Journal of Empirical Finance 59 (2020) 278–299

Contents lists available at ScienceDirect

Journal of Empirical Finance

journal homepage: www.elsevier.com/locate/jempfin

Artificial Intelligence Alter Egos: Who might benefit from robo-investing?☆

Catherine D'Hondt [a], Rudy De Winne [a], Eric Ghysels [a,b,c,*], Steve Raymond [d]

[a] *Louvain School of Management and Louvain Finance (LIDAM), Université catholique de Louvain, Belgium*
[b] *Department of Economics, University of North Carolina Chapel Hill, Department of Finance, Kenan-Flagler Business School, United States of America*
[c] *CEPR, United Kingdom*
[d] *Department of Economics, University of North Carolina Chapel Hill, United States of America*

A B S T R A C T

We use a unique data set covering brokerage accounts for a large cross-section of investors over a sample from January 2003 to March 2012, which includes the 2008 financial crisis, to assess the potential benefits of robo-investing. We explore robo-investing strategies commonly used in the industry, including some involving advanced machine learning methods. We shadow each of our investors with a robo-advisor to shed light on possible benefits the emerging robo-advising industry may provide to certain segments of the population, such as low income and/or low education investors.

1. Introduction

To assess the benefits of robo-investing we use a unique data set covering brokerage accounts for a large cross-section of 22,972 individual investors covering a sample from January 2003 to March 2012, and therefore includes the 2008 financial crisis. We have records of all trades, and in addition have detailed information about each individual investor's characteristics such as age, gender, education, annual net income, and most importantly, risk aversion assessed on the basis of responses to survey questions. Although we work with Belgian individual investors, most of their trading activities pertain to foreign stocks (86% are non-Belgian and roughly a quarter are US). Hence, our analysis pertains to international portfolio selection of stocks and ETFs.

We explore robo-investing strategies commonly used in the industry, including some involving advanced machine learning methods. The man versus machine comparison allows us to shed light on potential benefits the emerging robo-advising industry may provide to certain targeted segments of the population, such as low income and/or low education investors.[1]

To the best of our knowledge there has not been any assessment of the potential benefits of robo-investing over a long period of time for a heterogeneous panel of individual investors. In addition, our sample has a number of appealing features to study robo-investing. Many investment brokerage firms are now targeting individuals with modest savings as it is generally believed that smaller

☆ The authors are grateful to the online brokerage house for providing the data. Any errors are the full responsibility of the authors. We thank the editor, associate editor and referee for insightful comments which helped improve our paper as well as conference participants at the Chinese Economists Society, Dalian, the 2019 SoFiE Annual Meetings, Shanghai, and seminar participants at HEC Liège, the Toulouse School of Economics, Georgia State University, UC San Diego, University of Chicago, and the University of Southern California for helpful comments.

* Corresponding author at: Department of Economics, University of North Carolina Chapel Hill, Department of Finance, Kenan-Flagler Business School, United States of America.

E-mail address: eghysels@unc.edu (E. Ghysels).

[1] In the US, robo-advisor start-ups saw an eight-fold increase in their AUM in recent years on the back of some retirement savings shifting to robo-advisor accounts. Cost advantages have been creating significant momentum for the industry. In addition, the success of passive investment strategies in recent years has also been beneficial. It is therefore fair to say that robo-advisors are posing a challenge to traditional financial advisory services. One expects that some robo-advisory start-ups will probably end up in partnerships or be the subject of takeovers by established asset management firms or banks in the coming years. Moreover, the traditional asset managers themselves are also adopting robo-investing strategies. In that respect, robo-advising will become more mainstream.

https://doi.org/10.1016/j.jempfin.2020.10.002
Received 6 January 2020; Received in revised form 5 October 2020; Accepted 15 October 2020
Available online 5 November 2020
0927-5398/

investors do not get the investment advice they need. In fact, 71% or almost 90 million American families have investment account balances worth less than $100,000. The growth of automated investment advisory services is filling a need for such investors. Our data set consists of individual investors typically targeted by robo-advising. In terms of annual net income, approximately 70% of the investors in our sample declare an income between 20,000 and 75,000 euros. The mean portfolio value in our sample is 29,244 euros and the average investor is about 48 years old. This is different from existing papers in the literature. For example, Rossi and Utkus (2020) study the evolution of investment advice among a sample of more than 80,000 individuals who were previously self-directed investors and sign up for professional financial advice during a much shorter period (2015–2018). Their sample consists of investors with considerable portfolio wealth (median portfolio wealth is $282,000) and a willingness to take equity risk (median equity share 59%). This means their sample arguably suffers from selection biases. Instead, we have investors more commonly targeted by robo-advisors and we also observe many attributes regarding our investors, including a survey-based measure of risk aversion, which are typically not available in other studies.

Note that our paper does not directly address the effect on wealth management of adopting robo-advising, as studied by for example (D'Acunto et al., 2019). On the one hand, our data is richer in terms of details regarding the characteristics – such as income, education, gender, risk aversion, trading habits – for each individual investor. On the other hand, we study a sample where robo-advising was not adopted by the brokerage firm whose trading data we examine. Instead, we introduce the idea of shadow robo-investors to assess the potential benefits of robo-advising. Namely, we study various robo-investors that shadow the individuals in our data set and the novelty of our approach is that we know what the investors have done in reality versus what a robo-investor would have done instead. In that sense our analysis is a real-time experiment with real data.

We start with a setting involving robo-investors limited to the set of stocks and ETFs in each individual investor's history of trading — using a rolling 2-year sample. This constraint ties each robot to a specific investor in our sample via their trading history. Note that the robo-investors use *all* the stocks/ETFs individual investor i held in the past two years, but may have sold in the meantime. Hence, the rationale is that the investor knows about the stocks/ETFs held by the shadow robo-investor. We call these shadow robo-investors Artificial Intelligence Alter Egos, or AI Alter Egos.[2] The 2-year constraint is arguably ad hoc, but we prefer to set a universe of assets that errs on being perhaps too restrictive rather than the opposite, with the consequence of possibly understating the advantage that a robo-advisor might have in reality.

There is another important reason why we focus on the Alter Ego scheme. It is well-known that retail investors under-diversify their portfolios. Adding some assets at random will therefore most likely contribute to a better portfolio performance. With the Alter Ego approach we want to focus on something much more interesting, namely we take the universe of assets any individual investor in our sample knows (because of past trading) and we study how they could benefit from with AI-guided advice. This allows us to check for the benefits of robo-investing, beyond a "simple" improvement of diversification. In particular, our results show that our shadow robo-investors do better than individual investors. Such results indicate that robo-investing is not only a question of better diversification.

The notion of AI Alter Egos is not unique to finance, although we might be the first to coin the term. To illustrate, let us look at machine learning (ML) advances in other fields, such as literature and music. Today, a ML text mining algorithm can analyze the writings of a famous author and create entirely new literature in the style of the writer it was exposed to and trained on. The same can be done with music. For example, Franz Schubert started his Unfinished Symphony in B minor in 1882, but wrote only two complete movements, though he lived another six years. Now, deep-learning ML has produced a completed version of the entire symphony. We can characterize this as Schubert's AI Alter Ego composing a new score. Would Schubert have done better than his AI Alter Ego? We prefer to leave that debate to the musicologists, but it is fair to say it would probably be hard to address the question. Fortunately, it is much easier to apply the notion of AI Alter Egos in a setting where comparing the outcomes of human and AI alternatives is more straightforward — such as in financial investments.

Robo-investors generally perform asset allocation with mean–variance (MV) analysis or a variant of it. Unfortunately, some robo-investors do not disclose information on how they estimate variances and correlations, but from what is known in the public domain it is clear that they primarily rely on historical data to form these estimates. For example, industry leader (in terms of AUM) Betterment uses the Black–Litterman model, which requires users to specify a variance covariance matrix for all asset classes which is estimated using historical data combined with the Ledoit–Wolf shrinkage to reduce estimation error.[3] Wealthfront generates standard deviation estimates by considering each asset class's long-term and short-term historical standard deviation and the expected volatility of each asset class as implied by pricing in options markets.[4] Schwab Intelligent Portfolios also do a variation of the MV approach.[5] We consider two (Markowitz, 1952) MV investment strategies — all inspired by current industry practice. The two strategies differ in terms of the sophistication regarding the conditional mean and variance estimates. The first involves two-year rolling sample estimates for both the mean and variance. For the second we rev up the robot engines and replace the rolling sample estimators by respectively expected return predictions using machine learning algorithms and sophisticated conditional covariance estimators. More specifically for the conditional mean we use Elastic-Net, Random Forest, Neural Network, and model ensemble estimators. For the conditional covariance matrix – looking at a total of 683 stocks and 393 ETFs – we use the (Engle et al., 2019) nonlinear shrinkage method derived from random matrix theory to correct in-sample biases of sample eigenvalues. Finally, it is

[2] Since the robo-investor schemes go beyond machine learning, as they involve portfolio allocation rules, we use the more general term of artificial intelligence. In our case the AI pertains a set of computer-driven self-learning rules which determine portfolio allocations.

[3] More details appear on the Betterment Website, see for example http://support.betterment.com/customer/portal/articles/1295723-why-is-betterment-changing-the-portfolio.

[4] See *Wealthfront Investment Methodology White Paper,* available at https://research.wealthfront.com/whitepapers/investment-methodology/

[5] Further details appear in https://intelligent.schwab.com/public/intelligent/insights/whitepapers/asset-allocation.html.

important to note that robo-investors have the option to hold cash, i.e. decide to avoid market risk exposure. No short selling is allowed, however.

We focus exclusively on the quarterly rebalancing scheme (we also computed results with respectively monthly and annual rebalancing which are available on request from the authors). Note that robo-investors buy and hold at fixed sampling frequencies — end of quarter in the lead example. This is in contrast to the individual investors in our sample who execute their trades at any point in time.

Overall our findings are as follows. The AI Alter Ego robo-investors involving rolling sample mean and variance estimates perform poorly and are of little value to any of our investors.[6] In contrast the machine learning MV AI Alter Egos result in significant investment portfolio performance improvements for certain types of investors. In particular, low income and/or low education investors typically benefit greatly from following the robo-investor strategies. These results confirm the claims made by practitioners in the industry regarding the promises the use of AI hold for the future of the FinTech industry. More intriguing, and somewhat unexpected are our results pertaining to the performance during the financial crisis. Robo-investors outperform a large swath of investors.

As a by-product of our analysis, we also identify which machine learning methods perform well. While deep learning is often the best across a large cross-section of stocks, a close second-best is a much simpler linear prediction model with elastic net penalty based on the same set of predictor, namely those suggested by Welch and Goyal (2007), which consist of a mixture of firm-specific and macroeconomic covariates. Put differently, the gains from non-linear models is marginal at best.

The paper is organized as follows. In Section 2 we describe the brokerage data. Section 3 describes the various robo-investor schemes. Section 4 reports the empirical results. Section 5 concludes the paper.

2. A large panel of individual brokerage accounts

Our primary data set comes from a large Belgian online brokerage firm and consists of the trading accounts of 22,972 individual investors. This unique data spans about 10 years from January 2003 to March 2012, and therefore includes the 2008 financial crisis. We have detailed information about each trade, such as the instrument, the time-stamp, the trade direction, the executed quantity, the trade price, and explicit transaction costs. We focus on common stock investments as well as ETFs and exclude other financial instruments.[7] Trading of ETFs, mutual funds, options and warrants is more prevalent with high income/education investors. Trading of bonds is overall insignificant. Because we examine robo-advisors which are mean–variance investors we focus exclusively on stocks and ETFs which best fit the portfolio allocation model. For high income/education investors in particular this means we leave out to a certain degree other assets which we have available. After applying some filters described in the Appendix A, we end up with a sample of 1,590,199 (stocks) + 60,344 (ETFs) = 1,650,543 trades (and more than 13 billion euros traded in stocks and close to 1 billion euros in ETFs) over the 111-month period covering 683 stocks and 393 ETFs or 70% of all the investors' trading activity.

Using the trading data we build end-of-month portfolios for each investor and use historical market data to compute monthly portfolio market values (daily returns are also computed for the rolling sample risk estimation). Combining end-of-month portfolio market values with the corresponding monthly aggregate cash-flows, we calculate for each investor 110 (i.e. from February 2003 to March 2012) monthly portfolio gross and net returns (the latter net of transaction costs). In addition, using estimates for individual investor cash holdings at the end of each month (see Appendix A.5 for further details), we compute the weight of cash in the end-of-month portfolio, which enables us to calculate monthly individual portfolio returns adjusted for cash holdings. Since our robo-investors have the option to hold cash, individual investor returns adjusted for cash should deliver more meaningful comparisons.

Investors are included as robo-investor candidates if they satisfy the return criteria – sufficient time periods and returns with no extraordinary outliers – and minimum trade restrictions (see Appendix A.1 for further details). In particular, we drop investors with more than 106 missing values in their return series (i.e. at least 4 months of returns are required to keep an investor) and drop outliers as well. This decreases the sample to 20,622 investors (down from 22,972). To deal with transaction costs we take two approaches: (1) ignore transaction costs for the individual investment accounts and the robo-investors and (2) compute returns net of transaction costs. For the individual investors we observe transaction costs in the data, whereas for the robo-advisors we use a constant advisory management fee of the investor's AUM (details appear in Appendix A.6). Since robo-investors tend to trade less than the average/median investor, namely only once a quarter, ignoring transaction costs should yield conservative estimates of the robo-investing gains. Taking into account transaction costs will increase the advantage of robo-investors. We will cover both cases in detail in our empirical analysis.[8]

Our data set also includes an extensive set of individual investor characteristics, such as age, gender, education, annual net income and a risk aversion measure based on surveys (described in Appendix A.2). Although we work with Belgian individual investors, most of their trading activities involve foreign stocks (mainly the US and bordering countries France and Germany —

[6] Note that among the existing robo-investor practices there are a number which proclaim using MV allocations and most likely use some type of rolling sample scheme — although most white papers are rather vague on the actual implementation.

[7] The details of the data are described in the Appendix A. The majority of trading occurs in either equity or ETFs. In Appendix A.1 we document that 6,741 investors also traded options and warrants with an aggregate number of 602,833 trades and 6,665 investors traded mutual funds with an aggregate number of 260,120 trades. Only a few investors (i.e. 1,813) traded bonds with an aggregate number of 5,999 trades.

[8] Belgium does not have a capital gains tax during our sample period. As a result we ignore tax issue in our analysis.

see Appendix Table A.3 for details). The majority of stocks pertain to the technology sector (16.93%), financials (15.91%), and industrials (14.01%).

As noted in the Introduction, our data set consists of individual investors typically targeted by robo-advising. We have about 70% of the investors in our sample who declare an annual net income between 20,000 and 75,000 euros. Only a minority (3.36%) earns more than 150,000 euros per year.[9] The average investor is about 48 years old and executes monthly 2.76 trades across 2.05 different stocks for a volume of 18,237 euros. Consistent with the literature, investors in our sample are under-diversified; the average (median) investor holds a five-stock (three-stock) portfolio. The average end-of-month portfolio value is about 28,003 euros (with a median value of about 7,552 euros). As for risk aversion, the majority of investors seem to be risk tolerant since 65.33% of them declare a medium risk aversion and 27.88% of them even a low risk aversion.

In terms of performance, our investors earn an average monthly gross return of 0.42% on stocks and ETFs (median return of 0.13% - see Table A.9 in the Appendix A), with a volatility of 10.04%. This high average volatility of individual portfolio gross returns is not surprising given our sample period includes turbulent market conditions.[10]

3. Robo-investors

We start with a setting where the robo-investors are limited to the set of stocks and ETFs in each individual investor's history of trading — using a rolling 2-year sample. This constraint ties each robot to the trading history of each individual investor. We call these shadow robo-investors Artificial Intelligence Alter Egos. Robo-investors have the option to hold cash, i.e. decide to avoid market risk exposure, but no short-selling occurs in our sample nor is it allowed for in the design of the robots. The two-year window is arguably somewhat arbitrary. Our results hold for longer windows. Shorter windows are less appealing given the trading frequency of many investors, with only a median of 2 trades per month (see Table A.9 in the Appendix A). The portfolio allocations of robo-investors occur at fixed intervals, either monthly, quarterly or annually. We focus here exclusively on the quarterly results (all other results are available on request from the authors).

3.1. AI Alter Egos

We construct two types of AI Alter Ego robo-investors. As mentioned earlier, each only uses stocks and ETFs held by an individual investor over the past two years, not the entire universe of stocks. The table below provides two illustrative examples (with t end of quarter).

Initial holdings	Trading t - 1	Trading t	Investor holdings	Robo-investor potential holdings
Stocks 1 & 2	Sells all of 2	Buys stock 3	Stocks 1 & 3	Stocks 1, 2 & 3
Stock 1	Sells all of 1	Buys ETF 4	ETF 4	Stock 1 & ETF 4

The first line portrays an investor holding two stocks – say 1 and 2 – at time t - 2 (column called Initial holdings). At the end of t - 1, the investors sells all holdings of stock 2 and at the end of the subsequent period t buys stock 3. Hence, at the end of t she/he holds stocks 1 and 3. The robo-investor has stocks 1, 2 and 3 to form a portfolio. The second case is similar, but the investor only holds stock 1, sells all of it in t - 1 and buys ETF 4 in t. The robo-investors has two assets to select from. It is important to stress that the robo-investor may hold cash, i.e. decide not to put all the money in the stock market. This will be important as will become clear when discussing the empirical results.

To proceed, we need to introduce some notation. Let S_{it} be the set of stocks/ETFs investor i held over a two-year period up to time t. The above illustrative examples clarified that this does not mean that the investor holds these stocks/ETFs at the end of year/quarter/month t. It only means that the investor held these stocks/ETFs in the recent two-year history. We denote by T_i the duration of time (months/quarters/years whichever applies) investor i appears in the sample. Moreover, we denote by $N_{it} = \#S_{it}$, the number of stocks/ETFs in the set. We only consider investors with $N_{it} \geq 2 \ \forall \ t = 1, \ldots, T_i$. This ensures that the investment opportunity set contains a minimally sufficient set of stocks/ETFs for the robo-investors. This leaves us with 20,622 investors who satisfy this criteria.

The robo-investors buy at the end of t and hold until end of $t + 1$, i.e. for a month, quarter or full year.[11] We then compute holding period returns for the robo-investor, $r^{ae}_{i,t+1}$, and compute the alter-ego-less-investor's realized return spread as $r^{s}_{i,t} = r^{ae}_{i,t} - r_{i,t}$.

We mentioned that there are two types of robo-investors. The first type relies on the mean–variance (Markowitz, 1952) strategy and uses a simple rolling sample estimator for expected returns and the linear shrinkage estimators for the second conditional moments. For the second type, machine learning for expected returns and more advanced conditional covariance estimators are used. Details will be discussed later.

[9] The income measure reported in our data is recorded at inception, when the investor complete the MiFID tests. Therefore the income classification may be stale over the 10 year sample period.

[10] To calculate portfolio returns, we use as an approximation the Modified Dietz Method, aiming at providing a return close to the money-weighted rate of return (e.g., Shestopaloff and Shestopaloff (2007)). See Appendix A.2 for further details.

[11] In between rebalancing periods, the portfolio weights adjust according to the performance of an individual asset relative to the performance of the portfolio as a whole. In particular when $t + 1$ is not a rebalancing period, $w_{i,t+1} = w_{i,t}(1 + r_{i,t+1})/[\sum_{i=1}^{N} w_{i,t}(1 + r_{i,t+1})]$.

Rolling sample markowitz. The mean–variance optimal portfolio is constructed as the maximum Sharpe ratio subject to the short-sale constraint and the individual's investment opportunity set. Investor i selects from the set S_{it} of stocks. Critical to the optimal portfolios are estimates of conditional expected returns (μ_t^i) and the conditional covariance matrix of returns (Σ_t^i) for the stocks in the set S_{it}. The robo-investor solves for $\hat{w}_{i,t}$ selecting among these stocks according to:

$$\max_{w_{i,t}} w'_{i,t}\mu_t^i - \frac{\gamma}{2} w'_{i,t}\Sigma_t^i w_{i,t} \tag{1}$$

$$\text{s.t. } w_{i,t} \geq 0, \tag{2}$$

where γ is often interpreted as a risk aversion parameter which we set equal to one as it maximizes the Sharpe ratio. We estimate μ_t^i with two-year rolling-window historical averages, $\hat{\mu}_t^i = \frac{1}{k}\sum_{j=0}^{k-1} r_{t-j}^d$, where r_t^d is an $N_{it} \times 1$ vector of daily returns and k is the number of days in the two-year historical sample. For covariance, we also use rolling sample estimator with linear shrinkage as in Ledoit and Wolf (2004) based on daily returns over the same time span. These estimates form our naive benchmark.

Machine learning and nonlinear shrinkage. Continuing with the Markowitz allocation scheme, we explore whether increasing the complexity of the rolling sample estimators translates into improved robo-investor performance. We assume that each investor's Alter Ego robo-investor has access to a common set of models that replace the rolling sample schemes. For expected return predictions we use machine learning algorithms applied to each of the 1076 assets (683 stocks and 393 ETFs) and the Alter Ego robo-investor picks the prediction pertaining to the stocks in the sets S_{it}. More specifically for the conditional mean estimates we use Elastic-Net (Zou and Hastie, 2005), Random Forest (Breiman, 2001), Neural Network (Friedman et al., 2016, Chap. 11), and model ensemble estimators (Friedman et al., 2016, Chap. 16). For the conditional covariance matrix – looking at a total of 1076 assets – we use the (Engle et al., 2019) nonlinear shrinkage method derived from random matrix theory to correct in-sample biases of sample eigenvalues.

3.2. Machine learning expected returns

What we have in mind is a situation where the robo-investors rely on a modeling department within the brokerage house to provide them with estimates of conditional means and conditional covariances for the entire universe of stocks/ETFs and supplying the Alter Ego investor associated with each individual investor with the estimates μ_t^i and Σ_t^i for the stocks in the set S_{it}. The modelers estimate a wide class of models and use out-of-sample performance metrics to determine the most appropriate panel of conditional means and conditional covariances to supply to the robo-investors. Our goal here is to provide a simple approximation to the comprehensive conditional modeling process that such a brokerage research group would undertake. In terms of expected returns models our analysis shares some of the methods also considered by Gu et al. (2018).

For the purpose of our analysis, let $r_{i,t-} = (r_{i,t}, \ldots, r_{i,t-k+1})'$ be the $k \times 1$ vector of own-lagged stock returns for stock i. We have $N = 1076$ stocks/ETFs to consider and $T = 110$ monthly periods. We use 70% of the data for training, 20% of the data as a validation sample (for hyperparameter tuning), and 10% of the sample for testing out-of-sample performance. To maximize the use of our unique data set, we start building our models using returns data from January 1993 to December 2002 — namely a 10-year sample prior to the start of our individual investor data.

We augment the panel of monthly stock/ETF returns with the five Fama–French monthly factors (Mkt, SMB, HML, RMW, CMA) as well as their momentum factor (see Ken French website for definitions), and Welch and Goyal (2007) predictors: div. price ratio, div. yield, earnings price ratio, div. payout ratio, stock variance, BM DJ stocks, net equity expansion, TBill, long-term yield, term spread, default yield spread, inflation (see their paper for definitions). We understand that a true data engineering group would likely create a much larger and more robust set of data sources. Our goal is not to replicate the true data-source generating process, but to provide a simple approximation to the set of all useful signals for prediction. Let x_t represent an $M \times 1$ vector of these predictors.

In each model class we estimate individual models for each stock/ETF separately, rather than pooling across stocks/ETFs, in order to allow as much heterogeneity as possible in model parameter estimates. The common modeling objective is to estimate $E_t[r_{i,t+1}]$, where $E_t[r_{i,t+1}] = f_i(z_{i,t})$. The modelers therefore employ different approaches to estimate $f_i()$, and also work to curate the best possible set of covariates $z_{i,t}$.

We separate our conditional mean models into (1) linear and (2) nonlinear model sets. Within the linear models, we consider OLS and elastic-net models. For nonlinear models we consider random forests of regression trees and shallow feed-forward neural networks. Hence, we consider two popular nonparametric and parametric machine learning models designed to introduce nonlinear interactions between covariates: random forests of regression trees (nonparametric) and artificial neural networks (parametric). Finally we consider a simple model ensemble across all models.

Linear models. The linear models we estimate for each stock i across time periods $t = k, \ldots, T-1$ are of the form:

$$r_{i,t+1} = \beta_{i,0} + \beta_{i,r} r_{i,t} + \beta_{i,x} x_t + \epsilon_{i,t+1} \tag{3}$$

In addition to estimating this model with OLS, we fit sets of linear models per stock i using Elastic Net involving two tuning parameters (α, λ) that we optimize over the validation sample.

$$\mathcal{L}_i(\theta) = \frac{1}{T-k}\sum_{t=1}^{T} \epsilon_{i,t+1}^2 + \alpha\lambda \sum_{m\in\#\beta} |\beta_m| + \frac{1}{2}(1-\alpha)\lambda \sum_{m\in\#\beta} \beta_m^2 \tag{4}$$

where $\beta = (\beta_{i,0}\beta_{i,r}\beta_{i,x})'$

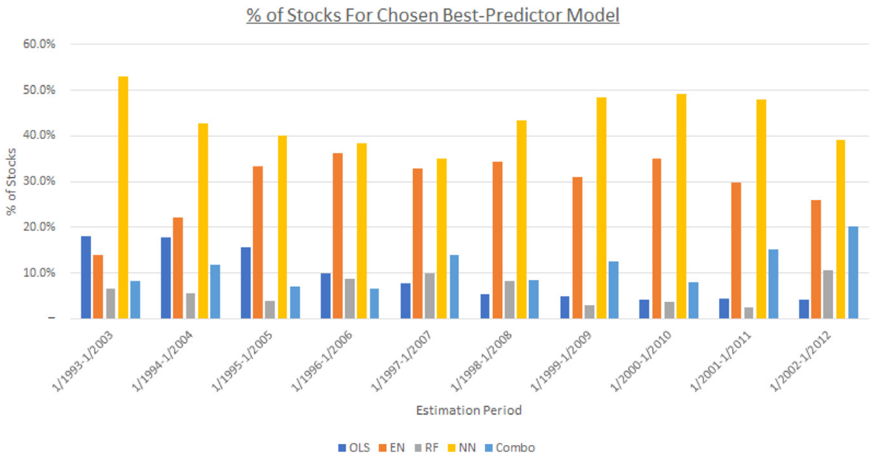


Fig. 1. Bar charts for 10-year rolling samples are displayed, where we only display yearly snapshots. The first covers the sample Jan 1993–Jan 2003 and the last Jan 2002–Jan 2012. For each of the 10 rolling samples the relative performance of the competing models (only looking at equities) is displayed. The bars add up to 100% for each of the 10 rolling samples. The out-of-sample (OOS) performance is measured in terms of MSE and the height of each bar represents the percentage a particular model has the lowest MSE in predicting the cross-section of returns for all the stocks in the sample. The models are OLS, Elastic Net (EN), Random Forest (RF), Neural Net (NN) and Ensemble (Comb). We use 70% of the data for training, 20% of the data as a validation sample (for hyperparameter tuning), and 10% of the sample for testing OOS performance. The bar charts pertain to the OOS performance.

Nonlinear models. We consider two popular nonparametric and parametric machine learning models designed to introduce nonlinear interactions between covariates: random forests of regression trees (nonparametric) and artificial neural networks (parametric). We employ the algorithm of Breiman (2001) to estimate random forest models and we use stochastic gradient descent to minimize an ℓ_2 objective function with regularization terms in order to train the neural networks. In both cases our estimation techniques are standard. Again we estimate the model on the training data and optimize all respective tuning parameters on the validation set.

Random forest. A random forest is a combination of individual regression trees. It is a bootstrapping method that seeks to avoid both overfitting and decrease correlation among trees by using random subsets of predictors at each branch of a given tree. Each tree can be classified as having K terminal nodes (called "leaves") with a depth of L. The prediction of a given tree then can be stated as:

$$h(z_{i,t};\beta,K,L)=\sum_{k=1}^{K}\beta_k 1\{z_{i,t}\in P_k(L)\} \tag{5}$$

where $P_k(L)$ is the kth partition that has at most L different branches that it considers. A set of branches for a given partition can be represented as a product of indicators for sequential branches. For a given partition, then $\hat{\beta}_k$ is the average of the returns for all members of that given partition. A standard greedy search algorithm is used to maximize the information gained at each split. The recursive binary splitting algorithm continues until a set of stopping criterion are met, which typically rely on the maximal additional information gained from a split being less than a threshold, or a max number of leaves and/or depth of a tree being reached.

For the random forest models, the key tuning parameters are the number of bootstrapped trees, the depth of each tree, and the random subset of predictors that are considered at each potential split within a tree. The random forest prediction is then the bootstrapped average at any prediction point across trees.

Neural network. Our neural network architecture is two hidden layers with 10 neurons per layer, sigmoid transfer functions in the input and hidden layers, and a linear transfer function in the output layer. We use stochastic gradient descent to minimize an ℓ^2 objective function with regularization terms in order to train the neural networks. In both cases our estimation techniques are standard. Again we estimate the model on the training data and optimize all respective tuning parameters on the validation set.

Finally we consider a model ensemble of the above linear and nonlinear estimators, restricting ourselves to an equal-weighting scheme across predicted expected returns as to limit introducing additional estimation uncertainty.

Fig. 1 displays a set of 10 bar plot clusters. Each displays end-of-year (last quarter) snapshots of forecasting performance. The 10 rolling samples displayed, each pertaining to a 10-year sample of return data to estimate, validate and forecast returns. For each of the 10 rolling samples the relative performance of the competing models (only looking at equities) is displayed. The out-of-sample performance is measured in terms of MSE and the height of each bar represents the percentage a particular model has the lowest MSE in predicting the cross-section of returns for all the stocks in the sample. For each cluster the height of the bars add up to 100% and each represents the fraction a particular class of models provides the best return prediction for the 683 stocks in the cross-section. We note that neural network models represent the most successful class of models, typically being the best for between 40 and 50 percent of the assets in the cross-section. Often a close second is the class of Elastic Net models. All other methods are less successful, although there is quite some variation across time.

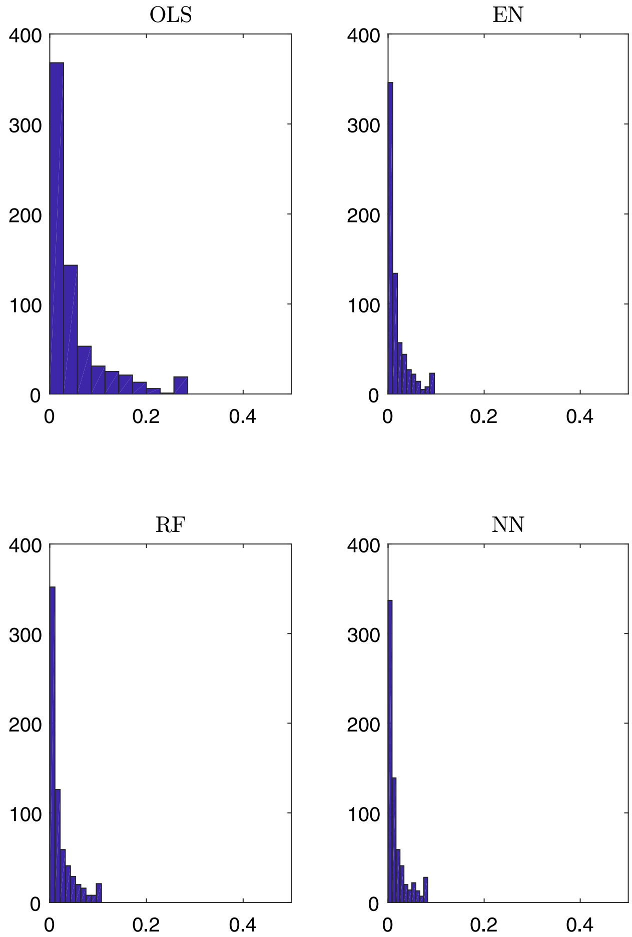


Fig. 2. The models are OLS, Elastic Net (EN), Random Forest (RF), and Neural Net (NN). The out-of-sample (OOS) performance is measured in terms of MSE. Histograms pertain to MSE for 1994–2004 subsample pertaining to the cross-section of returns. We use 70% of the data for training, 20% of the data as a validation sample (for hyperparameter tuning), and 10% of the sample for testing OOS performance. The histograms pertain to the cross-section of OOS MSE performance.

The results displayed in Fig. 1 may leave the impression that neural network models are dominant. Let us turn our attention to Fig. 2 which sheds perhaps a different light on this result. It provides a collection of four histograms documenting for a particular 10-year rolling sample, the 1994–2004 period, the cross-section of out-of-sample MSE's. The models are OLS, Elastic Net (EN), Random Forest (RF), and Neural Net (NN). The out-of-sample (OOS) performance is measured in terms of MSE. Eyeballing the four histograms we see that OLS is clearly worse than EN (Elastic Net), RF (random forest) and NN (neural network), but the differences among the three ML methods is not as clear. The cross-section of MSE's appears indeed similar. Further evidence of this appears in Table 1 which reports the average MSE and MAE of out of sample forecasts across all assets and rolling sample schemes. It shows that the elastic-net and neural net models deliver the lowest out-of-sample MSE when aggregating performance across stocks/ETFs. However, the differences between EN and NN are very small, indicating that while NN perhaps provides the best predictions, EN is typically a close second and arguably much easier to implement. Moreover, the EN is a linear model, whereas the NN is nonlinear. The presence of nonlinearities does not seem to substantially pay off.

All the models/estimators have dimensions on where they could be refined, but ultimately the modeling group delivers a set of conditional mean estimates by stock/ETFs to the robo-investors. Each of these chosen conditional mean estimates come from the model with the lowest out-of-sample MSE. A common model need not be chosen across stocks/ETFs, and indeed we can see that even in a few cases, OLS with all covariates included is the model with the best out-of-sample performance. The final panel of $(\hat{E}_t[r_{i,t+1}])_{i,t}$ is used in the robo-investors' optimal portfolio problems.

4. Empirical results

The empirical results focus on answering a number of questions: (a) who gains from robo-advice, (b) how does robo-investing perform during a major financial crisis, and (c) how do AI Alter Egos compare to passive investment schemes? A subsection is devoted to each of these questions.

Table 1
Out-of-sample MSE across stocks.

	Cross-sectional MSE				
	OLS	EN	RF	NN	Comb
Mean	0.0207	0.0104	0.0114	0.0100	0.0101
Median	0.0147	0.0072	0.0081	0.0066	0.0070

Notes: Cross-sectional average and median MSE's on the out-of-sample testing data for: OLS, Elastic-Net (EN), Random Forest (RF), Neural Network (NN), and ensemble (Comb).

4.1. Who gains from robo-advise?

Using quarterly rebalancing and MV ML/Rolling Variance robo-investors, we report in Table 3 the median, Q1, Q3 as well as confidence interval for the median of the cross-sectional distributions of the spreads between AI Alter Ego and individual investor returns, considering the entire universe of 683 stocks and 393 ETFs.[12] We also report the differences between Sharpe ratios of the AI Alter Ego and individual investors. Panel A covers individual investor realized returns while Panel B refers to realized returns adjusted for cash holding. Panel C covers results with returns adjusted for cash holdings and transaction costs. Summary statistics are computed for separate samples with low/high education, low/high risk aversion and low/high income classification for investors.

Let us start with high versus low risk aversion investors in Panel A. High risk averse median individual investors stand to gain 5.14 percent from robo-investor shadow Alter Egos. Their low risk aversion counterparts only gain 3.29. Both clearly benefit, since the confidence intervals for either type of investor indicates that the median spreads are significantly different from zero. In addition, the 95% confidence interval for the difference in medians is $[0.5546, 3.1111]$, and therefore excludes zero. Hence, the median high risk averse investor gains statistically significantly more from robo-investing than the median low risk averse investor does. A similar pattern emerges for high/low income, with the median low income investor gaining roughly two-thirds more (4.13 percent versus 2.76) than the high income median investor. Low and high education differences are not as pronounced, with a wedge of 63 basis points. The inference indicates, however, that the high/low median spreads for income and education are not statistically significant. Needless to say that a spread between median 2.76 (high income) and 4.13 (low income) percent return per year is economically quite substantial.

When we examine the differences between the Sharpe ratios, we note that the gains in terms of risk-adjusted returns are typically above 30 basis points for the median and more than 20 basis points in the left tail. The differences in median Sharpe ratios across the three classifications are not statistically significant, although for low versus high income investors it is close to 10 basis points per annum.

The results appearing in Panel B, which account for investor cash holdings discussed in Appendix A.5, are slightly different. The difference in medians between high and low risk averse investors is lower (42 basis points) and no longer significant. As far as income is concerned, the median low income investor is still gaining more from robo-investing than the median high income investor does (2.92% versus 1.65%). By contrast with what we observe in Panel A, such a spread of 1.27 percent is now statistically significant and is still economically substantial. Similarly, low and high education differences are also statistically significant for cash-adjusted realized returns. They reveal that low education investors gain more than high education investors, with a wedge of 1.14 percent.

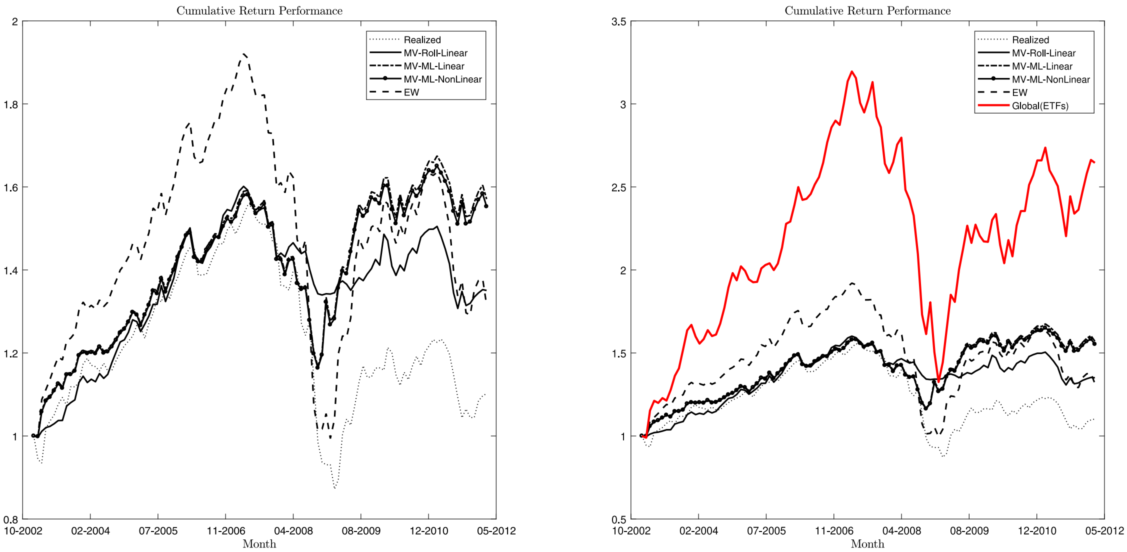
Turning our attention to difference in Sharpe ratios appearing in Panel B, we find the differences typically slightly less than 30 points. This time, however we see statistically significant improvements for low education and low income investors.

Panel C reports spreads for cash-adjusted returns net of transaction costs based on the calculations documented in Appendix A.6. We find a further reduction in spreads, but many of the patterns remain. Notably low income investors with a spread of 1.77% benefit more than high income who earn less than half that spread. It is interesting to note that for high income investors, the median spread is the only case where the null of a zero spread is not rejected. The differential impact of the robo-advisors helping low versus high education investors now increases when accounting for transaction costs. The median spread are significantly different from each other as well. Low education investors earn a 2.47% extra return when following the Alter Ego robo-advisor versus less than half that for high education ones. The results for the risk aversion classification are close to those in Panel B, but different from those with raw returns in Panel A. The Sharpe ratio spreads are on the order of 20 basis points — and for education the differential is small but statistically significant.

4.2. How does robo-advising perform during a major financial crisis?

In Table 4 we report returns and Sharpe ratios for individual investors and their Alter Egos as they relate to the 2008 Great Recession (before, during and post). The subsamples are benchmarked using the NBER chronology identifying the crisis period as

[12] To construct confidence intervals for aggregate summary statistics we do the following. We first randomly sample individuals according to an individual bootstrap method whereby each investor is assumed independent of each other investor, and sample the entire time-series path of each investor to maintain the dependence structure. For each bootstrap repetition we compute the relevant statistic per individual and aggregate the per-individual statistics over all of the sampled investors. Let $\{\tilde{\theta}_r\}_{r=1}^{R}$ be the constructed statistic over R bootstrap repetition, and let $\hat{\theta}$ be the point estimate of interest. Let $\tilde{\theta}_{(\alpha/2)}$ and $\tilde{\theta}_{(1-\alpha/2)}$ represent the $\alpha/2$ and $1-\alpha/2$ percentiles of the bootstrap statistic. We then construct pivotal $1-\alpha$ confidence intervals according to $[2\hat{\theta}-\tilde{\theta}_{(1-\alpha/2)}, 2\hat{\theta}-\tilde{\theta}_{(\alpha/2)}]$.



(a) **Realized versus AI Alter Egos** (b) **Realized vs AI Alter Egos vs Global ETF**

Fig. 3. The lines in panel (a) correspond to (1) cash-adjusted realized cumulative return of median investor, (2) median AI Alter Ego returns using MV Rolling Mean/Rolling Variance scheme (3) median AI Alter Ego returns using MV ML/Rolling Variance scheme (4) median AI Alter Ego returns using MV ML/Nonlinear Variance and finally (5) the median EW robo-investor. All start out with one unit of investment at the beginning of the sample and median returns are compounded. Panel (b) contains those very same cumulative returns — although on a different scale to allow comparison with the one series added, Global ETF Robo-investor which supercedes and dominates all other investment strategies.

12/2007 - 6/2009.[13] The focus is again on the MV ML/Rolling Variance AI Alter Ego scheme. For each of the subsamples we compute the median, Q1 and Q3 realized returns and Sharpe ratios along with the same statistics for the AI Alter Ego returns. Note that, since the median of a spread is not the difference in median returns, we are not inferring something directly related to the spreads reported in prior tables. We focus on the returns (and Sharpe ratios) instead in order to highlight a very important finding.

Prior to the crisis we note that the median investor had an annual return of 9.26%, almost double the return of the median AI Alter Ego (4.17%). We also note though that the inter-quartile spread for investors is twice as large as the same statistic for robo-investors using ML. For individual investors the Q1–Q3 ranges from −4.70 to 20.98 percent, whereas the AI Alter Egos feature a better Q1 of minus two percent, and a lower Q3 of almost eleven percent. The comparison is however less dramatic when considering individual investor realized returns adjusted for cash holding or adjusted for both cash holding and transaction costs. Adjusted for cash holdings, the median investor earned an annual cash-adjusted return of 6.35% and the corresponding inter-quartile spread is unsurprisingly narrower (from −0.93 to 13.57 percent). While accounting for cash holdings, investors still outperform the MV ML/Rolling Variance AI Alter Ego scheme, but to a lesser extend. Including transaction costs completely erases the advantage of individual investors as we note a median return of only 2.40%. When we adjust the median AI Alter Ego of 4.17% for transaction costs, using the 30 basis points fee as explained in Appendix A.6, there is a significant difference in returns. Compared to the Alter Egos the inter-quartile range is also much worse for individual investors. Turning our attention to Sharpe ratios we actually observe little variation across the three configurations, except when we adjust for both for cash and transaction costs.

During the crisis things take a dramatic turn. The median robo-investor has zero return — meaning the median AI Alter Ego holds cash. In contrast, for individual investors the median is a loss, i.e. -29% or −16.42% when investor cash holdings are included (or −17.39% if transaction costs are added). With or without cash holdings, the Q3 investor still has a negative annual return (either −3.59% or −0.82%), compared with 23.81 percent return for Q3 of the AI Alter Egos. Adding transaction costs yield results similar (but somewhat worse as expected) to the cash holding adjusted findings. The Sharpe ratios obviously reflect the dismal return results.

After the crisis, things reverse to the pattern observed prior to the crisis — namely the median investor does better than the median AI Alter Ego, but once cash holdings and transaction costs are taken into account this is no longer true. In fact, throughout all subsamples the Alter Ego investor outperforms the individual investors once transaction costs and cash holdings are considered.

A more striking picture emerges when we turn our attention to Panel (a) of Fig. 3. The five lines correspond to (1) cash-adjusted realized cumulative returns of median investor, (2) median AI Alter Ego returns using MV Rolling Mean/Rolling Variance scheme (3) median AI Alter Ego returns using MV ML/Rolling Variance scheme (4) median AI Alter Ego returns using MV ML/Nonlinear

[13] We also examined the more specifically targeted Belgian crisis dates related to the severe difficulties of the country's financial sector. The results are broadly speaking similar and not reported here.

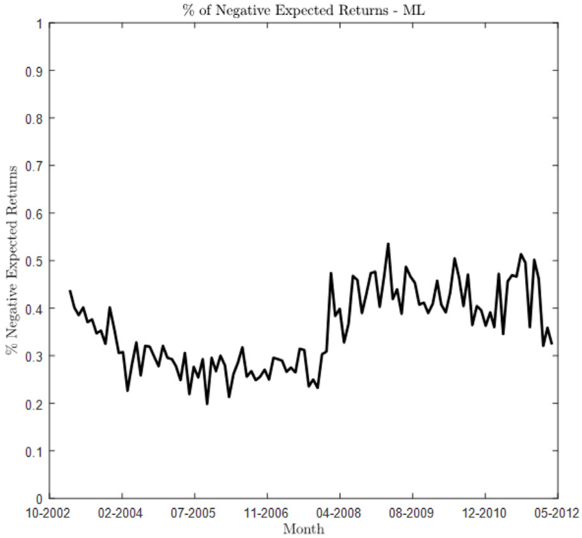


Fig. 4. Time series plot of the fraction among the cross-section of stocks with negative expected returns, according to the best machine learning model — see Fig. 1 for details.

Variance and finally (5) the median EW robo-investor.[14] One word of caution: these medians do not represent the same investor or AI Alter Ego through time, so this is not the performance of a specific individual or robot. Each line starts out with one unit of investment at the beginning of the sample and the median returns are compounded subsequently. Prior to the crisis, the median investor reaches roughly 1.5. This means that the initial capital is increased by 50% over a five year span from 2002 until 2007. By the time the devastation of the crisis took its toll, the median investor is under water by about 15 percent and finally ends up with a meager 10 percent return over a 10-year period. It is remarkable that even the EW robo-investor achieves a higher return at the end of the sample. The best overall performance is obtained from the MV ML/Rolling Variance median robo-investor (again not shadowing always the same investor across time) with a 60 percent overall return. This median robo-investor has a relatively slow start but already slightly over-performs prior to the crisis. In addition, it features small losses during the tumultuous market conditions. Note also that the MV ML/Nonlinear Variance AI Alter Ego is almost identical to the ML/Rolling Variance scheme. Finally, the MV Rolling Mean/Rolling Variance scheme tracks the ML performance very closely until the financial crisis.

To shed further light on this we turn our attention to Table 2 displaying the ranking of the regressors based on their ℓ_2 contribution across stocks for the Elastic Net regressions defined in Eqs. (3)–(4). We focus on the EN regressions as they provide a fairly simple regression-based interpretation. In addition, it is often the best or nearly the best prediction model. The ranks are computed for 10-year rolling samples starting with 93-03 and ending with 02–12. Of particular interest is the crisis period spanning across the 96–06 through 99–09 samples. The top ranked predictor in all but the last of rolling samples is dfy namely the default yield spread. Another top-ranked series during the crisis is lty or the long term yield. Looking across all samples we also see $svar$ stock variance, $ntis$ net equity expansion and $infl$ inflation. Interestingly, the usual Fama–French regressors rarely appear among the top-ranked regressors. This should not perhaps come as a surprise, since the Fama–French factors are meant to price the cross-section of returns.

In Fig. 4 we provide a time series plot of the fraction with negative expected returns among the cross-section of stocks, according to the best machine learning model. Early in the sample we see that typically between 20 and 30% of the stocks featured negative expected returns. The fraction shoots up above 50% in 2008 and goes as high as 60%. As a result, the majority of stocks featured negative expected returns, which explains why the AI Alter Egos have a propensity to move out of the market.

Finally, in Table 4 we also reports on a variation of the Alter Ego scheme. So far the MV involved maximizing the Sharpe ratio across all investors. As an alternative, we examine risk aversion bespoke Alter Egos. More specifically, the bespoke Alter Ego schemes are optimal portfolio returns with γ tailored on the basis of the survey-based risk aversion measure with Low, Mid, and High involving $\gamma = 1$ (low), 5 (mid) and 10 (high) for the MV ML scheme (the Mean–Variance with Machine Learning Mean/Rolling Variance scheme). Before the crisis, Alter Egos fine-tuned for risk aversion do outperform the generic counterpart, although only to a minor degree, and actually under-performs in terms of Sharpe ratios. During the crisis the bespoke Alter Ego setup has a significant impact, however. Low risk averse Alter Egos have a median positive return of 4.15% whereas high risk averse ones have a return of 1.59%, both higher than the cash zero return. While the difference in median is remarkable, a word of caution is in order when we also take into account the Q1 and Q3 results. The inter-quartile range is huge with more than 30% losses for both low and mid risk averse AI Alter Egos while the Q3 is 104.93% is only true for low risk averse Alter Egos. After the crisis, the bespoke scheme falls flat. Hence, overall the findings regarding Alter Egos tailored to investor risk aversion characteristics are mixed.

[14] For comparison purposes, we also endow the robo-investor with a (DeMiguel et al., 2007) equal weighting strategy. In particular, for each individual i; the Alter Ego buys and holds at time t all the stocks in the set $S_i t$ with equal allocations $1/N_{it}$. Henceforth we will refer to this as the EW portfolio rule.

Table 2
Ranked variables based on relative ℓ_2 contribution across stocks.

Rank	93–03	94–04	95–05	96–06	97–07	98–08	99–09	00–10	01–11	02-12
1	SMB	svar	ntis	dfy	dfy	dfy	lty	Mkt-RF	ntis	svar
2	dfy	lty	lty	lty	svar	ntis	dfy	lty	svar	dfy
3	lty	infl	infl	svar	SPvwx	lty	svar	svar	Mkt-RF	lty
4	RMW	dfy	tbl	infl	infl	svar	tbl	dfy	infl	infl
5	infl	Mkt-RF	svar	HML	Mkt-RF	tbl	Mkt-RF	SPvw	SPvw	ntis
6	ntis	SMB	HML	tbl	dy	Mkt-RF	infl	ntis	CMA	RMW
7	HML	RMW	RMW	Mkt-RF	lty	infl	ntis	tbl	bm	tbl
8	Mkt-RF	ntis	dy	bm	RMW	dy	SPvw	CMA	lty	SPvwx
9	svar	HML	bm	SMB	ntis	SPvwx	RF	bm	ltr	SPvw
10	SPvw	bm	dfy	SPvw	dp	dp	SPvwx	RMW	SMB	Mkt-RF
11	Mom	Mom	Mkt-RF	RF	SMB	SPvw	HML	SMB	SPvwx	CMA
12	tbl	ltr	ep	ntis	tbl	bm	CMA	HML	dfy	HML
13	ltr	dy	RF	dy	bm	SMB	RMW	SPvwx	HML	SMB
14	de	SPvw	SMB	CMA	HML	RMW	bm	dy	dy	ltr
15	bm	CMA	dp	RMW	RF	HML	SMB	Mom	Mom	dp
16	SPvwx	de	CMA	SPvwx	SPvw	CMA	ltr	dp	dp	RF
17	RF	SPvwx	SPvw	Mom	Mom	RF	dy	ltr	tbl	Mom
18	CMA	dp	Mom	ltr	de	de	dp	infl	RMW	bm
19	dy	tbl	ltr	ep	ep	ep	Mom	ep	ep	dy
20	ep	RF	de	dp	CMA	ltr	de	de	RF	ep
21	dp	ep	SPvwx	de	ltr	Mom	ep	RF	de	de

Notes: Elastic Net regressions defined in Eqs. (3)–(4) involve the following set of regressors: dp Dividend/Price, dy Dividend Yield, ep Earnings/Price, de Dividend Payout, $svar$ Stock Variance, bm Book-to-Market, $ntis$ Net Equity Expansion, tbl T-Bill Rate, lty Long Term Yield, ltr Long Term Return, dfy Default Yield Spread, $infl$ Inflation, $SPvw$ S&P 500, $SPvwx$ S&P 500 (excl. dividends), the following Fama–French factors $Mkt-RF$ Market, SMB, HML, RMW, CMA, RF, and Mom Momentum (see Ken French website for definitions), and Welch and Goyal (2007) for definitions.

4.3. AI Alter Egos versus passive investments

How do AI Alter Egos measure up against passive investment strategies, in particular buying and holding a market-wide ETF? To address this question we turn our attention to Table 5. We report summary statistics for both return and Sharpe ratio spreads with respect to two ETFs. One tracks the S&P 500 index and the other is the iShares MSCI Belgium ETF. Neither is ideal, but we did not find an index available throughout the entire sample period that mimics the basket of stocks held by the investors in the brokerage data set.[15] Unfortunately, the results reported in Table 5 depend on which ETF is selected.

When focusing on returns in Table 5, the right panel (first three columns) displays the results for investor realized (stocks+ETFs) returns minus the benchmark ETF spreads (either S&P 500 or Belgian ETF). Similarly, the left panel (next three columns) provides the results for AI Alte ego MV/ML/Nonlinear returns. Each panel contains the median, first and third quartile of the spreads. The full sample results appear in the top part of Table 5. Subsamples stratified according to NBER crisis dates appear in the lower part. The median investor has a spread of −8.50% against the S&P 500, meaning the median investor vastly under-performs the benchmark. For the Belgian ETF the results are not as dramatic, since the median investor does better with a positive spread of 1.37%. There is wide cross-sectional variation, although the third quartile for the US market index is only 2% (while 12% for the Belgian index). The AI Alter Ego spreads are better in both cases, although the US benchmark still yields a negative spread of −6.18%. Against the Belgian ETF, the AI Alter Ego has a positive median spread of almost 4 percent.

When we look at the pre-crisis sample we note that the median investor and AI Alter Ego have returns below the two benchmarks, more so for the Belgian ETF than its US counterpart. It is also worthwhile noting that the median AI Alter Ego performs worse when focusing on returns (Sharpe ratio spreads are still negative but higher for AI Alter Egos). The crisis period is a totally different story. The AI Alter Egos median investors vastly outperform the benchmark by respectively 18.32% (US benchmark) and 48.31% (Belgian benchmark). Moreover, the median investor does better than the Belgian ETF by a substantial margin of 21.24% but is 8.78% below the S&P 500 ETF. In both cases we see significant improvements from the Alter Ego schemes. Post-crisis things return back to the pre-crisis situation. Finally, the differences for Sharpe ratios provide overall consistent findings with those for returns and adjusting returns for cash holdings or transactions also yield comparable results.

5. Conclusions

Artificial intelligence enhancements are increasingly shaping our daily lives. Financial decision-making is no exception to this. We introduce the notion of AI Alter Egos, machine-driven decision makers which shadow a particular individual, and apply it in the area of robo-investing using a brokerage accounts data set rich in both cross-sectional and time series features.

The purpose of our analysis is to assess the highly touted benefits of robo-advising. Through the AI Alter Ego scheme we address a number of questions: (a) who gains from robo-advise, (b) how does robo-investing perform during a major financial crisis, (c)

[15] According to the results reported in Table A.3 investors hold 26% of US stocks and 14% of Belgian stocks.

Table 3
AI Alter Egos return and sharpe ratio spreads — education, risk aversion and income.

	Return spreads			CI Median		Sharpe ratio spreads		
	Median	Q1	Q3	C(2.5%)	C(97.5%)	Median	C(2.5%)	C(97.5%)
Panel A: Realized returns								
Education								
Low	2.83	−10.47	16.79	1.18	4.22	0.29	0.24	0.34
High	3.46	−10.15	17.26	3.01	3.85	0.31	0.30	0.33
Confidence interval difference in medians: [−0.5989, 2.1723]						[−0.0266, 0.0475]		
Risk Aversion								
Low	3.29	−10.86	17.43	2.39	4.38	0.31	0.29	0.32
High	5.14	−9.12	17.28	3.63	6.36	0.37	0.33	0.41
Confidence interval difference in medians: [0.5546, 3.1111]						[−0.0241, 0.0526]		
Income								
Low	4.13	−10.00	17.57	3.01	5.20	0.35	0.32	0.38
High	2.76	−11.70	15.01	0.82	4.91	0.26	0.21	0.32
Confidence interval difference in medians: [−3.1016, 0.6014]						[−0.0620, 0.0329]		
Panel B: Realized returns - cash adjusted								
Education								
Low	3.17	−4.78	12.54	2.35	3.94	0.30	0.27	0.33
High	2.03	−5.23	10.75	1.80	2.40	0.25	0.24	0.26
Confidence interval difference in medians: [−1.7511, −0.3788]						[−0.0573, −0.0028]		
Risk Aversion								
Low	2.56	−5.21	12.03	2.24	3.01	0.27	0.25	0.29
High	2.14	−5.26	10.26	1.35	2.78	0.26	0.22	0.30
Confidence interval difference in medians: [−1.1906, 0.2726]						[−0.0518, 0.0160]		
Income								
Low	2.92	−4.86	12.51	2.40	3.59	0.29	0.27	0.31
High	1.65	−4.58	8.71	0.92	2.82	0.25	0.21	0.30
Confidence interval difference in medians: [−2.1572, −0.2615]						[−0.0754, −0.0036]		
Panel C: Realized returns - adjusted for cash and transaction costs								
Education								
Low	2.47	−6.28	13.47	1.47	3.58	0.23	0.20	0.26
High	1.10	−7.12	11.03	0.75	1.43	0.19	0.18	0.21
Confidence interval difference in medians: [−2.5106, −0.6079]						[−0.0968, −0.0200]		
Risk Aversion								
Low	1.78	−6.55	12.44	1.22	2.32	0.22	0.20	0.23
High	1.44	−6.89	10.86	0.40	2.68	0.22	0.17	0.26
Confidence interval difference in medians: [−1.2788, 0.7945]						[−0.0446, 0.0409]		
Income								
Low	1.77	−6.60	12.20	1.07	2.48	0.21	0.18	0.24
High	0.85	−6.08	9.93	−0.38	2.14	0.22	0.16	0.27
Confidence interval difference in medians: [−1.9844, 0.1428]						[−0.0910, 0.0132]		

Notes: Entries in the first set of columns are median, first and third quartiles, as well as confidence interval for the median of the cross-sectional distributions of the spreads between AI Alter Ego and individual investor returns, considering the entire universe of 683 stocks and 393 ETFs. The second set of columns are the Sharpe Ratio spread medians and their corresponding 95% confidence intervals. The AI Alter Ego scheme is Mean–Variance (MV) with Machine Learning (ML) Mean/Rolling Variance — using the methods displayed in Table 1. Summary statistics are computed for separate samples with low/high education, low/high risk aversion and low/high income classification for investors. 95% confidence intervals for differences in medians are computed as described in footnote 12. The return and Sharpe ratio spreads are in percentages per year. The Sharpe ratios are computed with a zero risk-free rate.

how do AI Alter Egos compare to passive investment schemes? Overall, we find that investors displaying certain characteristics – in particular low education and low income – stand to gain significantly. Moreover, machine learning methods provide important portfolio return improvements. During the financial crisis, robo-investors have a greater propensity to cash out of the market, which contributes to their overall return superiority. Risk aversion bespoke Alter Egos outperform the generic Alter Ego before and during the financial crisis. Results based on Sharpe ratios broadly yield the same findings.

Compared to passive ETF investment, we find that the evidence is mixed, although during the financial crisis AI Alter Egos were vastly better than the passive strategy.

Table 4
Returns and sharpe ratios pre-crisis, crisis and post-crisis.

	Returns			Sharpe ratios		
	Median	Q1	Q3	Median	Q1	Q3
	Pre-crisis					
Realized	9.26	−4.70	20.98	0.47	−0.17	0.96
Realized (cash adjusted)	6.35	−0.93	13.57	0.50	−0.08	0.96
Realized (cash and transaction costs)	2.40	−6.56	10.15	0.22	−0.53	0.79
MV ML	4.17	−2.00	10.70	0.43	−0.21	0.93
Risk Aversion Bespoke Alter Ego						
MV ML Low	4.76	−1.51	13.08	0.22	−0.20	0.53
MV ML Mid	4.99	−1.82	13.08	0.23	−0.21	0.53
MV ML High	5.77	−0.93	14.91	0.24	−0.21	0.54
	During crisis					
Realized	−29.04	−45.87	−3.59	−0.69	−1.04	−0.09
Realized (cash adjusted)	−16.42	−30.89	−0.82	−0.69	−1.06	−0.10
Realized (cash and transaction costs)	−17.39	−33.27	−1.13	−0.74	−1.17	−0.10
MV ML	0.00	−13.81	23.81	0.00	−0.75	0.84
Risk Aversion Bespoke Alter Ego						
MV ML Low	4.15	−32.37	104.93	0.16	−0.65	0.83
MV ML Mid	3.70	−33.33	92.38	0.16	−0.65	0.84
MV ML High	1.59	−18.84	64.20	0.18	−0.66	0.85
	Post-crisis					
Realized	5.64	−6.47	15.30	0.25	−0.19	0.65
Realized (cash adjusted)	3.79	−1.26	9.80	0.33	−0.10	0.73
Realized (cash and transaction costs)	1.70	−5.17	8.73	0.15	−0.40	0.64
MV ML	2.05	−1.76	8.40	0.23	−0.17	0.64
Risk Aversion Bespoke Alter Ego						
MV ML Low	0.22	−33.00	57.44	0.10	−0.45	0.58
MV ML Mid	0.65	−29.15	53.64	0.10	−0.45	0.57
MV ML High	0.00	−11.50	23.11	0.08	−0.44	0.57

Notes: The subsamples are benchmarked based on the NBER Crisis Time Period 12/2007 - 6/2009. The Pre-crisis sample starts in 2002 and ends 11/2007, the post-crisis sample covers 7/2008 until end of sample, 2012. MV ML refers to the AI Alter Ego scheme is Mean–Variance with Machine Learning Mean/Rolling Variance — using the methods displayed in Table 1. The returns and Sharpe ratios are in percentages per year and the risk-free rate is equal to zero. The Risk Aversion Bespoke Alter Ego are optimal portfolio returns for risk-aversion specific models using the Mean–Variance with Machine Learning Mean/Rolling Variance scheme. Low, Mid, and High refer to the surveyed risk-aversion choice. The schemes fix $\gamma = 1$ (low), 5 (mid) and 10 (high) for the MV ML scheme.

CRediT authorship contribution statement

Catherine D'Hondt: Research, Writing of the paper. **Rudy De Winne:** Research, Writing of the paper. **Eric Ghysels:** Research, Writing of the paper. **Steve Raymond:** Research, Writing of the paper.

Appendix A. Data descriptions

A.1. Data sources

Our primary data set comes from a large Belgian online brokerage firm and consists of the trading accounts of 22,972 individual investors. This unique data spans about 10 years from January 2003 to March 2012, which includes the 2008 financial crisis. We have detailed information about each trade, i.e. the ISIN code of the instrument, the time-stamp, the trade direction, the executed quantity, the trade price, and the explicit transaction costs.[16] For the purpose of this research, we focus on common stocks and ETFs investments and exclude other financial instruments. Over the sample period, 6,741 investors also traded options and warrants for an aggregate number of 602,833 trades and 6,665 investors traded mutual funds for an aggregate number of 260,120 trades. Only a few investors (i.e. 1,813) traded bonds for an aggregate number of 5,999 trades. Overall, the individual investors trade across 12,818 different stocks but we count for most of the stocks only a few trades. We therefore filter the trades and keep the 683 stocks

[16] We also know the currency in which the trade is executed, which allows us to compute monetary volumes in euros using historical exchange rates from the European Central Bank and Bloomberg.

Table 5
AI Alter Ego return and sharpe ratio spreads vis-à-vis benchmark ETFs.

	Returns						Sharpe ratios					
	Realized stocks+ETFs minus ETF			AI Alter Ego minus ETF			Realized stocks+ETFs minus ETF			AI Alter Ego minus ETF		
	Median	25q	75q	Median	25q	75q	Median	C(2.5%)	C(97.5%)	Median	C(2.5%)	C(97.5%)
	Full sample											
S&P 500 ETF	−8.50	−21.05	2.00	−6.18	−13.99	1.32	−0.47	−0.48	−0.46	−0.14	−0.15	−0.14
Belgian ETF	1.37	−9.90	12.00	3.93	−4.89	13.10	−0.22	−0.22	−0.21	0.11	0.11	0.12
	Pre-crisis											
S&P 500 ETF	−2.22	−16.02	9.18	−7.20	−13.88	−0.41	−0.98	−1.01	−0.96	−0.77	−0.79	−0.77
Belgian ETF	−9.57	−21.28	1.45	−14.58	−21.33	−7.02	−1.03	−1.05	−1.01	−0.82	−0.83	−0.81
	During crisis											
S&P 500 ETF	−8.78	−27.02	12.48	18.32	−1.75	39.32	−0.01	−0.03	−0.00	0.44	0.42	0.45
Belgian ETF	21.24	1.23	40.35	48.31	25.60	71.02	0.31	0.30	0.33	0.76	0.74	0.78
	Post-crisis											
S&P 500 ETF	−11.37	−23.16	−1.78	−14.97	−19.62	−8.43	−0.75	−0.77	−0.75	−0.78	−0.79	−0.78
Belgian ETF	−1.48	−12.59	8.26	−4.82	−9.85	2.36	−0.11	−0.12	−0.10	−0.14	−0.15	−0.13

Notes: The AI Alter Ego scheme is the ML Mean/Nonlinear Smoothed Variance — using the methods displayed in Table 1. The returns and Sharpe ratios are in percentages per year and the risk-free rate is equal to zero. The subsamples are benchmarked based on the NBER Crisis Time Period 12/2007 - 6/2009. The Pre-crisis sample starts in 2002 and ends 11/2007, the post-crisis sample covers 7/2008 until end of sample, 2012. The benchmark ETFs are the SPDR ETF tracking the S&P 500 index and the iShares MSCI Belgium ETF.

Table A.1
Types of assets in the data base and prevalence of trading.

	Income		Risk aversion		Education	
	High	Low	High	Low	High	Low
%investors						
Stocks	100%	100%	100%	100%	100%	100%
ETFs	30%	18%	21%	22%	22%	14%
Mutual funds	37%	26%	28%	29%	31%	18%
Options	13%	8%	9%	13%	11%	4%
Warrants	27%	18%	20%	22%	20%	13%
Bonds	9%	7%	8%	7%	9%	4%

The table shows the types of assets in our database as well as the proportion of active investors — low/high income, risk aversion and education.

that are most traded over the sample period. This universe covers 71.5% of all the investors' trading activity on stocks. We end up with a sample of 1,590,199 trades on stocks (and 13,015,509,557 traded volume in euros) over the 111-month period. In addition to stocks, 4,693 investors (i.e. 20.45%) also trade ETF. Over the whole sample period, those ETF investments correspond to 60,344 trades on 393 different ETFs for a total monetary volume of 898,496,821 in euros.

Table A.1 shows the types of assets in our database as well as the proportion of active investors per type of investors — low/high income, risk aversion and education. We note that all investors trade equities. Trading of ETFs, mutual funds, options and warrants is more prevalent with high income/education investors. Trading of bonds is overall insignificant. Because we examine robo-advisors which are mean–variance investors we focus exclusively on stocks and ETFs which best fit the portfolio allocation model. For high income/education investors in particular this means we leave out to a certain degree other assets which we have available.

A.2. Portfolios and individual investor characteristics

Using the trades data, we build end-of-month portfolios for each investor and use historical market data to compute monthly portfolio market values. We also compute both monthly and daily asset returns.[17] Combining end-of-month portfolio market values with the corresponding monthly aggregate cash-flows, we calculate for each investor 110 monthly portfolio gross and net returns, i.e. from February 2003 to March 2012. To calculate portfolio returns, we opt for an approximation of the Modified Dietz Method, aiming at delivering a return close to the money-weighted rate of return (e.g., Shestopaloff and Shestopaloff, 2007). Specifically, we compute portfolio returns per investor and month assuming that all the purchases (sales) executed in a given month take place on the first (last) day of the month. Mathematically, it gives: $R_t = (EMV_t - EMV_{t-1} - P_t + S_t)/(EMV_{t-1} + P_t)$ where R_t is the portfolio gross return for month t, EMV_t is the end-of-month portfolio value at month t, EMV_{t-1} is the end-of-month portfolio value at month $t-1$, P_t and S_t are the aggregate monetary value of all purchases and sales executed during month t. When calculating net returns, we subtract from the numerator the aggregate monetary value paid on transaction costs during month t.

Robo-investors can hold cash. It is therefore of interest to comment on how investor cash holdings – which we do not observe – affect our return calculations. To that end, let us consider an illustrative example. Suppose an investor has 10 euros (units do not

[17] Historical price data come from both Eurofidai and Bloomberg.

Table A.2
Illustrative examples of cash flows and return computations.

	t-1	Trading	t	t-1	Trading	t	t-1	Trading	t
Stock 1	5	−5	0	5	−5	0	5	0	5
Stock 2	5	0	5	5	5	10	5	5	10
EMV	10		5	10		10	10		15
Cash-in		0			5			5	
Cash-out		5			5				
Return			0%			0%			0%

Table A.3
International coverage of stocks.

Countries	Stocks (%)
USA	26%
France	22%
Germany	18%
Belgium	14%
Netherlands	10%
Others	10%

Trades cover 12,818 different stocks. Those with few trades are discarded and we keep 683 stocks, which cover 71.5% of all the investors' trading activity on stocks. The table shows the main countries where the 683 stocks in our sample originate.

matter) at end of t - 1 with 5 invested in stock 1 and 5 in stock 2. To further simplify the example assume that this investor did not hold any other stocks in the past. The robo-investor decides the allocation between stock 1 and 2 - buying and holding the shares. Furthermore, the investor cashes out 5 euros in stock 1 shortly thereafter (assuming stock 1 does not appreciate). At time t investor only has stock 2 and holds a portfolio worth 5 euros. The robo-investor can buy stock 1 and 2 for a value of portfolio up to 5 euros but suppose expected returns are negative for stock 1 so only holds stock 2. In the end investor and shadow Alter Ego both hold the same portfolio. In this example, the time t + 1 returns from the robo-investor's strategy are going to be independent of the actions that the investor takes with respect to their non-negative (could be 0) allocation between stocks 1 and 2 at time t. The robo-investor only sees that the investment opportunity set has remained the same. Table A.2 provides a few scenarios – including the aforementioned one – which show that the return calculations are relatively robust to various cash inflow and outflow calculations.

Two final additional filters were applied on the resulting portfolio gross returns (prior to transaction costs). First, we drop any investors if they had more than 106 missing values in their return series (i.e. at least 4 months of returns are needed to keep an investor). Second, because of some outliers, we disregard investors for whom we observe at least one month where the absolute value of their return is greater than 100. These two filters decrease the sample from 22,972 to 20,622 investors.

In addition to the above information about trading activity, we have an extensive set of individual investor characteristics, including age, gender, education, but also risk aversion, annual net income, and amount invested in financial markets. Some of the individual characteristics are based on surveys and pertain to gauging attitudes towards risk, i.e. measure risk aversion. These survey-based individual investor measures of risk aversion are collected by the brokerage firm within the context of the MiFID regulation for all EU member states that came into effect in November 2007.[18] European regulation has made it compulsory for brokerage firms to collect specific information about their retail clients' needs and preferences. Accordingly, investment firms operating in the EU are obliged to submit questionnaires (referred to as "MiFID tests") to their clients in order to determine their level of knowledge and experience, their investment objectives as well as their financial literacy. MiFID tests can be viewed as regulated Investment Policy Statements (IPS) required when any retail investor asks for financial advice and/or portfolio management services. It is worth noting that the MiFID regulation does not impose standardized questionnaires. Each brokerage firm is free to devise and organize its own questionnaire(s) provided it abides by some general guidelines.[19]

A.3. Sample of stocks and ETFs

Table A.3 shows that the 683 stocks retained in our sample are international assets. Although we work with Belgian individual investors, most of their trading activities concern foreign stocks (mainly US and neighboring countries) despite the well-known

[18] MiFID stands for the Markets in Financial Instruments Directive. MiFID I (2004/39/EC) is known as the first version of this Directive while a review of it was recently implemented in January 2018 (known as MiFID II (2014/65/UE)). For more details, please visit the European Commission website (http://ec.eurosopa.eu/internal_market/securities/isd/mifid2/index_en.htm).

[19] For more details on the MiFID tests, please refer to Bellofatto et al. (2018).

Table A.4
Stock distribution across industry sectors.

Sectors	Stocks (%)
Technology	16.93
Financials	15.91
Industrials	14.01
Consumer Goods	10.07
Health Care	9.78
Consumer Services	9.63
Basic Materials	9.05
Oil and Gas	8.46
Utilities	3.50
Telecommunications	1.89

Trades cover 12,818 different stocks. Those with few trades are discarded and we keep 683 stocks, which cover 71.5% of all the investors' trading activity on stocks. The table reports the proportion of stocks across sectors based on the Industry Classification Benchmark (ICB) industry classification taxonomy. This information is available for 680 stocks in our sample.

Table A.5
Statistics about ETFs underlying asset (or basket of assets).

Underlying	Multiplier = 1	Multiplier ≠ 1
Panel A: # ETF		
Equities	227	29
Commodities	76	19
Fixed-income	31	0
Currencies	4	1
Real estate	6	0
Panel B: # trades		
Equities	23 053	24 624
Commodities	11 501	436
Fixed-income	512	0
Currencies	11	2
Real estate	205	0

Our sample is made of 393 different ETFs for a total number of 60,344 trades. The table reports aggregate statistics per type of underlying asset (or basket of assets). For each category, we distinguish ETFs based on their multiplier. When the multiplier is equal to one, the ETF simply tracks the underlying asset (or basket of assets). When the multiplier differs from one, the ETF can be leveraged, inverse, or even leveraged-inverse depending on the multiplier value. Panel A gives the number of ETFs while the corresponding number of trades are provided in Panel B.

home bias.[20] This feature of our data is mainly due to the small size of the Belgian stock market. About 150 stocks are listed on the Euronext Brussels Stock Exchange in comparison to respectively, 1,791 and 2,545 domestic companies listed on the New York Stock Exchange and the Nasdaq (WFE Annual Statistics Guide 2017). Industry sectors are listed in Table A.4. The top 3 consist of technology (16.93%), financial (15.91%), and industrial stocks (14.01%).

Table A.5 provides summary information about the ETFs in our sample. Most of them are equities or commodities-linked. Although the majority of ETFs are passive (344 out of 393, i.e. 87%), some use a multiplier different from one. In particular, such leveraged and/or inverse ETFs display the highest number of trades (i.e. 24,624) in the equities category. This is consistent with the top 5 provided in Table A.6. Among the five most traded ETF, four are equities-linked, and, more importantly, three of them are either leveraged, inverse or even leveraged-inverse. Our sample period, which covers the 2008 financial crisis, probably explains why such non-passive ETFs attracted some investors. Among the 4,693 investors who traded ETF, 3,225 focused only on passive ETF; the others (i.e. 30%) traded non-passive ETFs (most of the time, in addition to passive ETF). Table A.6 also reveals that the trading activity on ETFs is mainly concentrated on some ETF, since the top 5 accounts for about 39% of the whole activity on ETFs (in number of trades or in euros).

Table A.7 provides cross-sectional statistics about asset prices, monthly returns and risk factors. The latter are estimated only for stocks using daily prices and Fama–French data available online. Prices and market cap are in euros whereas returns are monthly. On average, the stocks in our sample have a one percent monthly return and 11 percent volatility. They are slightly right skewed and not surprisingly feature fat tails. For ETF, the average monthly return is 3.23% with an average volatility of 37.84%.

[20] Investors' nationality is not available in the data set but we can reasonably assume that most of the investors are Belgian or based in Belgium.

Table A.6
ETFs Top 5.

Name	# trades		Euros	
Lyxor CAC 40 Daily (−2x) Inverse	10,098	(17%)	177,819,698	(20%)
Gold Bullion Securities ETC	4,080	(7%)	62,324,680	(7%)
Lyxor CAC 40 Daily (2x) Leveraged	3,583	(6%)	61,342,462	(7%)
Lyxor Euro Stoxx 50 Daily (−1x) Inverse	2,987	(5%)	28,499,711	(3%)
Lyxor BEL 20 ETFs	2,866	(5%)	17,453,004	(2%)

The table lists the five most traded ETFs in our sample. For each one, we report the aggregate number of trades and the monetary value in euros. Corresponding percentages as a fraction of the whole trading activity on the 393 ETFs are provided in brackets.

Table A.7
Cross-sectional statistics for asset prices, monthly returns and risk factors.

	Mean	Q3	Median	Q1
Panel A: Stocks				
Price (euros)	39.74	34.41	19.49	9.92
Market cap (euros)	14,485.74	13,112.77	2,601.70	591.39
Return (%)	1.01	1.68	0.88	0.32
Volatility (%)	11.02	13.52	10.52	7.79
Skewness	0.1956	0.5368	0.1348	−0.2125
Kurtosis	2.3082	2.9150	1.3910	0.6234
Market	0.3923	0.5817	0.4183	0.2083
SMB	−0.3823	0.1498	−0.3426	−0.8722
HML	0.2931	0.5454	0.2690	−0.0138
RMW	−0.0569	0.1744	−0.1136	−0.5022
CMA	−0.3884	−0.1029	−0.3365	−0.6760
Momentum	0.1449	0.3312	0.1243	−0.0724
Panel B: ETF				
Price (euros)	74.57	88.82	35.75	19.25
Return (%)	3.23	0.95	0.35	−0.08
Volatility (%)	37.84	8.73	6.26	4.83
Skewness	−0.1216	0.1963	−0.1839	−0.4933
Kurtosis	1.5899	1.6593	0.6270	0.0904

The table reports the cross-sectional mean, median, lower and upper quartiles for asset prices and monthly returns. Panel A refers to stocks while Panel B refers to ETF. $Price$ is expressed in euros. $Market\ Cap$ is expressed in millions of euros and is based on the monthly number of shares outstanding available for the stocks in our sample. $Return$ is the monthly price return expressed in %. $Volatility$ is the standard deviation of monthly returns expressed in %. $Skewness$ and $Kurtosis$ computed on monthly returns are also provided. $Market$, SMB, HML, RMW and CMA refer to the estimated Fama–French 5 factors for stocks. These factors summarize the excess return on the market ($Market$), the performance of small stocks relative to big stocks (SMB, Small Minus Big), the performance of value stocks relative to growth stocks (HML, High Minus Low), the performance of robust stocks relative to weak stocks (RMW, Robust Minus Weak), and the performance of conservative stocks relative to aggressive stocks (CMA, Conservative Minus Aggressive). $Momentum$, which is the average return on the two high prior return portfolios minus the average return on the two low prior return portfolios, is also estimated using Fama–French available data.

A.4. Sample of investors

Table A.8 shows that 10.11% of investors are female. Moreover, our sample is mainly composed of highly educated investors since 73.59% of them have a university degree or equivalent. As for risk aversion, the majority of investors seem to be risk tolerant since 65.33% of them are classified as medium risk averse and 27.88% of them even featuring low risk aversion. When considering the annual net income over the entire sample, we observe that about 70% of the investors declare an income between 20,000 and 75,000 euros. Only a minority (3.36%) earn more than 150,000 euros per year. The income measure reported in our data is recorded once, when the investor completed the MiFID tests. Hence, the classification may therefore be noisy over the 10-year sample period, particularly for the early entries. However, when looking at risk aversion and income together, the income range 20,000–75,000 euros features the most investors with a low or medium aversion while high risk averse investors tend to have lower incomes.

Table A.9 provides cross-sectional statistics for investors' age, monthly trading activity and portfolio. The average investor is about 48 years old. When focusing on stocks only (Panel A), the average investor executes monthly 2.76 trades across 2.05 different stocks for a volume of 18,237 euros. Consistent with the literature, investors in our sample are under-diversified: the average (median) investor holds a five-stock (three-stock) portfolio. The average end-of-month portfolio value is about 28,003 euros (with

Table A.8
Investor characteristics joint percentages.

RA	Gender	Female			Male				Total	
	Income // Educ	No	HS	Univ	No	HS	Univ	Total		
Low	<20,000	0.05	0.19	0.16	0.19	0.82	1.49	2.90		13.86
	20,000–40,000	0.08	0.22	0.54	0.54	2.40	6.06	9.85		36.89
	40,000–75,000	0.06	0.16	0.62	0.45	1.56	6.42	9.28		33.02
	75,000–150,000	0.01	0.04	0.32	0.13	0.45	3.54	4.49		12.89
	>150,000	0.02	0.02	0.12	0.07	0.14	1.01	1.37	27.88	3.36
Mid	<20,000	0.14	0.41	0.66	0.74	2.06	4.01	8.01		
	20,000–40,000	0.22	0.74	1.72	1.56	5.10	15.66	25.01		
	40,000–75,000	0.14	0.30	1.45	1.20	2.90	16.53	22.52		
	75,000–150,000	0.07	0.07	0.54	0.34	0.45	6.45	7.94		
	>150,000	0.01	0.01	0.21	0.05	0.11	1.46	1.85	65.33	
High	<20,000	0.05	0.12	0.21	0.32	0.63	1.62	2.95		
	20,000–40,000	0.05	0.05	0.12	0.17	0.45	1.18	2.03		
	40,000–75,000	0.01	0.03	0.09	0.06	0.14	0.90	1.23		
	75,000–150,000	0.00	0.01	0.01	0.02	0.05	0.38	0.47		
	>150,000	0.00	0.01	0.01	0.01	0.01	0.09	0.13	6.80	
	Total	0.92	2.39	6.81	5.84	17.27	66.79			
Total				10.11			89.90			
		6.76	19.66	73.59						

The table reports investor characteristics joint percentages computed on survey data. For education ($Educ$), 'No' refers to investors without any degree, 'HS' to investors with a secondary/high school degree, and 'Univ' to investors with an university degree or equivalent. For risk aversion (RA), 'High' refers to investors who state a high risk aversion (i.e. they do not want to take any risks and prefer safe investments), 'Mid' refers to investors who state a medium risk aversion, and 'Low' to investors who state a low risk aversion (i.e. they would invest even more when facing a sharp loss of 20% to reduce their average purchase prices). For annual net income ($Income$), five categories are available: lower than 20,000 euros, 20,000–40,000 euros, 40,000–75,000 euros, 75,000–150,000 euros, and larger than 150,000 euros.

a median value of about 7,552 euros). For comparison, Kumar and Lee (2006) and Korniotis and Kumar (2013) document that a typical investor at a major U.S. discount brokerage house over the period 1991–1996 holds a four-stock portfolio (with a median of three) with an average size of 35,629 USD (with a median of 13,869 USD). Using a more recent sample, Leal et al. (2017) find that a typical investor at a Portuguese brokerage holds 2.34 different stocks over the period 2003–2007.

As a proxy of experience, we look at the investor stock portfolio holding period (# $months$): our average investor holds stocks for 54.12 months out of 111. However, this average hides different holding patterns: 21,865 investors (i.e. 95%) hold stocks at some point over the sample period. In particular, 1.9% of them hold throughout, 1.5% hold then leave at some point, 81.5% enter and hold throughout and 15.1% enter and exit.

In terms of performance, our average investor earns a monthly gross return of 0.28% on stocks, with a volatility of 10.63%. This relatively high average volatility of individual stock portfolio gross returns is not surprising given our sample period that includes market ups and downs.

Panel B of Table A.9 reports similar cross-sectional statistics for the subsample of investors who traded both stocks and ETF. Not surprisingly, these investors exhibit a higher monthly trading activity, either in number of trades or in euros. Combining stocks and ETF, these investors hold more diversified and wealthy portfolios: the typical investor holds a six-asset portfolio (with a median of 4.5) with an average size of 49,961 euros (with a median of 14,696 euros). This better diversification appears profitable since the average investor earns a monthly gross return of 0.42%, with a volatility of 10.04%.

We should point out that almost all the variables in Table A.9 are positively skewed and display a large heterogeneity. This is consistent with usual tremendous variations in behavior and in outcomes across individual investors (Barber and Odean, 2013). Finally, Table A.10 provides statistics on ETF usage in our sample. They show that ETF users are somewhat older and that they are more educated, report lower risk aversion, have higher income, invest more money in financial markets. In addition, they are more active investors on stocks (based on both the number of trades or the monetary volume traded).

Table A.11 tells us something about trading frequencies. The median investor trades two assets a month, with a slightly higher mean of 2.8. There is very little variation in the median across the various cross-sectional sample splits. The means of trading frequencies are also relatively flat across the various sub-populations, with the exception perhaps for income, where high income investors trade more often.

A.5. Individual investor returns adjusted for cash holding

Our robo-investors have the option to hold cash. When anticipating a bear market, they can reduce (or even avoid) market risk exposure by holding (only) cash. Our retail investors are also allowed to hold cash but, unfortunately, cash holdings are not available in our data set. Although our monthly portfolio returns are relatively robust to various cash inflows and outflows (see Appendix A.2), they are computed assuming that the weight of the cash account is zero. This can negatively (positively) affect individual investor returns during bear (bull) market periods, which can lead to some unfair comparisons with Alter-Egos that may

Table A.9
Cross-sectional statistics for investors' age, monthly trading activity and portfolio.

	Mean	Q3	Median	Q1
Age	48.57	58	48	38
Panel A: Stocks only				
# trades	2.76	3	2	1.4
Volume (euros)	18,237.29	13,278.06	5,584.25	2,488.75
# ISIN	2.05	2.36	1.68	1.25
# assets	4.43	5.66	2.75	1.27
Portfolio value (euros)	28,003.08	22,350.32	7,552.07	2,484.01
# months	54.12	81	50	27
Return (%)	0.28	0.79	0.03	−0.70
Volatility (%)	10.63	11.66	8.02	5.89
Panel B: Stocks and ETF				
# trades	3.88	4.22	2.66	1.88
Volume (euros)	29,211.35	20,936.43	9,115.29	4,430.90
# ISIN	2.73	3.15	2.18	1.6
# assets	6.68	9.17	4.5	2
Portfolio value (euros)	49,961.58	41,461.24	14,696.78	4,925.21
# months	57.47	88	59	29
Return (%)	0.42	0.72	0.13	−0.42
Volatility (%)	10.04	9.38	6.72	5.35

The table reports the cross-sectional mean, median, lower and upper quartiles for investors' age as well as trade-based and portfolio-based variables. Panel A refers to stocks only and Panel B refers to both stocks and ETF. Statistics in Panel A are computed across the entire sample (i.e. 22,972 investors) while those in Panel B are computed across the subsample of investors who traded ETFs in addition to stocks (i.e. 4,693 investors). *Age* is the difference between 2012 and the investor's year of birth. # *trades* is the monthly number of trades executed. *Volume* is the corresponding monthly monetary volume in euros. # *ISIN* is the monthly number of different assets traded. # *assets* is the monthly average number of assets held portfolio computed over the holding period. *Portfolio value* is the corresponding monthly average market portfolio value in euros (without cash holdings). # *months* refers to the portfolio holding period. *Return* is the monthly average gross portfolio return expressed in % (without cash holdings). *Volatility* is the standard deviation of monthly gross portfolio returns expressed in %.

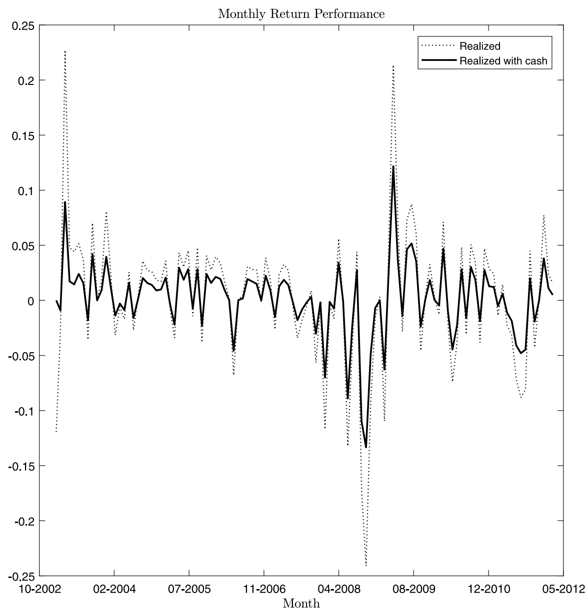


Fig. A.1. This figure provides monthly realized returns with/without the cash adjustment for the median portfolio over the sample period.

Table A.10
Statistics for ETFs usage.

	ETFs users	non-ETFs users
Age	51.40***	47.85
% Females	9.14**	9.91
Education (%)		
No	4.54***	7.39
HS	16.51***	20.60
Univ	78.95***	72.01
Risk aversion (%)		
Low	30.79***	27.27
Mid	62.20***	65.81
High	7.01***	6.92
Income (%)		
Low	12.21***	14.62
Mid	82.87***	82.43
High	4.92***	2.94
Funds invested (%)		
Low	35.86***	53.20
Mid	34.88***	29.77
High	29.26***	17.03

The table reports the mean age and characteristics percentages computed on survey data for investors who traded ETFs ('ETF users') versus those who did not ('non-ETF users'). 'Age' is defined as the difference between 2012 and the investor's year of birth. For the level of education ('Education'), 'No' refers to investors without any degree, 'HS' to investors with a secondary/high school degree, and 'Univ' to investors with an university degree or equivalent. For the level of risk aversion ('Risk aversion'), 'High' refers to investors who state a high risk aversion (i.e. they do not want to take any risks and prefer safe investments), 'Mid' refers to investors who state a medium risk aversion, and 'Low' to investors who state a low risk aversion (i.e. they would invest even more when facing a sharp loss of 20% to reduce their average purchase prices). For annual net income ('Income'), 'High' refers to investors who report they earn more than 150,000 euros, 'Mid' to investors who earn between 20,000–150,000 euros, and 'Low' to investors who state earning less than 20,000 euros. For the amount invested in financial markets ('Amount invested'), 'High' refers to investors who report they invest more than 250,000 euros, 'Mid' to investors who invest between 50,000–250,000 euros, and 'Low' to investors who invest less than 50,000 euros. *, **, *** indicate whether percentages (or means) differ between ETF users and non-ETF users at the level of 10%, 5%, and 1%, respectively.

benefit from the opportunity to invest in a risk-free asset. In order to overcome this hurdle, we decide to estimate a proxy for the cash holding of each investor. This estimation relies on the following assumptions. (1) The cash account of any investor is empty at the beginning of our sample period. (2) The cash account depends on the investor's trading activity: it increases by the monetary value of each sale and decreases by the monetary value of each purchase. (3) Cash holdings cannot be negative, so that the cash account is automatically fed for any purchase whose monetary value is exceeding the cash available. (4) The cash account is not rewarded (return is 0%) as it is for our robo-advisors. Using our estimation of the cash holdings at the end of each month, we are then able to compute the weight of cash ($wCash$) in the investor end-of-month portfolio and calculate the monthly return adjusted for cash by multiplying the portfolio return by $(1 - wCash)$.[21]

Fig. A.1 compares monthly individual investor returns with/without the cash adjustment for the median portfolio over the sample period. As expected, taking into account investor cash holdings reduces the range of returns. This reduction is especially substantial during the 2008 financial crisis, indicating that some investors reduce their exposure to market risk.

[21] Of course, one could argue that a large amount of cash could be used by an investor to invest in real estate or whatever. It is true that our proxy will not capture this kind of decision as it will not capture a cash contribution that is slow to be invested in the portfolio. However it seems more reasonable to use this proxy than to assume a zero cash account.

Table A.11
Summary statistics monthly trading frequency stocks and ETFs.

	Mean	SD	Median	Obs
Global	2.80	3.22	2	22 972
Education				
Low	2.80	3.25	2	1 563
High	2.76	3.28	2	16 868
Low-High pval	*0.5760*			
Risk Aversion				
Low	3.13	3.33	2.23	6 430
High	3.18	3.43	2.16	1 594
Low-High pval	*0.6060*			
Income				
Low	2.79	2.83	2	3 246
High	3.52	5.49	2.35	769
Low-High pval	*0.0004*			

Notes: Summary statistics for trading frequency, defined as the average number of monthly trades, for each investor. Entries to the table are the averages, standard deviations, median and sample sizes. The p-values (pval) of the differences in mean between high and low tests within education, risk aversion and income investors are reported.

A.6. Transaction costs

To handle the transaction costs of robo-advisors we mimic the industry standard of charging a flat fee based on AUM. We picked a 0.30% fee on AUM as a reasonable number given current industry practice.[22] To assess the impact of such transaction costs, let us denote by α the management advisory fee that the brokerage firm charges the investor (or, in our specific case, each AI Aler Ego shadowing a given individual investor in our sample), which is based on a percentage of either the investor's ending or beginning period AUM.

Let $R_{i,t}$ be the $N\times1$ vector of gross returns for the N assets, and let $r^p_{i,t}$ be the portfolio return for investor i where $r^p_{i,t} \equiv w'_{i,t-1}R_{i,t}-1$, and $w_{i,t-1}$ is the $N \times 1$ vector of portfolio weights. Let the realized investor returns be $\bar{r}^p_{i,t}$ with transaction costs.

We start with a case where transaction costs are computed at the end of the investment period (i.e. quarter):

$$A_{i,t} = A_{i,t-1}w'_{i,t-1}R_{i,t} - \alpha A_{i,t} = \frac{1}{1+\alpha}A_{i,t-1}w'_{i,t-1}R_{i,t}$$
$$\bar{r}^p_{i,t} \equiv \frac{A_{i,t}}{A_{i,t-1}} - 1 = \frac{1}{1+\alpha}w'_{i,t-1}R_{i,t} - 1 = r^p_{i,t} - \frac{\alpha}{1+\alpha}$$

If the fee is levied at the beginning of the period, then similar calculations yield: $\bar{r}^p_{i,t} = r^p_{i,t}$ - α. As we can see from both cases, the fee amounts to a constant being subtracted from the pre-fee returns.

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jempfin.2020.10.002.

References

Barber, B.M., Odean, T., 2013. The behavior of individual investors. In: Constantinides, G.M., Harris, M., Stulz, R.M. (Eds.), Handbook of the Economics of Finance, Volume 2, Part B. pp. 1533–1570.
Bellofatto, A., D'Hondt, C., De Winne, R., 2018. Subjective financial literacy and retail investors' behavior. J. Bank. Financ. 92, 168–181.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
D'Acunto, F., Prabhala, N., Rossi, A.G., 2019. The promises and pitfalls of robo-advising. Rev. Financ. Stud. 32, 1983–2020.
DeMiguel, V., Garlappi, L., Uppal, R., 2007. Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? Rev. Financ. Stud. 22, 1915–1953.
Engle, R.F., Ledoit, O., Wolf, M., 2019. Large dynamic covariance matrices. J. Bus. Econ. Stat. 37, 363–375.
Friedman, J., Hastie, T., Tibshirani, R., 2016. The Elements of Statistical Learning, second ed. Springer.
Gu, S., Kelly, B., Xiu, D., 2018. Empirical Asset Pricing Via Machine Learning. Discussion paper. National Bureau of Economic Research.
Korniotis, G.M., Kumar, A., 2013. Do portfolio distortions reflect superior information or psychological biases? J. Financ. Quant. Anal. 48, 1–45.
Kumar, A., Lee, C., 2006. Retail Investors Sentiment and Returns Comovements, Vol. 61. pp. 2451–2486.
Leal, C.C., Loureiro, G., Armada, M.J.R., 2017. Selling winners, buying losers: Mental decision rules of individual investors on their holdings. Eur. Financial Manag. 24, 362–386.

[22] See for example: https://www.nerdwallet.com/best/investing/robo-advisors. It should be noted that this website lists even a fair number of US offerings that do not charge a management fee at all.

Ledoit, O., Wolf, M., 2004. Honey, I shrunk the sample covariance matrix. J. Portf. Manag. 30, 110–119.
Markowitz, H., 1952. Portfolio selection. J. Finance 7, 77–91.
Rossi, A.G., Utkus, S.P., 2020. Who benefits from robo-advising? Evidence from machine learning. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3552671.
Shestopaloff, Y., Shestopaloff, A., 2007. A hierarchy of methods for calculating rates of return. J. Perform. Measur. 12, 39–52.
Welch, I., Goyal, A., 2007. A comprehensive look at the empirical performance of equity premium prediction. Rev. Financ. Stud. 21, 1455–1508.
Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. J. R. Stat. Soc. Ser. B Stat. Methodol. 67, 301–320.