Who Benefits from Robo-advising? Evidence from Machine Learning

Alberto G Rossi
University of Maryland

Stephen Utkus
Vanguard

September 24, 2018

Preliminary and Incomplete
Please do not Cite or Circulate Without Permission

Abstract

We study the effects of the largest US robo-adviser, Vanguard Personal Advisor Services (PAS), on investor performance. Across all clients, PAS reduces investors holdings in money market mutual funds and increases bond holdings. It reduces the holdings of individual stocks and US active mutual funds, and moves investors towards low-cost indexed mutual funds. Finally, it increases investors' international diversification and investors' overall risk-adjusted performance. From sign-up, it takes approximately six months for PAS to adjust investors' portfolios to the new allocations. We use a machine learning algorithm, known as Boosted Regression Trees (BRT), to explain the cross-sectional variation in the effects of PAS on investors' portfolio allocation and performance. The investors that benefit the most from robo-advising are the clients with little investment experience, as well as the ones that have high cash-holdings and high trading volume pre-adoption. Clients with little mutual fund holdings and clients invested in high-fee active mutual funds also display significant performance gains.

The paper has benefited from the comments made at presentations at the Vanguard Group. We are grateful to Francesco D'Acunto, Cindy Pagliaro, and YinYin Yu for comments and suggestions. Send correspondence to Alberto Rossi at arossi@rhmith.umd.edu.

Robo-advisers have surged in popularity in recent years as investors seek low-cost, automated investment opportunities. Robo-advisers allow investors to set up customized, diverse portfolios and can give access to wealth management services previously reserved for wealthy individuals like tax-loss harvesting and financial planning. In addition to be comparably inexpensive, robo-advisers have the potential to be superior to human financial advisers, as the latter have been shown to display behavioral biases and cognitive limitations (see Linnainmaa, Melzer, and Previtero, Forthcoming). As a result, robo-advisers are quickly attracting attention from investors at all levels.

This paper provides the first comprehensive analysis of a major US robo-adviser, the Vanguard Personal Advisor Services (PAS). PAS is currently the largest robo-adviser in the world—with $112 Billions AUM—and is almost 4 times larger than the second largest competitor, Schwab Intelligent Portfolios ($33 Billions). Other well-known robo-advisers are Betterment ($14 Billions), Wealthfront ($10 Billions), and Personal Capital ($7.5 Billions). PAS provides personalized investment solutions for clients at low costs. At sign-up, PAS clients are profiled on the basis of their financial objectives, risk-tolerance, investment horizons and demographic characteristics. They are then proposed a comprehensive financial plan. Clients are officially enrolled into PAS only after accepting the proposed plan and agreeing to move forward with the service. From that moment, PAS places trades automatically on behalf of the investor to reach the desired portfolio allocation. Investor positions are revisited quarterly by the algorithm and trades are placed if portfolio weights deviate substantially from target weights.

We first explore the effects of PAS across all clients. PAS operates significant changes on investors' portfolios. It increases investors' bond-holdings from 24% to 40% and decreases investors' cash holdings from 22% to 1%. On the other hand, we find little to no changes in equity holdings. PAS also operates very large changes on the investment vehicles held by investors. The proportion of wealth invested in mutual funds increases from 72% pre-PAS to 96% post-PAS. The increase in mutual fund holdings is mainly financed by reducing holdings in individual stocks, money market mutual funds and ETFs.

Not only PAS places investors into mutual funds, but it also affects the type of mutual funds clients are invested in. The percentage of wealth in indexed mutual funds almost doubles: it increases from 47% to 84%. Investors' international diversification increases threefold: the percentage of wealth in international mutual funds increases from 10% to 33%. Moving investors into indexed mutual funds results into lower fees paid by investors. Investors' average expense ratios are more than halved—from 19 to 9 basis points.

Turning to trading activity, PAS increases investors' trading volume for five-to-six months and it decreases it thereafter. The average monthly trading volume before signing up for PAS is $78,000 while it is only $12,000 twelve months after signup. In the intermediate period, however, the trading volume increases substantially. This is because it takes almost 6 months for PAS to achieve the clients' target portfolio allocations.

Finally, we estimate whether PAS affects clients' investment performance. As a measure of performance, we use the annualized abnormal Sharpe Ratio, that is, the difference between the realized Sharpe ratio of each client across all accounts and the realized Sharpe ratio of the market portfolio, where the latter is computed as the value-weighted returns on the NYSE/AMEX/NASDAQ CRSP portfolio. Irrespective of the horizon and the specification, we find investors' performance increases after PAS-adoption. For example, at the 6-month horizon, the post-PAS annualized Sharpe ratios average 0.115, statistically significant at the 1% level, while the pre-PAS Sharpe ratios average -0.067, significantly different from zero at the 1% level. As a result the difference in performance equals 0.182, statistically significant at the 1% level.

The average results computed across all clients hide considerable cross-sectional heterogeneity. In an effort to understand which customers are more likely to benefit from robo-advising, we explore the cross-section of clients using a machine learning algorithm known as Boosted Regression Trees (BRT). BRTs allow us to analyze what investor characteristics are valuable in explaining the cross-sectional variation in the changes in portfolio allocations as well as the changes in investment performance pre- and post-PAS.

For portfolio changes, the three most important client characteristics are 1) the proportion of wealth held in equities by the client at sign-up, 2) the age of the client, and 3) the proportion of wealth held in cash. We find a very strong and negative relation between the change in the equity share and the fraction of wealth in equities at sign-up. Those investors with no wealth in equities experience an increase in the share of equities of by PAS of 30%. On the other hand, PAS decreases by

30% the share of equities for those investors with 100% of their wealth in equities. Finally, PAS does not change the equity positions for those investors that already have roughly a 60-40 split between equities and bonds.

PAS systematically increases the equity exposure of the clients that have less than 55 years of age and decreases the equity exposure of those clients that have more than 55 years of age. The reduction in equity exposure is significant. It equals approximately -12% for the clients over 60 years old and almost +15% for those under 40 years of age.

Finally, those investors with no wealth in money market mutual funds experience a reduction in their equity share. At the other extreme, those investors with 100% of their wealth in money market mutual funds, experience an increase in equity share.

We find that a large number of clients' characteristics related to the change in performance pre- and post-PAS. Among them, we highlight that the cash share and the traded volume at sign-up are positively related to the improvement in performance post-PAS, indicating that those investors that were trading a lot and/or where holding a very large portion of their wealth in cash, benefit more from PAS.

Other economically important relations are those associated with clients' tenure, the percentage of wealth in mutual funds, and the percentage of mutual fund holdings in index funds. In all cases, BRTs uncover a negative relation, suggesting that those clients that were not holding a lot of indexed funds and were not holding a lot of their funds in mutual funds, are the ones benefitting more from signing up for PAS. Finally, BRTs uncover that less experienced individuals are the ones that benefit the most from PAS.

BRTs uncover also markedly non-linear and non-monotonic relations between the change in risk-adjusted performance and clients' age, the mutual funds' management fees and the number of assets held. For the first two covariates, the relation is U-shaped. The results suggest that the clients benefitting the most are the ones in their forties and mid-fifties, and the very senior citizens, while there is a negative relation between age and change in post-PAS performance for clients in the second half of their fifties and their sixties. This is because PAS increases the equity exposure of the clients in their forties and mid-fifties, decrease it for clients in the second half of their fifties and their sixties, and leave them unchanged for the clients in their seventies.

The relation between fees and change in performance is also U-Shaped, indicating that those customers investing in very expensive active funds as well as those investing in money market mutual funds—that charge close to zero management fees—benefit the most from PAS.

The relation between number of assets and performance change has instead an inverse U-shaped relation. This is because individuals with few assets are likely to be holding mutual funds. Associated with a higher number of assets are instead those clients that invest in individual equities. These customers do benefit from PAS as it increases their diversification. Those individuals that instead had 25 or 30 assets where likely to be already rather diversified, even if they were holding individual stocks. They therefore do not benefit much from adopting the robo-adviser.

Finally, the last covariate and most relevant covariate is the share of equities held (relative influence of 27%). The positive monotonic relation suggest that PAS increases investors' performance more for those with higher equity shares, indicating that PAS invests in a portfolio of mutual funds with higher risk-return trade-offs, compared to the average investor.

BRTs uncover—in many cases— strong non-linearities between regressand and covariates. To show that these are not the result of over-fitting, we perform an out-of-sample cross-validation exercise. We show that BRTs do not overfit the training sample and that they provide superior in- and out-of-sample performance, compared to linear models that use the same covariates. In fact, BRTs perform so much better than linear models in our context that the *out-of-sample* performance of BRTs is superior than the *in-sample* performance of linear models.

1 Related Literature

TBC

2 Data Construction

The study uses proprietary data from Vanguard as well as data from a variety of data sources. The four Vanguard data tables used in our study are named *Trades*, *Positions*, *Client Demographics*, *Client-Advisor Mapping and Appointment*. We provide a brief description of each data table below.

Trades

The Trades table includes the record of all the trades made by account holders that have interacted with PAS over the period January 2015 through December 2017. This comprises 356,416 account holders, but we work with a random sample of 50,000 account holders. The Trades table has approximately 4,944,019 trades and each observation contains the following information: SPOID, a client identifier; ENTRPRISE_ACCT_ID, a unique account identifier; CUSIP_NO, the cusip of the security traded; TCKR_SYM, the ticker of the security traded; HIGH_LEVL_TXN_TYP_CD, a code indicating the type of transaction; GROSS_AM, the gross amount of the transaction; PROCS_DT, the process date of the transaction.

Positions

The Positions table includes monthly holdings for the 50,000 clients in our sample. The file has 15,123,616 observations and the following variables: SPOID, a client identifier; ENTRPRISE_ACCT_ID, a unique account identifier; MONTH_END_DT, the end-of-month date associated with the holding; CUSIP_NO, the cusip of the security traded; TCKR_SYM, the ticker of the security traded; POSN_BAL_AM, the balance amount as of the month end date; MANAGED_FL, a flag to indicate whether the account is managed with Vanguard's PAS program, or not (self-managed).

Client Demographics

The *Client Demographics* table contains information on the characteristics of the clients. SPOID, a client identifier; CLNT_SGMNT_CD, the client segment; GENDR_CD, the client gender; CLNT_ENTRY_DT, the date on which the client joined Vanguard (not necessarily as a PAS client); MRTL_STATUS_CD, client marriage status as of Dec 2017; AGE, client age as of Dec 2017; STATE, client state of residence as of Dec 2017; INIT_DT, date the client initiated PAS; ENROLL_DT, date the client actually enrolled in PAS; IMPLM_DT, date when the investment recommendations are implemented; OFFBRD_DT, date when the client is no longer in PAS and is designated a RMAT status.

Client-Advisor Mapping and Appointment

The *Client-Advisor Mapping and Appointment* table contains information on the relation between clients and advisors. It contains the following variables: ASGN_ADVSR_PO_ID, id of client's advisor if the client has a designated advisor; ASGN_ADVSOR_EFFTV_BGN_DT, date the client-advisor relationship began if the client has a designated advisor; ASGN_ADVSOR_EFFTV_END_DT, date the client-advisor relationship ended if the client has a designated advisor; CREW_DIVISN_NUM, division number of advisor for designated advisors only; CREW_DEPT_NM, department name of advisor for designated advisors only; CREW_DPT_ORG_ID, department organization id of advisor for designated advisors only; ASSGN_ADVSR_SUPV_PO_ID: id of advisors' supervisor for designated advisors only; APPT_ID, the appointment identifier; APPT_SCHDLD_DT, the date when the appointment is scheduled; APPT_STRT_DT, the date of the appointment; APPT_STRT_TM, the start time of the appointment; APPT_END_TM, the end time of the appointment; APPT_STATUS_CD, the appointment status, categorized into "Scheduled',' "Complete," "No Show," "Rescheduled," "Reassigned," and "Canceled;" SCHDL_BY_INTRNL_FL, flag indicating whether the appointment was scheduled by the crew or the client; CMNT_TXT, comments from the client; APPT_STATUS, a long text description of appointment status; AVT_DESC_TX, a string describing the type of appointment; MTG_DURTN_MIN_QY, the meeting duration.

Additional Data Sources

Stock market information such as prices, returns and trading volumes – among others – is obtained from CRSP, and CRSP Mutual Funds. In addition, the CRSP Mutual Funds database contains information regarding mutual fund fees, turnover, expense ratios, investment allocations, degree of indexation and the mutual fund classification provided by Lipper.

3 Robo-Advising and Portfolio Characteristics

In this section, we first present demographic and portfolio characteristics of robo-adviser investors before they sign-up for the PAS service. We then present how the portfolio characteristics of PAS investors change over time after they sign up for the robo-advising service. Finally, the last part of this section analyzes the type of assets PAS and non-PAS account-holders are invested in.

We compute the main results at the client level, but occasionally present results the account-level to highlight how investors and PAS behave differently in taxable and non-taxable accounts.

3.1 Demographic and Portfolio Characteristics Pre-PAS

We start by reporting demographic and portfolio characteristics of the investors that sign up for PAS, computed the month before the investors sign up for the service. The results are reported in Table 1, where for every variable we report mean, standard deviation and various percentiles of the distribution—ranging from the 1^{st} to the 99^{th} percentile. Panel A focuses on the demographic characteristics. The average investor is 63 years old and the median is 65; 53% of the users are males and 35% of the customers are married. The tenure at vanguard varies a lot. It ranges from half a year at the first percentile to almost 36 years at the 99% percentile. For comparison, the average customer age is 51 in Gargano and Rossi (2017) and Barber and Odean (2001). The percentage of women, which equals 46%, is larger in Vanguard compared to both both Gargano and Rossi (2017), 27%, and Barber and Odean (2001), 21%. At approximately, 14 years, average client tenure is also longer compared to other brokerage account datasets in the literature. Average client tenure in Gargano and Rossi (2017) is less than 9 years.

Panel B of Table 1 reports results for portfolio allocation. Clients' wealth is substantial. It averages \$580,815 and is heavily skewed to the right. The median invested wealth in each account is \$279,065 and it exceeds 4 million dollars at the 99-*th* percentile. The number of assets per account is 7.7 and the median is 5. It may seem that these Vanguard investors are heavily under-diversified, but this is really not the case, because many of these investors are very heavily invested in mutual funds. On average, 72% of the wealth is invested in mutual funds rather than individual stocks, so investors are likely to be very diversified, even if they hold only 5 assets.

The average investor has 54% of his/her portfolio invested in equities, followed by 24% in bonds, and 22% in cash—mainly money market mutual funds. These averages hide a very large cross-sectional variation, with almost 10% of the investors almost completely invested in equities and 15% of the investors invested only in bonds and/or cash. Stocks and bonds are not held directly, but mainly through mutual funds. In fact, 70% of the wealth is invested in mutual funds, followed by cash at 20%. Interestingly, only 3% of investors' wealth is held in individual stocks and 3% in ETFs. Finally, only a negligible number of clients have direct exposure to corporate bonds and options (not reported in the table).

Mutual fund holdings can be decomposed according to the fund strategies. We isolate indexed mutual funds using the "IndexFlag" from the CRSP mutual fund database. We also identify the funds with international exposure as the ones classified as either "international" or "global" by the Lipper classification. Finally, we identify the emerging market funds using the "emerging" Lipper classification. As reported at the bottom of Panel B, 47% of mutual fund holdings are in indexed mutual funds, while 10% of mutual fund holdings are in funds that invest internationally. Finally, only a negligible number of customers invest in mutual funds with emerging market exposure.

Panel C focuses on fees and transactions. Starting from mutual fund fees, the average management fee is 14 basis point, but some investors spend as much as 58 basis points a year in management fees. The expense ratio results are similar. The average is 0.19, the median is 0.14, and some investors have expense ratios close to 1% per year. The third row of Panel C focuses on the turnover ratio of the mutual funds held, that averages 0.32. In terms of active transactions, investors place on average 3 transactions per month, for an average of $79,000 dollars. As we show below, these quantities do not represent the steady-state level of investor activity, because investors make more transactions in the months immediately preceding PAS sign-up. They generally transact in an effort to consolidate their accounts before enrolling into PAS.

Tables Online I and Online II repeat the analysis of 1, but conducts the analysis at the account-level and break down the analysis into taxable (Table Online I) and non-taxable (Table Online II) accounts. We highlight the differences between the two account-types below. The demographics characteristics in Panel A are virtually identical across the two tables. The portfolio characteristics in Panel B is where we start seeing some differences. Average wealth is higher, at $400K and with tails close to $3.8M dollars, for brokerage account holders, compared to IRAs. IRAs' average balance is $208K and is rarely exceeds $1M—the 99-*th* percentile is $1.4M. Brokerage accounts also hold more assets than IRAs, 5 versus 3.5.

The equity share is similar across the two account types at 55%, but taxable account holders have a lower percentage of their wealth in bonds, 20%, compared to 26% for IRA accounts. As a result, the percentage of wealth in money market mutual funds (cash) is higher for taxable account holders than IRA accounts, 24% versus 20%. In terms of investment vehicles, 62% of investors' wealth is in mutual funds. It is instead 74% for IRA accounts. Brokerage account holders are also more likely to hold directly money market mutual funds, 23% of wealth compared 18% for IRA accounts, and brokerage accounts also tend to have more individual stock holdings, 9% compared to 3% for IRA accounts. Finally, taxable accounts have more ETFs than IRAs—5% of wealth versus 3%—and the degree of indexation is greater in IRA accounts, compared to brokerage accounts—49% versus 40%—testimony that clients take more active management positions in taxable account than non-taxable accounts.

Finally, with almost 2 transactions per month, taxable accounts have more trades than non-taxable accounts that average a little over one transaction per month. Consistent with the number of active trades, also the trading volume is higher for taxable accounts: $43K versus $30K for IRAs.

3.2 Demographic and Portfolio Characteristics post-PAS

In this section we report the effect of PAS on investors' portfolio allocation after signing up for the service. We also show that the portfolio changes do not occur overnight, but take several months. The results are reported in Table 2 and Figures 1 and 2.

In Table 2, we compute the same quantities of Table 1, but focus on the 12 months after PAS adoption. In Figures 1 and 2, we instead focus on those investors for which we have at least 12 months of portfolio allocation before and after signing up for PAS. We then track their portfolio changes every month. Focusing on the results in Tables 1 and 2 has the advantage of maximizing the number of observations. However, the results mix the effect of PAS with the sample composition effect, as we do not have as many clients 12 months after signing up for PAS compared to the month before signing up. On the other hand, focusing on the results in Figures 1 and 2 have the advantage of purging the results from any composition effect, as we only track investors that survive over the 24-month window of PAS adoption. The drawback is that the number of clients used to compute the results is lower.

The demographic characteristics such as age, tenure, proportion of males and married people are unchanged by construction. We report these quantities for completeness in Panel A of Table 2. Panel B reports the portfolio allocation results. At $700K, average wealth is higher than in Table 1. This is the result of stock market appreciation and clients' contribution to their portfolio. The number of assets in each account decreases slightly from 7.66 to 7.45. The percentile distribution shows that PAS shrinks dramatically the number of stocks held in the tails of the distribution. The 99-*th* of the number of assets held in each account drops from 39 in Table 1 to 24 in Table 2.

Continuing with the results in Panel B, portfolio allocation is where we observe strong changes, particularly in the allocation to bonds and money market mutual funds (cash). The percentage of bonds increases by 16 percentage points to 40%, while the allocation to cash decreases by 21 percentage points to only 1%. Finally, the equity share increases by 4 percentage points to 58%. The next four lines in Panel B of Table 2 focus on the investment vehicles used. Almost all of investors' wealth—96% of it—gets invested in mutual funds, with almost no share of wealth in money market mutual funds, ETFs, or individual stocks.

PAS has a very large effect on indexation and international diversification as well. Before PAS, the average investor has 47% of their wealth in index funds. This increases to 84% after joining PAS. We find a similar effect for investor's exposure to international markets, that increases from 10% to 33%. Interestingly, we do not find much of an effect in terms of emerging markets exposure that, if anything, declines after PAS. As we show later, this is mainly due to the fact that many international mutual funds (VTIAX, for example) have emerging market exposure that is not well captured by the Lipper classification.

Panel C of Table 2 shows PAS moves investors to mutual funds with lower fees and turnover ratios. Mutual fund fees are halved, from 14% to 7%, while the expense ratio is reduced by more than 50% as it drops from 0.19 to 0.09. The turnover ratio instead drop by approximately 20%, from 0.32 to 0.27. Finally, the average amount traded drops from $33,000 to only $3,000 per month a year after signing up for PAS.

The results in Table 2 are computed 12 months after PAS. In figures 1 and 2, we show the time-series behavior of the most important quantities before and after signing up for PAS. In each plot, the blue line represents average values while the red dashed lines are 95% confidence intervals. Time "0" represent the month before investors sign up for PAS.

Subfigures (a), (b), and (c) of Figure 1 show the time-series behavior of the changes for bond, cash and equity-holdings, respectively. Subfigures (d), (e), and (f) plot instead the results for wealth directly invested in mutual funds, ETFs, and individual stocks. We highlight several findings. First, it takes more than 6 months for PAS to fully converge to the new portfolio allocations. This has—potentially—important implications when it comes to evaluating investors' performance and characteristics pre- and post-PAS. Second, a close look at the plots also reveals that the confidence bands around the average values (the blue lines) shrink after signing up for PAS, indicating that PAS has a significant effect in homogenizing investors' portfolios. Finally, the results in Figure 1 are very much in line with the ones in Tables 1 and 2.

Figure 2 presents the results for indexation, international and emerging markets exposure, trading and fees. In all cases, the changes takes place over the course of 6 months and are in line with the results in Tables 1 and 2. The trading volume results are unique as they display marked non-monotonicities. Trading volume spikes for approximately two months after PAS enrollment. We then observe a gradual monotonic decline in trading volume that reaches a new steady-state after approximately 6 months.

3.3 Conditioning by Account type

The results in Section 3.2 are computed at the client level, and therefore combine taxable and non-taxable accounts. As we show in Section 3.1, however, taxable and non-taxable accounts display a number of differences when it comes to portfolio characteristics. In an effort to understand the differential effect of PAS on taxable and non-taxable accounts, we re-compute Table 2 using only brokerage and IRA accounts. The results are reported in Tables Online III and Online IV, respectively. We also recompute the results in Figures 1 and 2 for brokerage accounts (Figures Online I and Online II) and IRA accounts (Figures Online III and Online IV).

As can be seen by looking at figures Online I through Online IV, the results for taxable and non-taxable accounts are, in many respects, very similar. In all cases, we observe that the percentage of wealth in money market mutual funds decreases substantially. The same is true for the percentage of wealth in ETFs and individual stocks as well as for management fees and expense ratios. Likewise, we find that—across both taxable and non-taxable accounts—PAS is associated with an increase in the percentage of wealth in mutual funds and an increase in indexation.

There are, however, at least two major difference between taxable and non-taxable accounts. First, while the percentage of wealth in equities is similar across taxable and non-taxable accounts pre-PAS, PAS increases dramatically the equity positions of taxable accounts to approximately 80%. It instead lowers the equity positions of IRA accounts to approximately 50%. The percentage of wealth in bonds instead decreases to 15% for brokerage accounts and increases to 50% for IRAs. PAS therefore tends to increase portfolio risk for brokerage accounts and lower it for IRA accounts.

The other major difference between taxable and non-taxable accounts has to do with international diversification. While PAS increases international diversification for both types of accounts, it does it to a larger extent for IRA accounts, compared to brokerage accounts. As can be seen in Tables Online III and Online IV, the percentage of mutual fund wealth in funds with international exposure equals 19% for taxable account and 45% for IRA accounts.

Finally, Tables Online III and Online IV reveal that post-PAS taxable accounts have roughly a 80-20 equity-to-corporate bonds allocation, while post-PAS non-taxable accounts have roughly a 50-50 split. At the client level, this results in a 60-40 split with only a very small percentage of wealth in money market mutual funds.

3.4 Portfolio Allocation of PAS and non-PAS investors

To better understand how PAS invests the wealth of its investors, we study the main holdings of PAS and non-PAS investors. Overall, PAS lowers the number of assets held across investors. We show it in three way. First, we present the top tickers held across PAS and non-PAS investors in January 2017.[1] These are reported in Table 3. Starting with mutual funds in Panel A, PAS investors are invested in four mutual funds, i.e. VTSAX, VTIAX, VBTLX, and VTABX. Combined, these four mutual funds represent almost 75% of PAS wealth in January 2017. The mix of funds is designed to expose PAS investors to both US and international diversified stock and bond portfolios.

VTSAX is the Vanguard Total Stock Market Index Fund Admiral Shares. It is the indexed mutual fund equivalent of the VTI ETF. Its benchmark is the CRSP US Total Market Index, which represents 100% of the CRSP US stock market index. The fees are extremely low: management fees are 0.03% per year and total annual operating expenses of only 0.04%. VTIAX is the Vanguard Total International Stock Index Fund Admiral Shares. The fund invests in European Equities (42%), Pacific Region (30%), Emerging Markets (21%) and North America (6.6%). The fees of this fund are also rather low, 0.08% management fees and 0.11% total annual operating expenses. The last two are bond funds. VBTLX is the Vanguard Total Bond Market Index Fund Admiral Shares and invests in public, investment-grade, taxable, fixed income securities in the US—including government bonds, corporate bonds, mortgage-backed and asset-backed securities. The fees are, once again, extremely low: 0.04% management fees and 0.05% total annual operating expenses. Finally, VTABX is the Vanguard Total International Bond Index Fund Admiral Shares. The fund invests in government, government agency, corporate, and securitized non-U.S. investment-grade fixed income investments, all issued in currencies other than the U.S. dollar and with maturities of more than one year. The fund features 0.09% management fees and 0.11% total annual operating expenses.

We see a similar effect in ETFs, where 30% of the PAS investor wealth in invested in VTI. In terms of stock holdings, we do not see any effect, possibly because the PAS does not do anything there. Finally, turning to cash, we find that the result all the wealth of PAS investors is concentrated

[1] The results are very similar if we compute the results for other dates.

in the money market funds: VMMXX, VMSXX, and VUSXX. Together, the three comprise 89% of investors' wealth.

Next, we analyze whether the concentration of wealth across tickers has varied over time. We also measure the extent to which wealth is dispersed across few or many assets in the cross-section. The results for the first exercise is reported in Figure 3, where we report—for each asset class—the proportion of wealth allocated to the top 5 tickers— across all Vanguard investors. The red dotted line represent results for PAS client while the black solid line the results for non-PAS clients. The results in Figure 3 are very much in line with those in Table 3. The cross-sectional results are reported in Figure 4. The figure plots the cumulative wealth invested across tickers, from the more to the least purchased. The red dotted line represent results for PAS client while the black solid line the results for non-PAS clients. The vertical red line represents the total number of either mutual funds, ETF, stocks, or money market funds held by each investor category, that is, PAS and non-PAS. The results show that PAS investors are invested in approximately 1,000 mutual funds in total, while non-PAS clients invest in almost 4,000 funds. You can also see that more wealth is concentrated in fewer funds for PAS clients. Largely, the results for ETFs and money market mutual funds are similar. PAS investors have more concentrated portfolios than non-PAS investors. The total number of ETFs held across non-PAS investors is close to 900, while the one for PAS investors is less than 300. Finally, PAS investors also invest in less individual stocks. Non-PAS clients collectively invest in as many as 4,500 individual stocks. The respective number of stocks for PAS investors is less than 1,000.

4 Average Performance before and after PAS

The portfolio allocation results reported so far suggest that PAS may improve investors' performance as it places account-holders in diversified US and international low-fee indexed mutual funds. It also reduces investors' cash holdings. In this section, we provide a comprehensive analysis of the pre- and post-PAS investment performance.

As a measure of performance, we use the annualized abnormal Sharpe Ratio, that is, the difference between the realized Sharpe ratio of each client across all accounts and the realized Sharpe ratio of the market portfolio, where the latter is computed as the value-weighted returns on the NYSE/AMEX/NASDAQ CRSP portfolio. For each account-holder, portfolio returns and volatilities are computed using beginning-of-month investment holdings. Furthermore, portfolio volatilities are computed as realized volatilities using squared daily returns. Performance is computed starting at adoption or starting 6 months before and after PAS is implemented in each account. The latter is to account for the fact that it takes approximately half a year for PAS to reach the new steady-state portfolio allocations after the investor signs up, as we show in Section 3.2.

The results computed across all clients are reported in Table 4. Panel A reports results for annualized Sharpe ratios computed at the 3-month horizon. Panels B and C reports results computed at the 6-, and 9-month horizons. Each column within each panel reports average abnormal Sharpe ratios, t-statistics testing the null hypothesis that the averages are equal to zero, and the number of observations used in the computations of the results. The first and second column of each panel report average abnormal Sharpe ratios after (column 1) and before (column 2) signing up for PAS, computed across all clients available. Columns 3 and 4 repeat the the exercise only for those accounts in the sample both before and after—we refer to them as "matched" clients. The last column reports the average performance difference after and before signing up for PAS for the matched clients.

Panel A shows the results at the 3-month horizon. Across all accounts, the average abnormal Sharpe ratio after signing up for PAS is positive, but not statistically different from zero. On the other hand, the abnormal Sharpe ratio before PAS sign-up is effectively. The matched results are stronger. The post-PAS abnormal Sharpe ratio is 0.059, significant at the 1% level. The pre-PAS Sharpe ratio is instead negative and not distinguishable from zero. The difference between the two equals 0.065 and is statistically significant at the 1% level.

The 6-month results in Panel B are stronger. Across all accounts the abnormal Sharpe ratio post-PAS is positive and significant at the 1% level. It is instead indistinguishable from zero pre-PAS. The matched results show that the post-PAS Sharpe ratios equal 0.115, statistically significant at the 1% level, while the pre-PAS Sharpe ratios equal -0.067, and are significantly different from zero at the 1% level. As a result the difference in performance equals 0.182, statistically significant at the 1% level.

The 9-month results in Panel C are again consistent. The abnormal Sharpe ratios post-PAS is positive and significant, the pre-PAS is either negative and significantly different from zero or insignificant.

The results starting the computations 6 months before and after PAS adoption are stronger, as expected. At the 3-, 6-, and 9- month horizons, we consistently find that the abnormal Sharpe ratio is positive and significant, with values ranging between 0.141 and 0.160, when computed across all clients. The pre-PAS Sharpe ratios are instead negative and significant or indistinguishable from zero, with values ranging between -0.033 and 0.001. The matched results aways show that the post-PAS performance is superior than the pre-PAS performance, with the highest value equaling 0.399 at the 9-month horizons, statistically significant at the 1% level.

5 Why Use Machine Learning to Assess the Effects of Robo-Advising

The previous sections analyzed how PAS changed the investment portfolios and the investment performance of the average investors. However, depending on the clients characteristics and the portfolio allocation of the client at the time of adoption, we can expect the robo-adviser to have a differential impact on the portfolio allocation and the performance of the investors.

As a motivating example, we work with the share of equities across investors. In Figure 1, we showed the average change in the equity share was rather small across investors. This result, however, hides a very large heterogeneity across investors. To illustrate the point, we report in Figure 6 portfolio changes for clients with low (less than 10%) and high (more than 90%) equity shares before signing up for PAS. In both cases, PAS resulted in a major portfolio overhaul. In the first, case PAS increased equity holdings from approximately 5% to almost 50%. In the second, it decreased it from 95% to approximately 65%.

The portfolio changes operated by PAS are largely a function of the investment portfolio of the investors at sign-up as well as investor preferences and demographic characteristics. For example, older individuals are likely to be assigned a lower share of their wealth in risky assets, while younger individuals a higher share. Investors' lifestyle may also play a role. Investors with higher projected expenses are likely to be assigned different investment portfolios, because of the investors' cash availability. Finally, investors' preferences such as risk-aversion play a role. The final portfolio allocation of each client is the result of numerous steps implemented by different divisions of the company. It is therefore difficult to know what factors play a role.

A standard way to analyze this problem would be to use linear regression, but it is not clear that investors' demographic and portfolio characteristics are linearly related to the changes in investors' portfolios. It is also not clear *ex-ante* what factors would be relevant. The result of running a kitchen-sink regression is that we would likely run the risk of overfitting the data and estimate spurious relations between regressors and regressand. Instead, we use a machine learning method known as Boosted Regression Trees. Boosted Regression Trees not only allows large conditioning information sets, but it also allows for non-linearities—all without overfitting or falling prey of the so-called curse of dimensionality. We provide a brief introduction of BRT below. Section 5.1 describes Regression Trees, Section 5.2 describes Boosting. Finally, Section 5.3 describes the implementation of BRT adopted in the paper.[2]

5.1 Regression Trees

Suppose we have P potential predictor ("state") variables and a single dependent variable over T observations, i.e. (x_t, y_{t+1}) for $t = 1, 2, ..., T$, with $x_t = (x_{t1}, x_{t2}, ..., x_{tp})$. Fitting a regression tree requires deciding (i) which predictor variables to use to split the sample space and (ii) which split points to use. The regression trees we use employ recursive binary partitions, so the fit of a regression tree can be written as an additive model:

$$f(x) = \sum_{j=1}^{J} c_j I\{x \in S_j\}, \tag{1}$$

where $S_j,\ j = 1, ..., J$ are the regions we split the space spanned by the predictor variables into, $I\{\}$ is an indicator variable and c_j is the constant used to model the dependent variable in each region. If the L^2 norm criterion function is adopted, the optimal constant is $\hat{c}_j = mean(y_{t+1}|x_t \in S_j)$.

[2] Our description draws on Friedman, Hastie, and Tibshirani (2001), who provide a more in-depth coverage of the approach.

The globally optimal splitting point is difficult to determine, particularly in cases where the number of state variables is large. Hence, a sequential greedy algorithm is employed. Using the full set of data, the algorithm considers a splitting variable p and a split point s so as to construct half-planes

$$S_1(p,s) = \{X|X_p \leq s\} \qquad and \qquad S_2(p,s) = \{X|X_p > s\}$$

that minimize the sum of squared residuals:

$$\min_{p,s}\left[\min_{c_1}\sum_{x_t\in S_1(p,s)}(y_{t+1}-c_1)^2 + \min_{c_2}\sum_{x_t\in S_2(p,s)}(y_{t+1}-c_2)^2\right]. \tag{2}$$

For a given choice of p and s the fitted values, $\hat{c}_1$ and $\hat{c}_2$, are

$$\begin{aligned}
\widehat{c}_1 &= \frac{1}{\sum_{t=1}^{T} I\{x_t \in S_1(p,s)\}}\sum_{t=1}^{T} y_{t+1} I\{x_t \in S_1(p,s)\}, \\
\widehat{c}_2 &= \frac{1}{\sum_{t=1}^{T} I\{x_t \in S_2(p,s)\}}\sum_{t=1}^{T} y_{t+1} I\{x_t \in S_2(p,s)\}.
\end{aligned} \tag{3}$$

The best splitting pair (p,s) in the first iteration can be determined by searching through each of the predictor variables, $p = 1,..,P$. Given the best partition from the first step, the data is then partitioned into two additional states and the splitting process is repeated for each of the subsequent partitions. Predictor variables that are never used to split the sample space do not influence the fit of the model, so the choice of splitting variable effectively performs variable selection.

Regression trees are generally employed in high-dimensional datasets where the relation between predictor and predicted variables is potentially non-linear. This becomes important in our context as it is not clear which variables may be more or less relevant *ex-ante*. Furthermore, it is difficult to know in our context whether there is a linear relation between predictor and predicted variables. On the other hand, the approach is sequential and successive splits are performed on fewer and fewer observations, increasing the risk of fitting idiosyncratic data patterns. Furthermore, there is no guarantee that the sequential splitting algorithm leads to the globally optimal solution. To deal with these problems, we next consider a method known as boosting.

5.2 Boosting

Boosting is based on the idea that combining a series of simple prediction models can lead to more accurate forecasts than those available from any individual model. Boosting algorithms iteratively re-weight data used in the initial fit by adding new trees in a way that increases the weight on observations modeled poorly by the existing collection of trees. From above, recall that a regression tree can be written as:

$$\mathcal{T}\left(x; \{S_j, c_j\}_{j=1}^{J}\right) = \sum_{j=1}^{J} c_j I\{x \in S_j\} \tag{4}$$

A boosted regression tree is simply the sum of regression trees:

$$f_B(x) = \sum_{b=1}^{B} \mathcal{T}_b\left(x; \{S_{b,j}, c_{b,j}\}_{j=1}^{J}\right), \tag{5}$$

where $\mathcal{T}_b\left(x; \{S_{b,j}, c_{b,j}\}_{j=1}^{J}\right)$ is the regression tree used in the b-th boosting iteration and B is the number of boosting iterations. Given the model fitted up to the $(b-1)-th$ boosting iteration, $f_{b-1}(x)$, the subsequent boosting iteration seeks to find parameters $\{S_{j,b}, c_{j,b}\}_{j=1}^{J}$ for the next tree to solve a problem of the form

$$\{\hat{S}_{j,b}, \hat{c}_{j,b}\}_{j=1}^{J} = \min_{\{S_{j,b}, c_{j,b}\}_{j=1}^{J}} \sum_{t=0}^{T-1} \left[y_{t+1} - \left(f_{b-1}(x_t) + \mathcal{T}_b\left(x_t; \{S_{j,b}, c_{j,b}\}_{j=1}^{J}\right)\right)\right]^2. \tag{6}$$

For a given set of state definitions ("splits"), $S_{j,b}$, $j = 1, .., J$, the optimal constants, $c_{j,b}$, in each state are derived iteratively from the solution to the problem

$$\begin{aligned} \hat{c}_{j,b} &= \min_{c_{j,b}} \sum_{x_t \in S_{j,b}} \left[y_{t+1} - (f_{b-1}(x_t) + c_{j,b})\right]^2 \\ &= \min_{c_{j,b}} \sum_{x_t \in S_{j,b}} \left[e_{t+1,b-1} - c_{j,b}\right]^2, \end{aligned} \tag{7}$$

where $e_{t+1,b-1} = y_{t+1} - f_{b-1}(x_t)$ is the empirical error after $b-1$ boosting iterations. The solution to this is the regression tree that most reduces the average of the squared residuals $\sum_{t=1}^{T} e_{t+1,b-1}^2$ and

$\hat{c}_{j,b}$ is the mean of the residuals in the jth state.

Forecasts are simple to generate from this approach. The boosted regression tree is first estimated using data from $t = 1, ..., t^*$. Then the forecast of y_{t^*+1} is based on the model estimates and the value of the predictor variable at time t^*, x_{t^*}. Boosting makes it more attractive to employ small trees (characterized by only two terminal nodes) at each boosting iteration, reducing the risk that the regression trees will overfit. Moreover, by summing over a sequence of trees, boosting performs a type of model averaging that increases the stability and accuracy of the forecasts.[3]

5.3 Implementation

Our estimations follow the stochastic gradient boosting approach of Friedman (2001) and Friedman (2002) with $J = 2$ nodes. The baseline implementation employs 10,000 boosting iterations, but we conduct a number of robustness checks to show that the results are not very sensitive to this choice.

We adopt two refinements to the basic boosted regression tree methodology. The first is *shrinkage*. As with ridge regression and neural networks, shrinkage is a simple regularization technique that diminishes the risk of over-fitting by slowing the rate at which the empirical risk is minimized on the training sample. We use a shrinkage parameter, $0 < \lambda < 1$, which determines how much each boosting iteration contributes to the overall fit:

$$f_b(x) = f_{b-1}(x) + \lambda \sum_{j=1}^{J} c_{j,b} I\{x \in S_{j,b}\}. \tag{8}$$

Following common practice we set $\lambda = 0.001$ as it has been found (Friedman (2001)) that the best empirical strategy is to set λ very small and correspondingly increase the number of boosting iterations.

The second refinement is *subsampling* and is inspired by "bootstrap aggregation" (bagging), see Breiman (1996). Bagging is a technique that computes forecasts over bootstrap samples of the data and averages them in a second step, therefore reducing the variance of the final predictions. In our context, the procedure is adapted as follows: at each boosting iteration we sample without replacement one half of the training sample and fit the next tree on the sub-sample obtained.

[3]See Rapach, Strauss, and Zhou (2010) for similar results in the context of linear regression.

5.4 Relative Influence Measures and Partial Dependence Plots

One criticism of machine learning algorithms is that they are "Black Boxes" that do not provide a lot of intuition to the researcher and the reader. This criticism is hardly applicable to Boosted Regression Trees that instead feature very useful and intuitive visualization tools.

5.4.1 Relative Influence measures.

The first measure commonly used is generally referred to as "relative influence" measures. Consider the reduction in the empirical error every time one of the covariates $x_{\cdot,l}$, is used to split the tree. Summing the reductions in empirical errors (or improvements in fit) across the nodes in the tree gives a measure of the variable's influence (Breiman et al. (1984)):

$$I_l(\mathcal{T}) = \sum_{j=2}^{J} \Delta e(j)^2 I(x(j) = l), \tag{9}$$

where $\Delta e(j)^2 = T^{-1} \sum_{t=1}^{T} \left(e_t(j-1)^2 - e_t(j)^2\right)$, is the reduction in the squared empirical error at the j'th node and $x(j)$ is the regressor chosen at this node, so $I(x(j) = l)$ equals one if regressor l is chosen and zero otherwise. The sum is computed across all observations, $t = 1, ..., T$ and over the $J - 1$ internal nodes of the tree.

The rationale for this measure is that at each node, one of the regressors gets selected to partition the sample space into two sub-states. The particular regressor at node j achieves the greatest reduction in the empirical risk of the model fitted up to node $j - 1$. The importance of each regressor, $x_{l,\cdot}$, is the sum of the reductions in the empirical errors computed over all internal nodes for which it was chosen as the splitting variable. If a regressor never gets chosen to conduct the splits, its influence is zero. Conversely, the more frequently a lag is used for splitting and the bigger its effect on reducing the model's empirical risk, the larger its influence.

This measure of influence can be generalized by averaging over the number of boosting iterations, B, which generally provides a more reliable measure of influence:

$$\bar{I}_l = \frac{1}{B} \sum_{b=1}^{B} I_l(\mathcal{T}_b). \tag{10}$$

This is best interpreted as a measure of relative influence that can be compared across regressors. We therefore report the following measure of relative influence, $\overline{RI}_l$, which sums to one:

$$\overline{RI}_l = \bar{I}_l / \sum_{l=1}^{L} \bar{I}_l. \tag{11}$$

5.4.2 Partial Dependence Plots.

The second visualization tool featured by BRT are *partial dependence plots*, that are defined as follows. Suppose we select a particular covariate, X_p, from the set of P predictor variables $X = (X_1, X_2, ..., X_P)$ and denote the remaining variables X_{-p}, i.e. $X_{-p} = X \backslash \{X_p\}$. We use the following measure of the average marginal effect of X_p on the dependent variable

$$f_p(X_p) = E_{X_{-p}} f(X_p, X_{-p}). \tag{12}$$

This is called the average partial dependence measure. It fixes the value of X_p and averages out the effect of all other variables. By, repeating this process for different values of X_p, we trace out the marginal effect this covariate has on the predicted variable.

An estimate of $f_p(X_p)$ can be computed by averaging over the sample observations

$$\bar{f}_p(X_p) = \frac{1}{T} \sum_{t=1}^{T} f(X_p, x_{t,-p}), \tag{13}$$

where $x_{t,-p} = \{x_{1,-p}, ..., x_{T,-p}\}$ are the values of X_{-p} occurring in the data.

6 Which Clients Experience the Biggest Portfolio Changes?

This section uses BRTs to decompose the effect of PAS on investors portfolio allocations. As a measure of portfolio change, we adopt the share of equity held by the investor. Intuitively, from Figure 1 we know that clients with a very high equity share before signing up for PAS are likely to experience a decrease in their equity share. Those with a low equity share are instead likely to experience an increase in their equity share.

In order to decompose and visualize how PAS changes investors' portfolio allocations after sign-up, we estimate a BRT model with 10,000 boosting iterations. The dependent variable is the change in the share of equities before and after signing up for PAS, that is:

$$\Delta\_Equity\_Share_i = Equity\_Share_{i,t+6} - Equity\_Share_{i,t}, \tag{14}$$

where i denotes each investor and t denotes the month in which each investor signs up for PAS. As conditioning variables we use a total of 15 regressors, divided into three groups. The first group contains demographic characteristics: Age, is the age of the client as of December 2017; $Male$, is the fraction of male clients; $Married$, is the fraction of married clients; $Tenure$, is the tenure of the client as of December 2017. The second contains regressors related to portfolio characteristics: $NumAssets$, is the number of assets held by the client across accounts; $PctEquityShare$, is the percentage of wealth in equities—held directly or through mutual funds; $PctCashShare$, is the percentage of wealth money market mutual funds—held directly or through mutual funds; $PctMutualFunds$, is the percentage of wealth directly invested in mutual funds; $PctStocks$, is the percentage of wealth directly invested in individual stocks; $PctETF$, is the percentage of wealth directly invested in ETFs; $PctIndex$, is the percentage of mutual fund wealth invested in index funds; $PctEmerging$, is the percentage of mutual fund wealth invested in emerging market funds—identified using the Lipper mutual fund classification. The third groups relates to variables related to transactions and fees paid: $MgtFees$, are the value-weighted management fees charged by the mutual funds held by the account-holders; $Transaction$, is the number of transactions directly initiated by the investors over the month before signing up for PAS; $Volume$, is the volume (in US dollars) traded by the investors over the month before signing up for PAS.

Out of the 15 predictor variables, only three variables have a relative influence higher than 1%. The variable $PctEquityShare$ has the highest relative influence measure, totaling 81.9%. This means that the splits based on $PctEquityShare$ contribute to 81.9% of the reduction in the empirical error of the model. The second variable is Age, that has a relative influence of 15.6%. Finally, the third variable is $PctCashShare$ that has a relative influence of 2.1%. The remaining covariates explain very little of the variation in clients' equity share post-PAS. This indicates that the change in risky share is mainly determined by the investors' positions when signing up for the PAS service and investors' age.

We report the univariate partial dependence plots for the three most important predictors in Panel A of Figure 7. The first subfigure reports results for the share of wealth in Equities pre-PAS. There is clearly a negative relation between the change in the equity share and the fraction of wealth in equities pre-PAS. Those investors with no wealth in equities experience an increase in the share of equities of by PAS of 30%. On the other hand, PAS decreases by 30% the share of equities for those investors with 100% of their wealth in equities. Interestingly, we find that that the partial dependence crosses the "0" on the y-axis, when the Equity Share equals approximately 0.6, indicating that PAS does not change the equity positions for those investors that already have roughly a 60-40 split between equities and bonds.

The second subfigure plots the partial dependence with respect to Age. Systematically, PAS increases the equity exposure of the clients that have less than 55 years of age and decreases the equity exposure of those clients that have more than 55 years of age. The reduction in equity exposure is significant. It equals approximately -12% for the clients over 60 years old and almost +15% for those under 40 years of age.

Finally, the third covariate is Cash share. The relation is once again positive. Those investors with no money in money market mutual funds experience a reduction in their equity share. At the other extreme, those investors with 100% of their wealth in money market mutual funds, experience an increase in equity share.

Panel B of Figure 7 present the bivariate dependence plots. The first plots the partial dependence of the change in equity share with respect to both the pre-PAS equity and cash share. The bivariate plot shows the joint negative relation between both regressors and the changes in the Equity share. It also shows that the relation between age and the change in equity share is monotonic, but not linear. The second plot instead displays the partial dependence plots of the change in equity share with respect to the equity share and the cash share. The plot that that, jointly, the change in equity share is negatively related to pre-PAS equity share and positively related to the cash share.

One interesting finding of the partial dependence plots is that the both the bond and cash share are linearly related to the changes in equity share operated by PAS. The only regressor that instead displays a monotonic, but not linear, relation is Age. As a result, it could well be that a linear regression could work as well as BRT in this case. This turns out to be the case, as we show in Section 8.

The change in portfolio share is a rather easy quantity to model as it is likely to be a deterministic function of investors' demographic and portfolio characteristics at the time of PAS sign-up, as well as their risk-preferences, liquidity needs, and employment characteristics, which we cannot observe. A more challenging question is whether we can use investors' observable characteristics to predict which clients are likely to benefit the most from Robo-advising. We undertake this analysis next.

7 Which Clients Benefit the Most from Robo-advising?

In this section, we explore whether we can explain the cross-section of changes in risk-adjusted performance pre- and post-PAS using investors' characteristics at the time of sign-up. We use the same setup and covariates as in Section 7, the only difference being that we replace the share of bonds with the share of equities as regressor, to ease the interpretation of the results.

The dependent variable is the change in the abnormal Sharpe ratio before and after signing up for PAS, that is:

$$\Delta\_Abn\_Sharpe_i = Abn\_Sharpe_{i,t+6} - Abn\_Sharpe_{i,t}, \tag{15}$$

where i denotes each investor and t denotes the month in which each investor signs up for PAS.

The task in this section is much more challenging that then one in the previous section, for at least two reasons. First, abnormal Sharpe ratios are noisier, as they are computed on the basis of only 6 months of returns and realized volatilities. Second, the change in abnormal Sharpe ratios is likely to not only be driven by the equity-bond allocation decision, but also by the characteristics of the individual securities held.

The relative influence results are reported in Figure 8, while the partial dependence plots for the top 9 covariates by relative influence are reported in Figure 9. Among the top 9 covariates, some of the relations are immediately intuitive. For example, the cash share (relative influence of 11%) and the traded volume (relative influence of 3.4%) at sign-up are positively related to the improvement in performance post-PAS. This is simply saying that those investors that were trading a lot and/or where holding a very large portion of their wealth in cash, benefit more from PAS.

Other economically intuitive partial dependence relations are those associated with clients' tenure (relative influence of 13%), the percentage of wealth in mutual funds (relative influence of 5.4%), the percentage of mutual fund holdings in index funds (relative influence of 2.2%). In all cases, BRTs uncover a negative partial relation. This suggests that those clients that were not holding a lot of indexed funds and were not holding a lot of their funds in mutual funds, are the ones benefitting more from signing up for PAS. The partial dependence with respect to clients' tenure suggest instead that less experienced individuals are the ones that benefit the most from PAS.

BRTs uncover also markedly non-linear and non-monotonic relations between the change in risk-adjusted performance and clients' age (relative influence of 16.5%), the mutual funds' management fees (13.3%) and the number of assets held (relative influence of 3.6%). For the first two regressors, the relation is U-shaped. The results suggest that the clients benefitting the most are the ones in their forties and mid-fifties, and the very senior citizens, while there is a negative relation between age and change in post-PAS performance for clients in the second half of their fifties and their sixties. The relation is probably due to the fact that PAS tends to increase the equity exposure of the clients in their forties and mid-fifties, decrease it for clients in the second half of their fifties and their sixties, and leave them unchanged for the clients in their seventies (see the third Subfigure in Panel A of Figure 7).

The relation between fees and change in performance is also U-Shaped, indicating that those customers investing in very expensive active funds as well as the few very cheap funds are the ones that benefit the most. The result is probably driven by the fact that over our sample period, actively managed funds have performed relatively well, compared to passive funds.

Finally, the relation between number of assets and performance change has an inverse U-shaped relation. This is due to the fact that individuals with few assets are likely to be holding mutual funds. Associated with a higher number of assets are instead those clients that invest in individual equities. These customers do benefit from PAS is it increases their diversification. Those individuals that instead had 25 or 30 assets where likely to be already rather diversified, even if they were holding individual stocks. They therefore do not benefit much from adopting the robo-adviser.

Finally, the last covariate and most relevant covariate is the share of equities held (relative influence of 27%). The positive monotonic relation suggest that PAS increases investors' performance more for those with higher equity shares, indicating that PAS invests in a portfolio of mutual funds with higher risk-adjusted profiles, compared to the average investor.

8 In- and Out-of-sample Performance of BRTs

One of the main criticisms against non-parametric models is that they tend to overfit the training dataset. One could be worried that the non-linearities and non-monotonicities uncovered in Figure 9 and described in Section 7 are the result of BRTs fitting noise rather than the structural relation between the covariates and the dependent variable. We show here that this is not the case. Crucially, we show that the most important free parameter, i.e. the number of boosting iterations, does not significantly affect the out-of-sample performance of the method.

To asses whether BRTs are overfitting the training dataset, we perform the following cross-validation analysis. We take the original dataset on which we estimate our BRT results and exclude half of the observations. We then estimate the BRT model on one half of the data and test its performance on the other half of the data. We repeat the analysis 100 times. On every iteration, we store the in- and out-of-sample performance of BRTs for boosting iterations that range from 100 to 20,000. For every iteration, we also store the in- and out-of-sample performance of a linear model that uses the same regressors as BRT. Finally, we report two figures. Figure 10 reports the in- and out-of-sample performance of BRTs —averaged across all cross-validation rounds— for different boosting iterations. Figure 11 plots the density of the out-of-sample performance of the BRT model and the linear model across all cross-validation rounds.

To show how the performance changes depending on the setting, we report the results for the change in the equity share for the investment portfolio as well as the change in the performance before and after signing up for PAS.

8.1 Change in Portfolio Allocation

As mentioned in Section 6, the out-of-sample prediction of the changes in portfolio allocation is likely to be not very challenging, because the change in the risky share is likely to be some deterministic function of investor demographic and portfolio characteristics (which we observe) as well as well as investor preferences and tolerance of risk (which we do not observe).

As highlighted in Section 6, the partial dependence plots show that the relations between the regressors and the independent variable are mostly linear, with the exception of *Age* that appears to be monotonic, but not linear. As a result, we should expect the linear model to perform rather well compared to BRTs. This is indeed what we find.

Panel A of Figure 10 show the average in- and out-of-sample performance of BRTs and the linear model for portfolio changes. As the number of boosting iterations increases, the fit of BRTs improves and rises to almost 60%, as shown by the black line. For comparison, note that the linear model has an in-sample R^2 of only 57.2%—green line. The out-of-sample performance BRT improves as the number of boosting iterations raises from 100 all the way to approximately 12,000, as shown by the red line. The out-of-sample fit then asymptotes and stabilizes at around 57.25%, a value greater than the in-sample R^2 of the linear model. The out-of-sample fit of the linear model is instead worse, equalling 56.7%—blue line.

The results in Figure 10 reports averages across simulation rounds. To show how the out-of-sample performance of BRTs and the linear model compare, we report in Panel A of Figure 11 the density of the out-of-sample R^2 across simulation rounds. Consistent with the findings in Panel A of Figure 10, BRTs consistently outperform the linear model out-of-sample.

8.2 Change in Investment Performance

Explaining the changes in investment performance is likely to be rather challenging, because the realized Sharpe ratios are estimated using only 6 months of daily data and are therefore very noisy. Also, over such short period of time, it is possible that certain stocks or portfolios will deliver very low or large returns for idiosyncratic reasons.

As shown in Section 7, the partial dependence plots show that the relation between the regressors and the independent variable is certainly non-linear and in some cases also strongly non-monotonic. As a result, we should expect the linear model to perform rather poorly compared to BRTs. This is indeed what we find.

Panel B of Figure 10 show the in- and out-of-sample performance of BRTs and the linear model. As the number of boosting iterations increases, the fit of BRTs improves and rises to almost 5.91%, as shown by the black line. For comparison, note that the linear model has an in-sample R^2 of only 1.70%—green line. The out-of-sample performance BRT improves as the number of boosting iterations raises from 100 all the way to approximately 10,000, as shown by the red line. The out-of-sample fit then asymptotes and stabilizes at around 1.92%, a value greater than the in-sample R^2 of the linear model. The out-of-sample fit of the linear model is instead rather poor, equalling 0.06%.

In Panel B of Figure 11 we present the density of the out-of-sample R^2 across simulation rounds. Consistent with the findings in Panel B of Figure 10, BRTs consistently outperform the linear model out-of-sample.

9 Conclusions

We study the largest US hybrid robo-adviser by assets under management, Vanguard Personal Advisor Services (PAS).

Across all clients, PAS reduces investors holdings in money market mutual funds and increases bond holdings. It reduces the holdings of individual stocks and US active mutual funds, and moves investors towards low-cost indexed mutual funds. Finally, it increases investors' international diversification and investors' overall risk-adjusted performance.

From sign-up, it takes approximately six months for PAS to adjust investors' portfolios to the new allocations. We use a machine learning algorithm, known as Boosted Regression Trees (BRT), to explain the cross-sectional variation in the effects of PAS on investors' portfolio allocation and performance.

The investors that benefit the most from robo-advising are the clients with little investment experience, as well as the ones that have high cash-holdings and high trading volume pre-adoption. Clients with little mutual fund holdings and clients invested in high-fee active mutual funds also display significant performance gains.

References

Barber, B. M., and T. Odean. 2001. Boys will be boys: Gender, overconfidence, and common stock investment. *Quarterly Journal of Economics* 116:261–92.

Breiman, L. 1996. Bagging predictors. *Machine learning* 24:123–40.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. 1984. *Classification and regression trees.* Statistics/Probability Series. Belmont, California, U.S.A.: Wadsworth Publishing Company.

Friedman, J., T. Hastie, and R. Tibshirani. 2001. *The elements of statistical learning*, vol. 1. Springer series in statistics New York, NY, USA:.

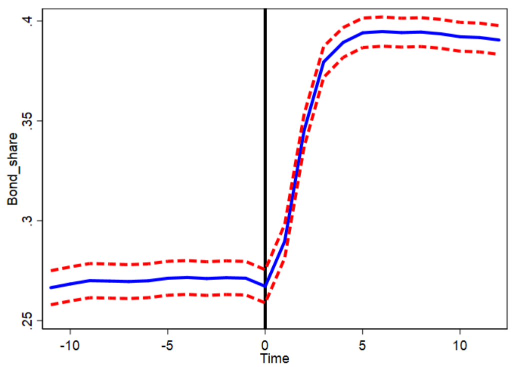
Friedman, J. H. 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics* 1189–232.

———. 2002. Stochastic gradient boosting. *Computational Statistics & Data Analysis* 38:367–78.

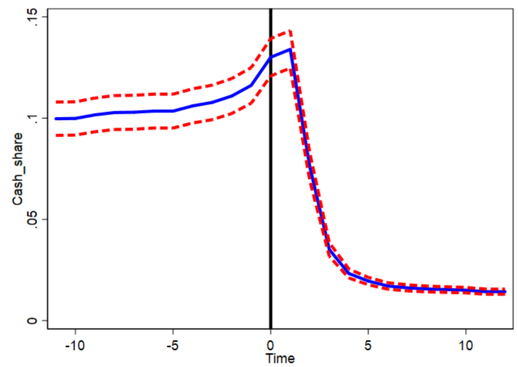
Gargano, A., and A. G. Rossi. 2017. Does it pay to pay attention? *Working Paper* .

Linnainmaa, J. T., B. Melzer, and A. Previtero. Forthcoming. The misguided beliefs of financial advisors. *Journal of Finance* .

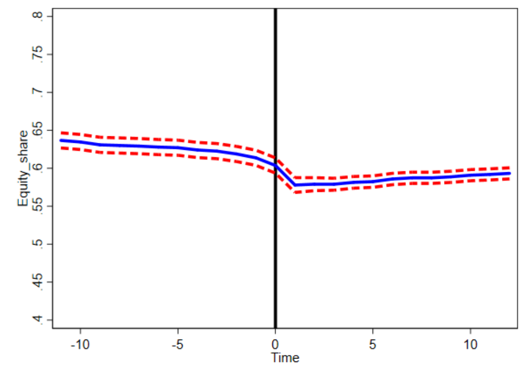
Rapach, D. E., J. K. Strauss, and G. Zhou. 2010. Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *The Review of Financial Studies* 23:821–62.



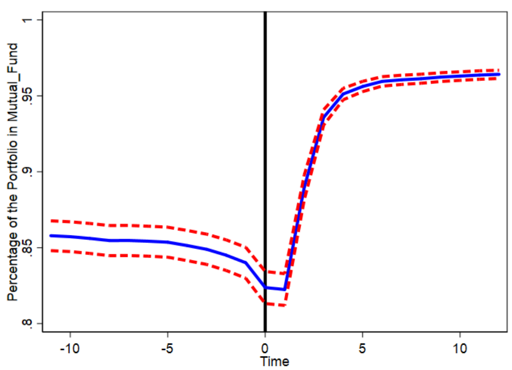
(a) Percentage of Bonds



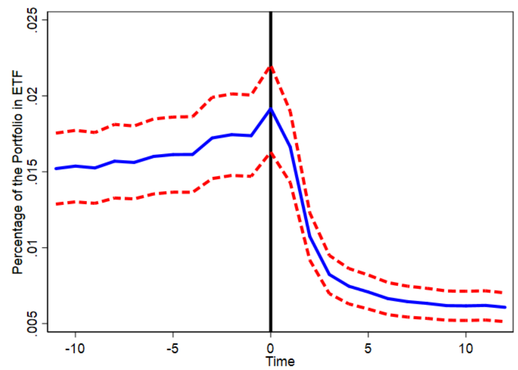
(b) Percentage of Cash



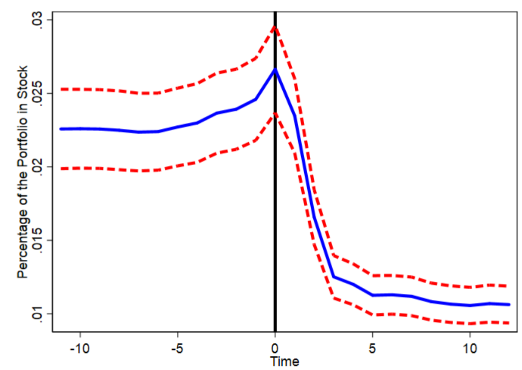
(c) Percentage of Equities



(d) Percentage in Mutual Funds

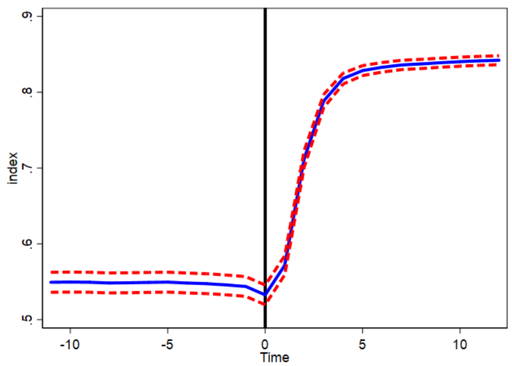


(e) Percentage in ETFs

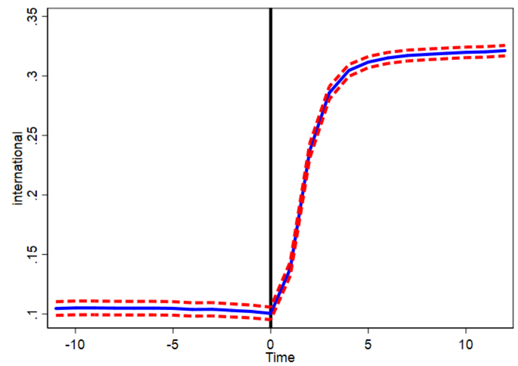


(f) Percentage in Individual Stocks

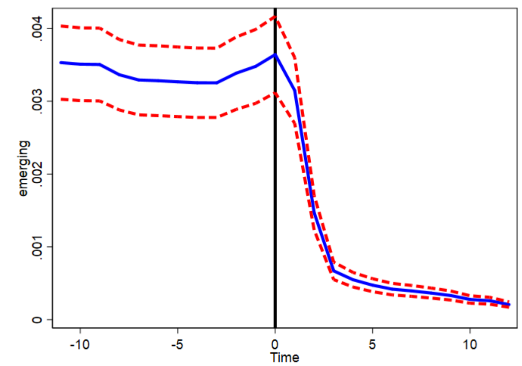
Figure 1. Portfolio Allocation before and after PAS: All Clients. This figure reports results for portfolio characteristics of Vanguard clients before and after signing up for PAS. The results are computed at the client level and include all account types, that is, taxable and non-taxable (IRA) accounts. Subfigures (a), (b), and (c) report results for the percentage of wealth held—directly or through mutual funds—in corporate bonds, cash and equities, respectively. Subfigures (d), (e), and (f) report results for the percentage of wealth invested in mutual funds, ETFs, and individual stocks, respectively. In each subfigure, time "0" represent the month before investors sign up for PAS. Results are computed using only investors that are in the sample for at least twelve months before and after signing up for PAS. The blue line denotes average values, while the red dashed lines are 95% confidence intervals.



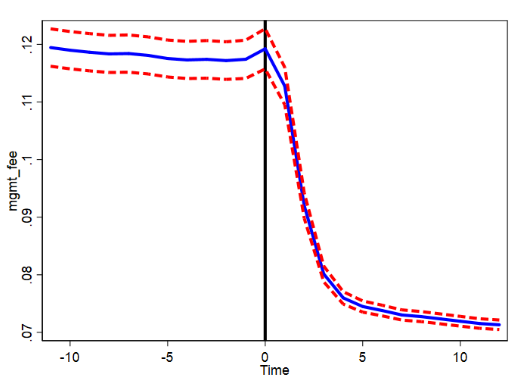
(a) Percentage of Indexed Funds



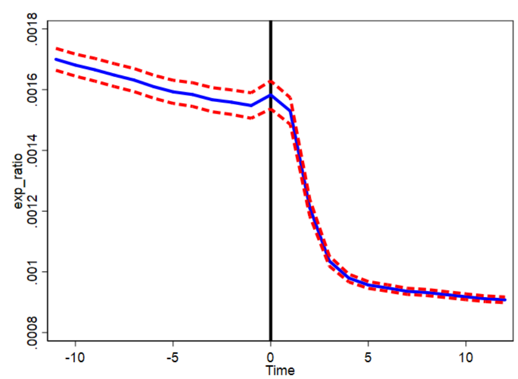
(b) Percentage of International Funds



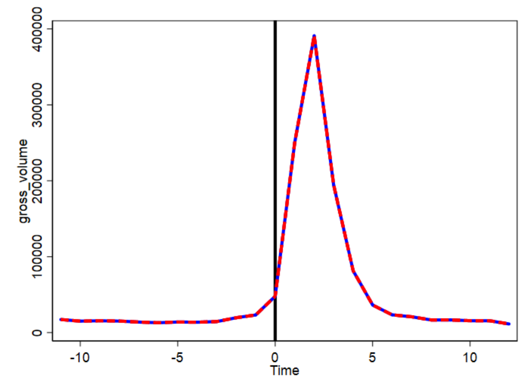
(c) Percentage of Emerging Markets Funds



(d) Management Fees

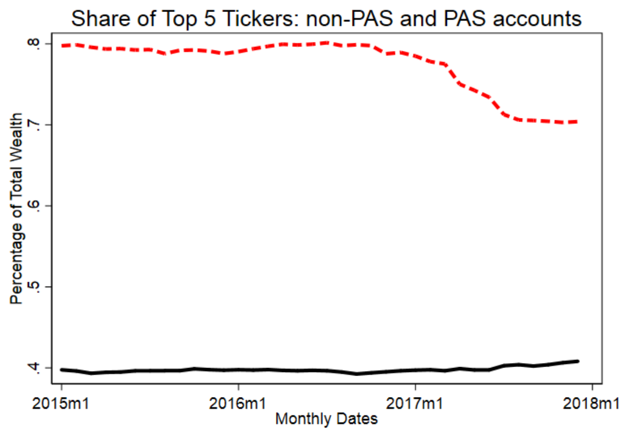


(e) Expense Ratios

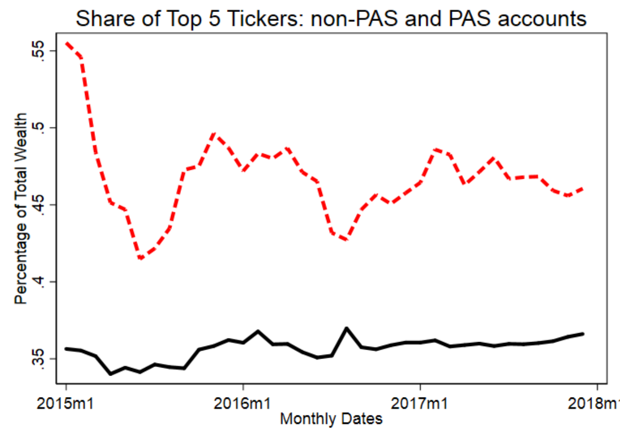


(f) Trading Volume

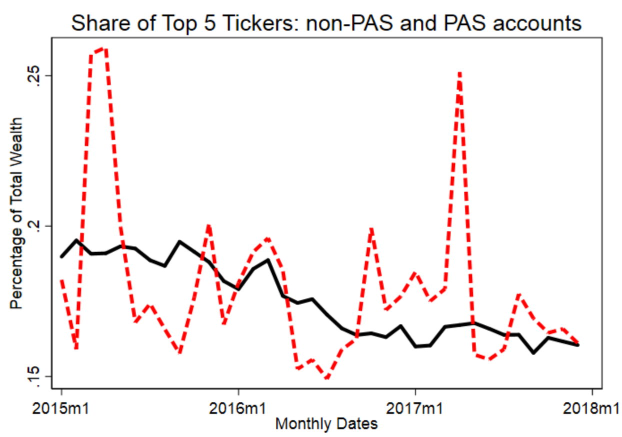
Figure 2. Indexation, international diversification and fees before and after PAS: All Clients. This figure reports results for degree of indexation, international diversification, trading volume and fees of Vanguard clients before and after signing up for PAS. The results are computed at the client level and include all account types, that is, taxable and non-taxable (IRA) accounts. Subfigure (a) reports results for the percentage of mutual fund wealth invested in index funds; Subfigure (b) the percentage of mutual fund wealth invested in global or international funds—identified using the Lipper mutual fund classification; and Subfigure (c) the percentage of mutual fund wealth invested in emerging market funds—identified using the Lipper mutual fund classification. Subfigure (d) presents results for the value-weighted management fees charged by the mutual funds held by the account-holders; and Subfigure (e), the value-weighted expense ratio charged by the mutual funds held by the account-holders. Finally, Subfigure (f) shows results for the monthly trading volume, in US dollars. In each subfigure, time "0" represent the month before investors sign up for PAS. Results are computed using only investors that are in the sample for at least twelve months before and after signing up for PAS. The blue line denotes average values, while the red dashed lines are 95% confidence intervals.



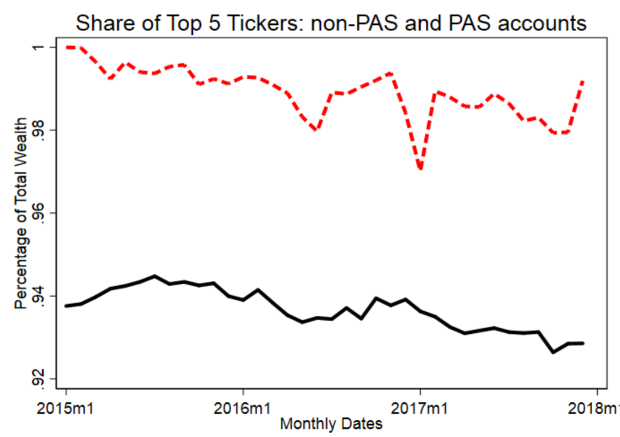
(a) Mutual Funds



(b) ETF

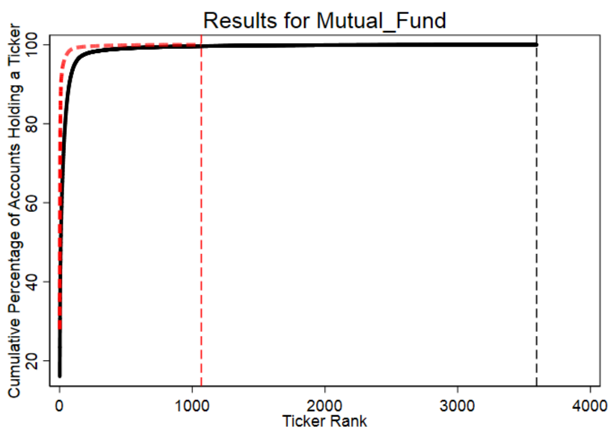


(c) Stocks

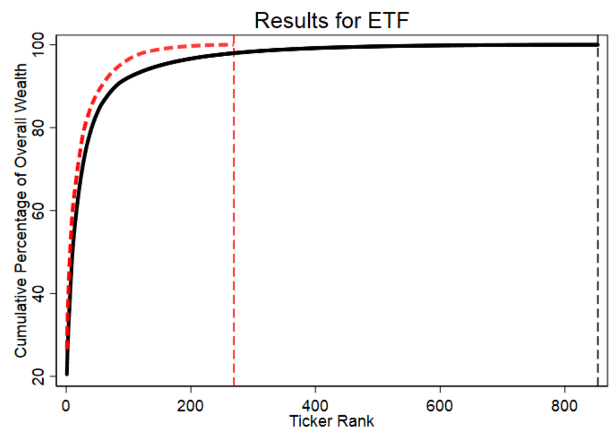


(d) Cash

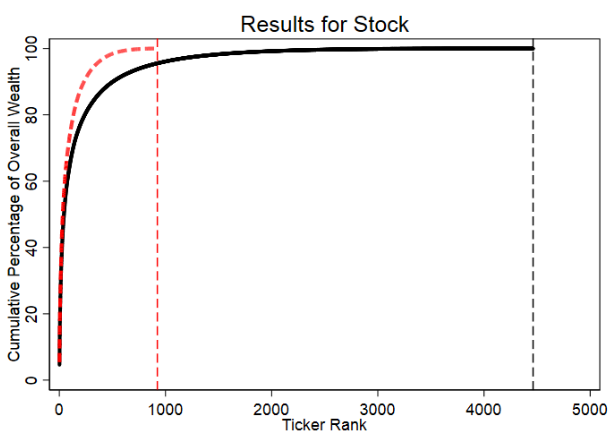
Figure 3. Percentage of Wealth in Top 5 Tickers by Asset Class Over Time. This figure reports the percentage of wealth in top 5 tickers over time, computed for non-PAS and PAS investors. The results are computed at the account-holder level and include all account types, that is, taxable and non-taxable (IRA) accounts. Subfigure (a) reports the results for mutual funds. Subfigure (b), (c), and (d) reports results for ETFs, individual stocks and money market mutual funds, respectively. In each subfigure, the black solid line presents the results for non-PAS account holders, while the red dashed-line presents the results for PAS account holders.



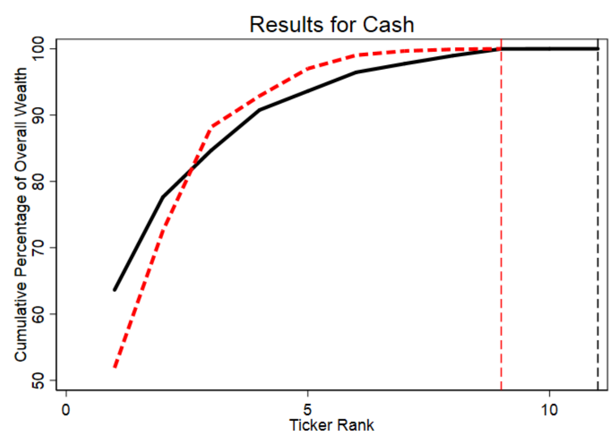
(a) Mutual Funds



(b) ETF



(c) Stocks



(d) Cash

Figure 4. Cumulative Percentage of Wealth Across Tickers in the Cross-section. This figure reports the cumulative percentage of wealth across all the tickers held in each asset class in January 2017, across non-PAS and PAS investors. The results are computed at the account-holder level and include all account types, that is, taxable and non-taxable (IRA) accounts. Subfigure (a) reports the results for mutual funds. Subfigure (b), (c), and (d) reports results for ETFs, individual stocks and money market mutual funds, respectively. In each subfigure, the black solid line presents the results for non-PAS account holders, while the red dashed-line presents the results for PAS account holders.

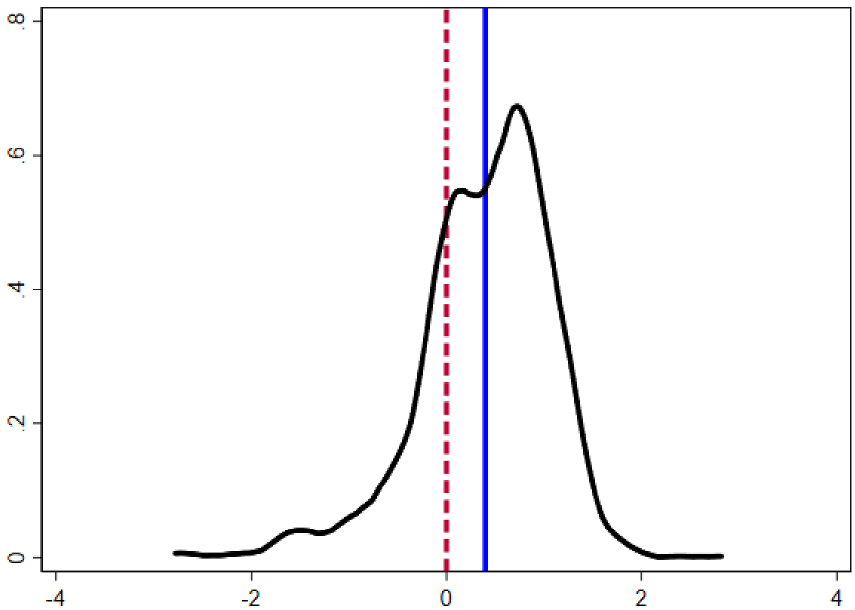
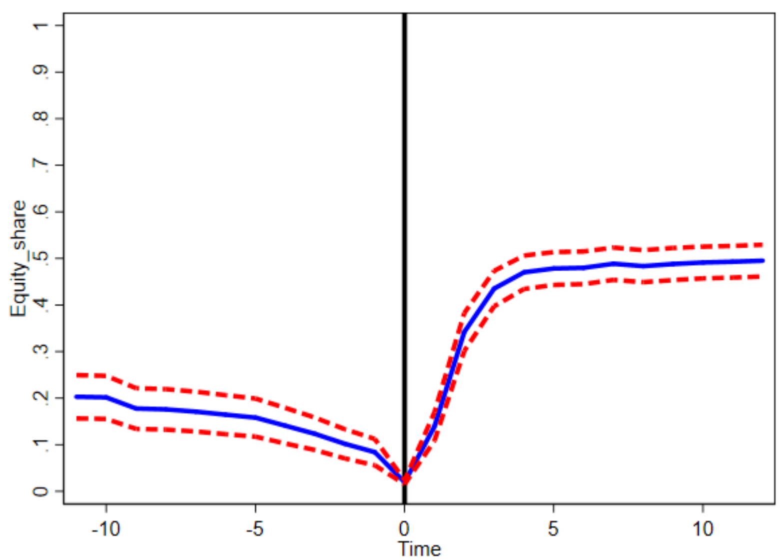
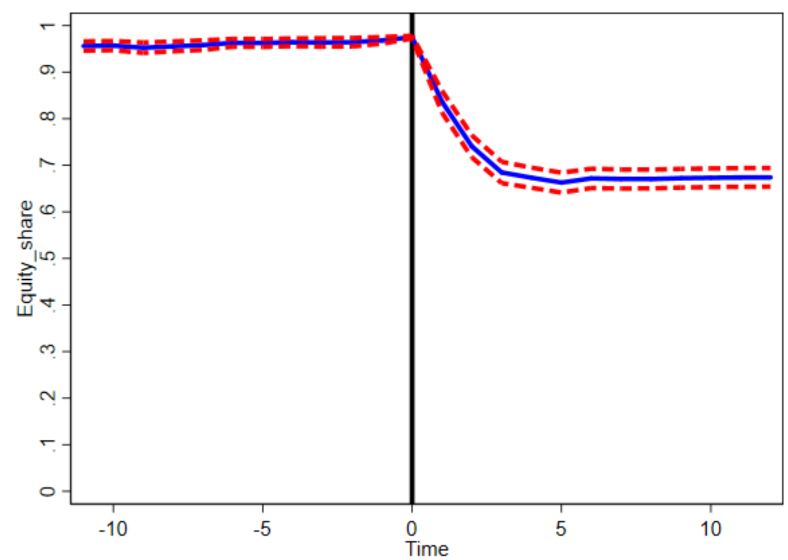


Figure 5. Pre- and post-PAS Performance difference distribution. Client Level. This figure reports the change in investment performance across Vanguard clients after signing up for PAS, compared to before signing up for PAS. The results are computed at the client level and are computed across all accounts, As a measure of performance, we use the abnormal Sharpe Ratio, that is, the difference between the realized Sharpe ratio of each account-holder and the realized Sharpe ratio of the market portfolio, where the latter is computed as the value-weighted returns on the NYSE/AMEX/NASDAQ CRSP portfolio. For each account-holder, portfolio returns and volatilities are computed using beginning-of-month investment holdings. Furthermore, portfolio volatilities are computed as realized volatilities using squared daily returns. The figure uses all the accounts that have been in the sample for at least 6 months before and after signing up for PAS and computes abnormal Sharpe ratios using 9-month windows. The black solid line denotes the density of the difference in post- and pre-PAS performance, computed at the client level. The red dashed line marks the value "0" of the x-axis, while the blue solid line denotes the mean of the distribution.



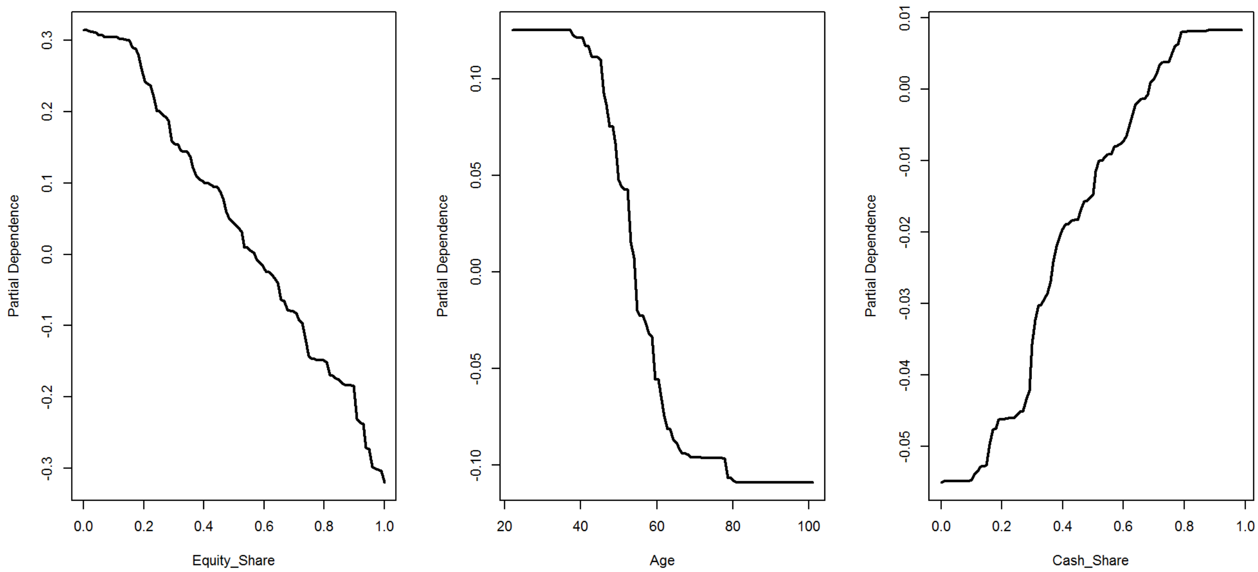
(a) Low Equity Share and PAS Sign-up



(b) High Equity Share and PAS Sign-up

Figure 6. Portfolio Changes for Clients with low and high Equity shares pre-PAS. This figure reports results for the changes in equity shares for Vanguard clients, before and after signing up for PAS. The results are computed at the client level and include all account types, that is, taxable and non-taxable (IRA) accounts. Subfigure (a) reports results for clients with less than 10% of their portfolio in equities pre-PAS Subfigure (b) reports results for clients with more than 90% of their portfolio in equities post-PAS. In each subfigure, time "0" represent the month before investors sign up for PAS. Results are computed using only investors that are in the sample for at least twelve months before and after signing up for PAS. The blue line denotes average values, while the red dashed lines are 95% confidence intervals.

Panel A. Univariate Partial Dependence Plots



Panel B. Bivariate Partial Dependence Plots

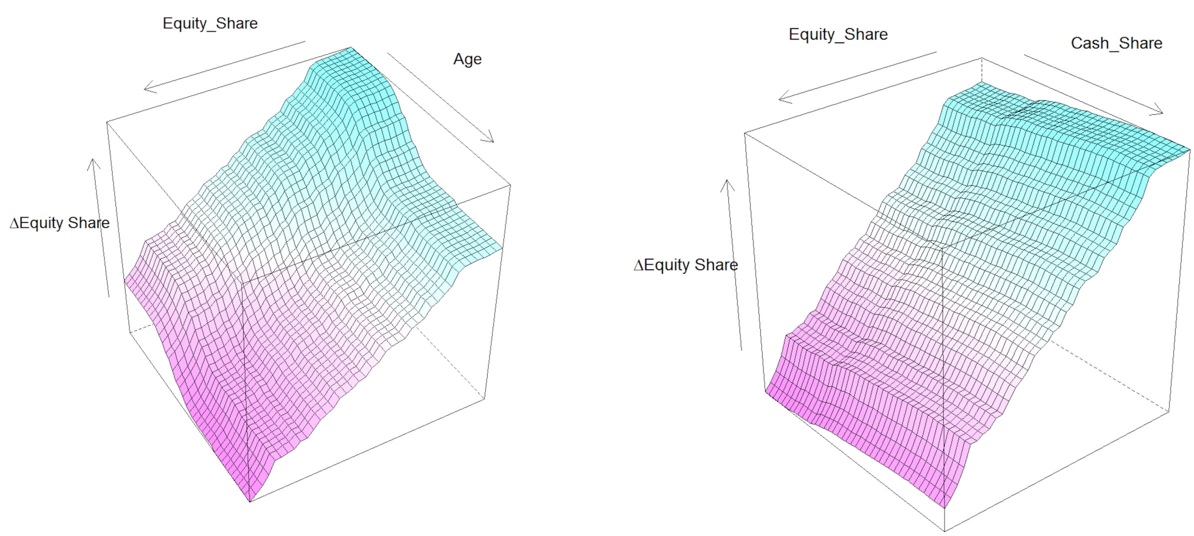


Figure 7. Partial Dependence Plots for the Change in Equity Share post-PAS. This figure presents the partial dependence plots for the change in equity share as a function of a total of 15 regressors described in Section 6. In Panel A of the figure, we report partial dependence plots for the three predictor variables with the highest relative influence: *PctEquityShare*, the share of wealth in Bonds (relative influence of 81.9%); *Age*, the age of the client (relative influence of 15.6%); and *PctCashShare*, the share of wealth in cash (relative influence of 2.1%). In Panel B, we report bivariate partial dependence plots for *PctEquityShare* and *Age*, and for *PctEquityShare* and *PctCashShare*. The horizontal axis covers the sample support of each predictor variable, while the vertical axis tracks the change in the equity share as a function of each individual predictor variable.

Relative Influence Measures

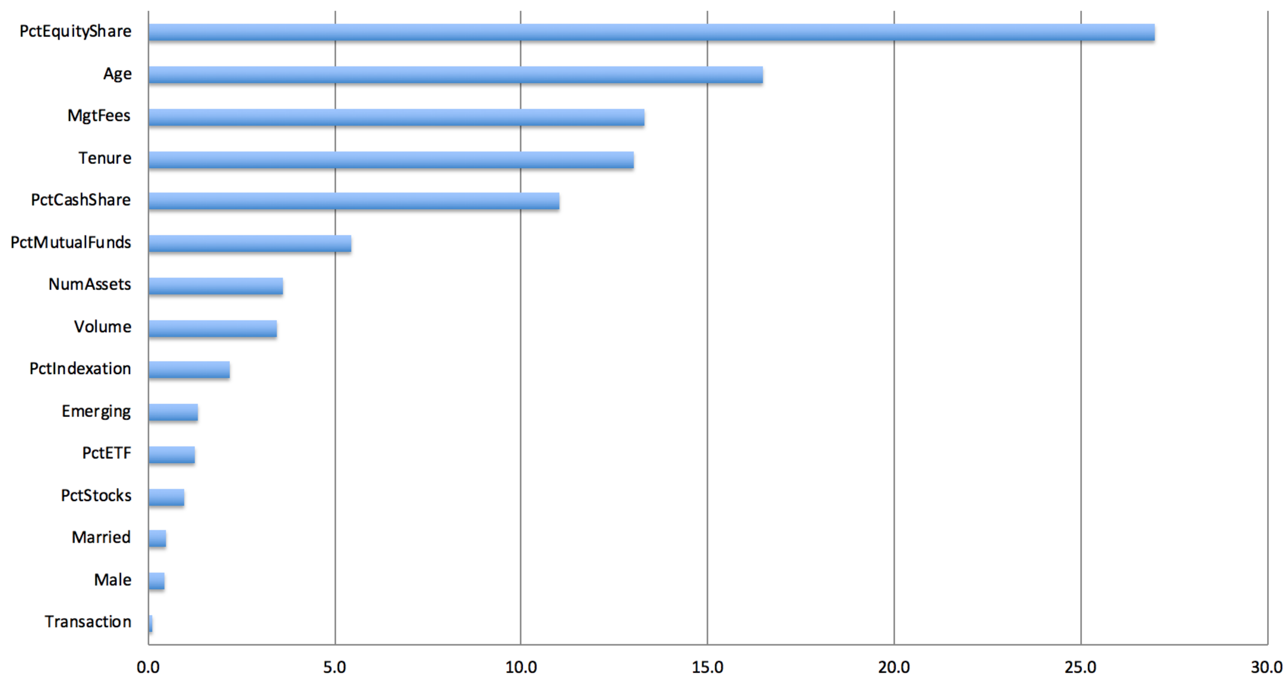


Figure 8. Relative Influence Plots for the Change in Performance post-PAS. This figure presents the relative influence plots for the change performance post-PAS adoption as a function of a total of 15 regressors described in Section 6. The relative influence value associated with each regressor corresponds the relative importance of the covariate in explaining the changes in performance post-PAS. By construction, the sum of the relative influences across all the covariates sums to 100.

Partial Dependence Plots for the 9 Most Relevant Predictors

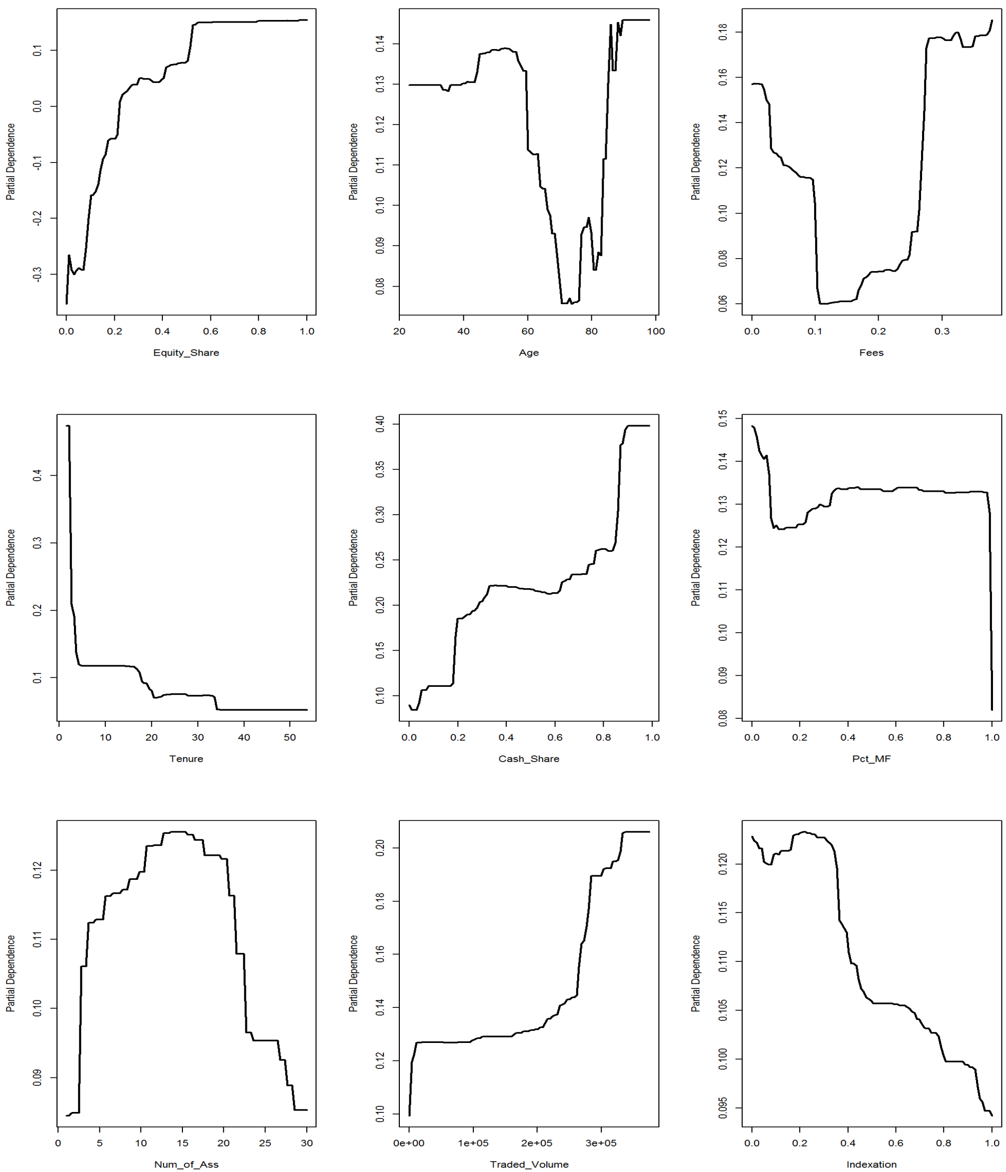
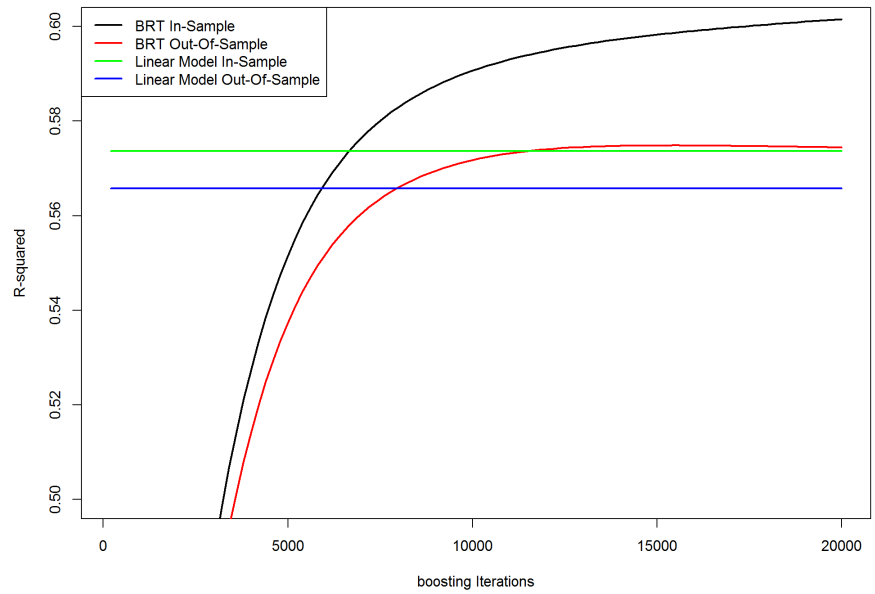


Figure 9. Partial Dependence Plots for Investment Performance Changes post-PAS. This figure presents the partial dependence plots for the change in abnormal Sharpe Ratio as a function of a total of 15 regressors described in Section 6. We report partial dependence plots for the 9 predictor variables with the highest relative influence: $PctEquityShare$, the share of wealth in Equities (relative influence of 27%); Age, the age of the client (relative influence of 16.5%); $MgtFees$, the value-weighted management fees charged by the mutual funds held by the client (relative influence of 13.3%); $Tenure$, the tenure of the client as of December 2017 (relative influence of 13.0%); $PctCashShare$, is the percentage of wealth money market mutual funds—directly or through mutual funds (relative influence of 11.0%); $PctMutualFunds$, is the percentage of wealth directly invested in mutual funds (relative influence of 5.4%); $NumAssets$, is the number of assets held by the client across accounts (relative influence of 3.6%); $Volume$, is the volume (in US dollars) traded by the investors over the month before signing up for PAS (relative influence of 3.4%); $PctIndex$, is the percentage of mutual fund wealth invested in index funds (relative influence of 2.2%). The horizontal axis covers the sample support of each predictor variable, while the vertical axis tracks the change in the equity share as a function of each individual predictor variable.

Panel A: Results for Portfolio Changes



Panel B: Results for Performance Changes

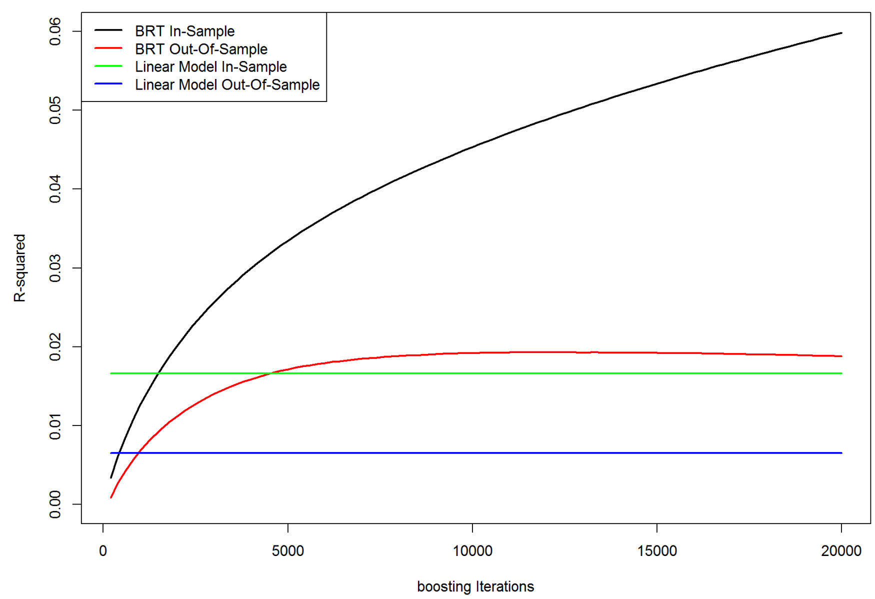
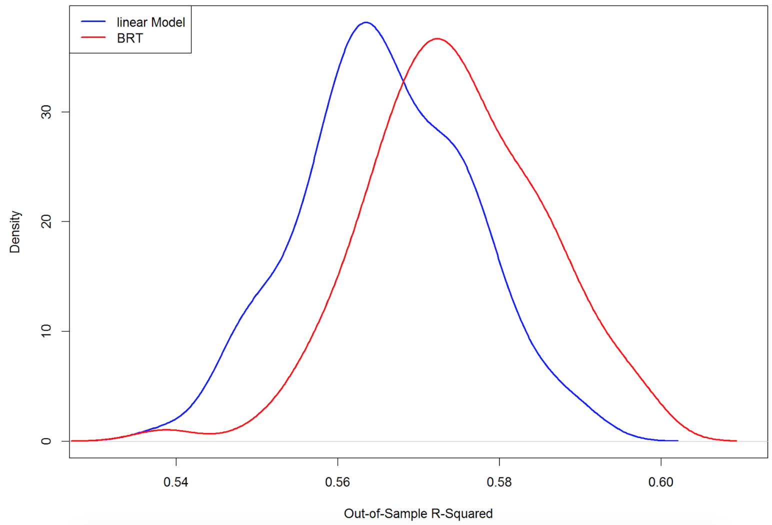


Figure 10. In- and Out-of-Sample Average BRT Performance Across Boosting Iterations. This figure plots the in- and out-of-sample performance, across boosting iterations, for a Boosted Regression Trees model and a linear regression model that uses the same covariates. Panel A reports the results for Portfolio Changes while Panel B the results for Performance Changes. In each panel, the BRT in-sample performance is denoted by a black line; the BRT out-of-sample performance is denoted by a red line; the linear model in-sample performance is denoted by a green line; and the linear model out-of-sample performance is denoted by a blue line.

Panel A: Results for Portfolio Changes



Panel B: Results for Performance Changes

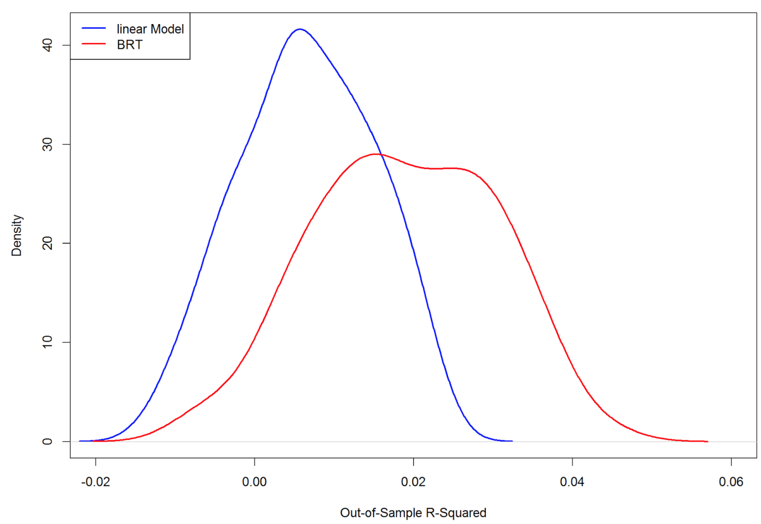


Figure 11. Out-of-Sample Performance of BRTs across Monte Carlo Samples. This figure plots densities of out-of-sample performance for a Boosted Regression Trees model with 20,000 boosting iterations and a linear regression model that uses the same covariates. Panel A reports the results for Portfolio Changes while Panel B the results for Performance Changes. In each panel, the BRT performance is denoted by a red line while the linear model performance is denoted by a blue line.

Table 1. Demographic and Portfolio Characteristics of PAS investors at Sign-up

CLIENT LEVEL

Panel A. Demographic Characteristics

	N	mean	sd	p1	p10	p25	p50	p75	p90	p99
Age	11,324	63.23	12.78	30.00	45.00	56.00	65.00	71.00	78.00	90.00
Male	11,596	0.53	0.50	0.00	0.00	0.00	1.00	1.00	1.00	1.00
Married	11,596	0.35	0.48	0.00	0.00	0.00	0.00	1.00	1.00	1.00
Tenure	11,595	14.10	9.32	0.42	2.00	5.00	14.08	20.50	26.08	35.67

Panel B. Portfolio Allocation

	N	mean	sd	p1	p10	p25	p50	p75	p90	p99
Wealth	11,596	580,815	802,728	5,880	41,760	105,116	279,581	698,568	1,443,591	4,081,860
NumAssets	11,596	7.66	7.89	1.00	1.00	2.00	5.00	10.00	17.00	39.00
PctEquityShare	11,501	0.54	0.31	0.00	0.00	0.33	0.58	0.78	0.96	1.00
PctBondShare	11,501	0.24	0.23	0.00	0.00	0.00	0.20	0.40	0.58	0.88
PctCashShare	11,501	0.22	0.35	0.00	0.00	0.00	0.02	0.27	1.00	1.00
PctMutualFunds	11,573	0.72	0.37	0.00	0.00	0.49	0.95	1.00	1.00	1.00
PctCash	11,573	0.20	0.34	0.00	0.00	0.00	0.01	0.22	0.99	1.00
PctStocks	11,573	0.03	0.09	0.00	0.00	0.00	0.00	0.00	0.10	0.47
PctETF	11,573	0.03	0.10	0.00	0.00	0.00	0.00	0.00	0.05	0.53
PctIndex	11,596	0.47	0.37	0.00	0.00	0.05	0.47	0.82	1.00	1.00
PctInternational	10,822	0.10	0.14	0.00	0.00	0.00	0.02	0.17	0.29	0.60
PctEmerging	10,822	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.09

Panel C. Transactions and Fees

	N	mean	sd	p1	p10	p25	p50	p75	p90	p99
MgtFee	10,806	0.14	0.12	0.00	0.03	0.06	0.10	0.17	0.28	0.58
ExpRatio($\times$100)	10,310	0.19	0.17	0.02	0.07	0.09	0.14	0.22	0.36	0.93
TurnRatio	10,252	0.32	0.27	0.03	0.07	0.14	0.25	0.40	0.63	1.40
Transaction	11,596	3.05	6.06	0.00	0.00	0.00	1.00	3.00	9.00	31.00
Volume	11,596	78,725	209,920	0.00	0.00	0.00	200	26,194	235,000	1,081,523

This table reports demographic characteristics and portfolio allocation behavior of Vanguard clients the month before signing up for PAS. The results are computed at the client level and include all account types, that is, taxable and non-taxable (IRA) accounts. Panel A reports demographic characteristics: *Age*, is the age of the client as of December 2017; *Male*, is the fraction of male clients; *Married*, is the fraction of married clients; *Tenure*, is the tenure of the client as of December 2017. Panel B focuses on portfolio characteristics: *Wealth*, is the account balance; *NumAssets*, is the number of assets held by the client across accounts; *PctEquityShare*, is the percentage of wealth in Equities—directly or through mutual funds; *PctBondShare*, is the percentage of wealth in corporate bonds—directly or through mutual funds; *PctCashShare*, is the percentage of wealth money market mutual funds—directly or through mutual funds; *PctMutualFunds*, is the percentage of wealth directly invested in mutual funds; *PctCash*, is the percentage of wealth directly invested in money market mutual funds; *PctStocks*, is the percentage of wealth directly invested in individual stocks; *PctETF*, is the percentage of wealth directly invested in ETFs; *PctIndex*, is the percentage of mutual fund wealth invested in index funds; *PctInternational*, is the percentage of mutual fund wealth invested in global or international funds—identified using the Lipper mutual fund classification; *PctEmerging*, is the percentage of mutual fund wealth invested in emerging market funds—identified using the Lipper mutual fund classification. Panel C focuses on transactions and fees paid: *MgtFees*, are the value-weighted management fees charged by the mutual funds held by the account-holders; *ExpRatio*, is the value-weighted expense ratio charged by the mutual funds held by the clients; *TurnRatio*, is the value-weighted turnover ratio of the mutual funds held by the clients; *Transaction*, is the number of transactions directly initiated by the investors over the month before signing up for PAS; *Volume*, is the volume (in US dollars) traded by the investors over the month before signing up for PAS. For each variable, we report the number of accounts used in the computations, the mean, the standard deviation, and various percentiles of the distribution: the 1^{st}, 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th}, and 99^{th}.

Table 2. Demographic and Portfolio Characteristics of PAS investors 12 Months after Sign-up

CLIENT LEVEL

Panel A. Demographic Characteristics

	N	mean	sd	p1	p10	p25	p50	p75	p90	p99
Age	11,324	63.23	12.78	30.00	45.00	56.00	65.00	71.00	78.00	90.00
Male	11,596	0.53	0.50	0.00	0.00	0.00	1.00	1.00	1.00	1.00
Married	11,596	0.35	0.48	0.00	0.00	0.00	0.00	1.00	1.00	1.00
Tenure	11,595	14.10	9.32	0.42	2.00	5.00	14.08	20.50	26.08	35.67

Panel B. Portfolio Allocation

	N	mean	sd	p1	p10	p25	p50	p75	p90	p99
Wealth	7,261	706,783	775,793	31,194	91,994	184,013	437,065	926,809	1,683,102	3,722,903
NumAssets	7,261	7.45	4.52	1.00	4.00	4.00	6.00	9.00	13.00	24.00
PctEquityShare	7,261	0.58	0.21	0.00	0.36	0.48	0.57	0.70	0.87	1.00
PctBondShare	7,261	0.40	0.21	0.00	0.12	0.28	0.41	0.50	0.61	1.00
PctCashShare	7,261	0.01	0.04	0.00	0.00	0.00	0.00	0.01	0.05	0.18
PctMutualFunds	7,261	0.96	0.08	0.65	0.87	0.97	1.00	1.00	1.00	1.00
PctCash	7,261	0.01	0.04	0.00	0.00	0.00	0.00	0.01	0.05	0.18
PctStocks	7,261	0.01	0.03	0.00	0.00	0.00	0.00	0.00	0.02	0.18
PctETF	7,261	0.01	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.15
PctIndex	7,261	0.84	0.16	0.38	0.60	0.75	0.87	1.00	1.00	1.00
PctInternational	7,261	0.33	0.13	0.00	0.19	0.28	0.34	0.37	0.42	0.78
PctEmerging	7,261	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01

Panel C. Transactions and Fees

	N	mean	sd	p1	p10	p25	p50	p75	p90	p99
MgtFee	7,260	0.07	0.02	0.04	0.06	0.06	0.06	0.08	0.10	0.16
ExpRatio(×100)	7,259	0.09	0.02	0.06	0.07	0.08	0.08	0.10	0.12	0.19
TurnRatio	7,258	0.27	0.14	0.03	0.08	0.18	0.28	0.34	0.41	0.67
Transaction	7,261	2.41	3.88	0.00	0.00	0.00	1.00	3.00	8.00	18.00
Volume	7,261	11,955	38,423	0.00	0.00	0.00	163	2,888	23,254	217,672

This table reports demographic characteristics and portfolio allocation behavior of Vanguard clients 12 months after signing up for PAS. The results are computed at the client level and include all account types, that is, taxable and non-taxable (IRA) accounts. Panel A reports demographic characteristics: Age, is the age of the client as of December 2017; $Male$, is the fraction of male clients; $Married$, is the fraction of married clients; $Tenure$, is the tenure of the client as of December 2017. Panel B focuses on portfolio characteristics: $Wealth$, is the account balance; $NumAssets$, is the number of assets held by the client across accounts; $PctEquityShare$, is the percentage of wealth in Equities—directly or through mutual funds; $PctBondShare$, is the percentage of wealth in corporate bonds—directly or through mutual funds; $PctCashShare$, is the percentage of wealth money market mutual funds—directly or through mutual funds; $PctMutualFunds$, is the percentage of wealth directly invested in mutual funds; $PctCash$, is the percentage of wealth directly invested in money market mutual funds; $PctStocks$, is the percentage of wealth directly invested in individual stocks; $PctETF$, is the percentage of wealth directly invested in ETFs; $PctIndex$, is the percentage of mutual fund wealth invested in index funds; $PctInternational$, is the percentage of mutual fund wealth invested in global or international funds—identified using the Lipper mutual fund classification; $PctEmerging$, is the percentage of mutual fund wealth invested in emerging market funds—identified using the Lipper mutual fund classification. Panel C focuses on transactions and fees paid: $MgtFees$, are the value-weighted management fees charged by the mutual funds held by the account-holders; $ExpRatio$, is the value-weighted expense ratio charged by the mutual funds held by the clients; $TurnRatio$, is the value-weighted turnover ratio of the mutual funds held by the clients; $Transaction$, is the number of transactions directly initiated by the investors over the month before signing up for PAS; $Volume$, is the volume (in US dollars) traded by the investors over the month before signing up for PAS. For each variable, we report the number of accounts used in the computations, the mean, the standard deviation, and various percentiles of the distribution: the 1^{st}, 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th}, and 99^{th}.

Table 3. Top Tickers by Asset Class in January 2017

Panel A. Mutual Fund

Rank	NON-PAS Ticker	NON-PAS Pct of Assets	PAS Ticker	PAS Pct of Assets
1	VTSAX	16%	VTSAX	28%
2	VFIAX	7%	VTIAX	18%
3	VBTLX	7%	VBTLX	16%
4	VTIAX	5%	VTABX	11%
5	VWIUX	4%	VFIDX	6%
6	VWENX	4%	VFSUX	4%
7	VGHAX	2%	VWIUX	2%
8	VWIAX	2%	VFIAX	2%
9	VTABX	2%	VMLUX	1%
10	VITSX	2%	VEXAX	1%

Panel B. ETF

Rank	NON-PAS Ticker	NON-PAS Pct of Assets	PAS Ticker	PAS Pct of Assets
1	VTI	20%	VTI	27%
2	VOO	6%	VEU	8%
3	VYM	3%	VXUS	4%
4	VWO	3%	SPY	4%
5	SPY	3%	VOO	4%
6	VXUS	3%	ISHG	3%
7	VEU	3%	IGOV	3%
8	VNQ	3%	BND	3%
9	VIG	3%	VIG	3%
10	VHT	2%	VUG	2%

Panel C. Stocks

Rank	NON-PAS Ticker	NON-PAS Pct of Assets	PAS Ticker	PAS Pct of Assets
1	AAPL	5%	BRK B	5%
2	BRK B	3%	AAPL	5%
3	XOM	3%	MSFT	4%
4	GE	3%	T	2%
5	T	3%	VZ	2%
6	JNJ	2%	MRK	2%
7	CVX	2%	GE	2%
8	MSFT	2%	GOOGL	2%
9	BRK A	1%	JNJ	2%
10	VZ	1%	HD	2%

Panel D. Cash

Rank	NON-PAS Ticker	NON-PAS Pct of Assets	PAS Ticker	PAS Pct of Assets
1	VMMXX	64%	VMMXX	52%
2	VMSXX	14%	VMFXX	21%
3	VUSXX	7%	VMSXX	16%
4	VMFXX	6%	VMRXX	5%
5	VMRXX	3%	VCTXX	4%
6	VCTXX	3%	VUSXX	2%
7	VYFXX	1%	VPTXX	1%
8	VPTXX	1%	VYFXX	0%
9	VNJXX	1%	VNJXX	0%
10	VOHXX	0%	–	–

This table reports the top 10 tickers held across NON-PAS and PAS investors. The results are computed at the account-holder level and include all account types, that is, taxable and non-taxable (IRA) accounts. Panel A reports the results for mutual funds. Panels B, C, and D reports results for ETFs, individual stocks and money market mutual funds, respectively. Within each panel, the left sub-panels report results for non-PAS investors, while the right sub-panels report results for PAS investors. Each sub-panel reports—from left to right—the holdings rank, the ticker, and the percentage of asset class wealth invested in the ticker, computed across account holders. The results are computed as of January 2017.

Table 4. Abnormal Sharpe Ratio at Robo-Advising Adoption: CLIENT LEVEL

Panel A. 3-Month Horizon

	All Accounts		**Matched Accounts**		
	After	**Before**	**After**	**Before**	**Difference**
Sharpe Ratio	0.017	0.002	0.059***	-0.006	0.065***
	(1.56)	(0.22)	(4.87)	(-0.58)	(3.92)
N	10318	8431	7786	7786	7786

Panel B. 6-Month Horizon

	All Accounts		**Matched Accounts**		
	After	**Before**	**After**	**Before**	**Difference**
Sharpe Ratio	0.064***	0.010	0.115***	-0.067***	0.182***
	(8.77)	(1.25)	(11.90)	(-8.06)	(13.02)
N	9484	7361	6088	6088	6088

Panel C. 9-Month Horizon

	All Accounts		**Matched Accounts**		
	After	**Before**	**After**	**Before**	**Difference**
Sharpe Ratio	0.068***	-0.025***	0.058***	0.004	0.054***
	(11.36)	(-3.78)	(6.28)	(0.63)	(4.25)
N	8421	6487	4490	4490	4490

This table reports investment performance across all Vanguard clients before and after signing up for PAS. The results are computed at the client level. As a measure of performance, we use the abnormal Sharpe Ratio, that is, the difference between the realized Sharpe ratio of each account-holder and the realized Sharpe ratio of the market portfolio, where the latter is computed as the value-weighted returns on the NYSE/AMEX/NASDAQ CRSP portfolio. For each account-holder, portfolio returns and volatilities are computed using beginning-of-month investment holdings. Furthermore, portfolio volatilities are computed as realized volatilities using squared daily returns. Panels A, B, and C report results for Sharpe ratios computed at the 3-, 6-, and 9-month horizons. Each column within each panel reports average abnormal Sharpe ratios, *t*-statistics testing the null hypothesis that the averages are equal to zero, and the number of observations used in the computations of the results. The first and second column report statistics for all clients available before and after signing up for PAS. The third and fourth column uses only the accounts present before and after. Finally, the last column reports column reports the average performance difference after and before signing up for PAS, computed at the client level. In all cases, performance is computed from PAS signup.

Table 5. Abnormal Sharpe Ratio at Six Months Before and After Robo-Advising Adoption: CLIENT LEVEL

Panel A. 3-Month Horizon

	All Accounts		**Matched Accounts**		
	After	**Before**	**After**	**Before**	**Difference**
Sharpe Ratio	0.151***	-0.018*	0.144***	0.066***	0.078***
	(15.53)	(-1.73)	(9.78)	(6.81)	(4.52)
N	9118	6805	5037	5037	5037

Panel B. 6-Month Horizon

	All Accounts		**Matched Accounts**		
	After	**Before**	**After**	**Before**	**Difference**
Sharpe Ratio	0.160***	-0.033***	0.162***	0.113***	0.050***
	(20.12)	(-4.11)	(10.89)	(13.39)	(3.15)
N	7869	5869	3299	3299	3299

Panel C. 9-Month Horizon

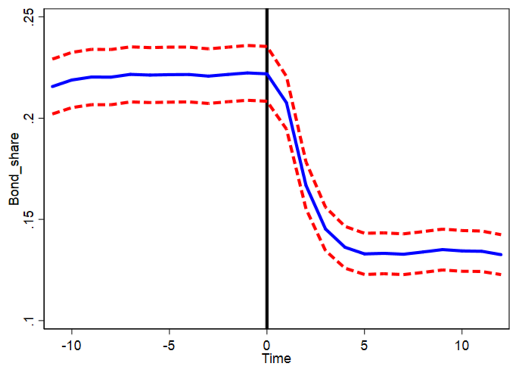
	All Accounts		**Matched Accounts**		
	After	**Before**	**After**	**Before**	**Difference**
Sharpe Ratio	0.141***	0.001	0.488***	0.090***	0.399***
	(20.29)	(0.18)	(32.79)	(9.62)	(23.73)
N	6698	4916	1573	1573	1573

This table reports investment performance across all Vanguard clients before and after signing up for PAS. The results are computed at the client level. As a measure of performance, we use the abnormal Sharpe Ratio, that is, the difference between the realized Sharpe ratio of each account-holder and the realized Sharpe ratio of the market portfolio, where the latter is computed as the value-weighted returns on the NYSE/AMEX/NASDAQ CRSP portfolio. For each account-holder, portfolio returns and volatilities are computed using beginning-of-month investment holdings. Furthermore, portfolio volatilities are computed as realized volatilities using squared daily returns. Panels A, B, and C report results for Sharpe ratios computed at the 3-, 6-, and 9-month horizons. Each column within each panel reports average abnormal Sharpe ratios, t-statistics testing the null hypothesis that the averages are equal to zero, and the number of observations used in the computations of the results. The first and second column report statistics for all clients available before and after signing up for PAS. The third and fourth column uses only the accounts present before and after. Finally, the last column reports column reports the average performance difference after and before signing up for PAS, computed at the client level. In all cases, performance is computed starting from 6 months before and after PAS is implemented for the client.

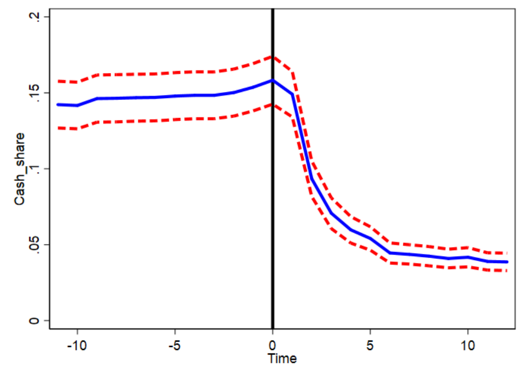
Online Appendix for the Paper:

Who Benefits from Robo-Advising? Evidence from Machine Learning

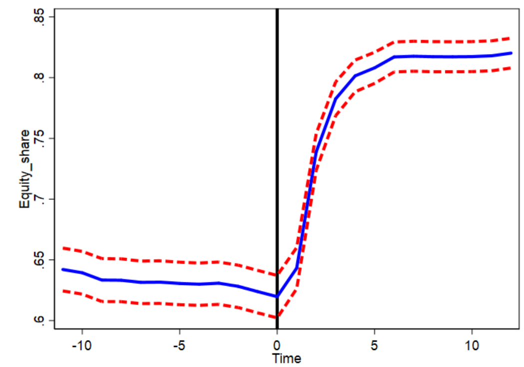
(Not for publication)



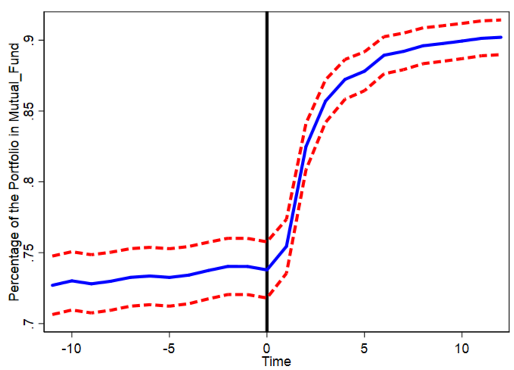
(a) Percentage of Bonds



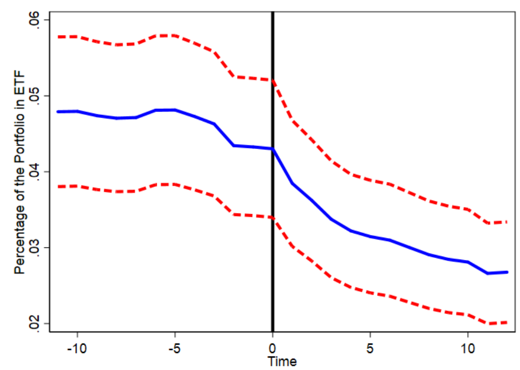
(b) Percentage of Cash



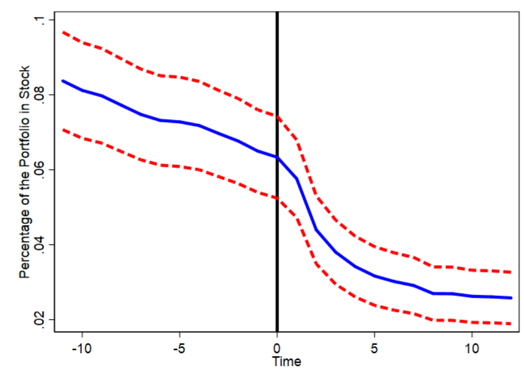
(c) Percentage of Equities



(d) Percentage in Mutual Funds

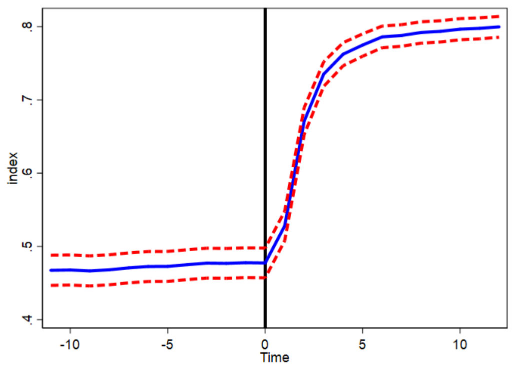


(e) Percentage in ETFs

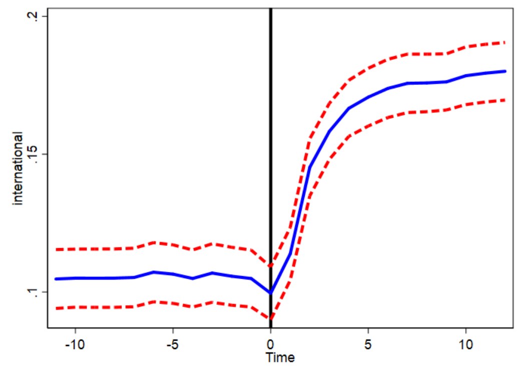


(f) Percentage in Individual Stocks

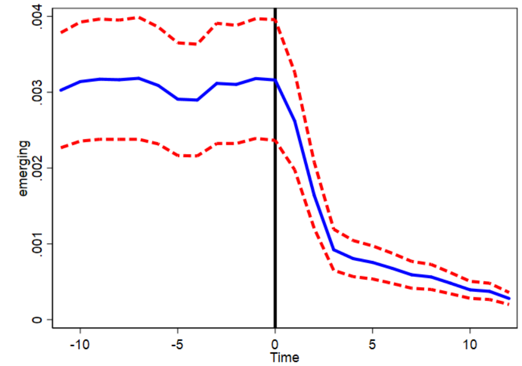
Figure Online I. Portfolio Allocation before and after PAS: Taxable accounts. This figure reports results for portfolio characteristics of Vanguard taxable account holders before and after signing up for PAS. The results are computed at the account-holder level and include only taxable accounts. Subfigures (a), (b), and (c) report results for the percentage of wealth held—directly or through mutual funds—in corporate bonds, cash and equities, respectively. Subfigures (d), (e), and (f) report results for the percentage of wealth invested in mutual funds, ETFs, and individual stocks, respectively. In each subfigure, time "0" represent the month before investors sign up for PAS. Results are computed using only investors that are in the sample for at least twelve months before and after signing up for PAS. The blue line denotes average values, while the red dashed lines are 95% confidence intervals.



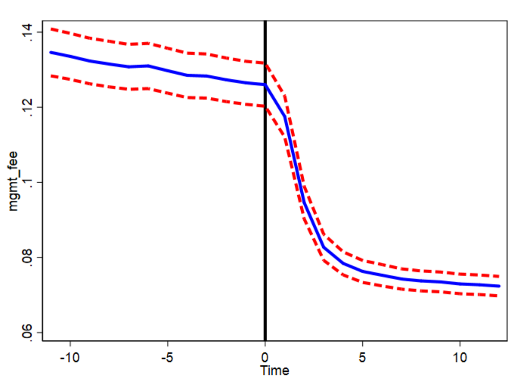
(a) Percentage of Indexed Funds



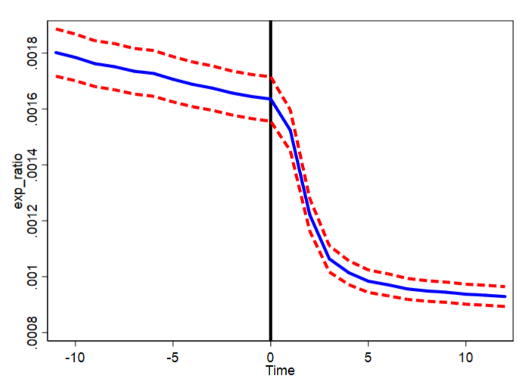
(b) Percentage of International Funds



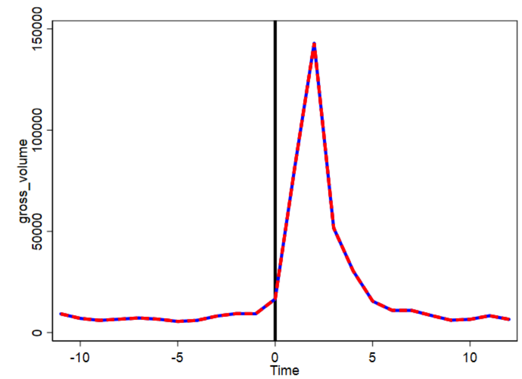
(c) Percentage of Emerging Markets Funds



(d) Management Fees

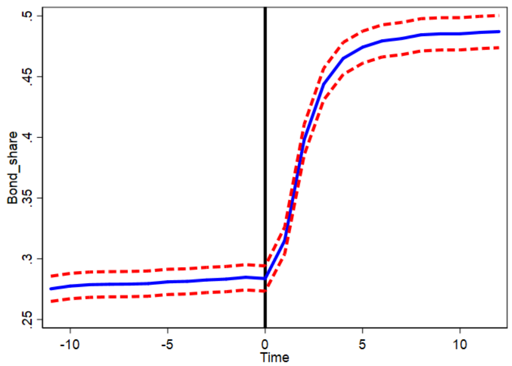


(e) Expense Ratios

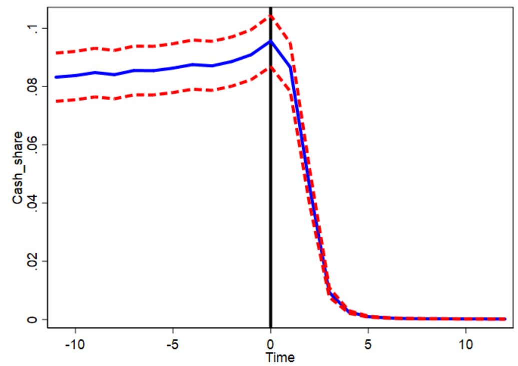


(f) Trading Volume

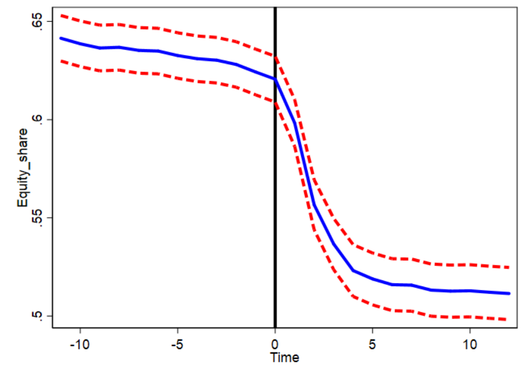
Online II. Indexation, international diversification and fees before and after PAS: Taxable accounts. This figure reports results for degree of indexation, international diversification, trading volume and fees of Vanguard taxable account holders before and after signing up for PAS. The results are computed at the account-holder level and include only taxable accounts. Subfigure (a) reports results for the percentage of mutual fund wealth invested in index funds; Subfigure (b) the percentage of mutual fund wealth invested in global or international funds—identified using the Lipper mutual fund classification; and Subfigure (c) the percentage of mutual fund wealth invested in emerging market funds—identified using the Lipper mutual fund classification. Subfigure (d) presents results for the value-weighted management fees charged by the mutual funds held by the account-holders; and Subfigure (e), the value-weighted expense ratio charged by the mutual funds held by the account-holders. Finally, Subfigure (f) shows results for the monthly trading volume, in US dollars. In each subfigure, time "0" represent the month before investors sign up for PAS. Results are computed using only investors that are in the sample for at least twelve months before and after signing up for PAS. The blue line denotes average values, while the red dashed lines are 95% confidence intervals.



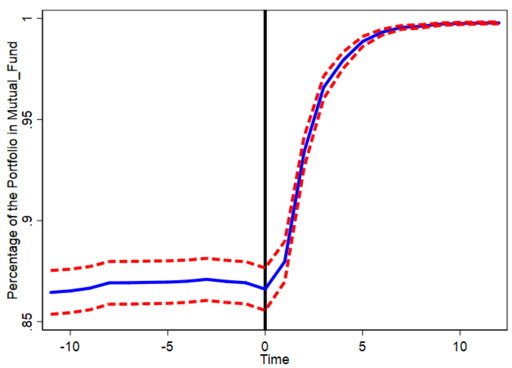
(a) Percentage of Bonds



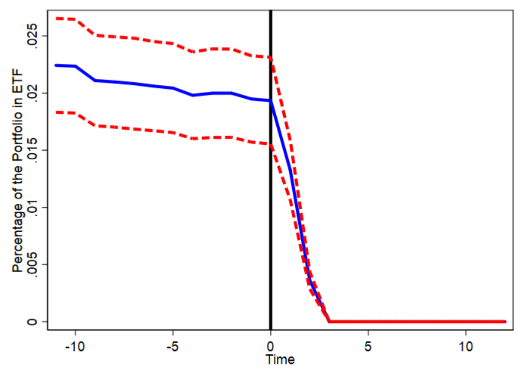
(b) Percentage of Cash



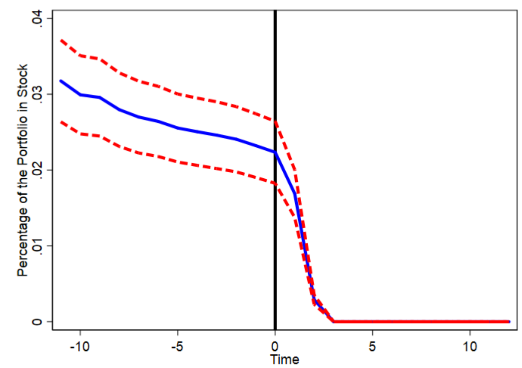
(c) Percentage of Equities



(d) Percentage in Mutual Funds

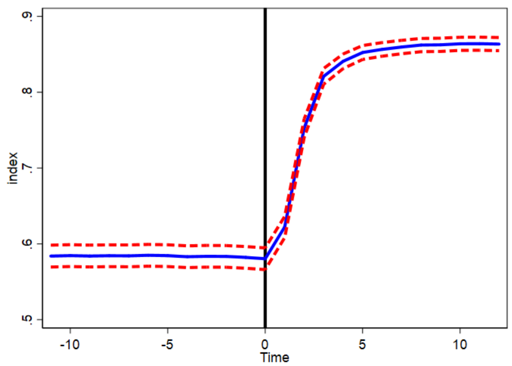


(e) Percentage in ETFs

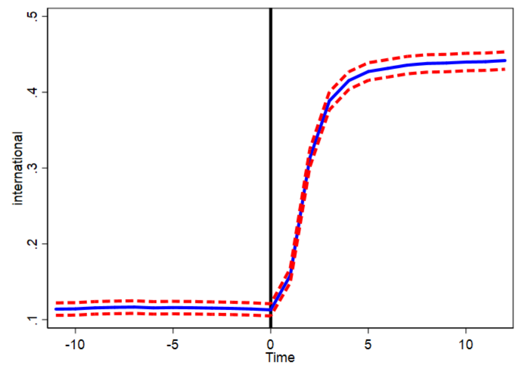


(f) Percentage in Individual Stocks

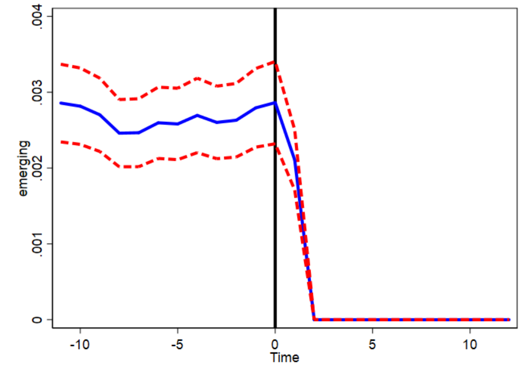
Figure Online III. Portfolio Allocation before and after PAS: IRA accounts. This figure reports results for portfolio characteristics of Vanguard non-taxable account holders before and after signing up for PAS. The results are computed at the account-holder level and include only non-taxable (IRA) accounts. Subfigures (a), (b), and (c) report results for the percentage of wealth held—directly or through mutual funds—in corporate bonds, cash and equities, respectively. Subfigures (d), (e), and (f) report results for the percentage of wealth invested in mutual funds, ETFs, and individual stocks, respectively. In each subfigure, time "0" represent the month before investors sign up for PAS. Results are computed using only investors that are in the sample for at least twelve months before and after signing up for PAS. The blue line denotes average values, while the red dashed lines are 95% confidence intervals.



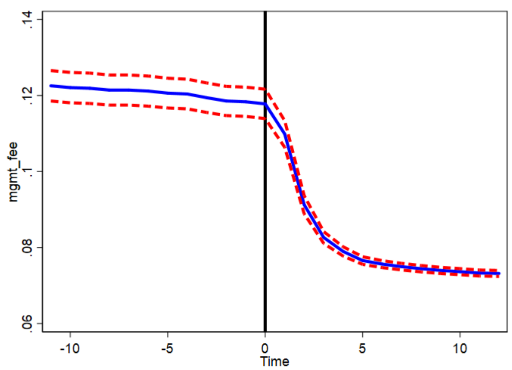
(a) Percentage of Indexed Funds



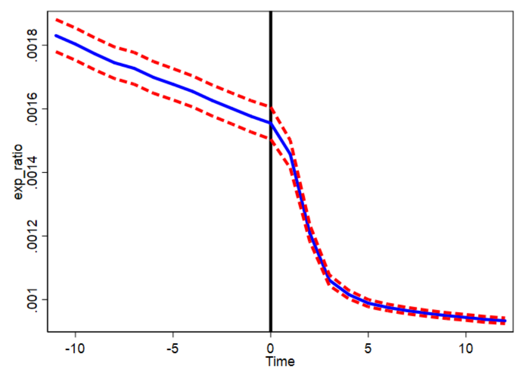
(b) Percentage of International Funds



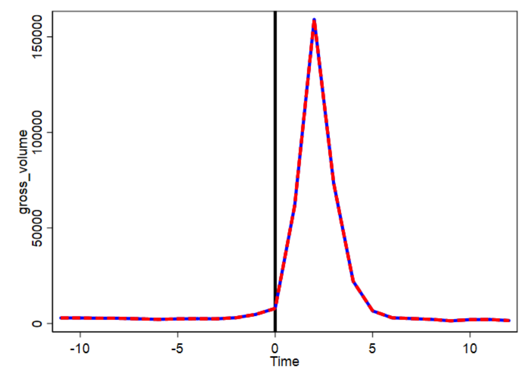
(c) Percentage of Emerging Markets Funds



(d) Management Fees



(e) Expense Ratios



(f) Trading Volume

Figure Online IV. Indexation, international diversification and fees before and after PAS: IRA accounts. This figure reports results for degree of indexation, international diversification, trading volume and fees of Vanguard non-taxable account holders before and after signing up for PAS. The results are computed at the account-holder level and include only non-taxable (IRA) accounts. Subfigure (a) reports results for the percentage of mutual fund wealth invested in index funds; Subfigure (b) the percentage of mutual fund wealth invested in global or international funds—identified using the Lipper mutual fund classification; and Subfigure (c) the percentage of mutual fund wealth invested in emerging market funds—identified using the Lipper mutual fund classification. Subfigure (d) presents results for the value-weighted management fees charged by the mutual funds held by the account-holders; and Subfigure (e), the value-weighted expense ratio charged by the mutual funds held by the account-holders. Finally, Subfigure (f) shows results for the monthly trading volume, in US dollars. In each subfigure, time "0" represent the month before investors sign up for PAS. Results are computed using only investors that are in the sample for at least twelve months before and after signing up for PAS. The blue line denotes average values, while the red dashed lines are 95% confidence intervals.

Table Online I. Demographic and Portfolio Characteristics of PAS investors at Sign-up: TAXABLE ACCOUNTS

Panel A. Demographic Characteristics

	N	mean	sd	p1	p10	p25	p50	p75	p90	p99
Age	7,428	64.28	13.78	29.00	45.00	57.00	65.00	73.00	81.00	91.00
Male	7,764	0.54	0.50	0.00	0.00	0.00	1.00	1.00	1.00	1.00
Married	7,764	0.37	0.48	0.00	0.00	0.00	0.00	1.00	1.00	1.00
Tenure	7,763	15.35	9.70	0.50	2.42	6.00	16.08	21.92	27.67	36.83

Panel B. Portfolio Allocation

	N	mean	sd	p1	p10	p25	p50	p75	p90	p99
Wealth	7,764	405,217	716,719	625.00	9,907	39,529	126,854	419,790	1,057,171	3,772,967
NumAssets	7,764	4.96	5.53	1.00	1.00	1.00	3.00	6.00	11.00	29.00
PctEquityShare	7,674	0.56	0.37	0.00	0.00	0.19	0.63	0.93	1.00	1.00
PctBondShare	7,674	0.20	0.27	0.00	0.00	0.00	0.06	0.34	0.59	1.00
PctCashShare	7,674	0.24	0.38	0.00	0.00	0.00	0.01	0.32	1.00	1.00
PctMutualFunds	7,762	0.62	0.44	0.00	0.00	0.00	0.92	1.00	1.00	1.00
PctCash	7,762	0.23	0.37	0.00	0.00	0.00	0.00	0.27	1.00	1.00
PctStocks	7,762	0.09	0.25	0.00	0.00	0.00	0.00	0.00	0.30	1.00
PctETF	7,762	0.05	0.19	0.00	0.00	0.00	0.00	0.00	0.04	0.99
PctIndex	7,764	0.40	0.40	0.00	0.00	0.00	0.30	0.81	1.00	1.00
PctInternational	6,752	0.10	0.18	0.00	0.00	0.00	0.00	0.15	0.30	0.87
PctEmerging	6,752	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.10

Panel C. Transactions and Fees

	N	mean	sd	p1	p10	p25	p50	p75	p90	p99
MgtFee	6,748	0.15	0.14	0.00	0.04	0.06	0.11	0.17	0.32	0.64
ExpRatio(×100)	6,536	0.20	0.20	0.05	0.06	0.09	0.14	0.20	0.41	1.00
TurnRatio	6,034	0.28	0.31	0.03	0.03	0.08	0.19	0.36	0.62	1.60
Transaction	7,764	1.71	3.49	0.00	0.00	0.00	0.00	2.00	5.00	18.00
Volume	7,764	43,220	137,585	0.00	0.00	0.00	0.00	7,181	100,000	780,881

This table reports demographic characteristics and portfolio allocation behavior of Vanguard taxable account holders the month before signing up for PAS. The results are computed at the account-holder level and include only taxable accounts. Panel A reports demographic characteristics: Age, is the age of the account-holder as of December 2017; $Male$, is the fraction of male account-holders; $Married$, is the fraction of married account-holders; $Tenure$, is the tenure of the account as of December 2017. Panel B focuses on portfolio characteristics: $Wealth$, is the account balance; *NumAssets*, is the number of assets held in the account; *PctEquityShare*, is the percentage of wealth in Equities—directly or through mutual funds; *PctBondShare*, is the percentage of wealth in corporate bonds—directly or through mutual funds; *PctCashShare*, is the percentage of wealth money market mutual funds—directly or through mutual funds; *PctMutualFunds*, is the percentage of wealth directly invested in mutual funds; *PctCash*, is the percentage of wealth directly invested in money market mutual funds; *PctStocks*, is the percentage of wealth directly invested in individual stocks; $PctETF$, is the percentage of wealth directly invested in ETFs; *PctIndex*, is the percentage of mutual fund wealth invested in index funds; $PctInternational$, is the percentage of mutual fund wealth invested in global or international funds—identified using the Lipper mutual fund classification; *PctEmerging*, is the percentage of mutual fund wealth invested in emerging market funds—identified using the Lipper mutual fund classification. Panel C focuses on transactions and fees paid: $MgtFees$, are the value-weighted management fees charged by the mutual funds held by the account-holders; *ExpRatio*, is the value-weighted expense ratio charged by the mutual funds held by the account-holders; *TurnRatio*, is the value-weighted turnover ratio of the mutual funds held by the account-holders; $Transaction$, is the number of transactions directly initiated by the investors over the month before signing up for PAS; $Volume$, is the volume (in US dollars) traded by the investors over the month before signing up for PAS. For each variable, we report the number of accounts used in the computations, the mean, the standard deviation, and various percentiles of the distribution: the 1^{st}, 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th}, and 99^{th}.

Table Online II. Demographic and Portfolio Characteristics of PAS investors at Sign-up: IRA ACCOUNTS

Panel A. Demographic Characteristics

	N	mean	sd	p1	p10	p25	p50	p75	p90	p99
Age	15,928	63.15	12.01	31.00	46.00	57.00	64.00	71.00	77.00	89.00
Male	15,937	0.54	0.50	0.00	0.00	0.00	1.00	1.00	1.00	1.00
Married	15,937	0.38	0.48	0.00	0.00	0.00	0.00	1.00	1.00	1.00
Tenure	15,935	15.50	9.19	0.58	2.58	7.67	16.08	21.42	27.00	36.33

Panel B. Portfolio Allocation

	N	mean	sd	p1	p10	p25	p50	p75	p90	p99
Wealth	15,937	208,329	296,555	1,654	8,914	27,090	87,109	255,400	576,470	1,396,192
NumAssets	15,937	3.48	3.68	1.00	1.00	1.00	2.00	4.00	8.00	18.00
PctEquityShare	15,657	0.55	0.38	0.00	0.00	0.16	0.61	0.91	1.00	1.00
PctBondShare	15,657	0.26	0.31	0.00	0.00	0.00	0.14	0.42	0.72	1.00
PctCashShare	15,657	0.20	0.37	0.00	0.00	0.00	0.00	0.12	1.00	1.00
PctMutualFunds	15,933	0.74	0.41	0.00	0.00	0.51	1.00	1.00	1.00	1.00
PctCash	15,933	0.18	0.35	0.00	0.00	0.00	0.00	0.07	1.00	1.00
PctStocks	15,933	0.03	0.15	0.00	0.00	0.00	0.00	0.00	0.00	0.87
PctETF	15,933	0.03	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.85
PctIndex	15,937	0.49	0.43	0.00	0.00	0.00	0.48	1.00	1.00	1.00
PctInternational	14,527	0.11	0.23	0.00	0.00	0.00	0.00	0.13	0.36	1.00
PctEmerging	14,527	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.11

Panel C. Transactions and Fees

	N	mean	sd	p1	p10	p25	p50	p75	p90	p99
MgtFee	14,378	0.15	0.15	0.00	0.00	0.05	0.11	0.19	0.33	0.67
ExpRatio	12,974	0.21	0.23	0.00	0.05	0.09	0.15	0.24	0.46	1.10
TurnRatio	13,185	0.33	0.29	0.03	0.04	0.11	0.25	0.46	0.72	1.37
Transaction	15,937	1.12	2.71	0.00	0.00	0.00	0.00	1.00	3.00	14.00
Volume	15,937	29,954	96,152	0.00	0.00	0.00	0.00	1,000	70,000	521,327

This table reports demographic characteristics and portfolio allocation behavior of Vanguard non-taxable account holders the month before signing up for PAS. The results are computed at the account-holder level and include only non-taxable (IRA) accounts. Panel A reports demographic characteristics: Age, is the age of the account-holder as of December 2017; $Male$, is the fraction of male account-holders; $Married$, is the fraction of married account-holders; $Tenure$, is the tenure of the account as of December 2017. Panel B focuses on portfolio characteristics: $Wealth$, is the account balance; $NumAssets$, is the number of assets held in the account; $PctEquityShare$, is the percentage of wealth in Equities—directly or through mutual funds; $PctBondShare$, is the percentage of wealth in corporate bonds—directly or through mutual funds; $PctCashShare$, is the percentage of wealth money market mutual funds—directly or through mutual funds; $PctMutualFunds$, is the percentage of wealth directly invested in mutual funds; $PctCash$, is the percentage of wealth directly invested in money market mutual funds; $PctStocks$, is the percentage of wealth directly invested in individual stocks; $PctETF$, is the percentage of wealth directly invested in ETFs; $PctIndex$, is the percentage of mutual fund wealth invested in index funds; $PctInternational$, is the percentage of mutual fund wealth invested in global or international funds—identified using the Lipper mutual fund classification; $PctEmerging$, is the percentage of mutual fund wealth invested in emerging market funds—identified using the Lipper mutual fund classification. Panel C focuses on transactions and fees paid: $MgtFees$, are the value-weighted management fees charged by the mutual funds held by the account-holders; $ExpRatio$, is the value-weighted expense ratio charged by the mutual funds held by the account-holders; $TurnRatio$, is the value-weighted turnover ratio of the mutual funds held by the account-holders; $Transaction$, is the number of transactions directly initiated by the investors over the month before signing up for PAS; $Volume$, is the volume (in US dollars) traded by the investors over the month before signing up for PAS. For each variable, we report the number of accounts used in the computations, the mean, the standard deviation, and various percentiles of the distribution: the 1^{st}, 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th}, and 99^{th}.

Table Online III. Demographic and Portfolio Characteristics of PAS 12 Months after Sign-up

TAXABLE ACCOUNTS

Panel A. Demographic Characteristics

	N	mean	sd	p1	p10	p25	p50	p75	p90	p99
Age	5,984	64.60	13.58	30.00	45.00	57.00	66.00	74.00	81.00	91.00
Male	6,267	0.54	0.50	0.00	0.00	0.00	1.00	1.00	1.00	1.00
Married	6,267	0.38	0.48	0.00	0.00	0.00	0.00	1.00	1.00	1.00
Tenure	6,266	16.06	9.39	1.50	3.17	7.25	16.92	22.17	27.92	37.17

Panel B. Portfolio Allocation

	N	mean	sd	p1	p10	p25	p50	p75	p90	p99
Wealth	4,155	398,826	553,205	2,555	18,241	61,460	175,478	486,398	1,092,161	2,643,665
NumAssets	4,155	4.31	2.94	1.00	1.00	2.00	4.00	6.00	8.00	13.00
PctEquityShare	4,131	0.81	0.25	0.00	0.45	0.66	0.94	1.00	1.00	1.00
PctBondShare	4,131	0.15	0.21	0.00	0.00	0.00	0.00	0.29	0.49	0.73
PctCashShare	4,131	0.03	0.10	0.00	0.00	0.00	0.00	0.00	0.08	0.59
PctMutualFunds	4,155	0.89	0.27	0.00	0.56	0.95	1.00	1.00	1.00	1.00
PctCash	4,155	0.03	0.09	0.00	0.00	0.00	0.00	0.00	0.07	0.51
PctStocks	4,155	0.04	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.99
PctETF	4,155	0.03	0.15	0.00	0.00	0.00	0.00	0.00	0.00	0.88
PctIndex	4,155	0.79	0.30	0.00	0.30	0.68	0.94	1.00	1.00	1.00
PctInternational	4,036	0.19	0.20	0.00	0.00	0.00	0.17	0.31	0.38	0.93
PctEmerging	4,036	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01

Panel C. Transactions and Fees

	N	mean	sd	p1	p10	p25	p50	p75	p90	p99
MgtFee	4,036	0.08	0.06	0.04	0.04	0.05	0.06	0.08	0.12	0.37
ExpRatio(×100)	4,029	0.09	0.08	0.05	0.05	0.06	0.08	0.10	0.14	0.56
TurnRatio	4,002	0.12	0.13	0.03	0.03	0.03	0.06	0.17	0.33	0.56
Transaction	4,155	1.36	2.50	0.00	0.00	0.00	0.00	1.00	4.00	12.00
Volume	4,155	6,209	21,034	0.00	0.00	0.00	0.00	1,156	10,183	120,313

This table reports demographic characteristics and portfolio allocation behavior of Vanguard taxable account holders 12 months after signing up for PAS. The results are computed at the account-holder level and include only taxable accounts. Panel A reports demographic characteristics: Age, is the age of the account-holder as of December 2017; $Male$, is the fraction of male account-holders; $Married$, is the fraction of married account-holders; $Tenure$, is the tenure of the account as of December 2017. Panel B focuses on portfolio characteristics: $Wealth$, is the account balance; $NumAssets$, is the number of assets held in the account; $PctEquityShare$, is the percentage of wealth in Equities—directly or through mutual funds; $PctBondShare$, is the percentage of wealth in corporate bonds—directly or through mutual funds; $PctCashShare$, is the percentage of wealth money market mutual funds—directly or through mutual funds; $PctMutualFunds$, is the percentage of wealth directly invested in mutual funds; $PctCash$, is the percentage of wealth directly invested in money market mutual funds; $PctStocks$, is the percentage of wealth directly invested in individual stocks; $PctETF$, is the percentage of wealth directly invested in ETFs; $PctIndex$, is the percentage of mutual fund wealth invested in index funds; $PctInternational$, is the percentage of mutual fund wealth invested in global or international funds—identified using the Lipper mutual fund classification; $PctEmerging$, is the percentage of mutual fund wealth invested in emerging market funds—identified using the Lipper mutual fund classification. Panel C focuses on transactions and fees paid: $MgtFees$, are the value-weighted management fees charged by the mutual funds held by the account-holders; $ExpRatio$, is the value-weighted expense ratio charged by the mutual funds held by the account-holders; $TurnRatio$, is the value-weighted turnover ratio of the mutual funds held by the account-holders; $Transaction$, is the number of transactions directly initiated by the investors over the 12th month after signing up for PAS; $Volume$, is the volume (in US dollars) traded by the investors over the 12th month after signing up for PAS. For each variable, we report the number of accounts used in the computations, the mean, the standard deviation, and various percentiles of the distribution: the 1^{st}, 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th}, and 99^{th}.

Table Online IV. Demographic and Portfolio Characteristics of PAS 12 Months after Sign-up

IRA ACCOUNTS

Panel A. Demographic Characteristics

	N	mean	sd	p1	p10	p25	p50	p75	p90	p99
Age	17,372	63.13	12.24	31.00	46.00	57.00	64.00	71.00	78.00	89.00
Male	17,434	0.54	0.50	0.00	0.00	0.00	1.00	1.00	1.00	1.00
Married	17,434	0.37	0.48	0.00	0.00	0.00	0.00	1.00	1.00	1.00
Tenure	17,432	15.24	9.34	0.42	2.25	6.92	15.83	21.33	27.00	36.33

Panel B. Portfolio Allocation

	N	mean	sd	p1	p10	p25	p50	p75	p90	p99
Wealth	9,541	259,217	342,512	3,586	13,631	38,409	116,353	334,817	708,155	1,554,805
NumAssets	9,541	3.23	1.97	1.00	1.00	1.00	3.00	4.00	6.00	8.00
PctEquityShare	9,540	0.50	0.39	0.00	0.00	0.03	0.52	0.99	1.00	1.00
PctBondShare	9,540	0.49	0.39	0.00	0.00	0.00	0.48	0.96	1.00	1.00
PctCashShare	9,540	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
PctMutualFunds	9,540	1.00	0.02	0.90	1.00	1.00	1.00	1.00	1.00	1.00
PctCash	9,540	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PctStocks	9,540	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PctETF	9,540	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PctIndex	9,541	0.86	0.25	0.00	0.58	0.81	1.00	1.00	1.00	1.00
PctInternational	9,485	0.45	0.33	0.00	0.00	0.24	0.37	0.66	1.00	1.00
PctEmerging	9,485	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Panel C. Transactions and Fees

	N	mean	sd	p1	p10	p25	p50	p75	p90	p99
MgtFee	9,413	0.07	0.02	0.04	0.05	0.06	0.07	0.08	0.10	0.16
ExpRatio(×100)	9,405	0.09	0.03	0.05	0.06	0.08	0.09	0.11	0.12	0.18
TurnRatio	9,411	0.32	0.25	0.03	0.03	0.06	0.30	0.49	0.70	0.84
Transaction	9,541	0.69	1.65	0.00	0.00	0.00	0.00	1.00	2.00	8.00
Volume	9,541	1,604	6,324	0.00	0.00	0.00	0.00	12.44	1,771	37,910

This table reports demographic characteristics and portfolio allocation behavior of Vanguard non-taxable account holders 12 months after signing up for PAS. The results are computed at the account-holder level and include only non-taxable (IRA) accounts. Panel A reports demographic characteristics: *Age*, is the age of the account-holder as of December 2017; *Male*, is the fraction of male account-holders; *Married*, is the fraction of married account-holders; *Tenure*, is the tenure of the account as of December 2017. Panel B focuses on portfolio characteristics: *Wealth*, is the account balance; *NumAssets*, is the number of assets held in the account; *PctEquityShare*, is the percentage of wealth in Equities—directly or through mutual funds; *PctBondShare*, is the percentage of wealth in corporate bonds—directly or through mutual funds; *PctCashShare*, is the percentage of wealth money market mutual funds—directly or through mutual funds; *PctMutualFunds*, is the percentage of wealth directly invested in mutual funds; *PctCash*, is the percentage of wealth directly invested in money market mutual funds; *PctStocks*, is the percentage of wealth directly invested in individual stocks; *PctETF*, is the percentage of wealth directly invested in ETFs; *PctIndex*, is the percentage of mutual fund wealth invested in index funds; *PctInternational*, is the percentage of mutual fund wealth invested in global or international funds—identified using the Lipper mutual fund classification; *PctEmerging*, is the percentage of mutual fund wealth invested in emerging market funds—identified using the Lipper mutual fund classification. Panel C focuses on transactions and fees paid: *MgtFees*, are the value-weighted management fees charged by the mutual funds held by the account-holders; *ExpRatio*, is the value-weighted expense ratio charged by the mutual funds held by the account-holders; *TurnRatio*, is the value-weighted turnover ratio of the mutual funds held by the account-holders; *Transaction*, is the number of transactions directly initiated by the investors over the 12th month after signing up for PAS; *Volume*, is the volume (in US dollars) traded by the investors over the 12th month after signing up for PAS. For each variable, we report the number of accounts used in the computations, the mean, the standard deviation, and various percentiles of the distribution: the 1^{st}, 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th}, and 99^{th}.