

Algorithm Aversion: Theory and Evidence from Robo-Advice*

Fiona Greig¹ Tarun Ramadorai² Alberto G. Rossi³
Stephen Utkus³ Ansgar Walther²

¹Vanguard, ²Imperial College, ³Georgetown University

April 2, 2023

Abstract

Automation can lower costs and democratize access to many consumer services, but human discomfort with automation can pose barriers to technology adoption. We build a structural model of psychological “algorithm aversion,” which features ongoing disutility of dealing with an algorithm, pessimism about the algorithm’s ability, and uncertainty about the algorithm’s performance; all three components can be assuaged by human interaction. We estimate model parameters using unique data from a “hybrid” robo-advising service in which portfolio management is automated, but clients are randomly matched with human advisors who provide different standards of support. Algorithm aversion is mainly driven by ongoing disutility and uncertainty, and human advice is especially important in retaining investors in robo-advice during market downturns.

*We thank Brian Martin at Vanguard for useful conversations on institutional details, and Paige Weisman for excellent research assistance.

Introduction

Technological automation has displaced human employment in a variety of tasks over the past few decades (Acemoglu and Restrepo, 2019). More recently, automation has progressed to highly-skilled tasks in industries such as financial advice, portfolio management, accountancy, and legal services.

In principle, automation could either substitute for or complement the skills of human experts (Autor, 2015). In the setting of financial services, for example, financial advisors may be prone to biases and misconduct (Egan, Matvos, and Seru, 2019). Cheaper, automated “robo-advisors” are a potentially appealing substitute that may be able to lower costs and democratize access to high-quality financial management and advice—services routinely delivered by humans and often accessible only to the wealthy (Gomes, Haliassos, and Ramadorai, 2021; D’Acunto and Rossi, 2022). On the other hand, trusted human experts might help in the widespread adoption of such automated solutions, especially if consumers view automation with skepticism, uncertainty, or aversion. While such “algorithm aversion” has been observed repeatedly, there is as yet little consensus about the psychological underpinnings of this phenomenon (Burton, Stein, and Jensen, 2020).¹

In this paper, we attempt to understand whether human experts can complement the automated provision of services by assuaging ongoing discomfort with algorithmic solutions or by helping clients learn about the quality of such solutions. We unpack the different psychological channels that underpin algorithm aversion by developing a structural model which we combine with novel evidence from the economically important field setting of robo-advising.²

The model can be described in general terms. Clients in the model enroll in an automated service whose quality they do not perfectly observe, and they experience a fixed per-period cost/disutility from interfacing with the service. They are matched randomly with human experts who can affect both the prior mean and precision of their beliefs about the expected quality of the service; human experts can also help to assuage ongoing disutility. Clients learn about the quality of the automated service/algorithm from observing noisy realized performance and decide in each period whether to

¹For example, psychology experiments have shown that people prefer inferior human forecasters to superior algorithmic forecasters even after observing differences in performance (Dietvorst, Simmons, and Massey, 2015).

²Robo-advising is a uniquely useful venue to understand algorithm aversion because of its current scale and expected growth—estimated at over \$10 trillion under management over the next decade (see, e.g., Deloitte and Touche, “The expansion of robo-advisory in wealth management,” 2016).

stay in the service or quit.

We customize this broad framework to the specific robo-advice portfolio management setting by adapting a standard dynamic portfolio choice framework (Campbell and Viceira, 1999). In the context of robo-advising, the per-period disutility from investing in the robo-advisor can be interpreted as an ongoing psychological cost associated with dealing with automated portfolio management. In addition, the investor learns about quality, i.e., the expected return available from the service, from the return stream, which we model as a generic risky asset paying a positive risk premium. As time evolves, the investor updates her prior belief using Bayes rule applied to the sequence of returns generated by the algorithm. This focus on learning creates similarities between our work and Timmermann (1993), Brennan and Xia (2001), and Pástor and Veronesi (2003).

This model setup allows for three conceptually distinct sources of algorithm aversion: the ongoing *disutility* of dealing with the robo-advisor, investors' prior *pessimism* about expected returns generated by the algorithm, and investors' prior *uncertainty* about expected returns. All three facets of algorithm aversion can be affected by human advice. In particular, depending on their type, human advisors can mitigate the investor's ongoing disutility, shift the investor's prior expectations, and affect the investor's prior uncertainty. Through this effect on uncertainty, human advisors affect the investor's rate of learning about the algorithm's ability to perform the complex task of portfolio optimization. However, the human advisor cannot affect the "true" return distribution, which is in keeping with the standard institutional features of robo-advising and the empirical setting to which we map our model.

These model features have two possible interpretations. First, different prior beliefs across different types of human advisors could themselves be the result of rational Bayesian updating. More specifically, imagine that the investor starts with an initial (meta-)prior about the algorithm's ability to deliver high expected returns, and assume that a human advisor can convey credible information to explain the true ability of the algorithm. In this case, the starting points of investors advised by advisors with different characteristics can be thought of as investors' updated beliefs conditional on such information. Second, different prior beliefs can also arise in a behavioral model, in which human advisors can use salesmanship and persuasion to move investors' beliefs, even if they convey no hard information (Gennaioli, Shleifer, and Vishny (2015)).

In addition, it is worth highlighting that the value of (algorithmic) advice can be interpreted more broadly. In our formal model, we define it chiefly in terms of the expected returns it generates. In reality, there are many possible dimensions of value-added, for example, helping investors set goals for retirement, or providing behavioral coaching and “financial peace of mind” (Pagliaro and Utkus, 2019; Rossi and Utkus, 2019). Our model can be viewed as incorporating these benefits while converting them into (equivalent) units of expected return. That said, we do provide some evidence that uncertainty about returns is a dimension of algorithm aversion in the setting that we study.

Key outcomes in the model, such as investor retention in the robo-advising service and the fraction of the portfolio delegated to the algorithm, vary with the type of human advice. In theory, variation in these outcomes can therefore identify the effects of the “deep” model parameters that capture the different components of algorithm aversion. In our empirical setting, to accurately estimate the model’s parameters, we require that the advisor type is randomly assigned to the investor in the sense that advisor type assignment is orthogonal to shocks to performance as well as to the investor’s preferences.

The data on robo-advising that we employ come from Personal Advisor Services (PAS), a division of Vanguard. The service it provides is “hybrid,” complementing algorithmic robo-advice with human advisory input.³ In the service we analyze, the algorithm manages the investment portfolio, while the human advisor interacts with investors to help them understand what the algorithm does and advises them on additional auxiliary services such as cash flow, tax, and estate planning. The data provide information on trades, positions, demographic characteristics, and investor-advisor interactions for a large set of previously self-directed investors who participated in the robo-advisor between January 2014 and March 2018.

An ideal experiment for estimating algorithm aversion would randomly assign clients to identical portfolios, managed either by a robo-advisor (“treatment”) or a traditional human advisor (“control”). Our empirical setting contains exogenous variation that approximates this ideal experiment, and we use it to back out the model parameters. More specifically, after eliciting investors’ characteristics (age, income, investment horizon, and preferences), the robo-component of the robo-advisor (the algorithm)

³Hybrid robo-advisors allow investors differing levels of interaction with human financial advisors in addition to the algorithmic solution, as opposed to “pure” robo-advisors which simply provide access to the algorithmic solution.

formulates a financial plan and an associated investment portfolio. The sign-up process then requires the investor to schedule an appointment with a human financial advisor who explains the financial plan, completes onboarding, and provides ongoing support to investors. Critically, the assignment of investors to advisors follows mechanical rules. This assignment is driven by workload balancing imperatives rather than any assessment of advisor type. Once the current client “load” of a given advisor is accounted for, we confirm that there is no empirical relationship between the historical retention rate of clients assigned to a given advisor and the assignment of new clients to the advisor.

This quasi-random assignment allows us to use client retention rates to measure advisor type. We can think of clients randomly assigned human advisors with low client retention as closer to the “treatment” group in the ideal experiment, and clients randomly assigned a high-client-retention advisor as closer to the “control” group. Leveraging this reasoning, we use the treatment effect of random assignment to advisors of different types to draw inferences about the parameters of our structural model of algorithm aversion. We are careful when we do so to avoid any mechanical association by using a “leave-one-out” estimator of advisor type (see, e.g., [Collinson et al. \(2022\)](#)).

In reduced-form estimates, the data reveal that assignment to advisors of different types predicts future client retention. Put differently, assignment to low-retention advisors predicts a greater likelihood of exiting robo-advising than assignment to a high-retention advisor. In addition, we find that human advice is particularly effective during periods of market turbulence when extreme signals of returns hit investors. Indeed, high-retention advisors’ clients are less likely to quit robo-advising during periods of market distress than low-retention advisors’ clients.

These observed differences allow us to identify the different dimensions of algorithm aversion. We derive an estimation equation that maps these reduced-form estimates to the structural parameters of the model. Interpreting our empirical findings through this lens, we derive several further economic insights. First, the presence of a high-retention human advisor removes a large proportion of the uncertainty component of algorithm aversion. Indeed, our estimates are consistent with a model in which a high-retention human advisor removes over 90% of the effect of investors’ prior about the expected returns generated by the algorithm.

We also find that high-retention human advisors reduce clients’ propensity to quit in benign mar-

ket conditions by around 23%. In the model, this estimate can be the result of either the pessimism dimension of algorithm aversion (high-retention human advisors correct pessimistic priors about expected performance), or the disutility dimension (a completely automated service imposes a greater psychological burden on investors). To get a sense of the relative importance of these two effects, we employ an auxiliary prediction of the model, namely, that the effect of advice conditional on market performance should be weaker on experienced robo-advising clients, who have substantially converged on their posterior beliefs about the ability of the automated service. We confirm that this holds in the data, where experienced clients, regardless of their advisor type, react less to market turbulence. This finding is consistent with experienced clients having less to learn about the algorithm’s ability to deliver returns.

This finding leads us to a second economic insight. Having confirmed that experienced clients’ beliefs are close to convergence to the algorithm’s true ability, it is natural to argue that prior pessimism should not be important for these clients’ decisions going forward. In models of learning, the weight that investors place on their prior beliefs converges to zero as they become more experienced. We can therefore extract a cleaner estimate of the disutility component of algorithm aversion by studying the behavior of experienced clients. Repeating our baseline empirical estimation in a sample of experienced clients, we find that high-retention human advisors still reduce the baseline propensity to quit by about 21%. While a more comprehensive structural estimation of model parameters would shed additional light on the decomposition between the disutility and pessimism components of algorithm aversion, our estimates using this strategy suggest that the disutility component is relatively more important.

Overall, our empirical work provides evidence that human experts are complementary to automated services and that both disutility and uncertainty are the most salient dimensions of algorithm aversion in this important empirical setting. While we do map the model’s parameters to our empirical work, in this draft of the paper we do not engage in full-blown structural estimation of the parameters. We intend to pursue this as the next natural step in our agenda and use such structurally-estimated parameters to evaluate counterfactual mechanisms to better understand the most effective approach to use human experts to complement the broader rollout of automated services.

Our work contributes to several strands of the finance and economics literatures. First, we con-

tribute to the literature on labor displacement and reinstatement in the face of automation (see, for example, [Autor \(2015\)](#); [Acemoglu and Restrepo \(2019\)](#)). Second, we contribute to the growing literature about the costs and benefits of dealing with algorithms and the digital revolution (see, for example, [Allcott, Gentzkow, and Song \(2021\)](#); [Niszczoła and Kaszás \(2020\)](#); [Castelo, Bos, and Lehmann \(2019\)](#); [Fuster et al. \(2022\)](#)). Third, we contribute to the growing literature on robo-advising (see, for example, [D’Acunto, Prabhala, and Rossi \(2019\)](#); [Reher and Sun \(2019\)](#); [Rossi and Utkus \(2019, 2020\)](#)). This literature is now beginning to explore the role of trust in robo-advice, linked to the broader literature on the role of trust in financial institutions and stock market participation (see, e.g., [Guiso, Sapienza, and Zingales \(2008\)](#)). Fourth, we contribute to the household finance literature, which seeks ways to both incentivize households to participate in financial markets and do so in an efficient, well-diversified manner (see, e.g., [Campbell \(2006\)](#); [Guiso and Sodini \(2013\)](#); [Gomes, Haliassos, and Ramadorai \(2021\)](#)). Fifth, our work is related to the more general literature on the quality of financial advice (see, e.g., [Linnainmaa et al. \(2018\)](#); [Egan, Matvos, and Seru \(2019\)](#)). In this setting, our focus is primarily on the ability of advisers to assuage investors’ concerns about the algorithmic solution.

The remainder of this paper is organized as follows. Section 1 presents the model and its testable implications. Section 2 describes the institutional details of our empirical setting and the details of our identification strategy. Section 3 contains our main empirical results, and Section 4 concludes.

1 Model

In this section, we describe a general conceptual framework that models the effects of human expertise in an automated service. In the appendix, we show that this framework nests a model of automated or “robo” investment advice as a special case. We then use the application to investment advice to obtain structural estimates of the parameters governing the effects of human expertise.

1.1 Model Environment

We consider clients indexed by $i = 1, \dots, I$ who use an automated service for up to T periods. Clients interact with human experts/advisors indexed by $j = 1, \dots, J$. The automated service is described

by a parameter $\theta \in \mathbb{R}$, which measures its average quality, and which is not observed by clients. This assumption reflects the idea that automated professional services, such as robo-advice, can be somewhat opaque and difficult for a layperson to assess.

Client i enrolls in the service at an initial time 0, and matches randomly with one human expert j . She quits at an endogenously chosen time δ_i , where $\delta_i \in \{1, \dots, T\}$ denotes the length of her enrolment in discrete time. We assume that clients cannot re-enrol after quitting, and cannot switch between human experts at any time.

The assigned human expert affects both the client's utility from consuming the service and the client's beliefs about the quality parameter θ . Concretely, we assume client i , upon matching with advisor j , enjoys utility $u^j(\cdot)$, which we define rigorously below. The client believes that $\theta \sim N\left(m_0^j, \frac{1}{\tau_0^j}\right)$, where m_0^j denotes expected quality, and τ_0^j denotes the precision/inverse variance of beliefs about quality. The goal of this paper is to obtain structural estimates of the parameters $\{u^j, m_0^j, \tau_0^j\}$, which govern the value-added of human expertise for different types of experts.

After forming prior beliefs about quality θ , clients can learn from the service's realized output/performance. At each date $t \in \{1, \dots, \delta_i\}$ at which client i is enrolled, she observes a measure of output

$$y_t^i = \theta + u_t^i, \tag{1}$$

where $u_t^i \sim N\left(0, \frac{1}{\tau_y}\right)$ is a random shock with precision/inverse variance τ_y . In the appendix, we show that y_t^i can be micro-founded in our setting as the excess return on the investment portfolio managed by a robo-advisor, relative to the return on the client's portfolio outside robo-advice.

At date t , the client updates her beliefs about θ by Bayes' rule, using the history of output $\{y_1^i, \dots, y_t^i\}$. After updating at date t , the client believes that $\theta \sim N\left(m_t^{i,j}, \frac{1}{\tau_t^j}\right)$. The precision of τ_t^j beliefs after observing t outputs is simply $\tau_t^j = \tau_0^j + t\tau_y$. The client's expectation $m_t^{i,j}$ satisfies the Kalman filtering equation:

$$m_t^{i,j} = m_{t-1}^{i,j} + \frac{\tau_y}{\tau_{t-1}^j + \tau_y} \left(y_t^i - m_{t-1}^{i,j}\right). \tag{2}$$

Notice that expectations depend both on the identity of the client i and that of her assigned expert j . This is because different clients may experience different output, and different advisors may induce

different prior beliefs that also lead to different posteriors.

We write clients' preferences recursively. We let $V_t^j(m)$ denote the continuation value of a client who is still enrolled at date t and has formed a current expectation $m_t^{i,j} = m$ of service quality. Continuation values satisfy the Bellman equation

$$V_t^j(m) = \max \left\{ u^j(m) + \hat{\mathbb{E}}^j [V_{t+1}(m') | m], 0 \right\}. \quad (3)$$

The first term in this expression, $u^j(m)$, stands for the expected flow of utility for a client who matches with expert j , has formed current expectations m , and uses the service for one period. We show in the appendix that, in the context of automated portfolio choice, this utility function maps one-to-one to the Sharpe ratio of the robo-advisor's investment portfolio, as long as the client's utility is logarithmic in final wealth. The second term, $\hat{\mathbb{E}}^j [V_{t+1}(m') | m]$, measures the client's expected continuation value if she remains enrolled until $t + 1$, where m' denotes next period's expectation. The notation $\hat{\mathbb{E}}^j [\cdot]$ is used to emphasize that the expectation is computed using the beliefs induced by expert j .

When deciding whether to quit at date t , the client compares the sum of flow utility and expected continuation values to the utility she can obtain outside the service, which we normalize to zero. Therefore, conditional on being enrolled at t and having formed expectations m , it is optimal to quit with $\delta_i = t$ if and only if

$$u^j(m) + \hat{\mathbb{E}}^j [V_{t+1}(m') | m] \leq 0. \quad (4)$$

Accordingly, the continuation value defined in Equation (3) is equal to the maximum of the left-hand side of Equation (4) and zero. Finally, after the terminal date T , clients do not use the service by assumption, so that the continuation value at this point becomes $V_T^j(m) \equiv 0$.

For our structural estimates, we introduce an implementation error which generates cross-sectional heterogeneity between clients. We assume that clients correctly calculate their continuation values and expectations m , but then act as if their expectations are $m - \xi_t^i$, where $\xi_t^i \sim N\left(0, \frac{1}{\tau_\xi}\right)$ independently across time and individuals, and where ξ_t^i is uncorrelated with all performance shocks u_t^i . In this specification, clients quit with $\delta_i = t$ if and only if

$$u^j(m - \xi_t^i) + \hat{\mathbb{E}}^j [V_{t+1}(m') | m] \leq 0. \quad (5)$$

We introduce the implementation error ξ_t^i because it allows us to generate cross-sectional variation in quit rates conditional on performance. We find it convenient to measure ξ_t^i in the same units as beliefs, and to think of it as a negative shock (i.e., a high ξ_t^i makes the client more likely to quit), but neither of these choices is essential for our results.

This model yields particularly clear estimation equations in the three-period case where $T = 2$. Having enrolled at time 0, and knowing that she will not continue using the service at time 2, the client's only decision is whether to quit at date 1. Substituting $V_2^j \equiv 0$ in Equation (5), we find that the client quits at date 1 under the following condition:

$$w^j \left(m_1^{i,j} - \xi_1^i \right) \leq 0 \quad (6)$$

Let $\phi^j \equiv (w^j)^{-1}(0)$ denote the critical value for expected quality that makes the client indifferent between quitting and continuation at date 1. Intuitively, ϕ^j is the amount by which expected output needs to exceed the client's outside option to make continued enrolment worthwhile. This parameter can be interpreted as a fixed cost/disutility of participation that clients perceive when matched with expert j .

Substituting for $m_1^{i,j}$ from Equation (2), we can now rearrange Equation (6) to obtain the equivalent condition for optimal quitting at date 1:

$$\underbrace{\phi^j - \frac{\tau_0^j}{\tau_0^j + \tau_y} m_0^j}_{\text{Baseline quit rate}} - \underbrace{\frac{\tau_y}{\tau_0^j + \tau_y} y_1^i + \xi_1^i}_{\text{Sensitivity}} \geq 0 \quad (7)$$

The first two terms on the left-hand side of (7) capture the client's propensity to quit for a given shock ξ_t^i . The first term can be interpreted as a client's baseline propensity to quit when matched with advisor j . This propensity is high when the fixed cost parameter ϕ^j is large or when the prior expectation m_0^j is small (we later discuss how to separately identify these two subcomponents). Moreover, the weight on prior expectations is large when the precision τ_0^j of prior beliefs is large relative to the precision τ_y of the signal contained in realized performance. Intuitively, Bayesian clients put a lot of weight on their prior beliefs if they are either very confident in their prior, or if performance is a noisy signal.

The second term in (7) measures the client’s response to realized performance y_1^i . The sensitivity to performance is one minus the weight placed on prior beliefs. Therefore, as is standard in models with learning, the client reacts more strongly to performance when she is either uncertain a priori or if performance is a precise signal.

1.2 Empirical Estimation

In principle, one can estimate the baseline quit rate and sensitivity components in Equation (7) using a standard econometric discrete choice model, such as a linear probability model if the distribution of ξ_1^i is uniform, or a logistic regression model if it is a logistic/extreme value distribution. In such an econometric model, a dummy variable for quits would be the dependent variable, the baseline quit rate would appear as the constant term, and the sensitivity would appear as the coefficient on recent performance as an explanatory variable.

A challenge in this approach is that these coefficients in equation (7) depend on the identity j of the human expert that clients are assigned to. In principle, one could estimate the equation separately for each human expert in the data, but that approach does not yield reasonable results unless every advisor has a long history of clients, which is not the case in our data.

We therefore pursue the alternative strategy of grouping human experts into two types, which we label as high retention (H) and low retention (L) types. Empirically, as we describe in detail in the next section, we infer these types from each expert’s historical performance, which we measure as the fraction of clients they are able to retain. We write $\{\phi^H, m_0^H, \tau_0^H\}$ for the preference and belief parameters induced by a high type, and $\{\phi^L, m_0^L, \tau_0^L\}$ for a low type. We then infer these parameters by estimating Equation (7) separately for clients advised by high and low type experts in our data or, equivalently, by running regressions that have expert type, performance, and the interaction between the expert type and performance as explanatory variables.

This empirical approach requires us to make several assumptions/caveats. First, we must assume that assignment of clients i to human experts j in the data is independent of shocks $\{\xi_t^i, u_t^i\}$. For example, if clients with favorable performance or preference shocks were systematically assigned to high type experts, we would estimate spuriously low baseline quit rates for that type. Empirically, it

therefore becomes important to check that assignment to different experts is not systematically related to subsequent performance, nor to clients' other observable characteristics.

Second, since we use the same data to infer advisors' types and estimate the parameters associated with each type, we must be wary of mechanical relationships. Such relationships can arise, for example, if one uses the fact that client i never quit the service as an indicator that her advisor j is a high type, and then ran a regression of quit dummies on advisor types in a sample containing i . To avoid this issue, we use leave-one-out estimators throughout our analysis below, which we describe in detail after outlining the empirical setting.

2 Institutional Setting and Data

We use data from a large US-based hybrid robo-advisor called Personal Advisor Services (PAS), a division of Vanguard, to estimate algorithmic aversion. The service is "hybrid" as it complements algorithmic robo-advice with human advisory input. More broadly, robo-advisors are commonly classified as either pure or hybrid robo-advisors. Pure robo-advisors do not feature any substantive interactions between investors and human financial advisors, whereas hybrid robo-advisors allow investors differing levels of interaction with human financial advisors. In our setting, the algorithm manages the investment portfolio, while the human advisor interacts with investors to help them understand what the algorithm does and advises them on additional auxiliary services, such as opening IRA accounts and estate planning.

2.1 Characteristics of Robo-advised Investors

Our data contains information on trades, positions, demographic characteristics, and investor-human advisor mappings for previously self-directed investors who signed up for the robo-advisor service. The trade data include all trades placed by the robo-advisor over the period January 2014 through March 2018. The position data contain associated monthly holdings observations, i.e., the data track both investments and trades. The data also track the dates on which investors initiated, enrolled, implemented, and quit the advice service. The demographic characteristics include investor age, gender, and tenure with the asset manager. The investor-advisor mapping data track the dates,

times, length, and the initiator (advisor or investor) of all advisor-investor interactions, including meetings and phone calls.

Table 1 reports summary statistics (mean, standard deviation, and percentiles of the distribution) of different variables characterizing robo-advised investors in the sample recorded 12 months after investors signed up for the service. Panel A focuses on investors' demographic characteristics. The average investor is 64 years old, 60% of the users are male, and the average investor has had an account with the asset manager for almost 14 years. The robo-advised population comprises older, wealthier investors who are more gender-balanced than datasets commonly used to study trading behavior. In contrast with the demographics in these data, the average investor age is 51 in the brokerage trading data employed by both [Gargano and Rossi \(2017\)](#) and [Barber and Odean \(2001\)](#), and women constitute lower fractions in those datasets, with 27% in [Gargano and Rossi \(2017\)](#) and 21% in [Barber and Odean \(2001\)](#).

Table 1 Panel B shows details of investors' portfolio allocations. The average investor has \$758,378 invested with the robo-advisor, which is substantial and more than 50% larger than the median (\$478,929), reflecting significant right-skewness. There are 8 assets on average in each account, comprising mutual funds, stocks, bonds, and ETFs, with the bulk of the investors being invested in only six assets and only 25% of the investors having more than 9 assets in their portfolio. This reflects the different "glide paths" to which investors are assigned by the algorithm; we discuss this in greater detail below. Almost all of the investors' wealth (97.4%) is invested in Vanguard-affiliated products (mostly indexed mutual funds).

Table 1 Panel C shows that 95% of investors' wealth is in mutual funds, followed by cash (2%), individual stocks (1.4%) and ETFs (0.8%). Only a negligible number of investors have direct exposure to corporate bonds and options (not reported).

Given that the majority of investors' wealth is in mutual funds, in Panel D, we analyze the characteristics of the mutual funds held. The first row shows that 83% of mutual fund holdings are in indexed mutual funds. The average management fee across all mutual fund holdings is 7 bp. The expense ratios are also low, on average 9 bp with a median of 8 bp. Finally, the average turnover ratio of the mutual funds in the portfolio is 27%.

2.2 Robo-advisor Characteristics and Returns of Advised Clients

Over the sample period, the robo-advisor classifies investors into five risk glide paths based on their financial objectives, investment horizons, and demographic characteristics. Although confidentiality prohibits disclosing the details of the algorithm that generates the investors' portfolio allocations, we report facts that help to explain broad features of the allocations and the performance of the robo-advised portfolios.

We compute average annualized monthly portfolio returns for all PAS investors. PAS investors can be categorized into two tiers based on their investment commitment. Higher-tier investors have a dedicated investment advisor, while lower-tier investors receive support from a pool of human advisors. (Section 2.4 describes the (quasi-random) process of assignment of the dedicated human advisors to higher-tier clients; lower-tier investors receive support based on scheduling/availability from the pool). For the purposes of these summary statistics, we co-mingle the two groups because their investment performance is virtually identical. For comparison, we also compute the average returns for non-robo-advised clients in Vanguard who interacted with PAS at some point. This includes the returns of clients who later signed up for PAS and those who did not subsequently sign up after considering the service.

We show the cross-sectional distributions of both sets of average returns in Figure 1(a), with the average returns of robo-advised investors in blue and those of self-directed investors in red. The plot shows that robo-advised investors have higher mean returns than non-robo-advised investors, i.e., the blue distribution is shifted to the right relative to the red distribution.⁴

There is considerable cross-sectional dispersion in both distributions and no visible reduction in the variance of the average returns of robo-advised investors, which range from 0% to 20% per year. However, the substantially greater mean creates a clear increase in Sharpe ratios for robo-advised investors. In terms of our theoretical model in the previous section, this increase corresponds to the case where the true expected return θ achieved by the algorithm, defined as an excess return above clients' outside option, is above zero, which is an important assumption in our structural analysis.⁵

⁴The red distribution exhibits prominent bunching at zero as some self-directed investors have the entirety of their portfolio in cash.

⁵Our model is technically well-defined even in the case where true quality is $\theta < 0$, and investors may choose to participate in the service if their prior expectations about θ are (wrongly) optimistic. However, this case also implies

To better understand the sources of variation in the returns of robo-advised investors, we conduct the following three-step procedure. In the first step, following [Balasubramaniam et al. \(2022\)](#), we compute principal components (PCs) of the equity share of 1,700 investors for which we have 36 continuous months of robo-advised returns. The first PC explains 46% of the variation in equity shares, and the first five PCs explain 81% of this variation. These results are the first indication of a pronounced factor structure in the robo-advisor portfolio allocations, arising from the small number of glide-paths to which investors are assigned. In the second step, we regress the time series of the equity share for each of the 55,000 investors in our data on the first five PCs estimated in the first step. Finally, in the third step, we cluster investors' loadings in five groups using a k -nearest neighbors estimation procedure. We find that 82.37% of the investors are classified in the first group, followed by the second and third groups with 13.51% and 3.37% of the investors, respectively. Finally, the last two groups contain less than 1% of the advised investors. All in all, these results suggest that, while 5 risk glide paths exist, the majority of investors end up on two glide paths over our sample period.

We complement this simple analysis of equity shares by regressing monthly investor total portfolio returns on the market portfolio, investors' equity share, and the interaction of these two variables. This regression has an R^2 of over 75%, highlighting that the equity-bond allocation decision and the variation in aggregate equity returns together explain the major share of variation in returns seen in the data. We report realized monthly returns against the predicted returns from this regression in a binned scatterplot, Figure 1(b), which shows that the two quantities line up very closely along the 45° line.

Given the robo-advising glide paths, the cross-sectional variation seen in the equity share and returns is likely an outcome of investors' preferences and demographic characteristics, which the algorithm translates into different portfolio allocations. While we do not have information regarding investors' risk preferences, we do have investors' age. We, therefore, regress monthly client returns on the market portfolio, investors' age, and their interaction. This regression has an R^2 of over 71%, which is very close to the R^2 obtained using equity shares. The corresponding binned scatterplot of realized vs. fitted returns reported in Figure 1(c) suggests that any risk preference variation over and

that, as investors obtain more information about the true θ , 100% of investors will eventually quit robo-advice. For this reason, the case with $\theta > 0$ is more interesting, and indeed appropriate for the empirical setting in which a large proportion of clients do not quit.

above that which is correlated with age likely plays a minor role in determining the portfolio allocation implemented by the algorithm.

Figure 1(d) provides a different perspective. The black line shows the distribution of cross-sectionally demeaned average returns on all robo-advised portfolios. The red and blue solid lines show the residuals from regressions on the market and equity share, and the market and age, respectively. The substantial reduction in the variance of the distribution of these residuals shows that equity share and age explain a majority of the variation in average portfolio returns, which means that performance is in large part homogenous for investors with similar demographic characteristics.

Figure 1(d) also shows dashed red and dashed blue distributions, which include advisor fixed effects into the regressions of performance on equity share and age, respectively. If advisors were instrumental in affecting the portfolio allocations recommended by the algorithm, we would expect these two distributions to have a smaller variance than the corresponding solid distributions of residuals from models that do not include advisor fixed effects. Instead, we find that the dashed blue and red distributions are virtually identical to the respective solid distributions, suggesting that human advisors play virtually no role in affecting investors' portfolio allocation. In the remainder of the paper, we use human advisor-client interactions to assess the degree to which human advisors are complementary to the automated portfolio strategy. Along with the random assignment of human advisors to clients, Figure 1(d) provides evidence that this complementarity in the data arises from sources other than human advisors directly affecting portfolio strategy, aiding parameter estimation from the structural model.

2.3 Measuring Advisor Type

We seek to understand how humans are complementary to automated solutions in this setting. Measuring human financial advisor type is challenging for a number of reasons. First, success in this setting is likely the combination of many different but complementary traits, including client-specific assessments of advisor communication and relationship-management skills. Second, the skills needed to be a successful “plain vanilla” financial advisor may be different from those needed as a hybrid robo-financial advisor, so it is difficult to rely on “standard” personality assessments or financial com-

petence scores in this novel setting. Third, an important determinant of customer satisfaction in the financial domain is portfolio performance, which is controlled by the robo-advising algorithm rather than human advisors, so using performance-related measures to measure advisor type—a common strategy in financial economics—is not appropriate in this setting.

We, therefore, rely on revealed preference and measure advisor type using advisor-client-specific average retention rates. More specifically, we construct a measure of advisor retention for each investor using a leave-out estimator. For each client, we compute the average client retention of the human advisor they are assigned to using all clients assigned to the advisor except for the client in question. That is, the retention-rate of advisor j applied to investor i is the ratio between all clients of the advisor who remain with robo-advice over the full sample, and the total number of clients assigned to advisor j , excluding investor i . In our main specifications, in case investor i quits, we also exclude the period in which they quit, to eliminate any potential bias arising from contemporaneous correlation in quits across investors assigned to the same advisor. We then discretize the retention measure, categorizing advisors as type-1 (high retention) or type-0 (low retention) using the median retention rate as the breakpoint.

In our empirical work in this draft, we use advisor-specific client retention to measure advisors' type, without taking a stance on specific skills that increase retention. In the model, we map type-1 advisors to the three dimensions defined in the previous section: fixed-cost/disutility associated with ongoing participation, prior expectations of performance (which we term *pessimism*), or *uncertainty* about the automated solution (which appears in sensitivity to performance in equation (7)).

2.4 Assignment of New Investors to Advisors

During the sample period, the robo-advisor assignment and onboarding process began by eliciting investors' characteristics (age, income, wealth, investment horizon, and preferences). Using this information, the robo-component of the robo-advisor (i.e., the algorithm) formulated a financial plan and an associated investment portfolio. The sign-up process then required the investor to schedule a time to meet with a human financial advisor who explained the financial plan and completed onboarding.

The assignment of clients to advisors was quasi-random. Advisors were assigned based on the

match between a prospective client’s availability for an initial onboarding call and the availability of advisors in the system. Advisor availability was driven by the need to balance workload across advisors (i.e., driven by advisor capacity) rather than any assessment of advisor type (i.e., advisor skills, characteristics, or ability to retain clients). Put differently, the assignment of new clients to advisors during the sample period was random with respect to advisor type, depending on advisors’ capacity and availability but not their skills or characteristics.

To verify random assignment, Table 2 shows the main attributes of investors assigned to high- and low-attrition advisors the month before they sign up for PAS. The rows are separated into different sets of covariates: 1) demographic characteristics of the investors, such as age, gender, and the tenure of investors with Vanguard before signing up for the robo-advisory service; 2) portfolio-related characteristics, such as the total assets at Vanguard, the number of assets, and the percentage of Vanguard products (mutual funds and ETFs) in investors’ portfolios; 3) investors’ asset allocation characteristics, such as the percentage of assets in mutual funds, cash, ETFs, individual stocks and bonds at the time of sign up; 4) characteristics of the mutual funds held by the investors, such as degree of indexation, fees charged, expense ratios and turnover ratios; and 5) investment performance measures, such as the average annualized monthly returns realized by the investors before signing up for PAS as well as the month-by-month cross-sectionally demeaned version that controls for time-variation in returns.

Across all characteristics, the covariates are balanced in that investors assigned to high- and low-attrition advisors are similar in their portfolio sizes, portfolio allocation across asset classes, the types of mutual funds they invest in, and their returns before signing up for the service. The only covariates where we detect statistical differences are age, gender dummies, and tenure at Vanguard before signing up. For example, the average age of those assigned to high-retention advisors is 64.5, while it equals 65.6 for those assigned to low-retention advisors. In all cases, however, even when the differences are statistically significant, they are economically small.

Figure 2 shows the relationship between advisor load, advisor leave-out retention, and the assignment of clients to advisors. For each advisor, we compute the number of clients they advise at the beginning of each month and sort advisors into quintiles based on their current workload. Figure 2 (a)

shows the net increase in the number of clients allocated to advisors in each group on average every month, computed as the number of investors allocated to each advisor minus the number of investors lost by each advisor every month due to attrition, together with 95% confidence intervals. The plot shows that there is a ramp up in the rate of net additions from the first up to the third quintile of current capacity and a decline from then to the fifth quintile as advisors reach full capacity. The top group of advisors that are close to full capacity has monthly net average growth that is not statistically different from zero. Figure 2 (b) repeats the analysis sorted using deciles of current capacity instead of quintiles, with qualitatively similar results.

To estimate the parameters of the model, we need the assignment of clients to advisors to be independent of performance and preference shocks. Table 2 demonstrates that client characteristics are balanced across advisor types. Figures 2 (c) and (d) plot client assignment rates to advisors of different types. These figures show gross additions to advisors (i.e., ignoring attrition) and split advisors into high- and low-retention, conditional on current capacity ((c) splits into quintiles of current capacity and (d) deciles). Except for the very first group in both plots, where we find an economically small difference in the assignment rates to high-retention advisors (in blue) and low-retention advisors (in red), all other capacity groups show no differential in rates of assignment to advisors with different retention rates.⁶

3 Empirical Estimates

The random assignment of advisors to clients allows us to estimate the causal effect of human advisors of different types on retaining investors in the robo-advice service.

3.1 Cross-sectional Variation in Advisor Retention

Figure 3 (a) shows a histogram of the (leave-one-out estimated) advisor retention measures, scaled in such a way that the highest leave-out retention estimates in the data are set to 100. The figure shows substantial cross-advisor dispersion in the retention measure. Super-imposed on the histogram in red is a non-parametric (lowess) estimate of the retention rate of clients who are randomly assigned to

⁶We later eliminate the lowest current capacity group in our results to check robustness. We note that with ten deciles, finding one statistically significant difference is consistent with a 10% rate of significance.

advisors of different quality, measured by their scaled leave-one-out retention rates. The plot shows a clear near-linear relationship between the retention rate of clients that are randomly assigned to advisors and the quality of these advisors.⁷ One possible concern with Figure 3 (a) is that quality may be mismeasured because of outliers caused by small base effects—i.e., advisors who never got to manage a large set of clients. Figure 3 (b) therefore removes advisors in the bottom decile of clients under management throughout the sample period; and shows that the picture is virtually unchanged.

3.2 Non-Parametric Survival Estimates

We next estimate Kaplan-Meier survival functions for clients assigned to type-1 and type-0 advisors:

$$\widehat{S}(t) = \prod_{i:t_i \leq t} \left(1 - \frac{d_i}{n_i}\right), \quad (8)$$

where t_i is a time when at least one investor quits robo-advice, d_i is the number of clients quitting robo-advice at time t , and n_i is the number of individuals that stayed with robo-advice (i.e., neither quit nor censored) up to time t_i .

Figure 4 (a) shows the Kaplan-Meier survival functions computed using all advisors in the data. The blue lines (and associated 95% confidence intervals) show survival rates for clients that are assigned to type-1 advisors, and the red lines show survival rates for investors assigned to type-0 advisors. The figure shows that 96.2% of investors in the type-1 advisor group and 95.5% of investors in the type-0 advisor group stay with robo-advising at the one-year mark. These survival estimates diverge further for the two groups over time—at the three-year mark, 90.6% of the investors assigned to type-1 advisors are still with robo-advice. The corresponding value for type-0 advisors is 86.8%.

Figure 2 shows that high- and low-retention advisors are assigned very similar numbers of clients, depending on their current load. The only exception is the bottom decile—the advisors with the lightest load—where high-retention advisors are assigned more investors: 1.5 on average per month versus 0.8. To ensure these advisors do not drive the results Figure 4 (b) reports survival estimates that exclude this bottom decile; the results are virtually identical.

As an alternative, we estimate a Cox proportional hazard model for attrition from the robo-advising

⁷This figure is similar to Figure 5 in [Collinson et al. \(2022\)](#) who use a judge stringency instrument.

service. Figure 4 (c) first plots the smoothed hazards to check if we satisfy the assumptions required for the Cox model, and shows that they are parallel, with greater proportional hazards for investors assigned to type-0 advisors. For this group of investors, the hazard estimates (red line and confidence interval) increase and peak at 6% at the one-year mark, only to decline and reach their bottom of 2.8% at the three-year mark. The blue line (and confidence interval) shows hazard estimates for those assigned to type-1 advisors, which follow similar dynamics at lower levels: the hazard estimates peak at 4.2% at the one-year mark, declining and bottoming out at 1.9% in year three. The results in Figure 4 (d) that exclude the advisors in the bottom decile of capacity deliver virtually identical results.

We then estimate a Cox proportional-hazard model of the following form on the full set of advisors:

$$h(t|x_j) = h_0(t) \cdot \exp(x_j\beta),$$

where we estimate the model using a “type-1” dummy as the only regressor. The coefficient estimate $\beta = -0.293$, implying that the ratio of the hazard between those clients assigned to type-0 and type-1 advisors is $\exp(-0.293) = 0.746$. In other words, those assigned to type-1 human advisors have a 25.4% lower hazard than those assigned to type-0 human advisors.

We also use advisor retention as a continuous variable in a robustness exercise. Using this specification, we obtain a highly statistically significant coefficient $\beta = -0.045$. Economically, the estimate implies that the hazard ratio between those with a higher-retention advisor (a 1% higher retention rate) and a lower-retention advisor is $\exp(-0.045) = 0.956$. These estimates confirm the causal impact of human advisor type on investor participation in robo-advising.

3.3 Estimating the Effect of Human Advice Across Market Conditions

In the model, advisor quality can affect the rate at which clients learn about the automated service from realizations of performance. To check for this channel in the data, we evaluate how client quit rates vary with market conditions and advisor type. To do so, we estimate regressions of the following

form at the monthly frequency:

$$\begin{aligned}
\textit{Dummy_quit}_{i,t} = & \alpha + \beta I_{\{MKT_RET_{t-1} < 0\}} + \gamma I_{\{Type1_Advisor_i = 1\}} \\
& + \delta I_{\{MKT_RET_{t-1} < 0\}} \times I_{\{Type1_Advisor_i = 1\}} + \epsilon_{i,t},
\end{aligned} \tag{9}$$

where $\textit{Dummy_quit}_{i,t}$ is equal to 1 if investor i quits robo-advising in month t and 0 otherwise, $I_{\{MKT_RET_{t-1} < 0\}}$ is an indicator variable equal to 1 if the market return is negative in month $t - 1$ and 0 otherwise, and $I_{\{Type1_Advisor_i = 1\}}$ is equal to 1 if investor i is assigned to an advisor with retention above the median and 0 otherwise.

In equation (9), α measures investors' unconditional monthly quit probability, β estimates the conditional increase in attrition from robo-advice in periods in which the market has performed poorly, γ measures the differential quit rate associated with clients assigned to type-1 advisors, and δ measures the extent to which the differential quit rate across clients assigned to type-1 and type-0 advisors varies with market conditions.

Table 3 reports estimates of this equation, where all coefficient estimates are multiplied by 100 to express magnitudes in percentage points. In the first column, returns on the CRSP value-weighted index are used to capture variation in client portfolio performance. The α coefficient in this column reveals an unconditional 0.369% quit probability per month, roughly translating to an annual attrition rate of $0.369\% \times 12 = 4.43\%$, in line with the attrition rate in Figure 4. The β coefficient shows that attrition from robo-advising increases by 0.136 percentage points in poor market conditions, i.e., when the CRSP value-weighted index return is lower than zero. This increase is large, translating to a $0.136/0.369 = 37\%$ increase in attrition in such times. The γ coefficient shows that being assigned to type-1 advisors reduces attrition by 0.86% per month, confirming the results from our survival analysis. Finally, δ is negative and significant, showing that clients randomly assigned to type-1 advisors have lower attrition than those assigned to type-0 advisors during down markets.

The results in the first column of Table 3 use the returns on the CRSP value-weighted index to capture variation in client portfolio performance. In the second column, instead of using the CRSP returns, we directly use the returns on the portfolio of each client in the previous month and separate these returns into periods in which they are above and below zero. Overall, the results are very similar

to this change.

In Figure 5, we use the regression estimates from Equation (9) to plot attrition across different market conditions and advisor types. Investors assigned to type-1 advisors have an attrition rate of slightly below 0.3% per month; this does not vary greatly with movements in the market. Investors assigned to type-0 advisors, on the other hand, have an attrition rate of 0.37% per month when the market return is positive and a probability of attrition of 0.51% when it is negative. These results are consistent with type-1 advisors reducing attrition both conditionally and unconditionally; type-1 advisors are particularly important when the market is down.

3.3.1 Using Past Volatility as Conditioning Information

We also check whether advisor type is important during periods of market stress, re-estimating equation 9 replacing periods of negative market returns with periods of high volatility, where the latter are defined as periods where return volatility exceeds median volatility computed over the full sample. The results from this exercise are reported in the last two columns of Table 3. In column (3) of this table, monthly volatility is computed using daily returns on the CRSP value-weighted index. In column (4), investor volatility is computed using the realized daily returns of each individual client portfolio. Overall, the results are economically and statistically in line with those in the first two columns of Table 3. In all cases, we find that type-1 advisors reduce client attrition from robo-advising unconditionally as well as during periods of high return volatility.

3.4 Structural Implications of Empirical Estimates

Our reduced-form results in the previous section provide empirical estimates of the discrete choice model implied by our theoretical analysis. Recall that Equation (7) describes the baseline quit rates during the first period of enrolment, and the sensitivity to performance during this period, for different types of human advisor.

The constants and coefficients on `Type1_Advisor` in the first column of Table 3 imply the following structural estimates of annual baseline quit rates for high- and low-type advisors in Equation (7):

$$\begin{aligned}\phi^L - \frac{\widehat{\tau_0^L}}{\tau_0^L + \tau_y} m_0^L &= 0.369 * 12 = 4.43\% \\ \phi^H - \frac{\widehat{\tau_0^H}}{\tau_0^H + \tau_y} m_0^H &= (0.369 - 0.086) * 12 = 3.40\%\end{aligned}\tag{10}$$

These estimates imply that a high-type advisor, relative to a low-type one, reduces baseline quit rates by $1 - 3.4/4.43 = 23.25\%$. Economically, this can be the result of either differences in the fixed cost/utility dimension of human expertise (ϕ^L vs. ϕ^H), or in differences in prior expectations brought about by different types of human advisor (m_0^L vs. m_0^H).

The coefficients on Bad Market and the interaction term in Table 3 imply the following structural estimates of sensitivity of annual quit rates to performance:

$$\begin{aligned}\frac{\widehat{\tau_y}}{\tau_0^L + \tau_y} &= 0.136 * 12 = 1.63\% \\ \frac{\widehat{\tau_y}}{\tau_0^H + \tau_y} &= (0.136 - 0.123) * 12 = 0.16\%\end{aligned}\tag{11}$$

These estimates imply that a high-type advisor, relative to a low type, reduces the sensitivity to performance by $1 - 0.16/1.63 = 90.18\%$. Notice also that we obtain very similar orders of magnitude when we use clients' own portfolio returns instead of market returns as the measure of performance, as in the second column of Table 3. After a simple transformation,⁸ these estimates also imply that a high-type human advisor reduces the variance in clients' prior beliefs about service quality by around 90.3% relative to a low type.

In order to provide further insights, Table 4 reports results from an auxiliary empirical exercise. We take the universe of investors and split them into low and high experience groups, where experience is computed as the number of months investors have been enrolled in robo-advice. We then report the same estimates as in Table 3 for both groups. An intriguing result is that more experienced/long

⁸Let $g^j = \frac{\tau_y}{\tau_0^j + \tau_y}$ denote the model counterpart of the estimated sensitivities in Equation (11). We can rearrange this expression to get

$$\frac{1 - g_L}{g_L} \frac{g_H}{1 - g_H} = \frac{1/\tau_0^H}{1/\tau_0^L},$$

which is the ratio of prior variances induced by a high- versus low-type advisor in our model. This ratio of variances given our estimates is $60.35/624 = 0.097$, implying a 90.3% reduction.

tenure investors are much less sensitive to recent performance. Indeed, the coefficient on “Bad Market” for these investors is not statistically different from zero. In terms of our model, this result can be interpreted as saying that beliefs of experienced investors have already “converged” to their steady state values. Convergence further implies that the weight placed on prior beliefs by experienced investors is close to zero. Therefore, we can use the constants and coefficients on Type1_Advisor in the second column of Table 4 to obtain the following structural estimates of baseline quit rates for experienced investors:

$$\begin{aligned}
\phi^L - \underbrace{\frac{\widehat{\tau}_0^L}{\widehat{\tau}_0^L + \tau_y} m_0^L}_{\simeq 0 \text{ (convergence)}} &\simeq \widehat{\phi}^L = 0.417 * 12 = 5.00\% \\
\phi^H - \underbrace{\frac{\widehat{\tau}_0^H}{\widehat{\tau}_0^H + \tau_y} m_0^H}_{\simeq 0 \text{ (convergence)}} &\simeq \widehat{\phi}^H = (0.417 - 0.089) * 12 = 3.94\%
\end{aligned} \tag{12}$$

When interpreted in this way, the estimates in Table 4 imply that a high-type human advisor reduces the fixed cost/disutility of her clients by an average of $1 - 3.94/5 = 21.2\%$. These results indicate that most of the 23% reduction in baseline quit rates, which we reported in Equation (10), is due to the fixed cost/disutility channel, as opposed to prior expectations.

4 Conclusion

We study the extent to which human experts are complementary to technological innovation using a simple model which we structurally map to evidence from a large hybrid robo-advising service. In the model, investors experience a per-period fixed cost or disutility arising from their use or interaction with the technology, as well as a “learning” channel in which they refine their understanding of the performance of the technology over time and across states of the economy. In a unique dataset from a large US hybrid robo-advisor, we leverage random assignment of human advisors to clients and clients subsequent retention rates in the service, and map these patterns in the data back to the parameters of the structural model.

The robo-advisor we study automatically manages the investment portfolio using a set of codified rules, while the human advisor interacts with investors to help them understand what the algorithm does, as well as providing auxiliary advice on issues such as estate planning. A key feature of this setting is that the assignment of investors to advisors follows mechanical rules driven by workload balancing imperatives rather than any assessment of advisor type. This means that once the current “load” of a given advisor is accounted for, the assignment of new clients to this advisor is orthogonal to the historical client retention of the advisor (a useful proxy for advisor type).

This random assignment of clients to different types of advisors allows us to cleanly map our empirical estimates to the parameters of the model. We find that this measure of advisor type predicts the future retention rate of clients that are assigned to them. We also find that high-retention advisors’ clients are less likely to quit robo-advising during periods of market turbulence than lower-retention advisors’ clients. Finally, we find that experienced clients, regardless of their advisor type, react less to market turbulence. These facts, when mapped back to the model, deliver the insight that humans are complementary to automated services in two main ways. For one, the estimates imply that high-quality human advisors help to reduce the variance in clients’ prior beliefs about service quality, facilitating learning about the algorithm’s ability to deliver returns. Second, the behavior of the attrition rates of experienced clients shows that human advisors can also significantly attenuate ongoing disutility from the automated portfolio management solution.

Our results make the notion of algorithmic aversion more transparent, separating this puzzling phenomenon into both learning and disutility components. In future drafts, we intend to rigorously structurally estimate the parameters of the model and use these parameters to evaluate counterfactuals to better understand how human advice can help to complement automation and the use of algorithms to facilitate the broader scaling of customized solutions in household finance and other domains.

References

- Acemoglu, D., and P. Restrepo. 2019. Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives* 33:3–30.
- Allcott, H., M. Gentzkow, and L. Song. 2021. Digital addiction. Working Paper, National Bureau of Economic Research.
- Autor, D. H. 2015. Why are there still so many jobs? the history and future of workplace automation. *Journal of economic perspectives* 29:3–30.
- Balasubramaniam, V., J. Y. Campbell, T. Ramadorai, and B. Ranish. 2022. Who owns what? a factor model for direct stock holding. *Journal of Finance* .
- Barber, B. M., and T. Odean. 2001. Boys will be boys: Gender, overconfidence, and common stock investment. *Quarterly Journal of Economics* 116:261–92.
- Brennan, M. J., and Y. Xia. 2001. Stock price volatility and equity premium. *Journal of monetary Economics* 47:249–83.
- Burton, J. W., M.-K. Stein, and T. B. Jensen. 2020. A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making* 33:220–39.
- Campbell, J. Y. 2006. Household finance. *The Journal of Finance* 61:1553–604.
- Campbell, J. Y., and L. M. Viceira. 1999. Consumption and portfolio decisions when expected returns are time varying. *The Quarterly Journal of Economics* 114:433–95.
- Castelo, N., M. W. Bos, and D. R. Lehmann. 2019. Task-dependent algorithm aversion. *Journal of Marketing Research* 56:809–25.
- Collinson, R., J. E. Humphries, N. S. Mader, D. K. Reed, D. I. Tannenbaum, and W. Van Dijk. 2022. Eviction and poverty in american cities. *Quarterly Journal of Economics* forthcoming.
- D’Acunto, F., N. Prabhala, and A. Rossi. 2019. The promises and pitfalls of robo-advising. *Review of Financial Studies* .

- D'Acunto, F., and A. G. Rossi. 2022. Robo-advising. *Palgrave Macmillan Handbook of Technological Finance* .
- Dietvorst, B. J., J. P. Simmons, and C. Massey. 2015. Algorithm aversion: people erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General* 144:114–.
- Egan, M., G. Matvos, and A. Seru. 2019. The market for financial adviser misconduct. *Journal of Political Economy* 127:233–95.
- Fuster, A., P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther. 2022. Predictably unequal? the effects of machine learning on credit markets. *The Journal of Finance* 77:5–47.
- Gargano, A., and A. G. Rossi. 2017. Does it pay to pay attention? *Working Paper* .
- Gennaioli, N., A. Shleifer, and R. Vishny. 2015. Money doctors. *The Journal of Finance* 70:91–114.
- Gomes, F., M. Haliassos, and T. Ramadorai. 2021. Household finance. *Journal of Economic Literature* 59:919–1000.
- Guiso, L., P. Sapienza, and L. Zingales. 2008. Trusting the stock market. *the Journal of Finance* 63:2557–600.
- Guiso, L., and P. Sodini. 2013. Household finance: An emerging field. In *Handbook of the Economics of Finance*, vol. 2, 1397–532. Elsevier.
- Linnainmaa, J. T., B. Melzer, A. Previtro, and S. Foerster. 2018. Financial advisors and risk-taking. *Working Paper* .
- Niszczoła, P., and D. Kaszás. 2020. Robo-investment aversion. *Plos one* 15:e0239277–.
- Pagliari, C., and S. Utkus. 2019. Assessing the value of advice. *Vanguard Research* .
- Pástor, L., and P. Veronesi. 2003. Stock valuation and learning about profitability. *The Journal of Finance* 58:1749–89.
- Reher, M., and C. Sun. 2019. Automated financial management: Diversification and account size flexibility. *Journal of Investment Management* .

- Rossi, A. G., and S. Utkus. 2019. The needs and wants in financial advice: Human versus robo-advising. *Working Paper* .
- Rossi, A. G., and S. P. Utkus. 2020. Who benefits from robo-advising? evidence from machine learning. *Evidence from Machine Learning (March 10, 2020)* .
- Timmermann, A. G. 1993. How learning in financial markets generates excess volatility and predictability in stock prices. *The Quarterly Journal of Economics* 108:1135–45.

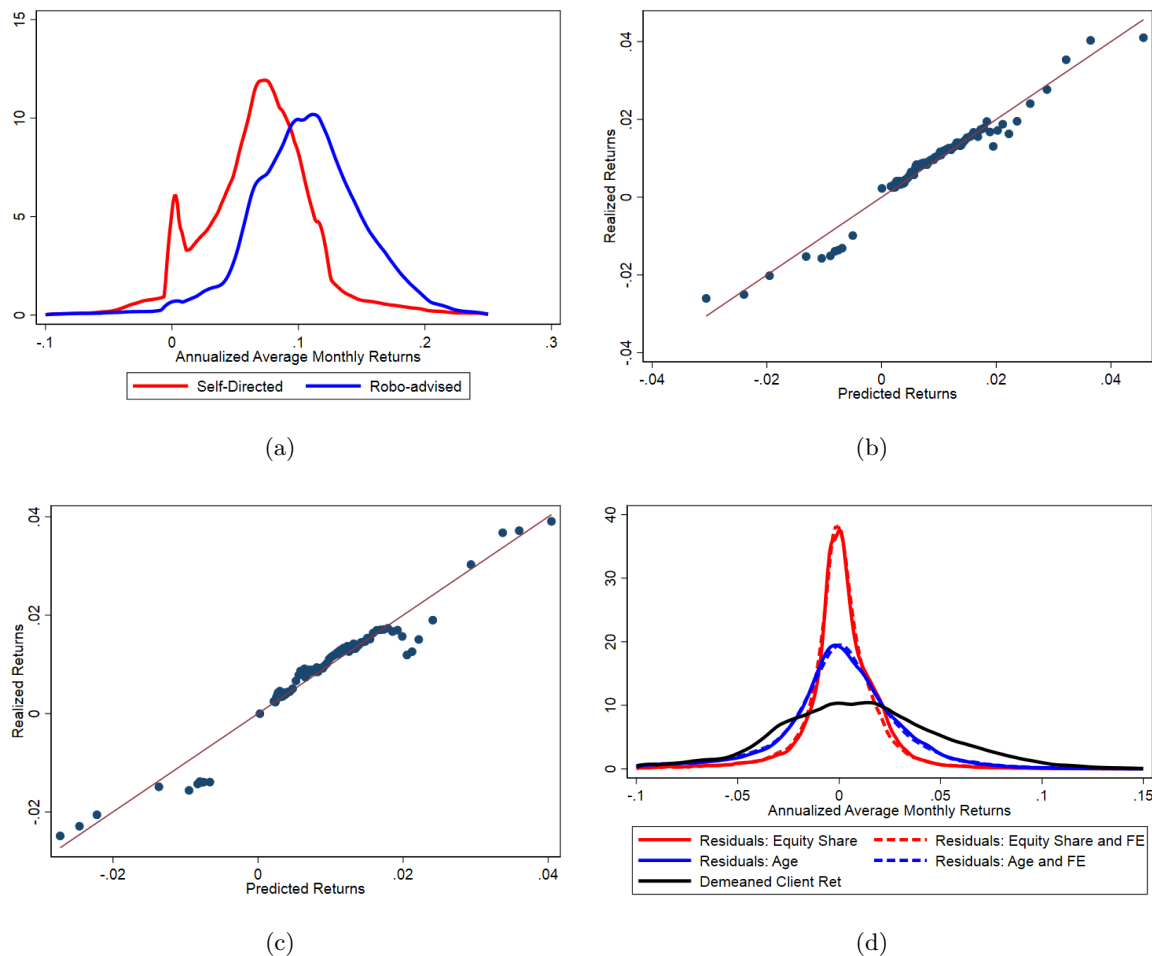
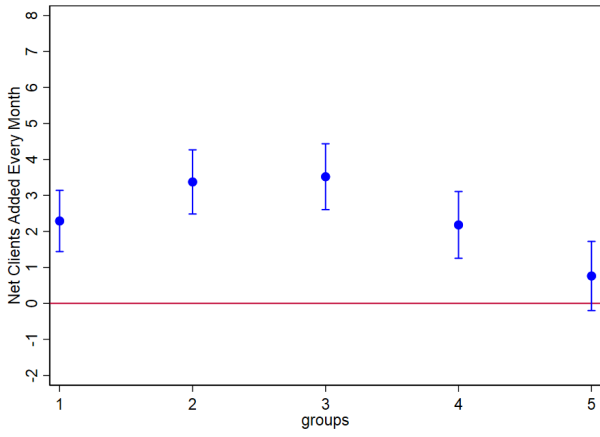
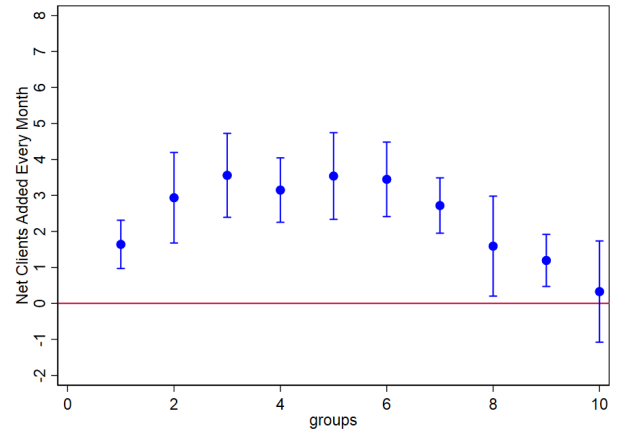


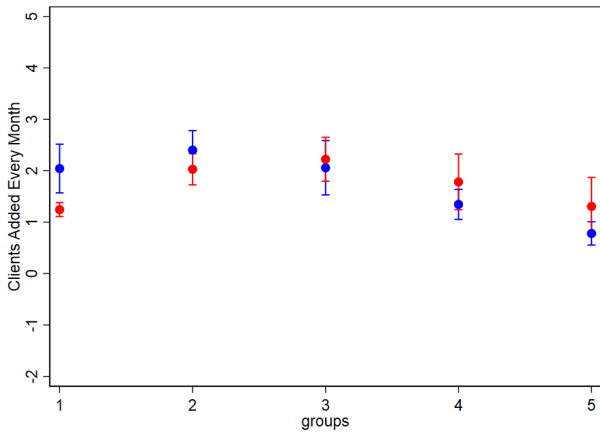
Figure 1: This figure reports results on investors' portfolio performance. We compute the cross-section of investment performance across robo-advised and non-robo-advised investors in Subfigure (a). Subfigures (b) and (c) relate monthly client returns to predicted returns on the basis of investors' equity share and age, respectively. Subfigure (d) shows the extent to which investor returns are determined by their equity share, demographic characteristics, and their assigned advisor.



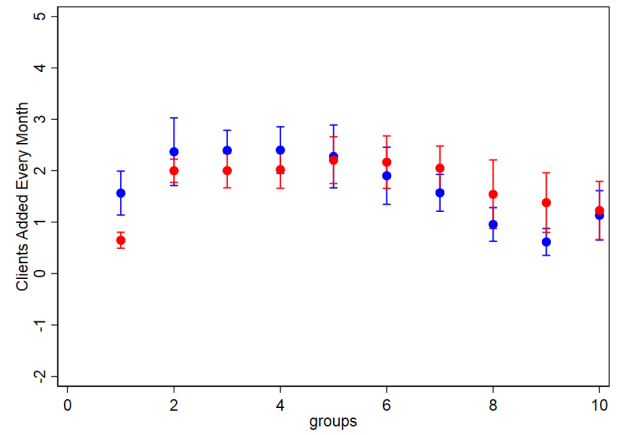
(a)



(b)

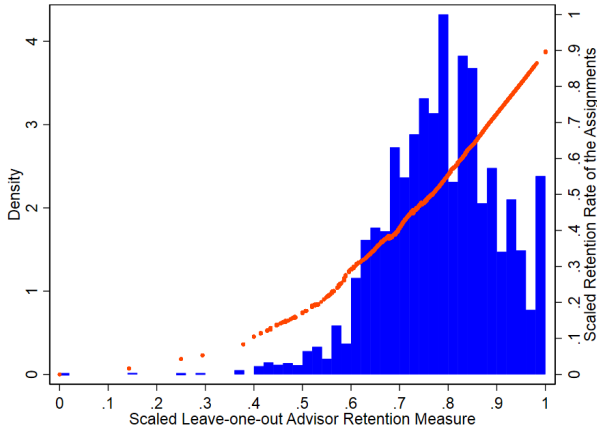


(c)

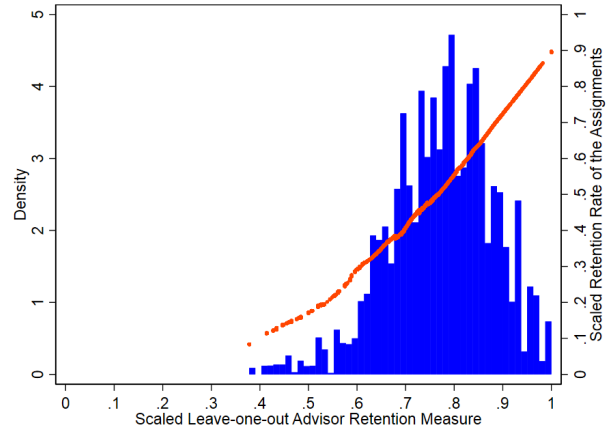


(d)

Figure 2: Subfigure (a) is constructed by computing, for every advisor, the number of investors they advise at the beginning of each month and sorting all the available advisors into quintiles based on their current workload. The figure then reports the average net increase in the number of clients allocated to advisors in each group every month, computed as the number of investors allocated to each advisor minus the number of investors lost by each advisor every month because of attrition, together with 95% confidence intervals. Subfigure (b) repeats the analysis using deciles instead of quintiles. Subfigures (c) and (d) repeat the exercise but focus only on investors' additions and split the advisors into high- and low-retention.

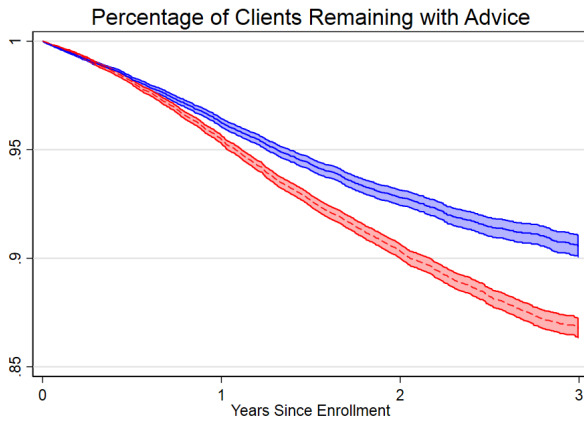


(a)

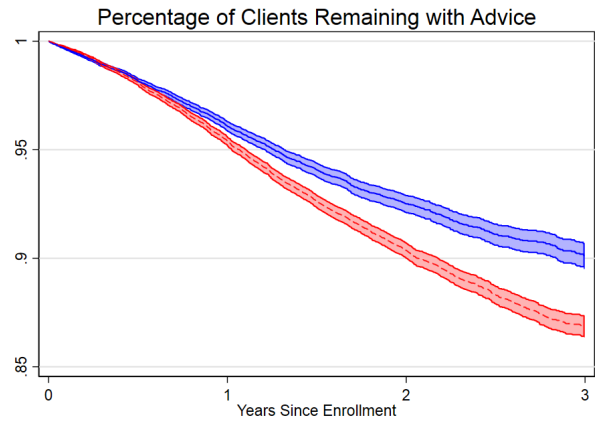


(b)

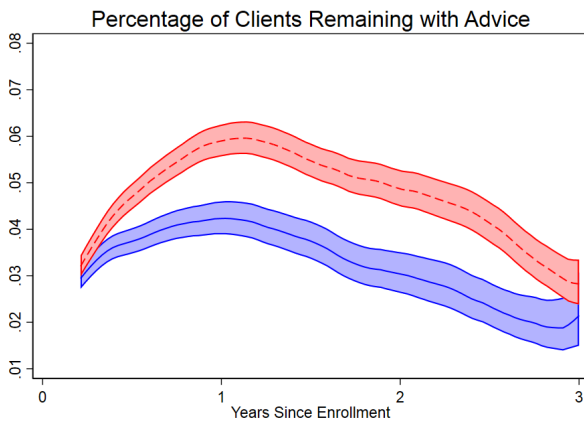
Figure 3: In Subfigure (a), the histogram reports the percentage of clients retained by each advisor-client pairing, scaled in such a way that the advisor-client pairings with the highest retention are assigned a value of 100. Super-imposed on the histogram, we report non-parametric estimates of the relation between the scaled leave-one-out retention measure of each advisor-client pair on the scaled retention rate of that specific client. Subfigure (b) repeats the exercise excluding the advisors in the bottom decile in terms of clients advised.



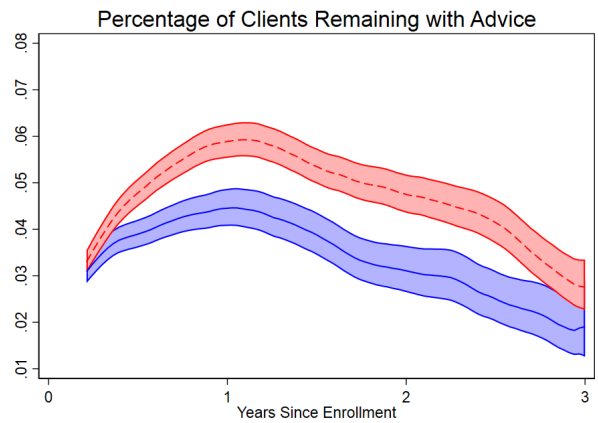
(a) All Advisors



(b) Excluding Lowest 10% of Advisors



(c) All Advisors



(d) Excluding Lowest 10% of Advisors

Figure 4: Subfigure (a) shows survival plots for investors assigned to advisors of different types. Subfigure (b) repeats the survival estimates excluding advisors in the bottom decile in terms of load. Subfigure (c) shows smooth hazards for investors assigned to advisors of different types. Subfigure (d) repeats the hazard estimates excluding advisors in the bottom decile in terms of load. Investors associated with type-1 advisors are in blue. Those with type-0 advisors are in red.

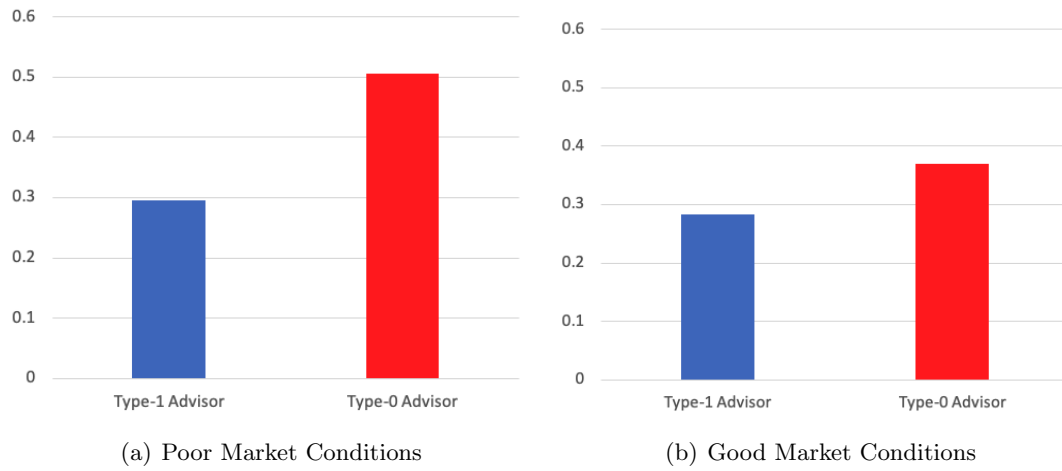


Figure 5: This figure reports the monthly attrition (in percentages) of investors assigned to advisors of different types, conditioning on different market conditions in subfigures (a) and (b).

Table 1. Demographic and Portfolio Characteristics of Advised Investors

Panel A. Demographic Characteristics						
	N	mean	sd	p25	p50	p75
Age	54,325	63.816	12.071	57.000	65.000	72.000
Male	54,744	0.598	0.490	0.000	1.000	1.000
Tenure	54,744	13.481	9.084	3.833	13.667	20.250
Panel B. Portfolio-Related Characteristics						
	N	mean	sd	p25	p50	p75
Wealth	54,744	\$758,378	\$821,029	\$210,800	\$478,929	\$981,330
NumAssets	54,744	7.952	4.910	5.000	6.000	9.000
PctVGProducts	54,744	0.974	0.069	1.000	1.000	1.000
Panel C. Asset Allocation Characteristics						
	N	mean	sd	p25	p50	p75
PctMutualFunds	54,744	0.952	0.102	0.960	1.000	1.000
PctCash	54,744	0.018	0.046	0.000	0.000	0.008
PctETF	54,744	0.008	0.030	0.000	0.000	0.000
PctStocks	54,744	0.014	0.045	0.000	0.000	0.000
PctBonds	54,744	0.000	0.002	0.000	0.000	0.000
Panel D. Characteristics of Mutual Funds Held						
	N	mean	sd	p25	p50	p75
AcctIndex	54,744	0.828	0.178	0.745	0.858	1.000
MgtFee	54,717	0.072	0.024	0.059	0.064	0.075
ExpRatio	54,707	0.093	0.027	0.078	0.083	0.096
TurnRatio	54,685	0.268	0.120	0.190	0.280	0.337

This table reports the demographic characteristics and portfolio allocation behavior of investors 12 months after signing up for advice. The results are computed at the investor level and include all account types, that is, taxable and non-taxable (IRA) accounts. Panel A reports demographic characteristics, Panel B focuses on portfolio-related characteristics, Panel C focuses on asset allocation characteristics, and Panel D focuses on the characteristics of the mutual funds held. For each variable, we report the number of accounts used in the computations, the mean, the standard deviation, and the 25th, 50th, and 75th percentiles of the distribution.

Table 2. Covariates Balancing Across Clients Assigned to High- and Low-Retention Advisors

	High Retention		Low Retention		Diff mean	Diff t-stat	N
	mean	N	mean	N			
Age	64.540	24,514	65.657	23,511	1.117***	(3.33)	48,025
Male	0.577	25,739	0.618	24,085	0.041***	(4.76)	49,824
Tenure	14.666	25,739	15.635	24,085	0.969***	(3.03)	49,824
Wealth	946,754	25,739	993,861	24,085	47,107	(0.39)	49,824
NumAssets	10.717	25,739	11.438	24,085	0.721	(1.69)	49,824
PctVGProducts	0.853	25,706	0.850	24,062	-0.003	(-0.81)	49,768
PctMutualFunds	0.666	25,706	0.672	24,062	0.006	(0.61)	49,768
PctCash	0.234	25,706	0.226	24,062	-0.007	(-0.67)	49,768
PctETF	0.035	25,706	0.034	24,062	0.000	(-0.27)	49,768
PctStocks	0.046	25,706	0.047	24,062	0.001	(0.84)	49,768
PctBonds	0.002	25,706	0.002	24,062	0.000**	(2.65)	49,768
AcctIndex	0.436	25,739	0.438	24,084	0.002	(0.14)	49,823
MgtFee	0.147	23,877	0.147	22,931	0.001	(0.28)	46,808
ExpRatio	0.209	23,299	0.206	22,396	-0.003	(-0.18)	45,695
TurnRatio	0.328	22,918	0.343	21,787	0.016**	(2.16)	44,705
Ret. Pre-PAS	0.051	22,040	0.045	20,884	-0.005	(-0.98)	42,924
Adj. Ret. Pre-PAS	-0.007	22,040	-0.009	20,884	-0.002	(-1.64)	42,924

This table reports balancing results for demographic characteristics and portfolio allocation behavior for investors 1 month before signing up for advice. For each characteristic, in the first four columns we report the mean and the number of observations for high- and low-retention advisors. In the last three columns we report the difference in means, the associated t-statistic and the total number of observations.

Table 3. The Effect of Human Advice Across Market Conditions

	CRSP Return	Investor Return	CRSP Volatility	Investor Volatility
Bad Market	0.136*** (2.98)	0.145*** (3.96)	0.125*** (3.20)	0.117*** (4.02)
Type1_Advisor	-0.086*** (-4.74)	-0.078*** (-4.03)	-0.051*** (-3.14)	-0.056*** (-4.41)
Interaction	-0.123*** (-3.77)	-0.123*** (-3.90)	-0.102*** (-3.21)	-0.091*** (-3.61)
Constant	0.369*** (17.70)	0.359*** (16.23)	0.325*** (17.78)	0.328*** (17.87)
Clustering	Date&User	Date&User	Date&User	Date&User
R-square	0.00011	0.00013	0.00014	0.00013
N	938,314	938,314	938,314	938,314

This table reports coefficient estimates of the following baseline regression estimated at the monthly frequency:

$$\begin{aligned}
\text{Dummy_quit}_{i,t} = & \alpha + \beta I_{\{BadMarket_{t-1}=1\}} + \gamma I_{\{Type1_Advisor_i=1\}} \\
& + \delta I_{\{BadMarket_{t-1}=1\}} \times I_{\{Type1_Advisor_i=1\}} + \epsilon_{i,t},
\end{aligned}$$

where $\text{Dummy_quit}_{i,t}$ is equal to 1 if investor i quits robo-advising in month t and 0 otherwise, $I_{\{BadMarket_{t-1}<0\}}$ is an indicator variable equal to 1 if market conditions are bad in month $t - 1$ and 0 otherwise, and $I_{\{Type1_Advisor_{i,t-1}=1\}}$ is equal to 1 if investor i is assigned to an advisor with a past retention above the median and 0 otherwise. We multiply all the coefficient estimates by 100 so they are expressed in percentages. The standard errors are clustered by user and date. We use CRSP return, Investor Return, CRSP Volatility and Investor Volatility as proxies for market conditions in columns (1) through (4).

**Table 4. The Effect of Human Advice Across Market Conditions
Long and Short Tenure Investors**

	Short Tenure	Long Tenure
Bad Market	0.219*** (7.69)	0.044 (1.36)
Type1_Advisor	-0.057*** (-3.29)	-0.089*** (-3.62)
Interaction	-0.158*** (-3.55)	-0.102** (-2.19)
Constant	0.302*** (23.77)	0.417*** (16.95)
Clustering	Date&User	Date&User
R-square	0.00020	0.00006
N	477,736	460,578

This table reports coefficient estimates of the following baseline regression estimated at the monthly frequency:

$$\begin{aligned}
 Dummy_quit_{i,t} = & \alpha + \beta I_{\{MKT_RET_{t-1} < 0\}} + \gamma I_{\{Type1_Advisor_i = 1\}} \\
 & + \delta I_{\{MKT_RET_{t-1} < 0\}} \times I_{\{Type1_Advisor_i = 1\}} + \epsilon_{i,t},
 \end{aligned}$$

where $Dummy_quit_{i,t}$ is equal to 1 if investor i quits robo-advising in month t and 0 otherwise, $I_{\{MKT_RET_{t-1} < 0\}}$ is an indicator variable equal to 1 if the market return is negative in month $t - 1$ and 0 otherwise, and $I_{\{Type1_Advisor_{i,t-1} = 1\}}$ is equal to 1 if investor i is assigned to an advisor with a past retention above the median and 0 otherwise. We multiply all the coefficient estimates by 100, so they are expressed in percentages. We split the sample of robo-advising users into two groups on the basis of their tenure and report the results for short tenure in column (1) and long tenure in column (2). In both cases, the standard errors are double-clustered by users and dates.

Online Appendix

(Not for publication)

Online Appendix A.1 Micro-Founded Theory of Portfolio Choice

We interpret the client in our general model as an investor who can allocate a fraction $\alpha_t \geq 0$ of her wealth each period to a robo-advisor, and the remaining $1 - \alpha_t$ of her wealth to an outside portfolio. The client's utility is given by $w_T = \ln(W_T)$, where W_T denotes her final wealth, and where we use lowercase letters to denote logs. In each period where she uses the robo-advisor, the client suffers a fixed cost/disutility f^j which, as in the general model, depends on the identity of her assigned human advisor.

We assume that the log return on the outside portfolio is deterministic and given by \bar{r} . We further assume that the robo-advisor, between dates t and $t+1$, invests in a portfolio that generates stochastic log returns given by $r_{t+1}^i = \bar{r} + \theta + u_{t+1}^i$. Thus, the performance measure $y_{t+1}^i = \theta + u_{t+1}^i$ in our general model now stands for the excess log return on the robo-advisor's portfolio, and the service quality parameter θ measures the expected excess return. The investor's beliefs about θ , as a function of her human advisor/expert j , are as in the general model.

The investor's log wealth, denoted w_t at date t , evolves according to the following approximate law of motion:

$$w_{t+1}^i - w_t \simeq \bar{r} + \alpha_t y_{t+1}^i + \frac{1}{2} \sigma^2 \alpha_t (1 - \alpha_t), \quad (13)$$

where $\sigma^2 = 1/\tau_y$ is the variance of the shock to excess returns each period. This law of motion holds exactly in continuous time. We present a derivation, which is similar to that in Campbell and Viceira (2002), below. Intuitively, the change in log wealth is equal to the return on the outside portfolio, captured by the first term, and the portfolio-weighted excess return on the robo-advisor's portfolio, captured by the second term. The third term, in addition, reflects Jensen's inequality: Because the log is a concave function, the log return on an average of two investments is greater than average of the log returns.

In the remainder of this appendix, we demonstrate that this setup is a special case of the general model we have considered in the paper and derive the specific functional of the client's utility function that maps to the portfolio choice problem. We continue to write m for the state variable measuring the investor's current expectation of θ , and we write w for a state variable measuring the investor's current log wealth. As usual, we let variables with primes (e.g., m' and w') denote one-period-ahead state variables.

We conjecture that the investor's continuation value if still enrolled at date $t < T$, is given by

$$F_t(w, m) = w + (T - t)\bar{r} + V_t(m). \quad (14)$$

The investor at date t has two options: To quit, in which case her final utility takes the deterministic value $w + (T - t)\bar{r}$, or to continue enrolling with an optimally chosen portfolio weight α_t . The investor's Bellman equation can therefore be written as

$$F_t(w, m) = \max \left\{ w + (T - t)\bar{r}, -f^j + \max_{\alpha \geq 0} \hat{\mathbb{E}}^j [F_{t+1}(w', m') | w, m, \alpha_t = \alpha] \right\} \quad (15)$$

Substituting our conjecture and the law of motion for wealth, we can evaluate the last term more explicitly as

$$\begin{aligned} \hat{\mathbb{E}}^j [F_{t+1}(w', m') | w, m, \alpha_t = \alpha] &= w' + (T - (t + 1))\bar{r} + \hat{\mathbb{E}}^j [V_t(m') | m] \\ &= w + (T - t)\bar{r} + \alpha m + \frac{1}{2}\sigma^2\alpha(1 - \alpha) + \hat{\mathbb{E}}^j [V_t(m') | m] \end{aligned} \quad (16)$$

since $m = \hat{\mathbb{E}}^j [y' | m]$ by definition. Therefore, the inner maximization problem in Equation (15) is solved by the optimal portfolio weight

$$\hat{\alpha} = \max \left\{ \frac{m + \frac{1}{2}\sigma^2}{\sigma^2}, 0 \right\},$$

and the value of (16) at this value is

$$\begin{aligned} E [F_{t+1} (w', m') | w, m, \alpha_t = \hat{\alpha}] &= w + (T - t) \bar{r} + \left(m + \frac{1}{2} \sigma^2 \right) \hat{\alpha} - \frac{1}{2} \sigma^2 \hat{\alpha}^2 + \hat{\mathbb{E}}^j [V_{t+1} (m') | m] \\ &= w + (T - t) \bar{r} + \hat{\mathbb{E}}^j [V_{t+1} (m') | m] + \frac{1}{2} [SR(m)]^2, \end{aligned}$$

where we have defined the (truncated) Sharpe ratio

$$SR(m) = \begin{cases} 0, & m + \frac{1}{2} \sigma^2 < 0, \\ \frac{m + \frac{1}{2} \sigma^2}{\sigma}, & \text{otherwise.} \end{cases}$$

Substituting this result into (15), along with our conjectured solution, we obtain the following simplified Bellman equation:

$$V_t(m) = \max \left\{ \underbrace{-f^j + \frac{1}{2} [SR(m)]^2}_{\equiv u^j(m)} + \hat{\mathbb{E}}^j [V_{t+1} (m') | m], 0 \right\}.$$

This Bellman equation maps exactly to our general model, as soon as we define the investor's utility function as follows:

$$u^j(m) = -f^j + \frac{1}{2} [SR(m)]^2.$$

Derivation of Intertemporal Budget Constraint

For clarity, we omit client i superscripts in this derivation. Consider the equivalent continuous-time economy in which the per-unit value P_t of the robo-advisor's portfolio follows

$$\frac{dP_t}{P_t} = \left(\bar{r} + \theta + \frac{1}{2} \sigma^2 \right) dt + \sigma dZ_t$$

where all parameters are defined as in our baseline model, and where Z_t is a standard Brownian motion. Notice that this model also implies the discrete-time representation of log returns

$$r_{t+1} \equiv \log \left(\frac{P_{t+1}}{P_t} \right) = \bar{r} + \theta + \sigma u_{t+1},$$

where we have defined $u_{t+1} = Z_{t+1} - Z_t$, so that it is indeed equivalent to our baseline formulation in Equation (1).

Further, assume that the per-unit value B_t^i of investor i 's outside portfolio evolves according to

$$\frac{dB_t}{B_t} = \bar{r}dt$$

The investor's wealth evolves according to

$$\begin{aligned} \frac{dW_t}{W_t} &= \alpha_t \frac{dP_t}{P_t} + (1 - \alpha_t) \frac{dB_t}{B_t} \\ &= \alpha_t \left[\left(\bar{r} + \theta + \frac{1}{2}\sigma^2 \right) dt + \sigma dZ_t \right] + (1 - \alpha_t) \bar{r}dt \end{aligned}$$

or, rearranging,

$$dW_t = \left[\alpha_t \left(\bar{r} + \theta + \frac{1}{2}\sigma^2 \right) + (1 - \alpha_t) \bar{r} \right] W_t dt + \alpha_t \sigma W_t dZ_t$$

Converting to log returns, and applying Ito's lemma to $f(W) = \log W$, we obtain

$$\begin{aligned} d \log W_t &= df(W_t) \\ &= f'(W_t) dW_t + \frac{1}{2} f''(W_t) (dW_t)^2 dt \\ &= \left[\alpha_t \left(\bar{r} + \theta + \frac{1}{2}\sigma^2 \right) + (1 - \alpha_t) \bar{r} \right] dt + \alpha_t \sigma dZ_t \\ &\quad - \frac{1}{2} (\alpha_t \sigma)^2 dt \\ &= \left[\alpha_t \left(\bar{r} + \theta + \frac{1}{2}\sigma^2 \right) + (1 - \alpha_t) \bar{r} + \frac{1}{2}\sigma^2 \alpha_t (1 - \alpha_t) \right] dt + \alpha_t \sigma dZ_t \end{aligned}$$

For our discrete time approximation, we set $dt = 1$ in the previous equation to get the budget

constraint:

$$\begin{aligned}\log W_{t+1}^i - \log W_t &= \alpha_t (\bar{r} + \theta) + (1 - \alpha_t) \bar{r} + \frac{1}{2} \sigma^2 \alpha_t (1 - \alpha_t) + \alpha_t \sigma u_{t+1} \\ &= \alpha_t (\bar{r} + \theta + \sigma u_{t+1}) + (1 - \alpha_t) \bar{r} + \frac{1}{2} \sigma^2 \alpha_t (1 - \alpha_t) \\ &= \alpha_t r_{t+1} + (1 - \alpha_t) \bar{r} + \frac{1}{2} \sigma^2 \alpha_t (1 - \alpha_t) \\ &= \bar{r} + \alpha_t (r_{t+1} - \bar{r}) + \frac{1}{2} \sigma^2 \alpha_t (1 - \alpha_t) \\ &= \bar{r} + \alpha_t y_{t+1} + \frac{1}{2} \sigma^2 \alpha_t (1 - \alpha_t),\end{aligned}$$

which establishes Equation (13), where we have again used $u_{t+1} = Z_{t+1} - Z_t$.