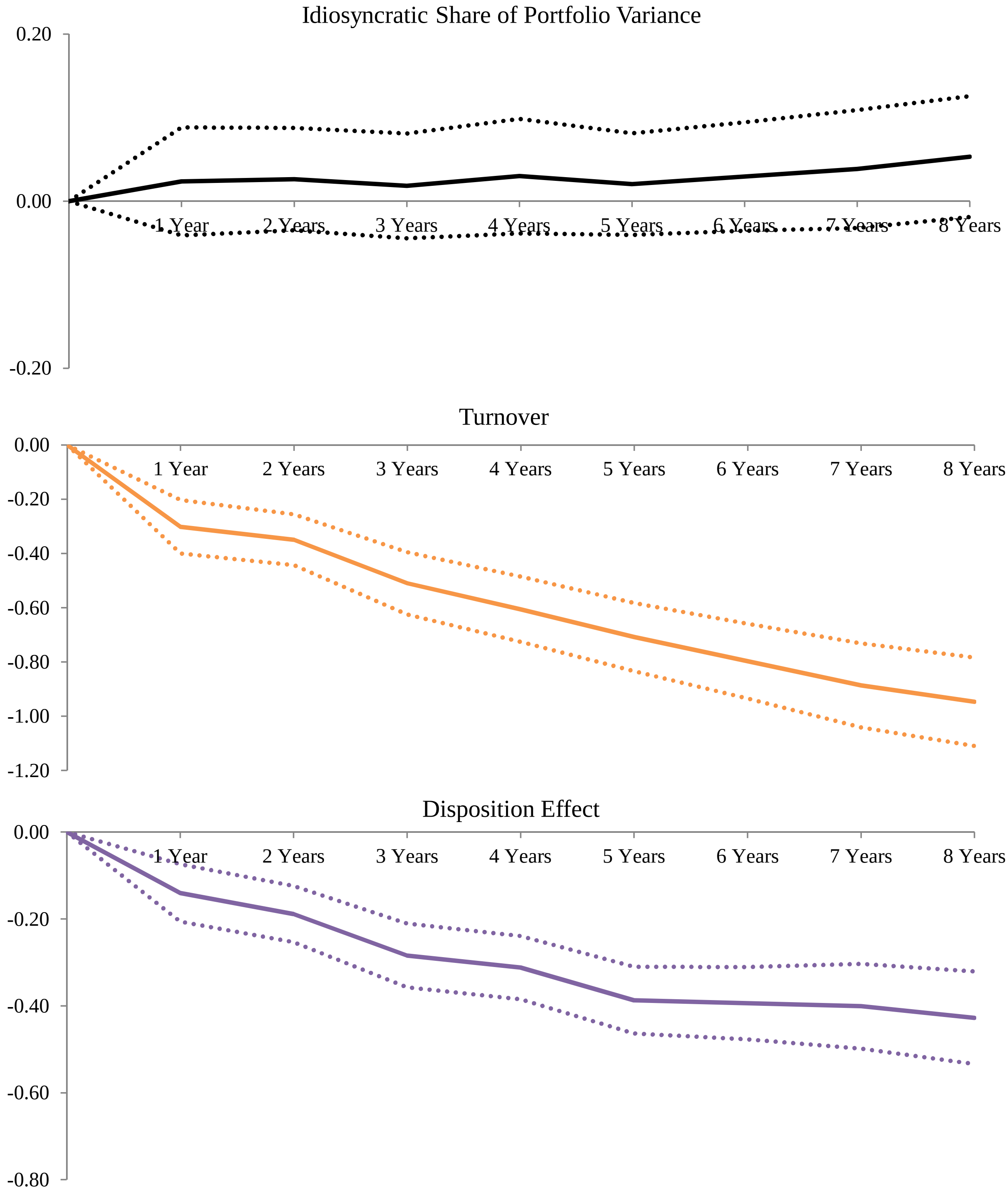


Figure A15: Account Age Effects on Investor Behaviors

Including Accounts Opened Prior to February 2002

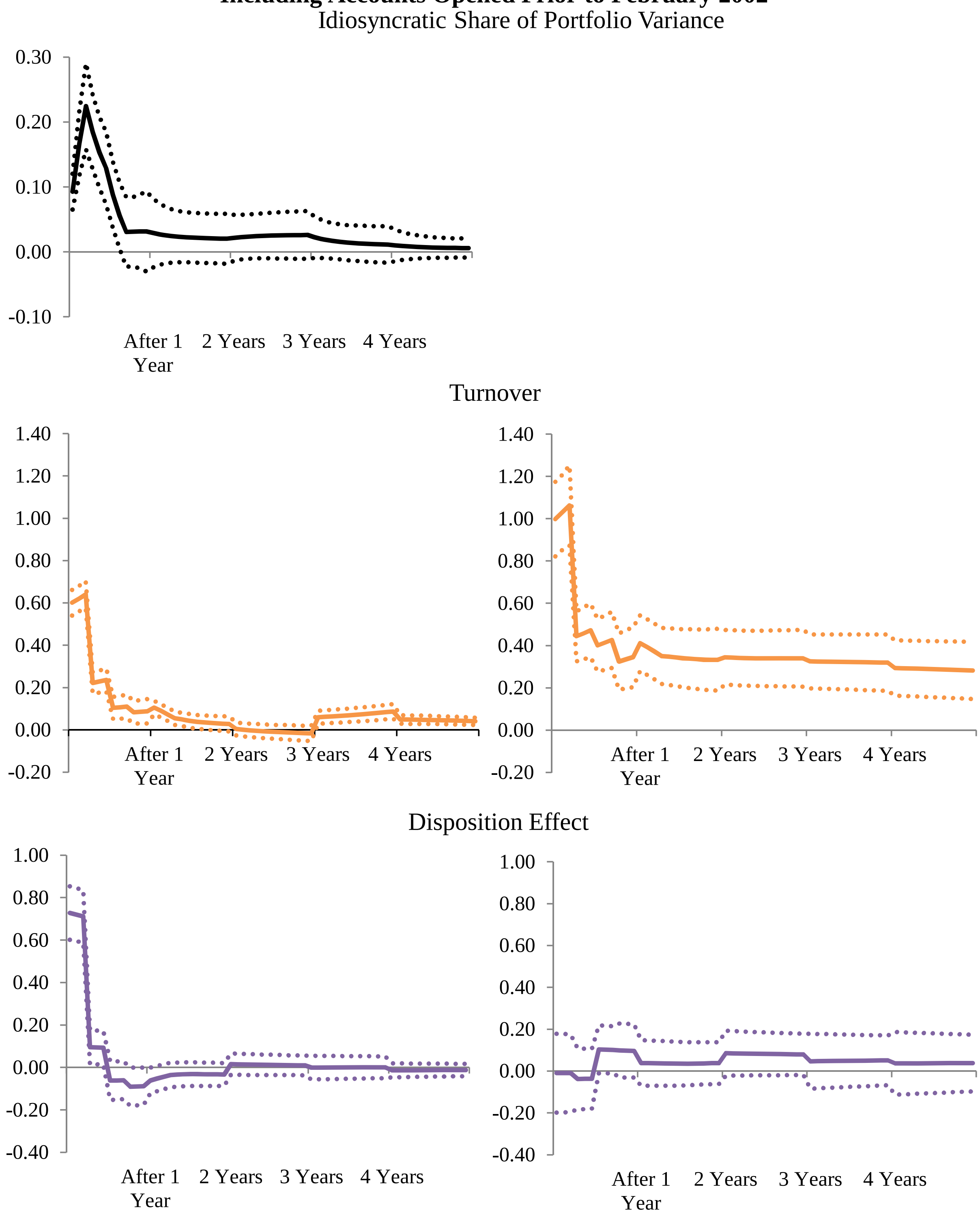


The plots above are analogous to those in Figure 1, but produced from data including accounts opened prior to February 2002.

Figure A16: Feedback Effects on Investing Behavior

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

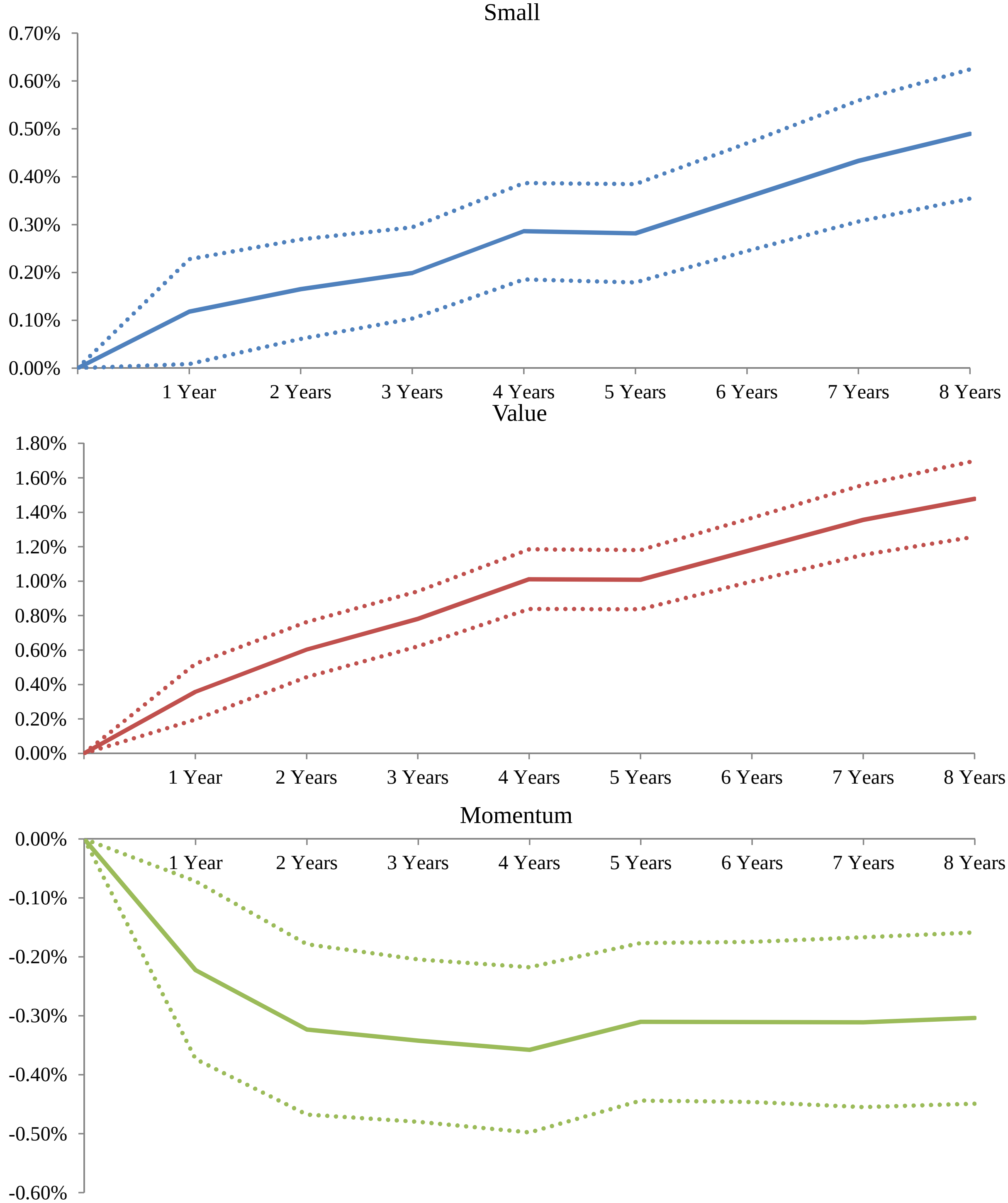
Including Accounts Opened Prior to February 2002



The plots above are analogous to those in Figure 2, but produced from data including accounts opened prior to February 2002.

Figure A17: Account Age Effects on Net Style Demand

Including Accounts Opened Prior to February 2002

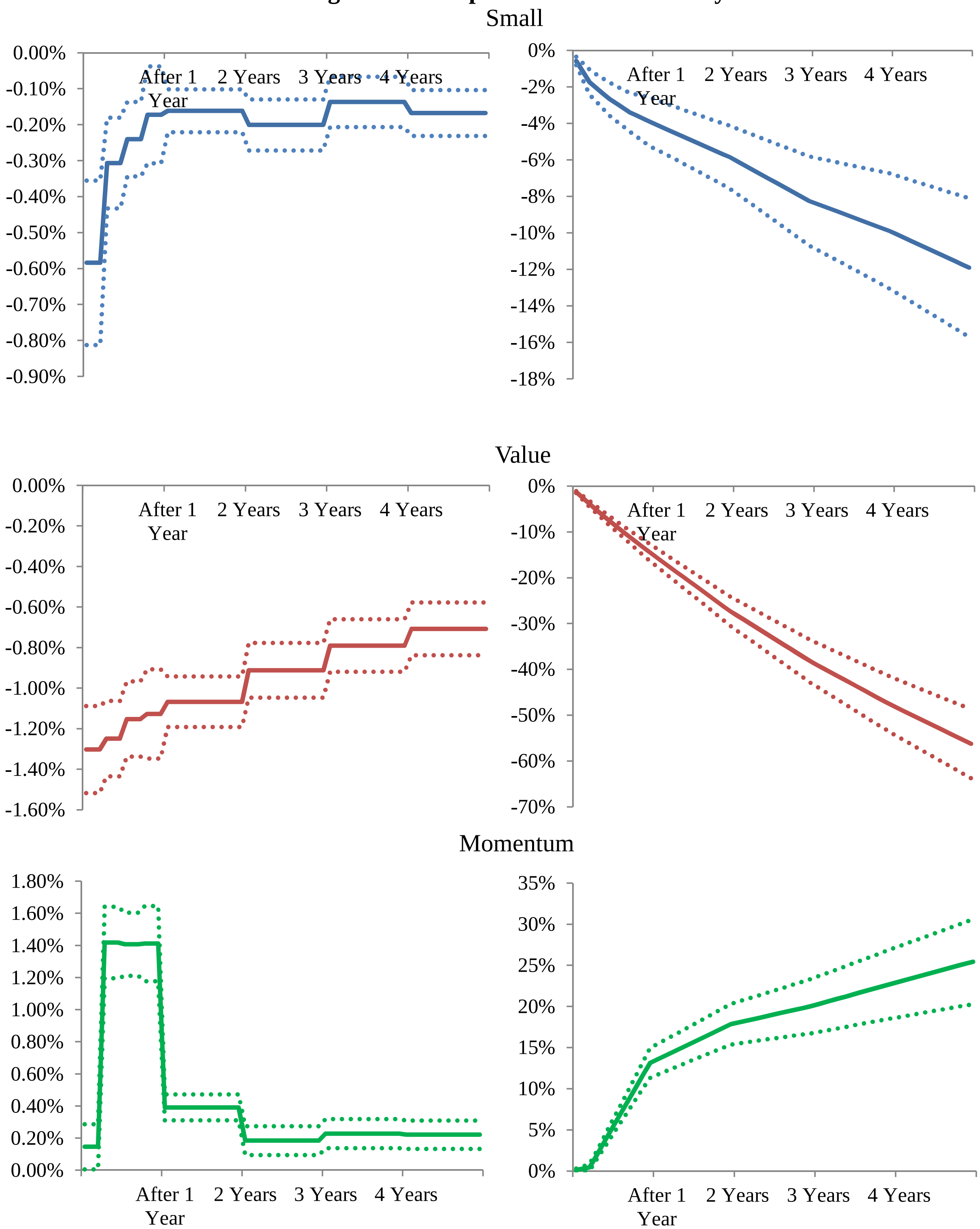


The plots above are analogous to those in Figure 4, but produced from data including accounts opened prior to February 2002.

Figure A18: Account Performance Feedback Effect on Net Style Demand

Response (Left) and Cumulative Response (Right)

Including Accounts Opened Prior to February 2002

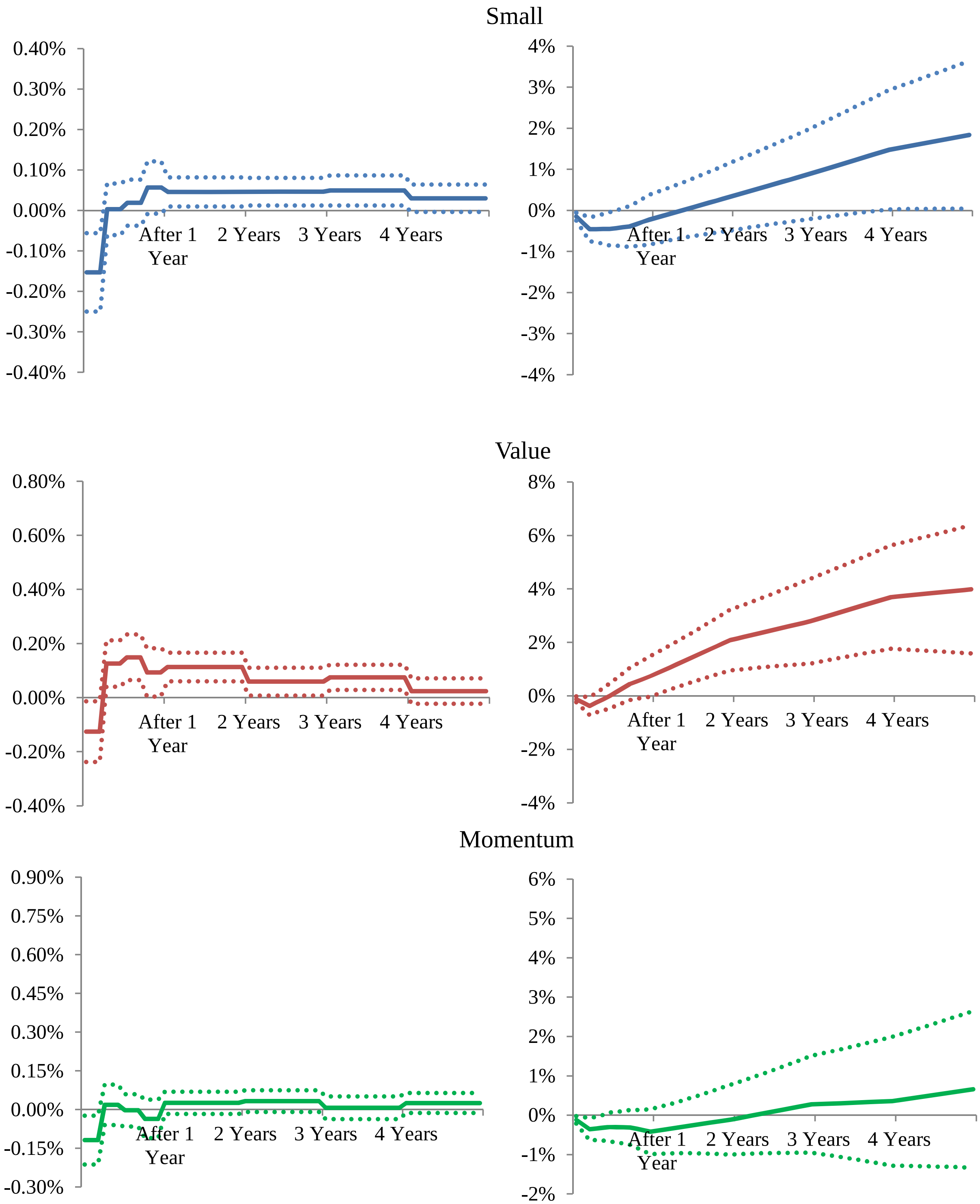


The plots above are analogous to those in Figure 5, but produced from data including accounts opened prior to February 2002.

Figure A19: Style Feedback Effect on Net Style Demand

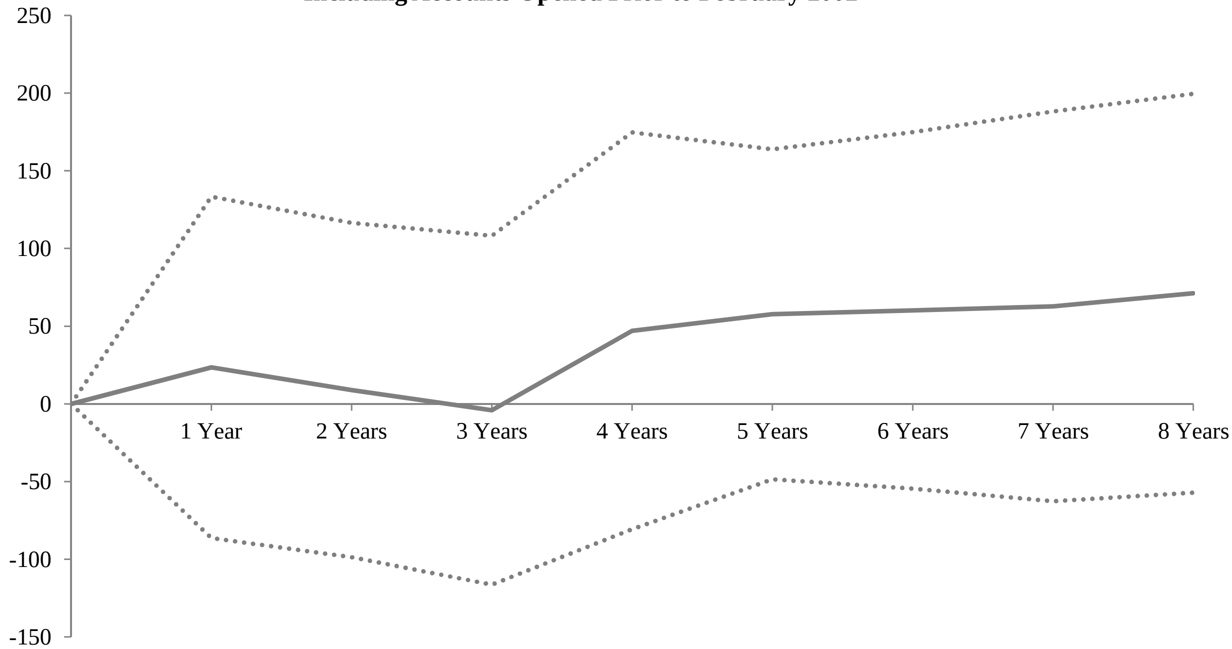
Response (Left) and Cumulative Response (Right)

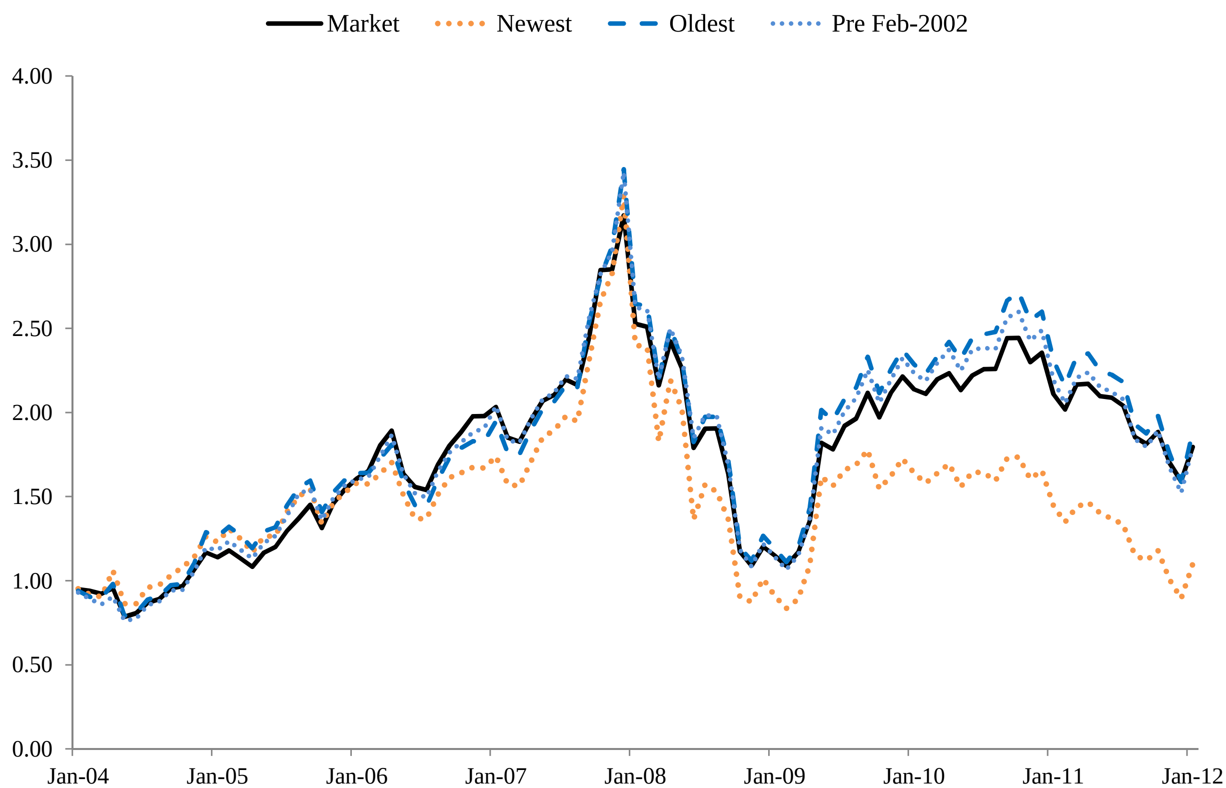
Including Accounts Opened Prior to February 2002



The plots above are analogous to those in Figure 6, but produced from data including accounts opened prior to February 2002.

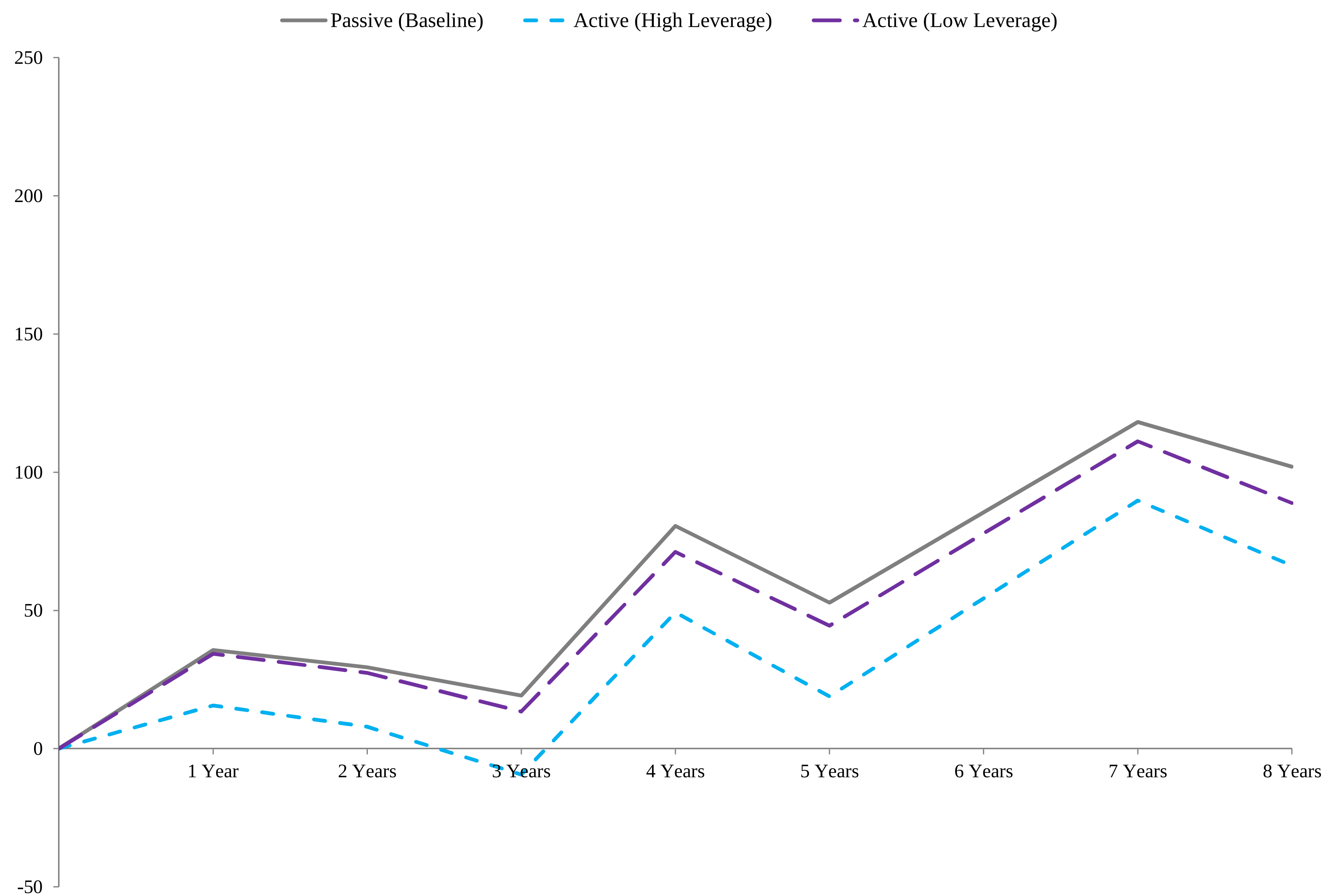
Figure A20: Top: Account Age Effects on Account Returns (bp/mo)
Bottom: Cumulative Excess Equity Returns to Old and New Accounts
Including Accounts Opened Prior to February 2002





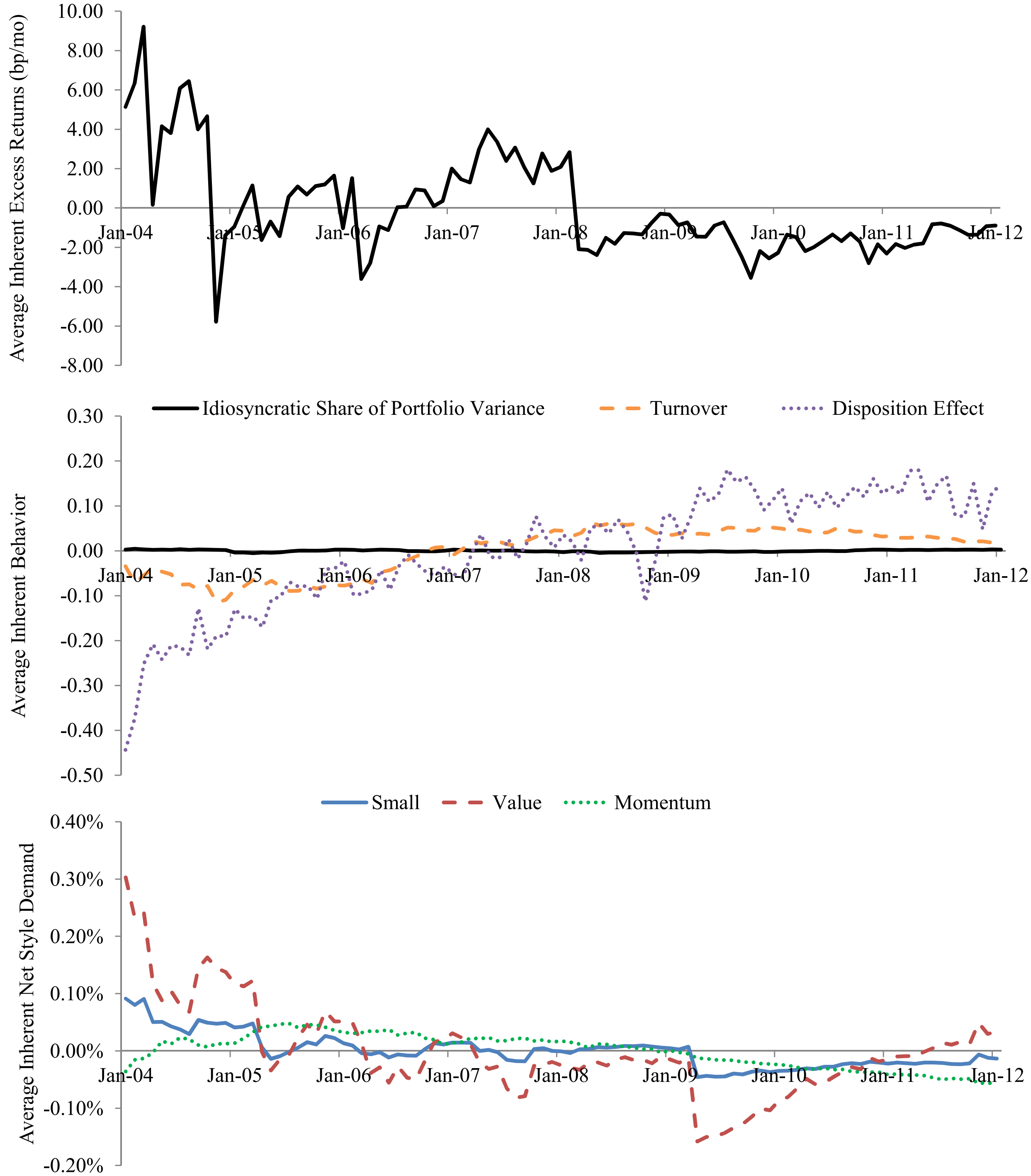
The plots above are analogous to those in Figure 8, but produced from data including accounts opened prior to February 2002.

Figure A21: Account Age Effects on Account Returns (bp/mo)
"Passive" (Baseline) versus "Active" Returns Used



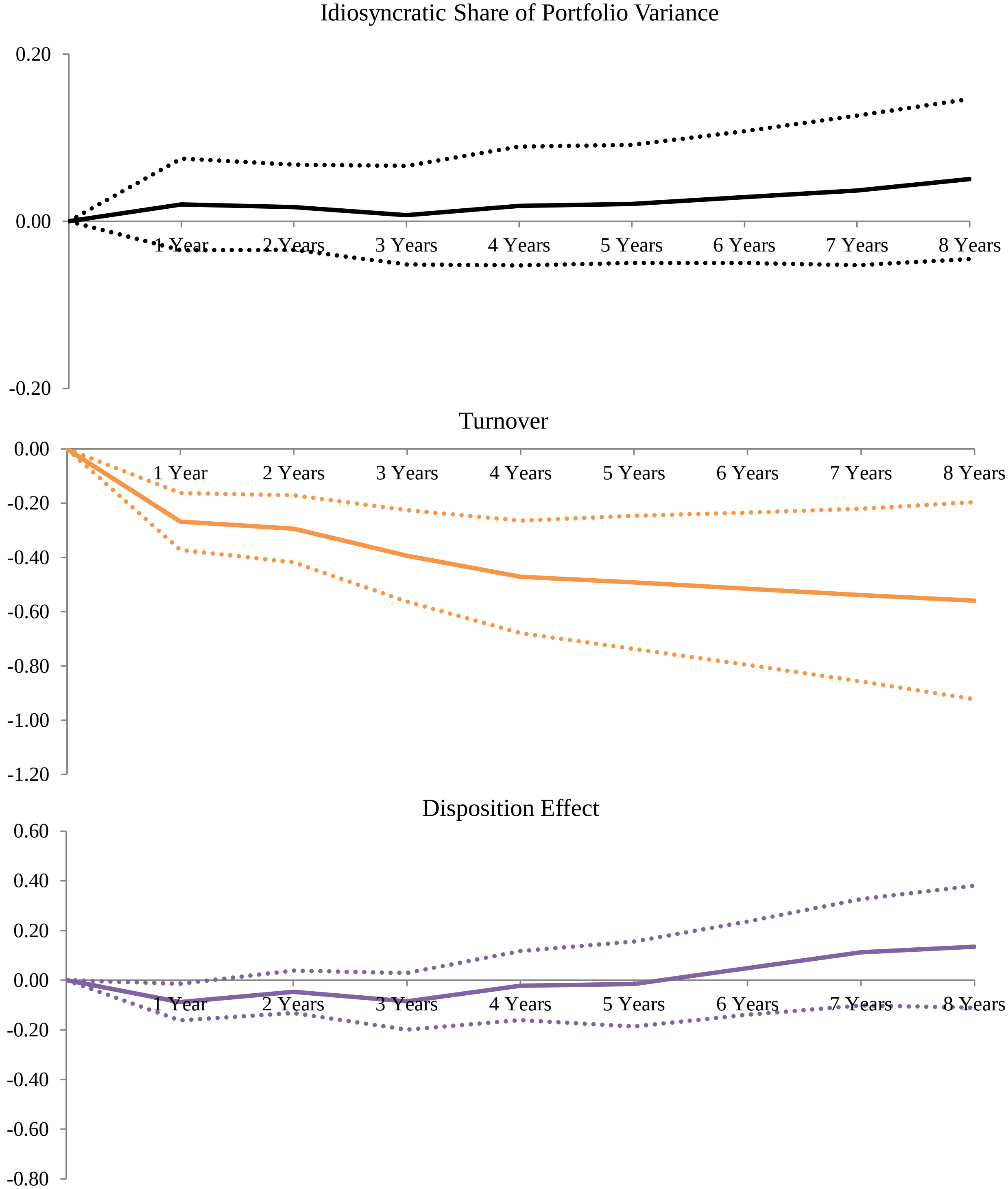
The plotted series replicates the top plot of Figure 8, as well as versions produced using both "low leverage" and "high leverage" active returns.

Figure A22: Estimated Average Inherent Returns/Behavior/Net Style Demand (αC_t from Equation 4)



The series above provide the de-meaned fitted values of αC_t estimated from regression equation (4). The fitted series represents predicted time-variation in the average investor's inherent returns/behavior/net style characteristic demand generated by time-variation in the inherent characteristics of investors in the market.

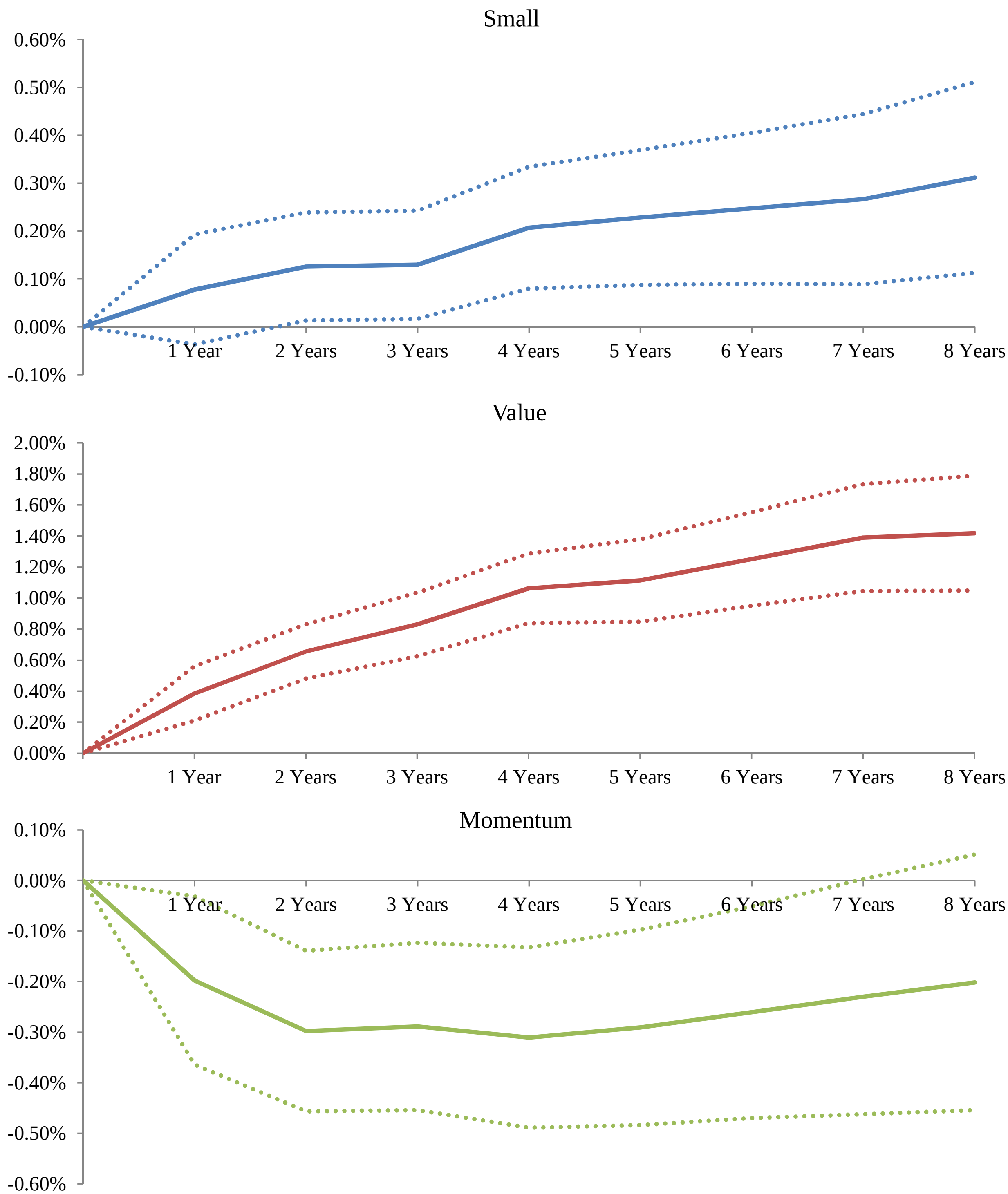
Figure A23: Account Age Effects on Investor Behaviors
Estimated Using Equation (4)



The plots above are produced in analogous manner to Figure 1 from investor behavior regressions which follow equation (4), including controls for the inherent behavior of investors present at each point in time .

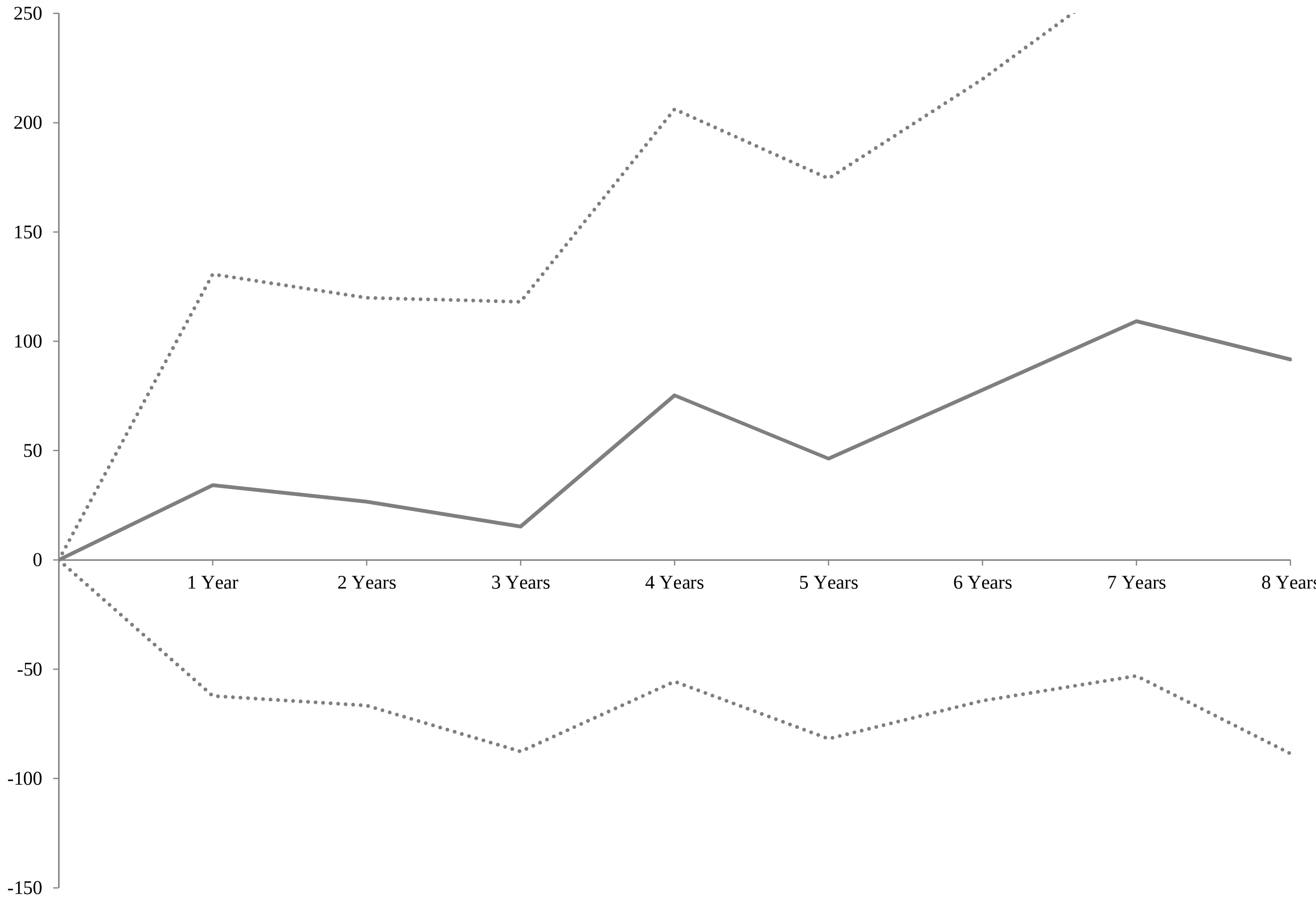
Figure A24: Account Age Effects on Net Style Demand

Estimated Using Equation (4)



The plots above are produced in analogous manner to Figure 4 from investor behavior regressions which follow equation (4), including controls for the inherent behavior of investors present at each point in time .

Figure A25: Account Age Effects on Account Returns (bp/mo)
Estimated Using Equation (4)

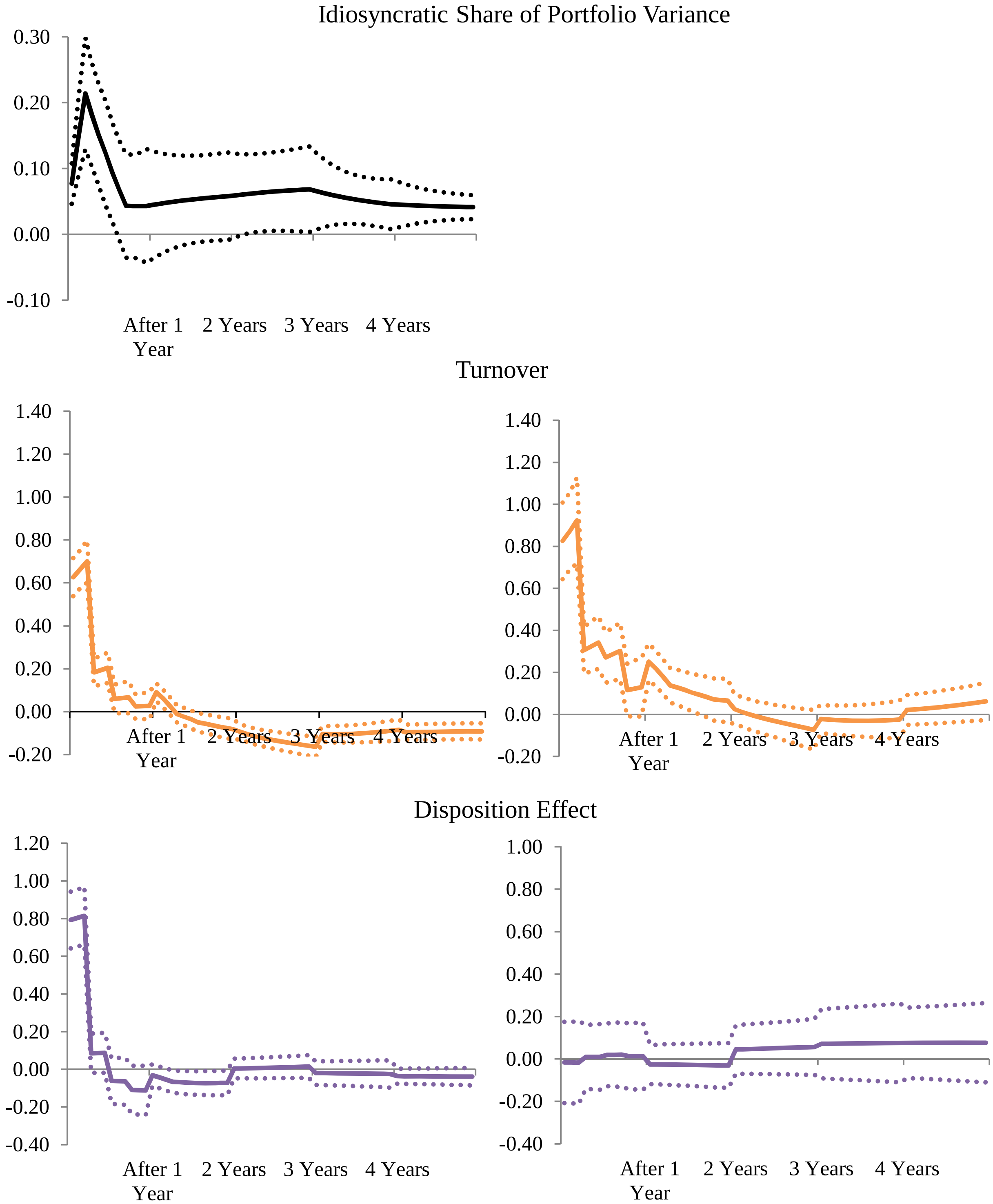


The plots above are produced in analogous manner to Figure 8 from investor behavior regressions which follow equation (4), including controls for the inherent ability of investors present at each point in time to earn excess returns.

Figure A26: Feedback Effects on Investing Behavior

Estimated Using Equation (5)

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

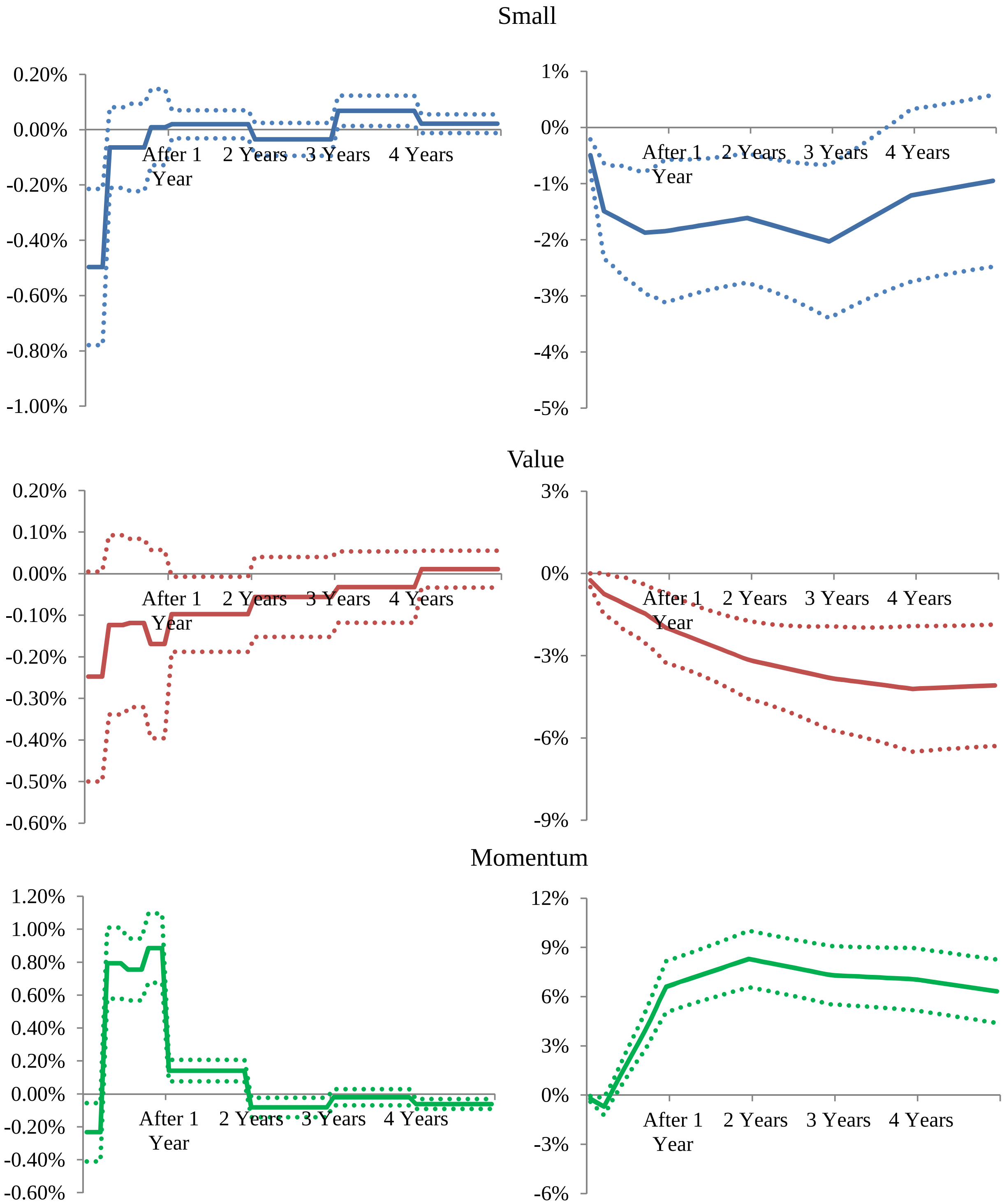


Plots are analogous to those in Figure 2, but generated using a specification using restricted individual effects, and time fixed effects (equation (5) in the text).

Figure A27: Account Performance Feedback Effect on Net Style Demand

Estimated Using Equation (5)

Response (Left) and Cumulative Response (Right)



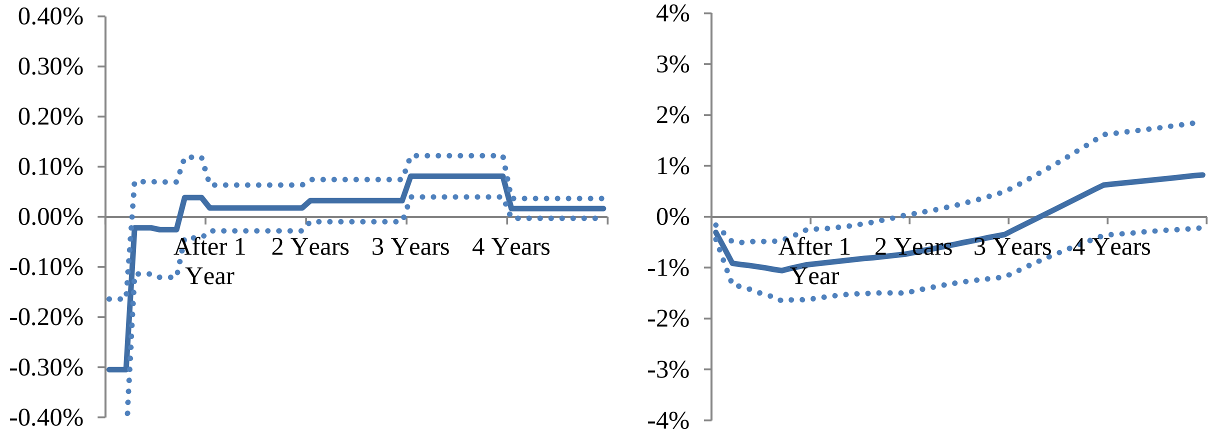
Plots are analogous to those in Figure 5, but generated using a specification using restricted individual effects, and time fixed effects (equation (5) in the text).

Figure A28: Style Feedback Effect on Net Style Demand

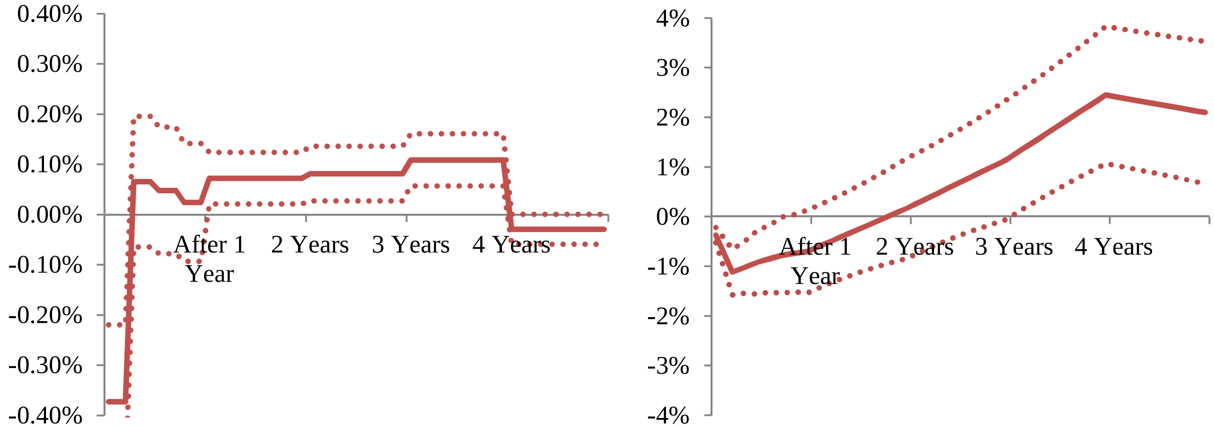
Estimated Using Equation (5)

Response (Left) and Cumulative Response (Right)

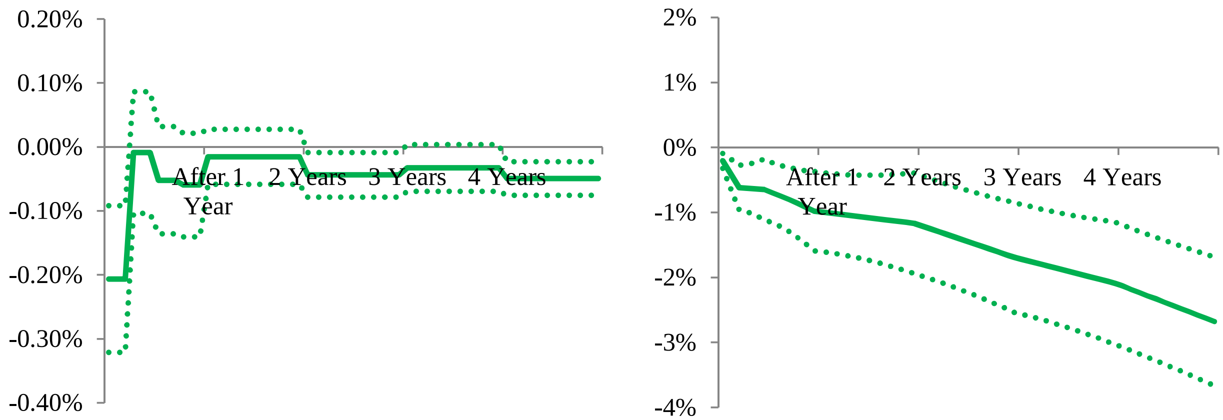
Small



Value



Momentum



Plots are analogous to those in Figure 6, but generated using a specification using restricted individual effects, and time fixed effects (equation (5) in the text).