Robo-Advising as a Human-Machine Interaction System

Agostino Capponi * Octavio Ruiz Lacedelli † Matt Stern ‡

July 6, 2018

Abstract

Robo-advising can substantially enhance a human’s efficiency in investment decisions. Since machines are unable to observe the preferences of the humans that they serve, it becomes important to establish an effective interaction protocol. While the investor’s and machine’s objectives are aligned, asymmetric information, along with heterogeneous sensitivities to risk by the investor and machine, make their joint optimization process a game with strategic interactions. We propose a framework based on risk-sensitive dynamic games, where the investor seeks to optimize her risk-sensitive criterion while the machine adaptively learns the investor’s preferences. We consider an investor with myopic mean-variance preferences and reduce the game to a partially observed Markov decision process (POMDP). The human-machine interaction protocol features a trade-off between allowing the robo-advisor to learn the investor’s preferences through costly communications and optimizing the investor’s objective relying on outdated information.

1 Introduction

Robo-advising can substantially enhance human efficiency in investment decisions by handling routine or cognitively challenging operations. It is crucial, however, that investors are able to efficiently communicate their preferences to the machine to optimize their risk criteria. A machine can only provide a useful or reliable service if its valuation of the costs and risks associated with each action are aligned with the investor that it serves.

In this paper, we propose a framework for the analysis of human-machine interactions in the context of robo-advising. The objectives of the investor and the machine are aligned, but there are informational asymmetries. The machine is unable to observe the investor’s preferences, and must infer them via a dynamic learning process by observing the effect of joint human-machine actions on the system’s state. Our model is designed

*Email: ac3827@columbia.edu, Department of Industrial Engineering and Operations Research, Columbia University, New York, NY, 10027
†E-mail: or2200@columbia.edu, Department of Industrial Engineering and Operations Research, Columbia University, New York, NY, 10027
‡Email: mns2141@columbia.edu, Department of Industrial Engineering and Operations Research, Columbia University, New York, NY, 10027
to capture a wide variety of situations in which an investor wants to utilize a machine to enhance the efficiency and effectiveness of the investment decisions. The machine is designed to serve a broad audience of humans, rather than tailored to a specific category. It is thus important for the machine to personalize itself to the investor and self-calibrate as the human reveals information regarding her risk preferences and objectives.

In the robo-advising context, we consider an investment firm that develops a tool to autonomously manage the investments for their clients. In each period, the robo-advisor must place the clients’ capital into one of several pre-constructed portfolios, each having a specific risk-return profile dynamically changing based on updated market information. For each investor, the robo-advisor finds the optimal portfolio from a catalogue of portfolios that lie on the efficient frontier. The client may elect to make the portfolio decisions herself over the recommendation of the robo-advisor, through the firm’s client communication channels. Overriding the portfolio choice of the robo-advisor, however, is burdensome to the client. In order to provide good service, the robo-advisor needs to understand the client’s preferences for risk and return. Through our framework, the robo-advisor can estimate the preferences of the client by observing her overriding investment decisions, or lack thereof. Additionally, the firm faces risk as aggressive portfolio choices by the robo-advisor may reflect badly on the firm if the estimates of the client’s preferences are incorrect and the client is thus burdened with frequent override investment decisions. The tolerance that the firm has towards the uncertainty over the investor’s preferences presents a separate form of risk-sensitivity that we define explicitly in our framework.

The distinguishing feature of our framework is the simultaneous handling of human-driven and context-driven risks. The uncertainty over the investor’s characteristics, such as her risk preferences, goals, and objectives, presents a human-driven risk to the machine. Depending on the machine’s attitude toward risk, it could, for example, operate to provide a good performance to the average investor. Alternatively, it could target humans whose characteristics belong to a specific quartile. We refer to the set of characteristics which uniquely identify the human’s behavior as her type. On the other hand, the unpredictable nature of market conditions in which the decisions need to be executed present context-driven risks to the human. The investor makes portfolio choices on the basis of her risk preferences, and in doing so it reveals information about her type to the robo-advisor. Both investor and machine share the cooperative goal of optimizing the human’s value. However, informational asymmetries and heterogeneous sensitivities to risk lead to strategic behavior of the agents, and make the joint minimization process of human’s costs a strategic game. In the absence of informational asymmetries, the objectives of the investor and the machine are perfectly aligned, so that the game becomes cooperative.

For models featuring both human-driven and context-driven risk, we introduce a new equilibrium concept, risk-sensitive Bayesian equilibria, a departure from the classical Nash equilibrium concept. Furthermore, we introduce decentralized optimization techniques to reduce the problem to a related single-agent, risk-sensitive, optimization problem. We remark that risk-sensitive optimization in the context of Markov decision processes has
been subject of considerable investigation recently (see Section 2 for additional details). In addition, our study paves the way for a systematic study of human-driven risk and its implications.

We develop a numerical study to demonstrate the power of our framework. We consider a myopic investor wishing to optimize her short-term risk-adjusted return, disregarding the impact of her decisions over the long-term. Examples include casual investors focusing on short-term gains and portfolio managers whose compensation package is dependent on their short-term performance. However, the robo-advisor firm wishes to keep the client for the long term, and therefore is optimizing the cumulative sum of the investor’s myopic objectives, accounting for the impact that current decisions might have on future investment opportunities. We illustrate the fundamental benefit/cost trade-off faced by the investor in communicating her risk preferences to the robo-advisor to obtain more tailored investment decisions. The investor is only willing to override the machines’ decision if the performance loss, defined as the difference between the risk-adjusted return of the optimal investor’s portfolio and that achieved by the robo-advisor, is higher than the overriding costs. If the performance gain from human’s intervention does not overcompensate for the overriding costs, then the investor would tolerate investment decisions that are suboptimal given her true risk-aversion parameter.

We also find that the human-machine system achieve a value of the risk function criterion that is lower than that of a human-only model, in which the investor chooses the portfolio herself, but incurring opportunity costs due to market research and frequent portfolio rebalancing. The avoidance of these opportunity costs is one of the major advantages of robo-advising, because the investor can delegate time-consuming activities to the machine and considerably reduce these costs.

The paper proceeds as follows. Section 2 puts our paper in perspective with existing literature. Section 3 develops the framework and presents solution concepts. Section 4 provides a numerical study for the human-machine interaction system in the context of robo-advising. Section 5 concludes the paper and discusses how the proposed framework opens the door to a new class of multi-agent decision making problems.

2 Contributions and Related Work

The proposed framework describes the cooperative decision making problem of a human and a machine, who are both sensitive towards risk. In a recent work, Hadfield-Menell et al. (2017) define a framework for human-machine interactions, based on the theory of inverse reinforcement learning (IRL). Both the machine and the human are risk-neutral agents and, as such, their framework does not capture human-driven or context-driven risk. They reduce the two-agent model to a joint optimization problem building on an earlier study of Nayyar et al. (2013a), and compare their solution concepts to existing IRL methods. In our study, we introduce a notion of risk-sensitive equilibrium to deal with risk aversion of human and machine, and both agents minimize a risk function.

One of the defining features of our framework is that both agents share the common goal of optimizing the human’s objective. Nayyar et al. (2013b) introduce a model
of decentralized stochastic control, where a team of agents work together to minimize a common objective. They show that this problem can be reduced to a POMDP by constructing a coordinator that determines strategies for the agents, based on the common information available in each period. Similar approaches have been employed by Vasal and Anastasopoulos (2016a, 2016b) to solve incomplete information games between agents with conflicting objectives. The coordinator technique is appealing because it reduces a game of multiple agents to a single-agent optimization problem. In our framework, we show that the solution of this single-agent problem corresponds to an equilibrium between the human and machine.

Our paper is related to existing literature on risk-sensitive Markov decision processes (MDP). Risk aversion in MDPs has been extensively studied. Earlier contributions focused on exponential utility (Howard and Matheson (1972)), mean-variance criteria (Sobel (1982)), and percentile risk criteria (Filar (1985)). Ruszczynski (2010) consider the class of risk measures, and show that these lead to tractable dynamic programming formulations. Recent contributions by Bäuerle and Rieder (2014) and Bäuerle and Ott (2011) solve the utility maximization process and the conditional value at risk criterion for a MDP. Haskell and Jain (2015) generalize these studies to a wider class of risk measures using a convex analytic approach. All these studies deal with the optimization of a single agent. In contrast, our framework features strategic interactions between agents, and employs risk-sensitive optimization to solve for a new class of equilibria corresponding to the optimal pair of human-machine actions.

The literature on robo-advising is still at its infancy. A popular approach in the wealth management industry is the goals-based investing. Investors specify quantifiable objectives such as guaranteeing the expected wealth to be above a certain threshold, given that the expected loss from returns outcomes falling below the threshold is smaller than a certain value. The goals-based approach is followed by Betterment, a leading robo-advisor firm, and has also been also investigated in the research literature by Das et al. (2018). Das et al. (2018) define a goals-based wealth management approach which restricts the efficient frontier to the subset of portfolios that achieve, with a specified probability, the investors’ chosen target wealth levels. In contrast, our approach elicits information about the investor’s risk preference, by offering a discrete catalogue of portfolios to the investors.

Another popular robo-advising firm, Wealthfront, estimates investors’ subjective risk tolerance by asking clients whether they are focused on maximizing gains, minimizing losses, or both equally. They construct a risk metric that is a weighted combination of subjective and objective risk measures, with a higher weight assigned to the component indicating higher risk aversion. The robo-advisor adopts a mean-variance optimization framework a-la Markowitz (1952) or variations of it. In this framework, the utility function of the investor trades off the expected return with the risk of the portfolio, weighted by the risk tolerance level of the investor. Thus, less risk-averse investors select portfolios with a higher risk and higher expected return as compared with risk-averse investors. Our approach to obtaining optimal portfolios is related to that used by Wealthfront in the short-term: in each period, the robo-advisor chooses from a catalogue
of portfolios on the efficient frontier. However, in our model the investor and the machine interact throughout the whole investment horizon, and the strategy reflects the machine’s learning process of the investor’s risk preferences based on the investor’s decisions. Our optimization criterion accounts for a long-term objective, given by the sum of single period mean-variance Markowitz utilities over the investment horizon.

Most recently, Dai et al. (2018) develop a dynamic mean-variance framework in the context of robo-advising. In their model, the input to the machine is the expected return of the investor, that uniquely identifies the mean-variance parameter. They argue that a quantitative asset allocation model should be based mainly on risk profile and investment horizon, while other factors such as age, labor, income etc can be captured in ad-hoc ways by the financial advisor after running the asset allocation model.

3 The Framework

In the proposed framework, both the human and the machine are modeled as risk-averse agents. This allows us to simultaneously capture context-driven and human-driven risk. We capture the risk preferences of human and machine through risk functions. We refer to Shapiro et al. (2009) for a treatment of single agent optimization using risk-functions as the optimization criterion. Consider a probability space \((\Omega, \mathcal{F}, P)\), and let \(L^\infty\) be the space of essentially bounded random variables.\(^1\) A risk function is a mapping \(\rho: L^\infty \to \mathbb{R}\) from an uncertain outcome \(Z\) onto the set of real numbers; see also Ruszczynski and Shapiro (2005). Risk functions can thus account for the entire probability distribution of an uncertain outcome, whereas expected utility functions can only depend on the realization of that outcome. We require the risk function to be monotone, i.e., that higher risk is associated with larger loss.\(^2\)

**Definition 1.** A human-machine interaction game is a \(T\) period dynamic game with asymmetric information played between two risk sensitive agents: a human, \(H\), and a machine, \(M\). The game is described by a tuple \(\langle S, \{A^H, A^M\}, \Theta, \{\rho^H, \rho^M\}, P, c, \pi_1 \rangle\), whose elements are defined as:

- \(S\) a set of system states: \(s \in S\);
- \(A^H\) a set of actions for \(H\): \(a^H \in A^H\);
- \(A^M\) a set of actions for \(M\): \(a^M \in A^M\);
- \(\Theta\) a set of possible risk parameters, only observed by \(H\): \(\theta \in \Theta\);
- \(\rho^H(\cdot)\) \(H\)'s risk function, parameterized by \(\theta\);
- \(\rho^M(\cdot)\) \(M\)'s risk function over a probability distribution on \(\theta\);
- \(P(\cdot, \cdot, \cdot)\) the probability transition function on the future state, given the current state and joint action: \(P(s'|s, a^H, a^M)\);
- \(c(\cdot, \cdot, \cdot)\) an instantaneous cost function that maps the system state and joint actions to a vector of real numbers: \(c: S \times A^H \times A^M \to \mathbb{R}\);

\(^1\)A random variable \(Z\) is essentially bounded if there exists \(M \geq 0\) such that \(P(|Z| > M) = 0\).

\(^2\)Risk Functions which satisfy the axioms of monotonicity, translation invariance and convexity are referred to as risk measures. See Artzner et al. (2015). In our framework, we only require the monotonicity assumption to study the risk-sensitive Bayesian equilibrium.
\[ \pi_1(\cdot) \) a common prior distribution over the risk parameters: \( \pi_1(\theta) \in \mathcal{P}(\Theta) \).

Above, we have used \( \mathcal{P}(\Theta) \) to denote the set of probability distributions on \( \Theta \). After each period \( t \), the human and the machine incur a common cost, \( c(s_t, a^H_t, a^M_t) \in \mathbb{R} \), depending on the current state of the system, and their joint action. Their incentives are partially aligned as both the human and the machine prefer to keep the total system costs low over the \( T \) period horizon. The human’s objective is to minimize the costs using her risk function as the optimization criterion \( \rho^H_\theta \), where \( \theta \) is the true type of the human. For example, the mean-variance risk function \( \rho^H_\theta = \theta_1 E \left[ \sum_{\tau=1}^{T} c(s_\tau, a^H_\tau, a^M_\tau) \right] + \theta_2 \text{Var} \left[ \sum_{\tau=1}^{T} c(s_\tau, a^H_\tau, a^M_\tau) \right] \) maps the random outcome for the total costs to a scalar quantity through the parameter vector \( \theta = [\theta_1, \theta_2] \in \mathbb{R}^2 \). In this case, the vector \( \theta \) not only describes the human’s risk sensitivity towards the costs, but it also describes the relative weight that the human assigns to each cost. The machine does not know the value of \( \theta \) at the initial stage of the game, but begins with a prior distribution \( \pi_1(\cdot) \in \mathcal{P}(\Theta) \).

Denote the set of public histories as
\[ H_t := \left( A^H \times A^M \right)^{t-1} \times \mathcal{S}, \]
where \( h_t = (s_1, a^H_1, a^M_1, \ldots, a^H_{t-1}, a^M_{t-1}, s_t) \in H_t \) for \( t > 1 \) and \( h_1 = s_1 \). A public history contains information that is observed by both the human and the machine, which includes the realization of the system’s states and the actions executed by both agents. The machine maintains the posterior distribution over the human’s type, \( \pi_t(x) := P(\theta = x|h_t) \), which we refer to as the machine’s belief in period \( t \).

A Markov strategy for the human \( \sigma^H = (\sigma^H_1, \ldots, \sigma^H_T) \) is a sequence of measurable maps \( \sigma^H_t : \mathcal{S} \times \mathcal{P}(\Theta) \times \Theta \to \mathcal{P}(A^H) \) so that
\[ \sigma^H_t(a|s_t, \pi_t, \theta) = P(a^H_t = a|s_t, \pi_t, \theta), \ \forall t \in \{1, \ldots, T\}, a \in A^H. \]

A Markov strategy for the machine \( \sigma^M = (\sigma^M_1, \ldots, \sigma^M_T) \) is a sequence of measurable maps \( \sigma^M_t : \mathcal{S} \times \mathcal{P}(\Theta) \to \mathcal{P}(A^M) \) so that
\[ \sigma^M_t(b|s_t, \pi_t) = P(a^M_t = b|s_t, \pi_t), \ \forall t \in \{1, \ldots, T\}, b \in A^M. \]

Notice that the human’s Markov strategy depends on the machine’s current beliefs because the action of the human is influenced by the action of the machine, which in turn depends on its belief over the human’s type.

The total (cumulative) cost is given by the random variable
\[ C_T := \sum_{\tau=1}^{T} c(s_\tau, a^H_\tau, a^M_\tau). \]

Given the conflicting objectives of both agents, the framework as presently defined is a two-player strategic game. As such, we define the corresponding risk-sensitive Bayesian
equilibrium as a pair of strategies \((\sigma^*H, \sigma^*M)\) and a belief profile \(\pi^* := (\pi^*_1, \ldots, \pi^*_T)\) such that
\[
\rho^H_0 \left( C_T|\sigma^*H, \sigma^*M, \pi^*_1, h_1 \right) \leq \rho^H_0 \left( C_T|\bar{\sigma}^H, \sigma^*M, \pi^*_1, h_1 \right), \\
\rho^M \left( \rho^H_0 \left( C_T|\sigma^*H, \sigma^*M, \pi^*_1, h_1 \right)|\pi^*_1 \right) \leq \rho^M \left( \rho^H_0 \left( C_T|\sigma^*H, \bar{\sigma}^M, \pi^*_1, h_1 \right)|\pi^*_1 \right),
\]
for all strategies \(\bar{\sigma}^H, \bar{\sigma}^M\). Furthermore, the machine’s belief profile \(\pi^*\) must be consistent with the strategies \((\sigma^*H, \sigma^*M)\) in that Bayes’ rule is used to update the beliefs whenever possible. Specifically, the machine’s belief on the true value of the human’s risk parameter \(\theta\) satisfy the standard nonlinear filter equation (Fudenberg and Tirole (1991))

\[
\pi^*_{t+1}(\theta) := \frac{\pi^*_t(\theta)\sigma^*H(a^*_t | s_t, \pi^*_1, \theta)}{\sum_{\bar{\theta}} \pi^*_t(\bar{\theta})\sigma^*H(a^*_t | s_t, \pi^*_1, \bar{\theta})},
\]
provided there exists a value of \(\bar{\theta}\) such that \(\pi^*_t(\bar{\theta}) > 0\) and \(\sigma^*H(a^*_t | s_t, \pi^*_1, \theta) > 0\). In period 1, the belief profile \(\pi^*_1\) is equal to the prior \(\pi_1\).

The first of the two inequalities in equation (1) indicates that the human has no incentive to unilaterally deviate from her action \(\sigma^*H\) to any other action \(\delta^H\) because her risk-adjusted total cost would increase. Similarly, the second inequality stipulates that the machine’s action yields the smallest risk-adjusted total cost, according to both the human’s type and the machine’s beliefs over the human’s type.

The canonical solution concept for dynamic games of incomplete information is the Bayesian equilibrium (BE). However, standard equilibrium concepts rely on maximizing the expectation of utility functions assigned to each player. A Bayesian equilibrium in our setup would require that both agents minimize the expected disutility of total system costs, rather than the general risk functions we present.

Context-driven risk in our model is captured by applying the risk function \(\rho^H\) to the total system cost. This allows us to capture a wide variety of cost criteria that depend on the statistical properties of the cumulative costs, including value at risk, conditional value at risk, and worst case measures. A special case of context-driven risk is when the human minimizes the expected disutility from costs, where disutility is quantified by a convex utility function. Human-driven risk is quantified by the risk function \(\rho^M\) over the distribution of human’s type. For example, if \(\rho^M\) is the expectation operator, then the machine aims for the best service to the average human type. On the other hand, if \(\rho^M\) represents the value at risk for some level of service \(\alpha\), then the machine aims to provide a good service for \(1 - \alpha\) percentage of the human’s types. Lastly, if the human’s type is revealed to the machine before \(T\), then there is no human-driven risk. In this case, \(\rho^M(\rho^H_0(C_T)) = \rho^H_0(C_T)\), so that the two inequalities in Eq. (1) coincide, and the game becomes cooperative.

The solution methodology that we propose to address the human-machine framework is to transform the strategic game to a single-agent problem by introducing a coordinator agent \(C\). The coordinator assigns a policy \(\sigma^C = (g^M, g^H)\) such that \(g^M\) is a strategy for the machine and \(g^H\) is a decision function, which prescribes the human’s strategy for each possible realization of \(\theta\). Hence, the coordinator is unaware of the human’s
risk parameter, but instead chooses a strategy for every possible type of human. The coordinator’s objective is to use these controls to minimize the machine’s risk function

\[
\min_{g^M, \theta^H} \rho^M \left( \rho^H \left( C_T, |\theta^H ; g^M, \pi_1, h_1 \right) |\pi_1 \right).
\]

The resulting problem is a partially-observable, risk-sensitive, Markov decision process (risk-POMDP). The following theorem connects the solution to the coordinator problem with the equilibrium concept for the human-machine interaction game.

**Theorem 1.** A solution to the coordinator problem is a risk-sensitive Bayesian equilibrium to the two-agent human-machine interaction game.

The proof of Theorem 1 is included in appendix A.

**Remark 1.** The risk-POMDP formulation can be reduced to a fully observable, risk-sensitive MDP, where the state space is enlarged to include the belief profile. The resulting single-agent problem can then be solved using existing risk optimization techniques described in Section 2.

4 Robo-Advising with Myopic Mean-Variance Preferences

We consider a \( T \) period investment framework in which a human investor hires a robo-advisor to select an investment portfolio at each period \( t \). The robo-advisor learns the human’s risk preference over time, and selects the risk-return profile of the portfolio to reflect the learned preferences. For instance, if the human’s tolerance for risk was known to be high, then the robo-advisor would choose a portfolio with a higher expected return, irrespective of its variance. Conversely, if the robo-advisor knew that the investor was sensitive to risk, then it would avoid portfolios with high variance even if they had higher expected return.

We specialize the general framework presented in Section 3 to capture decision making in robo-advising settings. The human \( H \) corresponds to the investor, and the machine \( M \) corresponds to the robo-advisor. Let the system states represent the market environment, assumed to be represented by the expected return and standard deviation of \( n \) available investment portfolios at each time \( t \). Formally, \( S = \{1, \ldots, n_s\} \) represent the set of economic scenarios. Portfolio \( i \) in state \( s_t \in S \) is assumed to have a known expected return \( \mu(s_t, i) \) and standard deviation \( \sigma(s_t, i) \). For example, \( s_t = 1 \) may correspond to a low return-low volatility market scenario, while \( s_t = n_s \) may represent a high return-high volatility scenario. Note that the expected return and standard deviation parameters of each portfolio are time invariant, i.e., they depend on the actual state \( s_t \), but not on the time period \( t \). The probability of a transition from state \( s \) to state \( s' \) is assumed to be independent of the human’s action, and is denoted by \( P(s' | s) \) for all \( s, s' \in S \). This means that the investor’s decisions cannot influence the market environment.

The set of actions available for \( M \) corresponds to the \( n \) available portfolios, i.e., \( A^M = \{1, \ldots, n\} \). An action \( a^M_t = i \in A \) corresponds to \( M \) choosing portfolio \( i \). In
addition, the human is allowed to override the decision of the robo-advisor, and therefore has a set of actions \( A_H = \{0, 1, \ldots, n\} \), where \( a_t^H = 0 \) corresponds to no-override (so that the investor keeps the portfolio selected by \( M \)), and \( a_t^H = i > 0 \) corresponds to an override decision of the machine that decides to invest in portfolio \( i \). The portfolio selected at time \( t \) is given by

\[
a_t := \begin{cases} 
a_t^M, & \text{if } a_t^H = 0 \\
 a_t^H, & \text{if } a_t^H > 0
\end{cases}
\]

Each override decisions \( a_t^H > 0 \) is costly, and we use \( \kappa(a_t^H) > 0 \) to denote its cost. The no-override decision has zero cost, i.e., \( \kappa(0) = 0 \). Costs capture the effort incurred by the investor in communicating her decision to the robo-advising firm, which result in lost opportunities because this time could be reserved to other activities.

The human’s risk-aversion levels are assumed to belong to a finite set \( \Theta \), such that \( |\Theta| = m \). \( H \) knows his risk-aversion parameter, while \( M \) does not know it and must estimate it using available information. The machine’s initial prior distribution over the risk-aversion levels is given by \( \pi_1 \in \mathbb{R}^m \).

The objective of the human-machine interaction system is to minimize a risk-adjusted expected cost for each period of the investment horizon. In particular, the time \( t \) cost is given by

\[
c(s_t, a_t^H, a_t^M) = \theta_t \sigma^2(s_t, a_t) - \mu(s_t, a_t) + \kappa(a_t^H),
\]

where \( a_t \) is given by (3). The above cost function weights the risk associated with the investment decision against the expected portfolio return, and accounts for the human’s cost of overriding the robo-advisor’s decision. Hence, the cost function penalizes the amount of risk undertaken (captured by the variance of the selected portfolio) according to the risk-aversion level of the investor. The total cumulative cost is then given by

\[
C_T := \sum_{t=1}^{T} c(s_t, a_t^H, a_t^M)
\]

The objective function therefore takes the form

\[
\min_{a_{1:T}^H, a_{1:T}^M} \mathbb{E} \left[ \sum_{t=1}^{T} \theta_t \sigma^2(s_t, a_t) - \mu(s_t, a_t) + \kappa(a_t^H) \right]
\]

where \( a_{1:T}^H := \{a_1^H, \ldots, a_T^H\} \) and, similarly, \( a_{1:T}^M := \{a_1^M, \ldots, a_T^M\} \). This choice of objective function reflects a non-myopic robo-advisor optimizing the risk-adjusted cost of a myopic investor. The rationale behind our modeling choice is driven by two main considerations. First, we model an investor who wishes to optimize his risk criterion on the short-term, without accounting for the impact of his decisions on later periods. An example of such an investor is a portfolio manager whose compensation package is contingent upon the short-term performance of his portfolio. Second, we think of the robo-advisor as making
decisions to minimize the long-term cost of the investor, so to maintain the long-term satisfaction of the investor high, and increase the changes that he keeping his business with this robo-advisor.

Assuming then that the human behaves myopically, as described above, we can explicitly write the human’s myopic policy \( a^H_t \). First, we denote by \( a_m \) the myopic optimal portfolio at time \( t \), i.e.,

\[
a_m := \arg\min_{a \in \mathcal{A} \setminus \{0\}} \theta_t \sigma^2(s_t, a) - \mu(s_t, a)
\]

Then the myopic human’s policy, after observing the machine decision \( a^M_t \), is given by

\[
a^H_t = \begin{cases} 
0, & \text{if } \theta_t \sigma^2(s_t, a^M_t) - \mu(s_t, a^M_t) + \leq \theta_t \sigma^2(s_t, a_m) - \mu(s_t, a_m) + \kappa(a_m) \\
\quad a_m, & \text{otherwise.}
\end{cases}
\]

so that the human will only override if the risk-adjusted cost of portfolio \( a^M_t \) is lower than the risk-adjusted cost of of the myopic optimal portfolio \( a_m \) plus the override cost.

The human decisions allow the machine to learn and track the human’s risk-aversion parameter via a standard Bayesian update, as described in equation (2).

From the form of the objective function and the learning capabilities of the robo-advisor, it is evident that the human faces a trade-off. On the one hand, the human would like to frequently communicate his risk preferences (through overriding actions) so that the machine is better informed to make investment decisions. On the other hand, the human would like to keep the costs low and not override, unless communication leads to significant improvements in the machine’s portfolio selection strategy. In other words, if the override costs \( \kappa(a^H_t) \), for \( a^H_t > 0 \), is large enough, the human would not have any incentive to override the machine’s decisions, even if they appear suboptimal under the current risk-aversion parameter. Under these circumstances, the robo-advisor will not be able to learn or track the risk-aversion of the human. On the other hand, if the override cost is sufficiently low, the human may find it optimal to communicate his preferences very frequently, and the machine will be able to learn the human’s risk preferences fast.

As discussed in Section 3, the optimization criterion of the robo-advising system may be formulated as a POMDP. It is well known that finding an exact solution to a POMDP, corresponding to an optimal policy, is generally computationally intractable. (Krishnamurthy (2016) presents a comprehensive review of POMDPs for the interested reader.) Therefore, most of the research in POMDPs focuses on efficient approximate solutions. Lovejoy (1991), Littman et al. (1995) and Kaelbling et al. (1998) review several algorithmic methods for approximating a solution and obtaining near-optimal policies, and outline the efficiencies and weaknesses of these procedures. Among the many standard heuristics in the POMDP literature, we consider a heuristic based on the greedy policy with respect to the so-called Q-function, given by equation (6) below, and is obtained from the solution to the fully-observed version of (4) via recursive formulas. This heuristic, in addition to being relatively intuitive and computationally fast (the complexity is the same as that of a fully-observed MDP), generates interpretable model
predictions. The $Q$-function of the fully-observed problem is defined as

$$Q_t(\theta_t, s_t, a_t) := -\mu(s_t, a_t) + \theta_t \sigma^2(s_t, a_t) + \kappa(a_t) + \mathbb{E}[V_{t+1}(\theta_{t+1}, s_{t+1}) \mid \theta_t, s_t]$$

(6)

$$V_t(\theta_t, s_{t+1}) := \min_{a_t} Q_t(\theta_t, s_t, a_t)$$

with boundary condition $V_{T+1} \equiv 0$. Then, the suboptimal policy to solve the unobservable version of (4) can be obtained from the $Q$-function by

$$a_t^M = \max_{a_t} \sum_{\theta \in \Theta} \pi_t(\theta) Q_t(\theta, s_t, a_t)$$

(7)

with $\pi_t$ representing the filtering distribution (or belief state) for the risk-aversion parameter at time $t$. We can then simulate the system under these heuristics to estimate a suboptimal value for the human-machine interaction problem, i.e., an upper bound for the actual optimal value of the human-machine minimization problem.

### 4.1 Numerical Results

This section develops a numerical study to analyze the rate at which the machine learns the human’s risk preferences, and to measure the value added by the robo-advising system over the stand-alone investor.

We use Monte Carlo simulation of the QMDP heuristic to estimate an upper bound for the solution of the human-machine interacting system (4). We fix the number of investment periods $T = 10$, and set the number of portfolios available to the investor at each time $t$ to $n = 4$. We also assume that there are $m = 40$ values that the investor’s risk-aversion parameter can take. Additional details on the numerical study are reported in Appendix B.

We start analyzing how the machine learns, over time, the investor’s risk aversion parameter $\theta_t$. Figure 1 shows the estimation process on one simulated path of the human-machine system. As the investor decides whether or not to override, the machine updates its belief on the investor’s risk-aversion parameter via Bayesian updating (2).

At time $t = 1$, the machine places a uniform prior distribution on a subset of possible initial risk-aversion parameters $\hat{\theta}_1$ of the human (see Appendix B for details). With time, the mass of the posterior distribution concentrates on the set of plausible values, i.e., those that are consistent with the human’s decisions so far.
Next, we analyze the value added by the machine in making decisions, as compared to an investor-only model who makes decisions without any support by the machine. To perform this comparison, we first calculate the approximate expected optimal value produced by the chosen heuristic in the risk-POMDP for a given override cost $\kappa(a^H_t) = k$ if $a^H_t > 0$. Then, we consider a human-only system by reducing the action space of the human to be $\mathcal{A}^H_0 := \mathcal{A}^H \setminus \{0\}$. This means that the human needs to choose her own portfolio at every period $t$ (or equivalently, she must always override the choice of the machine). For comparison purposes, in the human-only setting we assume any action $a^H_t \in \mathcal{A}^H_0$ has the same cost $\kappa(a^H_t) = k$. This cost may be interpreted as the effort incurred by the investor for choosing a portfolio. She needs to closely monitor financial markets, solving her own optimization problem, and communicating her choice to an asset manager. Moreover, the attention span required to make decisions on short time-scales is time subtracted to other activities, and thus represent an opportunity cost for the investor. Clearly, the human-only system corresponds to a fully-observed problem, because the investor knows her own risk-aversion parameter, and so (4) becomes a fully-observed Markov Decision Process (MDP) which can be solved to optimality. Figure 2 shows the approximate optimal value of the risk-POMDP human-machine interaction system, and compares it to the optimal human-only MDP system for a variety of cost levels $k$.

Figure 1: Example of the belief update of the risk-aversion parameter for one simulated path of the financial human-machine system.
Figure 2: Expected approximate minimum cost of Human-Machine system (blue) and expected minimum cost of the Human-Only system (yellow) for different decision costs. The decision cost corresponds to the override cost in the Human-Machine system, and to the decision cost in the human only system.

Figure 2 gives interesting insights into the value added by the robo-advisor. Assuming the override cost in the human-machine system and the decision cost in the human-only system are equal, we can observe that the human-machine interaction provides a lower expected cumulative cost over the investment horizon, compared to the human-only system. This difference can be explained by two main observations: first, the human will not incur costs if the robo-advisor selects the optimal portfolio for the true (unknown) risk-aversion parameter (i.e. there no override is needed), whereas the human-only system incurs these decision costs on every period. In a realistic setting, this is one of the advantages of robo-advising, that allows the human to delegate research on investment instruments, times for portfolio re-balancing, and other time-consuming activities to the machine. This delegating process may considerably reduce human’s costs. It also appears from figure 2 that, as the override / decision cost increases, the overall expected cumulative cost increases (i.e., it becomes less negative). However, this increase in cost is not reflected in a similar fashion by the human-only and the human-machine system. In the human-only system, we observe a linear increase in expected cumulative cost while in the human-machine system the increase in expected cumulative cost slows down for override costs greater than 2%. This can be explained by the previously mentioned trade-off faced by human when deciding on overriding: if the override cost is too high, the human will never choose to override and the robo-advisor does not efficiently learn the risk-aversion parameter of the human. As a result, it will make decisions that satisfy the average human, where average is taken with respect to the initial belief state on the risk aversion.
4.2 Model Extensions

4.2.1 Dynamic Risk-Aversion

In the framework presented in Section 3, the risk-aversion parameter is static. If \( \theta_1 \in \Theta \) is the true risk-aversion of the human at time \( t = 1 \), then \( \theta_t = \theta_1 \) for all \( t \geq 1 \). Our framework may be easily generalized to account for a dynamic risk-aversion parameter, so that the dynamics of \( \theta_t \in \Theta \) are governed by a transition model of the form

\[
\theta_{t+1} = f_\theta(\theta_t, a_t^H, a_t^M, s_t, s_{t+1}) \text{ for } t = 1, \ldots, T - 1. \tag{8}
\]

The transition function \( f_\theta \) may be designed by the modeler to reflect the typical behavior of an investor. Note that \( f_\theta \) is not only a function of the decisions \( a_t^H \) and \( a_t^M \), but also of the current state \( s_t \) and next state \( s_{t+1} \). Hence, risk-aversion parameter transitions are both impacted by investment decisions and changes in the market environment. For example, the modeler may believe that riskier choices, i.e. portfolios with a higher standard deviation, should have a higher impact on the risk-aversion parameter. More specifically, an investor could end up with a higher appetite for risk if the market moves in a favorable direction and the portfolio chosen was high-risk high-return, because the resulting capital and the investor’s optimism would then have increased. Similarly, if the high-risk high-return portfolio was chosen but the market moved in an adverse direction, the appetite for risk could be lower because both the resulting capital and her optimism would have taken a hit. Additionally, the magnitude of the change in risk preferences may also depend on the riskiness of the chosen portfolio, so that a high-risk high-return portfolio may have a higher impact on the capital and optimism than a low-risk low-return portfolio.

Equivalently, we can combine the transition function of the risk-aversion parameter \( \theta \) with the state transitions \( P(s_{t+1} \mid s_t) \), and obtain the risk-aversion transition probability function as

\[
P(\theta_{t+1} \mid \theta_t, a_t^H, a_t^M, s_t) := \sum_{s_{t+1} \in S} \mathbb{I}_{\{\theta_{t+1} = f_{\theta}(\theta_{t}, a_t^H, a_t^M, s_t, s_{t+1})\}} P(s_{t+1} \mid s_t) \tag{9}
\]

where \( \mathbb{I} \) denotes the indicator function. The corresponding Bayesian updating formula would then extend (2) to include the transition probabilities, so that

\[
\pi^*_t(\theta_{t+1}) := \frac{\sum_{\theta_t} \pi^*_t(\theta_t) \sigma^H(\theta_t^H \mid s_t, \pi_t^*, \theta_t) P(\theta_{t+1} \mid \theta_t, a_t^H, a_t^M, s_t)}{\sum_{\theta_t} \pi^*_t(\theta_t) \sigma^H(\theta_t^H \mid s_t, \pi_t^*, \theta_t) P(\theta \mid \theta_t, a_t^H, a_t^M, s_t)}. \tag{10}
\]

4.2.2 Imperfect Human Communication

In the previous sections, we have assumed that the investor always acts optimally, so that any override decision only happens if the portfolio chosen by the robo-advisor significantly differs from the optimal myopic portfolio of the investor. However, there can be situations in which an investor does not have the time or flexibility to override a suboptimal decision made by the robo-advisor. On the other hand, there can be situations in which the
investor incorrectly overrides an otherwise optimal decision made by the robo-advisor. This may happen if the investor incorrectly estimates her risk-aversion parameter, and consequently chooses a suboptimal portfolio. The frequency of these errors is higher for short time-scales, because the investor has a smaller amount of time at her disposal to make decisions.

We consider an extension of the model, which treats humans as imperfect agents. More specifically, we allow the human to commit two types of errors. The first type of error is a false override, which occurs when a human overrides a machine decision that would have been myopically optimal, i.e., recalling the definition of \( a_m \) from equation (5), then

\[
a_t^H = a_m > 0 \text{ if } D_t \geq 0
\]

where

\[
D_t := \theta_t \sigma^2(s_t, a_m) - \mu(s_t, a_m) + \kappa(a_m) - \left[ \theta_t \sigma^2(s_t, a^M_t) - \mu(s_t, a^M_t) \right].
\]

The second type of error is a missed override, in which the human fails to override a suboptimal decision taken by the machine, i.e.

\[
a_t^H = 0 \text{ if } D_t < 0.
\]

We assume both errors occur randomly, with the false override occurring with probability \( P_f \), conditional on \( D_t \leq 0 \), and the missed override occurring with probability \( P_m \), conditional on \( D_t > 0 \). These errors would confuse the learning process of the machine, and we thus expect the machine to take longer in learning the risk-aversion parameter of the human.

Figure 3 illustrates the estimation process on one simulated path of the dynamic risk-aversion parameter, for an imperfect human. The top of figure 3 corresponds to the robo-advisor tracking the risk-aversion parameter as it changes according to market movements and past decisions. The bottom of figure 3 shows the belief on the initial risk aversion parameter \( \theta_1 \), as the investor provides more information to the machine. A direct comparison between figure 1 and figure 3 illustrates that the presence of errors delays the learning process of the machine (after 10 periods, the mode of the belief for the perfect human system reaches 0.5, while the same quantity for the imperfect human system reaches 0.24). In addition, panel (a) of figure 3 shows that the mode of the filtering probability mass changes to reflect the actual dynamics of the risk-aversion parameter.
Figure 3: Updating of beliefs on the dynamic risk-aversion parameter for an imperfect human with $P_m = 0.4$ and $P_f = 0.1$. We illustrate the result on a sample path of the human-machine system. (a) Tracking of the human’s dynamic risk-aversion parameter via the bayesian filtering distribution. (b) Smoothed probability mass over the initial risk-aversion parameter $\theta_1$.

5 Conclusion and Future Work

In this paper, we presented a framework for human-machine decision making, accounting for both human-driven and context-driven risk. Due to the different risk sensitivities of the human and the machine, respectively, to the context in which the task is being executed and to the category of humans served, the optimal decision making problem may be formulated as a game with strategic interactions. We have introduced the concept of risk-sensitive equilibria to deal with the corresponding game, and introduced risk-sensitive optimization techniques to solve a related coordinator problem. The optimal solution to the risk-POMDP is then a risk-sensitive Bayesian equilibrium for the human-machine framework.

We have specialized our framework to capture the interactions between an investor and a robo-advising firm. Our numerical study highlights the trade-off between frequent communication of preferences by the investor and the costs of such a communication. If the investor intervenes frequently, the machine can learn the risk-aversion parameter of the investor faster, and therefore make more tailored portfolio decisions. On the other hand, each override decision of the investor is costly, and these total costs may exceed the performance gain stemming from more informed investment decisions by the machine. The robo-advising firm provides a service to the investor that may be superior to a stand-alone investor making all investment decisions on her own. Assuming that override
costs occurring in the human-machine system and market research costs occurring in the human-only system are equal, our numerical analysis suggests that the objective risk function is lower in the human-machine interacting system. More importantly, since human costs are incurred in all periods for the human-only system, an increase in these costs translates to a linear increase in the human-only expected cumulative costs. On the other hand, the cost increase in the human-machine system is bounded. This is because if the investor does not communicate her preferences, the robo-advisor will make portfolio decisions using its initial belief on the human’s risk aversion, without updating it. These decisions, however, will not be tailored to the specific risk-profile of the investor.

Future directions for this research include the development of new solution methods to integrate risk optimization techniques with concepts from game theory. A key refinement to equilibrium in dynamic games is the notion of subgame perfection. This enforces incentive compatibility for both agents in each subgame initiated at the start of each period. However, many commonly used risk functions are not time-separable, i.e. the risk over the entire horizon cannot be decomposed into a set of risks, each allocated to a different time period. Without time separability, the risk-POMDP no longer satisfies the Markov property.

Appendix

A Proof of Theorem 1

Given a solution to the coordinator problem, \( \sigma^* = (g_{\sigma^M}, g_{\sigma^H}) \), assume that the human’s true type is an arbitrary value \( \theta \in \Theta \). Then it is sufficient to confirm that the strategies \( \sigma^H = g_{\theta}^H \) and \( \sigma^M = g_{\theta}^M \) satisfy the three properties for the risk-sensitive Bayesian equilibrium:

(I) Machine’s incentive compatibility

\[
\rho^M \left( \rho^H \left( C_T | \sigma^H, \sigma^M, \pi_1, h_1 \right) | \pi_1 \right) \leq \rho^H \left( C_T | \sigma^H, \sigma^M, \pi_1, h_1 \right) | \pi_1
\]

(II) Human’s incentive compatibility

\[
\rho^H \left( C_T | \sigma^H, \sigma^M, \pi_1, h_1 \right) \leq \rho^H \left( C_T | \sigma^H, \sigma^M, \pi_1, h_1 \right)
\]

(III) The consistent belief profile

\[
\pi_{t+1}(\theta) := \frac{\pi_t^\theta(a_t^H | s_t, \pi_t^\theta, \theta) \sigma_t^H \left( s_t, \pi_t^\theta, \theta \right)}{\sum_{\theta'} \pi_t^\theta(a_t^H | s_t, \pi_t^\theta, \theta')}
\]

The machine’s incentive compatibility (I) is satisfied since the objective function of the coordinator is equal to the objective function of the machine. The consistent belief profile (III) follows directly from the formulation of the coordinator problem as a POMDP.
We provide further details on the numerical study conducted in Section 5.1. Recall that this characteristic is also observed in equity markets, and known as the leverage effect. We are assuming the time horizon causes an increase in expected return, the standard deviation corresponding decreases. Hence, portfolios feature the risk-return tradeoff. Riskier portfolios have a higher standard deviation for the state in the middle, i.e., for \( \mu, \sigma \) given by

\begin{equation}
\mu(s, i) = \mu(s, i) + 0.02(s - s_1), \quad \text{for } s \in S, \ i = 1, 2, 3, 4,
\end{equation}

\begin{equation}
\sigma(s, 1) = \sigma(s_1, 1) - 0.005(s - s_1), \quad \text{for } s \in S,
\end{equation}

\begin{equation}
\sigma(s, 2) = \sigma(s_1, 2) - 0.01(s - s_1), \quad \text{for } s \in S,
\end{equation}

\begin{equation}
\sigma(s, 3) = \sigma(s_1, 3) - 0.02(s - s_1), \quad \text{for } s \in S,
\end{equation}

\begin{equation}
\sigma(s, 4) = \sigma(s_1, 4) - 0.04(s - s_1), \quad \text{for } s \in S.
\end{equation}

As mentioned in Section 4, for each state \( s \in S \) and portfolio \( i \in \{1, 2, 3, 4\} \) we have expected return \( \mu(s, i) \) and standard deviation \( \sigma(s, i) \). We define these state and portfolio dependent parameters using a parametric specification. We specify expected return and standard deviation for the state in the middle, i.e., for \( s_1 = 11 \) we set \( \mu(s_1, 1) = 0.05, \mu(s_1, 2) = 0.10, \mu(s_1, 3) = 0.15, \mu(s_1, 4) = 0.20 \), and \( \sigma(s_1, 1) = 0.05, \sigma(s_1, 2) = 0.15, \sigma(s_1, 3) = 0.30, \sigma(s_1, 4) = 0.50 \). Hence, \( \mu(s, i) \) and \( \sigma(s, i) \) are increasing with respect to the index \( i \), with portfolio 4 being the riskiest and portfolio 1 being the safer. The remaining parameters are given by

\begin{equation}
P(s'|s) = \begin{cases} 
0.5, & \text{if } s < n_s \text{ and } s' = s + 1, \\
0.5, & \text{if } s = n_s \text{ and } s' = n_s, \\
0.5, & \text{if } s > 1 \text{ and } s' = s - 1, \\
0.5, & \text{if } s = 1 \text{ and } s' = 1, \\
0, & \text{otherwise} 
\end{cases}
\end{equation}

Then the monotonicity property implies that the coordinator’s objective function can be reduced using the strategy \( q^H_\theta = \tilde{\alpha}^H. \) Thus we arrive at the desired contradiction that the solution to the coordinator problem is optimal.

**B Details on the Numerical Study**

We provide further details on the numerical study conducted in Section 5.1. Recall that we are assuming the time horizon \( T = 10 \), and setting the number of portfolios to \( n = 4 \).

We take the state space to be the set of indices \( S := \{1, \ldots, n_s = 21\} \), with transitions given by

\begin{equation}
P(s'|s) = \begin{cases} 
0.5, & \text{if } s < n_s \text{ and } s' = s + 1, \\
0.5, & \text{if } s = n_s \text{ and } s' = n_s, \\
0.5, & \text{if } s > 1 \text{ and } s' = s - 1, \\
0.5, & \text{if } s = 1 \text{ and } s' = 1, \\
0, & \text{otherwise} 
\end{cases}
\end{equation}

This characteristic is also observed in equity markets, and known as the leverage effect.
The space $\Theta$ of risk-aversion parameters consists of equally spaced points on a grid of size $m = 40$ on the interval $[0, 1]$, i.e., $\Theta := \{0.025, 0.05, 0.075, \ldots, 0.95, 0.975, 1.0\}$. The transition function $f_\theta$ is given by

$$f_\theta(\theta_t, a_t^H, a_t^M, s_t, s_{t+1}) := \begin{cases} 
\min(\theta_t + 0.025 \times a_t, 1.0), & \text{if } s_t < s_{t+1} \\
\theta_t, & \text{if } s_t = s_{t+1} \\
\max(\theta_t - 0.025 \times a_t, 0.025), & \text{if } s_t > s_{t+1}
\end{cases}$$

where $a_t$ is given by (3). Note that this definition of $f_\theta$ captures the properties discussed in Section 4, i.e., changes in the risk-parameter are greater for riskier portfolios (as given by $a_t$), while the direction of the change in risk-aversion is determined by the sign of the state change, so that $\theta_{t+1} > \theta_t$ if $s_t < s_{t+1}$ and $\theta_{t+1} < \theta_t$ if $s_t > s_{t+1}$, as discussed in Section 4.

We assume that the initial distribution $\pi_1$ on $\theta_1$ is uniform on a subset of twenty values, given by $\Theta_s := \{0.275, 0.30, 0.325, \ldots, 0.70, 0.725, 0.75\} \subset \Theta$. Hence, $\pi_1(\theta) = 0.05$ for $\theta \in \Theta_s$. The rationale behind this choice is to allow a certain degree of uncertainty in the risk-aversion parameter, and to allow for the possibility of the risk-aversion parameter to take values outside $\Theta_0$, while at the same time decreasing the probability of hitting the boundaries of $\Theta$.

References


