Robo-Advising as a Human-Machine Interaction System

Agostino Capponi Octavio Ruiz Lacedelli Matthew Stern

Columbia, Columbia, Columbia

July 2018

Focus of Research

A framework for the analysis of human-machine interactions in the context of robo-advising.

- **Human**
  - Internalized preferences
  - "Soft" information
- **Machine**
  - Learns preferences
  - "Hard" information

Objectives:

- How does a machine learn the preferences of the human?
- What is the tradeoff between humans' communication costs and preference gains?
- What if risk preferences evolve? What if humans make mistakes?

Replacing soft information with hard information

- Traditional advisor / investor relationship
  - Periodic communication
  - Investor education
  - Long term portfolio planning
- Robo-advisor / investor relationship
  - Web portal access / electronic reminders
  - Investor-reported quantitative information
  - Database of investors with similar characteristics

Human-Machine System

- Human-Driven Risk - $\rho^{\mathbf{M}}$
  - Machine has uncertainty over humans' characteristics:
    - risk preferences
  - Machine can optimize for average human, top quartile, worst-case.
- Context-Driven Risk - $\rho^{\mathbf{H}}$
  - Market outcomes are random
  - Portfolio choices depend on human's attitude towards risk

Setup and Objectives

- Both human and machine share the common goal of optimizing investor's value
- Heterogenous risk sensitivity: human and machine employ different risk functions
- Informational asymmetries: human knows his type, but machine does not
- Competitive game
  - *Human*: Align objectives of machine with his own
  - *Machine*: Provide reliable service to an unknown human
- Objectives are aligned in the absence of informational asymmetries
  - Cooperative game

Related Fields of Study

- Single agent risk optimization:
  - **Expected utilities:** Bauerle and Rieder (2014, 2016)
  - **Dynamic risk measures:** Shapiro et al. (2009), Haskell and Jain (2015)
- Multi-agent optimization:
  - **Cooperative IRL:** Russell (2016)
  - **Decentralized optimization:** Nayyar et al. (2013), Vasal and Anastasopoulos (2016), Seuken and Zilberstein (2007)
- Robo-advising:
  - **Goals-Based Investing:** Das et al. (2000, 2018), Bettermont
  - **Dynamic mean-variance:** Dai et al. (2018)

Framework

- *Actions*:
  - Machine: $a_t^{\mathbf{M}} \in \mathcal{A}^{\mathbf{M}}$
  - Human: $a_t^{\mathbf{H}} \in \mathcal{A}^{\mathbf{H}}$
  - Total Costs: $C_T := \sum_{\tau=1}^{T} c\left(s_\tau, a_\tau^{\mathbf{H}}, a_\tau^{\mathbf{M}}\right)$
- *Risk Measure of Human*: $\rho_\theta^{\mathbf{H}}(C_T)$
  - $\theta \in \Theta$ known to Human but not to Machine
- *Risk Measure of Machine*: $\rho^{\mathbf{M}}\left(\rho_\theta^{\mathbf{H}}(C_T)\right)$
- *States*:
  - System state: $s_t \in \mathcal{S}$
  - Belief state over Human's preferences: $\pi_t(\theta) \in [0,1]$

Strategies

- Public histories
$$H_t := \left(\mathcal{A}^{\mathbf{H}} \times \mathcal{A}^{\mathbf{M}}\right)^{t-1} \times \mathcal{S}^t$$
- A Markov strategy for the Human $\sigma^{\mathbf{H}} = (\sigma_1^{\mathbf{H}}, \ldots, \sigma_T^{\mathbf{H}})$ is
$$\sigma_t^{\mathbf{H}}(a|s_t, \pi_t, \theta) = P(a_t^{\mathbf{H}} = a|s_t, \pi_t, \theta), \ \ \forall t \in \{1, \ldots, T\}$$
- A Markov strategy for the Machine $\sigma^{\mathbf{M}} = (\sigma_1^{\mathbf{M}}, \ldots, \sigma_T^{\mathbf{M}})$ is
$$\sigma_t^{\mathbf{M}}(b|s_t, \pi_t) = P(a_t^{\mathbf{M}} = b|s_t, \pi_t), \ \ \forall t \in \{1, \ldots, T\}$$
- Human's strategy depends on the Machine's current beliefs:
  - Human is influenced by the action of the Machine, which in turn depends on its belief over the Human's type
- Conflicting objectives: two-player strategic game.

Solution Methodology

- Transform the strategic game to a single-agent problem.
- Coordinator with objective: $\min_{\sigma^{\mathbf{C}}} \rho^{\mathbf{M}} \left(\rho_\theta^{\mathbf{H}} \left(C_T^{\sigma^{\mathbf{C}}} \right) \right)$
- Coordinator assigns a policy $\sigma^{\mathbf{C}} = (\sigma^{\mathbf{M},\mathbf{C}}, \sigma_\theta^{\mathbf{H},\mathbf{C}})$
- $\sigma_\theta^{\mathbf{H},\mathbf{C}}$, the human's strategy for each possible realization of θ.
- Theorem: A solution to the coordinator problem is a risk-sensitive Nash equilibrium to the two-agent Human-Machine interaction game.

Robo-Advising: Setup

Investor wants to delegate portfolio management to a robo-advisor

- **Benefits:**
  Investor delegates market research on investment instruments, times for portfolio re-balancing, and other time-consuming activities to the robo-advisor
- **Costs:**
  - If robo-advisor does not act in accordance with investor's risk preferences, the client may override the portfolio decisions of the robo-advisor
  - Overriding actions are burdensome to the client

Cost Functions

- Cost in period t

$$c_\theta(s_t, a_t^{\mathbf{H}}, a_t^{\mathbf{M}}) = \theta\sigma^2(s_t, a_t) - \mu(s_t, a_t) + \kappa(a_t^{\mathbf{H}}),$$

where a_t represents the actual portfolio chosen

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

- Cost weights the risk associated with the investment decision against the expected portfolio return
- Accounts for the human's cost of overriding the robo-advisor's decision, $\kappa(a_t^{\mathbf{H}})$
- The total cumulative cost is given by

$$C_T := \sum_{t=1}^{T} c_\theta(s_t, a_t^{\mathbf{H}}, a_t^{\mathbf{M}})$$

- The risk function of the human is then given by

$$\rho_\theta^{\mathbf{H}} := \mathbb{E}[C_T]$$

where the expectation is taken w.r.t. probability distribution of the state path $s_1, \ldots, s_T$

Human-Machine Objectives

- The risk function of the machine is given by

$$\rho^{\mathbf{M}} := \mathbb{E}[\rho_\theta^{\mathbf{H}}]$$

where the expectation is taken w.r.t. the machine's belief on the investor's risk-aversion parameter θ

- A non-myopic robo-advisor optimizes the risk-adjusted cost of a myopic investor

$$\min_{a_{1:T}^{\mathbf{H}}, a_{1:T}^{\mathbf{M}}} \rho^{\mathbf{M}} = \min_{a_{1:T}^{\mathbf{H}}, a_{1:T}^{\mathbf{M}}} \mathbb{E}\Big[\sum_{t=1}^{T} \theta\sigma^2(s_t, a_t) - \mu(s_t, a_t) + \kappa(a_t^{\mathbf{H}})\Big]$$

- The investor wishes to optimize her risk criterion on the short-term, without accounting for the impact of her decisions on later periods.
  - Portfolio manager whose compensation package is contingent upon the short-term performance of her portfolio
- Robo-advisor makes decisions to minimize the long-term cost of the investor, so to maintain the long-term satisfaction of the investor high

Machine Policy

- Requires finding an exact solution to a POMDP (computationally intractable)
- Consider a heuristic based on the Q-function

$$
\begin{aligned}
Q_t(\theta, s_t, a_t) &:= \theta\sigma^2(s_t, a_t) - \mu(s_t, a_t) + +\kappa(a_t) + \mathbb{E}\Big[V_{t+1}(\theta, s_{t+1}) \mid s_t\Big] \\
V_t(\theta, s_{t+1}) &:= \min_{a_t} Q_t(\theta, s_t, a_t) \\
V_{T+1} &:= 0.
\end{aligned}
$$

- Suboptimal policy for the human-machine objective is the greedy policy w.r.t. Q-function, i.e.,

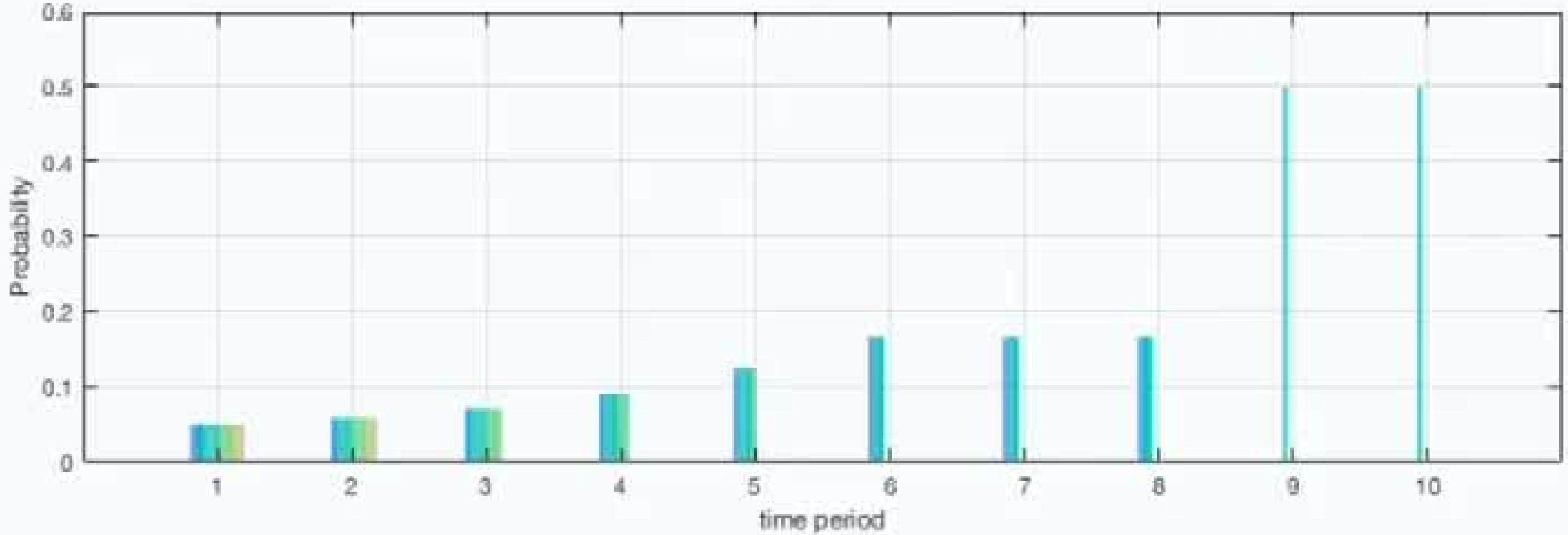
$$
a_t^{\mathsf{M}} = \max_{a_t} \sum_{\theta \in \Theta} \pi_t(\theta) Q_t(\theta, s_t, a_t)
$$

- Simulate the system under this heuristic policy to estimate an upper bound for the cost objective of the human-machine interaction problem

Numerical Study

- The learning trade-off:
  - Frequent overriding (human-machine interactions) help machine learn human's preferences faster
  - Each override action is costly
- Performance improvement:
  - Benchmark: **Human-Only** System
  - Investor makes all portfolio decisions
  - Incurs a cost κ at every period, equal to the override cost in Human-Machine system
  - κ captures effort, time, resources spent in market research
  - This fully-observed system can be solved to optimality

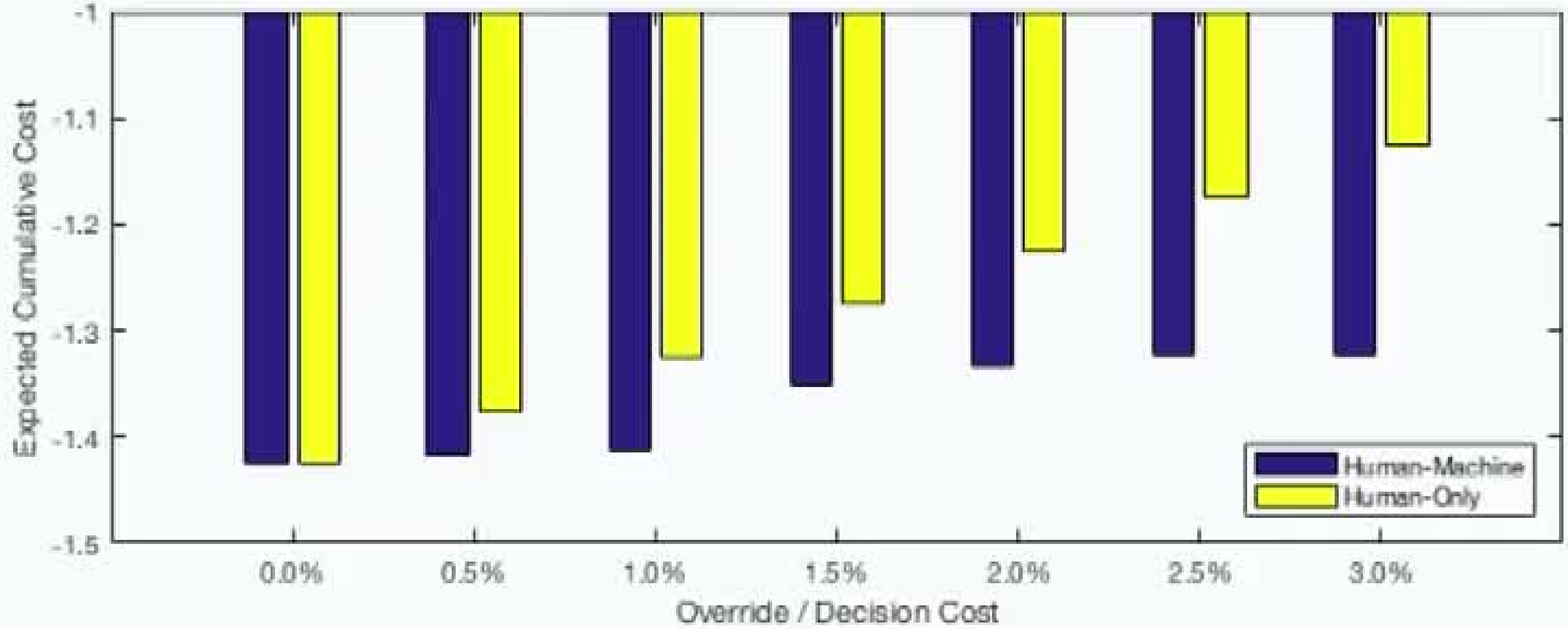
Learning Investor Preferences



Belief update of the risk-aversion parameter on one sample path.

- Assume $m = 20$ and $\Theta := \{0.05, 0.10, \ldots, 0.95, 1.0\}$
- Assume $\pi_1(\theta) = 0.05$ for $\theta \in \Theta$
- With time, beliefs concentrate on θ values that are consistent with the investor's decisions
- After 9 periods, machine learns θ can be two possible values (each w.p. 0.5)

Does the Machine add Value?

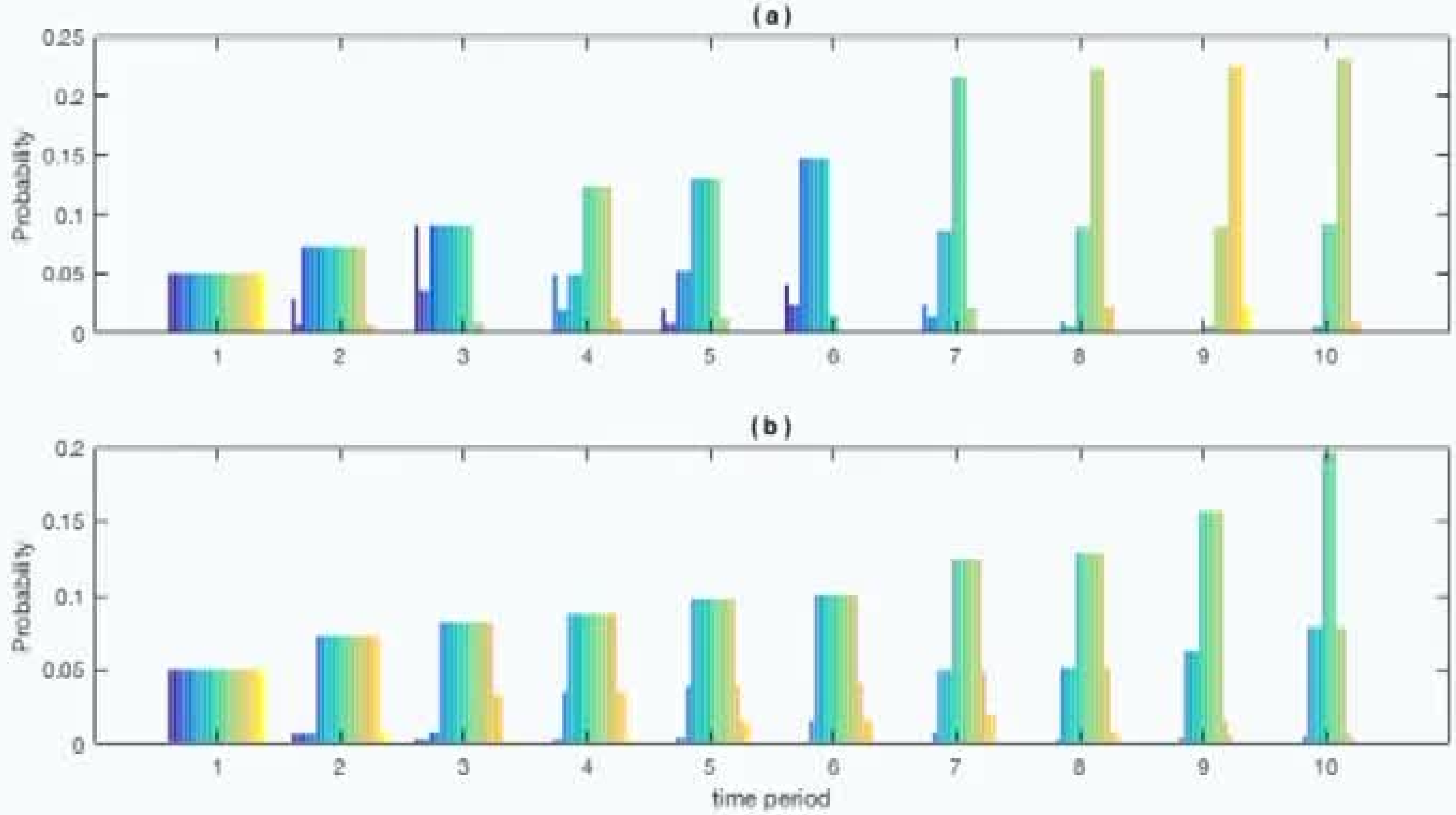


- Human-Only System: cost incurred every period by the investor
  - Expected cumulative cost increases linearly
- Human-Machine System: cost incurred when investor overrides
  - Increase in expected cumulative cost is limited
  - Override-Learning trade-off: if override cost is too high, investor does not override and machine does not learn investor's preferences.
  - Without overriding, the Robo-advisor's decisions satisfy the average investor (w.r.t. π_1)

Model Extensions: Imperfect Human

- Investors do not always act optimally
- Can generalize framework to include error prone investors:
  - **False Override**: investor overrides a machine decision that would have been myopically optimal
  - **Missed Override**: investor fails to override a suboptimal decision taken by the machine
- Assume errors occur randomly:
  - False Override occurs with probability P_f
  - Missed Override occurs with probability P_m
- These errors would confuse the machine and delay learning process
- Expect the machine to take longer for learning the risk-aversion parameter of the investor

Numerical Results: Model Extensions



- Belief updates of dynamic risk-aversion parameter for an imperfect human ($P_m = 0.4$ and $P_f = 0.1$) for one sample path
- (a) Tracking the risk-aversion parameter as it changes according to market movements and past decisions
- (b) Belief updates on the initial risk aversion parameter θ_1, as the investor provides more information to the machine