Human Edge, Machine Limits: AI-Human Competition in Financial Markets

November 18, 2025

Abstract

We develop a theoretical framework to study competition between AI-powered and human investors with heterogeneous sophistication. Human investors possess superior private information but are limited by bounded strategic reasoning, modeled through a cognitive-hierarchy structure. AI investors, in contrast, learn and trade through reinforcement learning that autonomously optimizes trading profits over time. We show that human investors can consistently outperform sophisticated AI investors because AI sophistication is constrained by the data it learns from, which reflects the behavior of the average rather than the most advanced human trader. Three forces limit AI profitability: (i) the advantage of human private information, (ii) the price-stabilizing actions of the most sophisticated human traders, and (iii) the growing price impact of AI trading as its market share expands. Together, these mechanisms reveal the limits of algorithmic superiority and provide a foundation for understanding AI–human competition in financial markets.

Keywords: Training Data Limitations, Bounded Strategic Reasoning, Level-K Reasoning, AI Dominance, Reinforcement Learning, Private Information, .

JEL Classification: C72, D80, G10, L19.

1 Introduction

Financial markets are undergoing a profound transformation driven by AI-powered trading that integrates algorithmic execution with reinforcement learning. These systems continuously interact with the market to learn and autonomously optimize their trading strategies. Their rapid expansion has heightened market complexity, challenged human investors, and raised pressing regulatory concerns.

A central question is whether AI algorithms can ultimately outperform human investors and erode their welfare, particularly among those with limited access to frontier AI trading technologies. Existing research has primarily examined the competitive dynamics, strategic interactions, and potential collusive behavior among oligopolistic AI-powered traders. In contrast, we focus on the direct competition between AI and human investors in financial markets. To this end, we develop a theoretical framework featuring a discrete-time, infinite-horizon trading environment with a continuum of human investors and a continuum of AI-powered investors, each seeking to maximize their own net trading profits.

The core of our framework lies in the distinct capabilities of each type of investor. Human investors possess an informational advantage: they can collect proprietary data, acquire valuable soft information, conduct fundamental research, and interact directly with firm management, which we model as receiving private signals about future asset payoffs that are difficult for AI algorithms to access. However, they are constrained by bounded strategic reasoning and may misperceive the behavior of other market participants, leading to misinterpretation of the information embedded in equilibrium prices. In contrast, AI investors start with no direct knowledge of the market environment because they do not know the distribution of asset payoffs or the strategies of other traders. Instead, they employ reinforcement learning algorithms that trade repeatedly, learn from realized profits and losses, and gradually optimize their strategies through experience. In practice, such reinforcement learning algorithms can be trained efficiently in synthetic trading environments that replicate realistic market dynamics and interactions at low cost.

Importantly, both types of investors base their demands on beliefs about price formation. Human investors apply Bayesian updating to combine private signals with information inferred from prices according to their perceived price formation model, forming posterior beliefs that guide their investment decisions. The AI investors' perceived price formation model is implicitly encoded within their reinforcement learning algorithms and evolves through experience. The market equilibrium is jointly determined by investors' demand functions and the market-clearing condition.

We model the bounded strategic thinking of human investors in the spirit of the Cognitive Hierarchy (CH) framework developed by Camerer, Ho, and Chong (2004). In this framework, some human investors hold incorrect and overconfident beliefs about the sophistication of others. Their strategies are formed through an iterated process of strategic thinking. We define the most naive investors as Step-0. These investors are unaware of the presence of AI investors and completely ignore the informational content of prices. A Step-k human investor is overconfident in the sense that they fail to recognize that other humans might be thinking at step k or higher. Instead, they believe the market consists only of a mix of human investors with fewer than k steps of reasoning, whose population follows a $(k-1)$-truncated Poisson distribution (right truncated at $k-1$). Step-k investors are aware of the AI investor but misperceive its strategy. They believe the AI has been trained against this simplified, $(k-1)$-truncated distribution of human investors (we term this the $(k-1)$-step AI).

Specifically, we assume that human investors believe the AI algorithm, after extensive training, can fully understand the environment it faces. The algorithm converges to an optimal, rational expectations strategy.[1] We provide both theoretical and simulation-based support for this assumption. Using Q-learning, a simple yet widely used RL algorithm, we demonstrate how the AI investor can solve the investment problem. We formally lay out the Q-learning algorithm and establish the necessary conditions for it to converge to the true action-value function, Q^*, and optimal demand, $x^*_{A,t}$. Simulation evidence further

[1] We abstract away from modeling human investors' specific beliefs about the dynamic training process of the RL algorithm.

confirms this convergence. We also show that the these results are not confined to Q-learning. They apply to a broad class of RL algorithms that have been proven, either theoretically or empirically, to converge to optimal strategies in such environments.

Our primary analysis centers on comparing the profitability of human and AI investors across different objective market environments. We define a k_{Env} objective environment as one that matches the training environment of a k_{Env}-step AI investor. Specifically, it is populated by human investors whose steps of thinking follow the k_{Env}-truncated distribution and a k_{Env}-step AI investor. A key finding is that human investors can consistently outperform AI-powered investors across a variety of environments. The average expected profit for human investors, even including those with naive strategies, is often higher than that of the sophisticated AI. Specifically, in an environment populated by the full spectrum of human sophistication levels (the $k_{Env} = \infty$ environment), we find that human investors at all sophistication levels outperform the corresponding Step-∞ AI. The AI investor can only outperform naive human investors in environments that consist solely of naive humans (e.g., only Step-0 and Step-1 humans) and only if the proportion of these investors is sufficiently high.

A key factor limiting AI's ability to outperform naive human investors is the quality of human investors' private information. The AI investor's outperformance is non-monotonic with respect to human signal precision and is constrained at both extremes of human signal quality. When human investors have very precise signals, they can outperform the AI due to this informational advantage. Even though the AI investor can correctly understand the information in equilibrium prices while naive humans misperceive it, the signals from prices are still noisier than the private signals of human investors. As the precision of human signals drops, this informational advantage decreases, and the AI begins to outperform humans. However, as signal precision drops further, the AI's competence deteriorates again. Market prices aggregate private information from human investors. Excessive noise in human signals makes prices less informative, limiting the AI's ability to profit from learning the price function. As human signals become infinitely noisy, neither humans nor the AI can obtain useful information, and their expected profit difference approaches zero. The AI's comparative advantage is its ability to learn from prices, but this information originates from the aggregation of human private signals. If this source of information becomes uninformative, the AI investor's primary advantage is nullified.

The presence of sophisticated human investors also constrains AI's profitability. In environments with more sophisticated humans (a higher k_{Env} environment), expected profits for both naive humans and AI decrease, but the effect is significantly more pronounced for AI. This occurs because the trading activity of sophisticated human investors tends to stabilize prices. The strategies of higher-step humans converge toward more moderate strategies. Their demand sensitivity to private information and price elasticity of demand are neither too high nor too low. As the proportion of these moderate investors increases, both information content of prices and overall price volatility decline. This price stabilization directly curtails AI's ability to extract profits from learning the price function and exploiting price fluctuations.

Finally, the relative size of the AI sector itself acts as a key constraint on its profitability. We find that as the proportion of AI investors in the market increases, the AI's profitability monotonically decreases compared to that of humans. A larger AI sector has a greater price impact, which inherently limits its ability to profit from its strategies against the remaining human investors.

Contributions and related literature. First, this paper contributes to the growing body of work on how AI market participants compete and influence the market environment. A closely related stream of this research investigates the impact of AI on financial markets, with a specific focus on the interactions among algorithms.[2] For instance, Dou, Goldstein, and Ji (2025) study informed speculators who use Q-learning algorithms. They find that

[2] Another stream of literature focuses on algorithmic collusion in retail markets, e.g., Calvano et al. (2020, 2021); Johnson, Rhodes, and Wildenbeest (2023); Waltman and Kaymak (2008); Hansen, Misra, and Pai (2021); Abada and Lambin (2023); Banchio and Mantegazza (2024).

these agents can autonomously learn to sustain collusive, supra-competitive profits without explicit communication, which harms competition and market efficiency. Colliard, Foucault, and Lovo (2022) focus on algorithmic market makers using Q-learning algorithms to set prices. They also show that these algorithms fail to learn competitive pricing strategies, a failure they attribute to limited experimentation and noisy feedback. Routledge (1999) and Routledge (2001) explore whether adaptive algorithms can converge to a rational expectations equilibrium within a repeated Grossman and Stiglitz (1980) model. They prove convergence for adaptive learning and provide examples for genetic algorithms, showing that both can converge to the rational expectations equilibrium. In these cases, their algorithms learn to make correct inferences about a signal from the market-clearing price. Other notable works studying the impact of AI algorithms on financial markets include Marimon, McGrattan, and Sargent (1990), Cartea et al. (2022), and Cartea, Chang, and Penalva (2022).

Our work differs from the existing literature in several important ways. First, while much of the literature focuses on the interaction among multiple reinforcement learning agents or the convergence of algorithms to rational expectations, we center our analysis on the competition between humans and AI. Our paper provides a framework for understanding the distinct comparative advantages of humans and AI and for analyzing how the market environment affects this human-AI competition. Second, whereas most existing works rely on simulation-based analysis, we provide an analytical characterization of the equilibrium and how changes in key factors affect competition and market outcomes. We also establish analytical conditions for AI algorithms to learn and converge to rational expectations in our setting. Third, unlike research that focuses on a single algorithm (e.g., Q-learning), our analysis is not restricted to one type but applies to a broad class of algorithms. We thereby offer a more general theoretical framework.

A related contemporaneous work by Banerjee and Szydlowski (2025) studies a market with rational investors and a single Q-learning trader. They find that the Q-learner's feedback-driven trading generates stochastic volatility and predictable returns, and can some-times improve overall investor utility despite increasing price volatility. While their focus on human-AI interaction is similar to ours, our paper differs and contributes in three significant ways. First, we incorporate information asymmetry, allowing humans and AI to possess distinct informational advantages. This asymmetry captures a key difference between humans and AI in financial markets and is crucial for understanding the competition between them. Second, our analysis applies to a wide range of reinforcement learning algorithms post-convergence to a stable policy. We abstract from the proprietary details of both the specific type and the training process of the algorithm. It is unrealistic for human investors to know the details of the algorithm being used by their AI competitors, such as the exact type of algorithm, specific hyperparameters, or its stage of training, all of which greatly affect how the algorithm evolves and converges. Our focus on the post-convergence phase therefore reflects a more realistic scenario where investors compete with stable, deployed trading strategies used by other market participants. Third, we introduce the concept of bounded strategic thinking for human investors, allowing them to have misperceptions about others' behavior when facing a complex market environment that includes AI traders. By using the CH framework, we provide a comprehensive framework to model human perception of AI. This approach more closely captures the diverging sophistication levels of real-world investors.

Furthermore, this paper contributes to the theory literature on imperfect competition and bounded strategic reasoning in financial markets. Our work introduces different levels of bounded strategic thinking into an imperfectly competitive financial market in the spirit of Kyle (1989).[3][4] We model human investors' strategic thinking using the Cognitive Hierarchy framework of Camerer, Ho, and Chong (2004), which posits that agents have iterated levels of reasoning about others. This approach is related to, but distinct from, the level-k

[3] See Crawford, Costa-Gomes, and Iriberri (2013) for a review on recent theory and evidence on strategic thinking and the applications of level-k models. Many other experimental papers show direct evidence of level-k thinking (Stahl and Wilson, 1994, 1995; Nagel, 1995; Costa-Gomes, Crawford, and Broseta, 2001; Costa-Gomes and Crawford, 2006).

[4] A recent literature studies the impact of bounded strategic thinking in macroeconomics (García-Schmidt and Woodford, 2019; Farhi and Werning, 2019; Angeletos and Lian, 2023).

thinking models used in papers like Zhou (2022). Unlike standard level-k models where a player believes all others are level-(k-1), the CH framework assumes a player best responds to a distribution of lower-level types. Zhou (2022) uses level-k thinking to model human speculators and focuses on how bounded strategic reasoning can generate momentum and contrarian trading strategies. In contrast, our application of the CH model provides a game-theoretic foundation for understanding how the distribution of human sophistication levels affects the competition between humans and AI. Other works study the effect of higher-order beliefs and the perception of information in asset prices on financial markets (Allen, Morris, and Shin, 2006; Han and Kyle, 2018; Eyster, Rabin, and Vayanos, 2019). Eyster, Rabin, and Vayanos (2019) studies how the presence of investors who do not fully invert prices to uncover others' information (cursed) affects the financial markets and considers the degree of cursedness as the measure of sophistication. This paper considers sophistication as the level of iterated reasoning, which incorporates human investors' reasoning about AI investors' behavior.

The rest of the paper is organized as follows. Section 2 lays out the benchmark model, characterizes the reasoning capacity and strategies of human investors, and derives market outcomes and investors' expected profits. Section 3 details the reinforcement learning algorithm and discusses the convergence properties of the AI's strategy. Section 4 shows the main results on the competition between AI and human investors. Section 5 concludes.

2 Benchmark Model

Within a rational expectations equilibrium framework, a key assumption is that investors accurately perceive the distribution of random variables and form expectations about equilibrium prices that are consistent with actual outcomes. However, the fast growth of AI trading algorithms has increased market complexity and poses challenges for market participants to form correct beliefs about others' behaviors and thus about price formation. These difficulties may cause the equilibrium to deviate from the rational expectations equilibrium. This deviation is important to understand the competition between AI and human investors.

Humans and AI exhibit distinct strengths and weaknesses in adapting to this environment. In Section 2.1, we present the framework of our benchmark model and describe the comparative competencies of human and AI investors. Section 2.2 defines the reasoning capacity of human investors and examines how the level of sophistication influences their investment strategies. Section 2.3 discusses the properties of the strategies of human investors with different levels of reasoning capacity and the strategies of different AI investors. Section 2.4 then defines the objective environment and describes the resulting market outcomes. Section 2.5 derives the expected profits of human and AI investors.

2.1 Model Setup

Assets. Time is discrete and infinite, indexed by $t = 1, 2, \ldots$. A single risky asset is traded in each period. The asset has a per capita supply of S. Its payoff, v_t, is realized at the end of the period and follows a normal distribution: $v_t \sim \mathcal{N}\left(\bar{v}, \sigma_v^2\right)$.

Human investors. There is a continuum of human investors of measure ψ in the market. They maximize the expected present value of net trading profits:

$$\mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \pi_{i,t}\right], \tag{1}$$

where $0 < \gamma < 1$ is the discount rate. The trading profit $\pi_{i,t}$ is determined by

$$\pi_{it} = (v_t - p_t) \cdot x_{it} - \frac{1}{2}\rho_M x_{i,t}^2.$$

p_t is the equilibrium asset price, and x_{it} is the number of shares of the risky asset that investor i purchases at the beginning of period t. The term $\frac{1}{2}\rho_M x_{i,t}^2$ represents transaction costs, and

ρ_M controls the level of these costs.[5] We assume that human investors are memoryless with respect to past prices.[6]

We assume that human investors possess informational advantages. By gathering data and conducting primary research to develop informed views about future fundamentals, they can generate more accurate forecasts of future asset payoffs.[7] At the beginning of each period t, each human investor i observes a private signal about the end-of-period payoff of the risky asset. The private signal is given by

$$\eta_{it} = v_t + e_{it},$$

where e_{it} are i.i.d. and $e_{it} \sim \mathcal{N}(0, \sigma_M^2)$.

Despite their informational advantages, human investors face bounded strategic reasoning capacity. They may hold mistaken beliefs about the behavior of other market participants, leading to misinterpretations of how equilibrium prices are formed. Within the k-level thinking framework, human investors perform a finite number of iterative reasoning steps to infer information from asset prices. Each investor i has a perceived model of price formation, $\mathcal{P}_i(\cdot)$, which is constrained by their reasoning capacity. The relationship between their reasoning capacity and their understanding of the price function, as well as the distribution of reasoning capacities across human investors, will be introduced in the next section. Investors submit limit orders, so x_{it} denotes a demand schedule that depends on both the price level and the private signal.

[5] The transaction cost term is similar to that in Gârleanu and Pedersen (2013). This setup simplifies the comparison of profitability between human and AI investors. Assuming investors have CARA or mean-variance utility over end-of-period wealth with a given level of precision of humans' private signals would yield similar results and not change our analysis in the following sections.

[6] This simplifies our analysis by ruling out collusive equilibria sustained by the price-trigger strategies of human investors. Alternatively, this assumption can be interpreted as investors being single-period lived and maximizing one-period trading profits.

[7] These advantages may stem from their ability to perform theory-guided analysis and conduct primary research. In practice, they combine qualitative analysis of regulated disclosures with primary research such as channel checks and store visits, interviewing customers, suppliers, and independent experts, and fielding bespoke surveys to form forward-looking views on firm fundamentals.

AI investors. We focus on an important class of AI investors that execute algorithmic trading via reinforcement learning (RL). There is a large, representative AI investor in the market, with its mass normalized to 1. The AI investor seeks to maximize the total expected discounted trading profits:

$$\mathbb{E}\left[\sum_{t=0}^{\infty}\gamma^{t}\pi_{A,t}\right]. \tag{2}$$

The net trading profit of the AI investor, $\pi_{A,t}$, is determined by

$$\pi_{At} = (v_t - p_t) \cdot x_{At} - \frac{1}{2}\rho_A x_{A,t}^2,$$

where ρ_A controls the level of the AI investor's transaction cost, and $x_{At}\left(\eta_{At}, p_t\right)$ represents the shares of the risky asset that it chooses to purchase at the beginning of period t.

Our analysis centers on AI agents specifically designed to trade and adapt autonomously to the market environment. To isolate this mechanism, we abstract from predictive AI algorithms used for information processing and assume the AI investor does not observe private signals about asset payoffs.

The AI investor begins with no direct knowledge of the market environment. It does not know the distributions of asset payoffs, nor does it understand the behaviors of other market participants. Instead, the AI investor employs an RL algorithm to trade. Through each interaction with the environment (trading) and the rewards received (realized trading profits in each period), the algorithm gradually learns about the environment and optimizes its trading strategy.

Let $\mathcal{P}_A(\cdot)$ denote the AI investor's perceived model of price formation. This mapping may be implicitly embedded within the algorithm or explicitly parameterized in certain algorithmic designs. In Section 3, we show that a broad class of RL algorithms can efficiently learn the environment and that their strategies converge to the optimal rational expectations strategy (as in Kyle (1989)). We use standard Q-learning as an illustrative example and provide the corresponding convergence conditions.

Noise traders. A unit measure of noise traders trade for non-informational reasons, such as hedging needs, estimation errors, or sentiment. Their aggregate demand is z_t shares of the risky asset, where $z_t \sim \mathcal{N}(0, \sigma_z^2)$. The demand z_t is independent of other shocks and across time.

Equilibrium. In the benchmark model, an equilibrium is a sequence of investors' demands $\{x_{it}\}$, $\{x_{At}\}$, and a sequence of prices $\{p_t\}$ such that:

1. Investors use Bayes' law to update their beliefs. They combine their prior with their private signal (if any) and information inferred from the price. For human investor i, the information set consists of the private signal η_{it} and information from the price, $\mathcal{P}_i^{-1}(p_t)$. For the AI investor, the information is inferred from the price, denoted as $\mathcal{P}_A^{-1}(p_t)$.

2. At the beginning of each period t, human investors choose to buy x_{it} shares of the risky asset to maximize their expected trading profits (1).

3. At the beginning of each period t, the AI investor chooses to buy x_{At} shares to maximize its expected profits (2).

4. Market clearing: At the end of each period t, the investors' total demand for the risky asset plus the noise traders' demand z_t equals the asset supply:

$$\int_0^{\psi} x_{i,t} di + x_{A,t} + z_t = S \quad \forall t. \tag{3}$$

2.2 Reasoning Capacity and Strategies of Investors

Human investors have bounded reasoning capacity. This limits their ability to fully understand the behavior of all market participants, including other human investors and the AI investor. In the spirit of the Cognitive Hierarchy (CH) model of Camerer, Ho, and Chong (2004), we assume that some human investors hold incorrect and overconfident beliefs about the sophistication of others. Their strategies are determined through an iterated process of strategic thinking. More sophisticated investors perform additional steps of reasoning about others' actions. Following Grossman and Stiglitz (1980) and Kyle (1989), we consider a linear market equilibrium where the strategies of both human and AI investors are linear in their private signals and prices.

2.2.1 Step-0 Human Investors

Step-0 (the most naive) human investors are unaware of AI investors and ignore the information in prices.[8] They rely solely on their private signals to form beliefs about the asset payoff and make investment decisions. Solving the first-order condition yields:

$$\begin{aligned} x_{it}^{0}\left(\eta_{it}, p_t\right) &= \frac{\mathbb{E}\left[v_t \mid \eta_{it}\right] - p_t}{\rho \operatorname{Var}\left(v_t - p_t \mid \eta_{it}\right)}, \\ &\equiv \beta_{M\eta}^{0} \eta_{it} - \beta_{Mp}^{0} p_t + \mu_M^{0}. \end{aligned} \tag{4}$$

The second line follows from the conjecture that investors' demand is linear in their private signals and in the price level.

Incorporating the private signal into the posterior belief, substituting the posterior into the optimal demand (4), and collecting terms give the strategies of step-0 human investors.

Proposition 1. *The strategies for step-0 human investors are characterized by:*

$$\begin{aligned} \beta_{M\eta}^{0} &= \frac{\sigma_v^2}{\rho_M \left(\sigma_v^2 + \sigma_M^2\right)}, \\ \beta_{Mp}^{0} &= \frac{1}{\rho_M}, \\ \mu_M^{0} &= \frac{\sigma_m^2 \bar{v}}{\rho_M \left(\sigma_v^2 + \sigma_m^2\right)}. \end{aligned} \tag{5}$$

[8] This assumption follows the spirit of Zhou (2022). We extend it by assuming that step-0 human investors also ignore the presence of AI investors in financial markets.

2.2.2 Step-k Human Investors

Beliefs about other human investors. A step-k human investor is overconfident. She does not recognize that other human investors may be thinking at step k or higher. Instead, she believes the market contains only a mix of human investors with fewer than k steps of reasoning. The full subjective distribution of thinking steps, $f(k)$, is assumed to follow a Poisson distribution:

$$f(k) = \frac{e^{-\tau}\tau^k}{k!}, \quad \forall k \geq 0.$$

A step-k human investor conjectures that other human investors are distributed across steps 0 through $k-1$. Let $g_{k-1}(h)$ denote a step-k human investor's belief about the proportion of step-h human investors. This belief is given by the right truncated Poisson distribution:

$$g_{k-1}(h) = \frac{f(h)}{\sum_{l=0}^{k-1} f(l)}, \quad \forall h \leq k-1,$$

and $g_{k-1}(h) = 0$ for all $h \geq k$. We refer to this distribution, $\{g_{k-1}(h)\}_{h=0}^{k-1}$, as the $(k-1)$-truncated distribution.

Perceived $(k-1)$-step AI investor. A step-k human investor is aware of the AI investor and correctly perceives the capabilities of AI algorithms. However, her misunderstanding of the market environment leads her to misinterpret the AI's strategy. Specifically, she believes a representative AI investor exists whose algorithm is trained against the $(k-1)$-truncated distribution of human investors.[9] In other words, she perceives that the AI investor's strategy converges to the optimal strategy in a benchmark equilibrium where the distribution of

[9] We deliberately abstract from modeling human investors' beliefs about the training process of RL algorithms, i.e., the dynamics of the AI algorithm before it converges to a steady strategy. To form beliefs about the future evolution of the algorithm, other investors would need to know its technical details and its exact stage in the training process. These details include, for example, its training and data usage (on-policy vs. off-policy), learning strategy (model-based vs. model-free), exploration vs. exploitation strategy, policy type (deterministic vs. stochastic), and the specific hyperparameters used, etc. Even with similar current trading strategies, the type of algorithm used or its stage in the training process will greatly affect its future evolution. In reality, investors keep information about their trading strategies confidential. It is more natural for other investors to conjecture that the algorithms have finished training before being deployed in the market.

human investors' thinking steps follows the $(k-1)$-truncated distribution. We refer to this perceived AI investor as the $(k-1)$-step AI investor. Its strategy is derived below.

The AI investor's optimal demand is given by:

$$\begin{aligned} x_{At}^{k-1}(\eta_{At}, p_t) &= \frac{\mathbb{E}\left[v_t \mid (\mathcal{P}_A^{k-1})^{-1}(p_t)\right] - p_t}{\rho_A + \lambda_A^{k-1}}, \\ &\equiv -\beta_{Ap}^{k-1} p_t + \mu_A^{k-1}. \end{aligned} \tag{6}$$

The second-order condition is $\lambda_A^{k-1} > -\rho_A$. This condition means that the AI investor's price impact must be greater than the negative of its transaction cost. Its "perceived" market-clearing condition is:

$$\sum_{j=0}^{k-1} \int_0^{\psi g_{k-1}(j)} x_{it}^j di + x_{At}^{k-1} + z_t = S,$$

which is equivalent to

$$\sum_{j=0}^{k-1} \left(\beta_{M\eta}^j \int_0^{\psi g_{k-1}(j)} \eta_{it} di - \psi g_{k-1}(j) \beta_{Mp}^j p_t + \psi g_{k-1}(j) \mu_M^j \right) + x_{At}^{k-1} + z_t - S = 0.$$

Let $\bar{\beta}_{M\eta}^{k-1}$, $\bar{\beta}_{Mp}^{k-1}$, and $\bar{\mu}_M^{k-1}$ denote the weighted averages of the demand coefficients for all lower-step human investors:

$$\begin{cases} \bar{\beta}_{M\eta}^{k-1} = \sum_{j=0}^{k-1} g_{k-1}(j) \beta_{M\eta}^j, \\ \bar{\beta}_{Mp}^{k-1} = \sum_{j=0}^{k-1} g_{k-1}(j) \beta_{Mp}^j, \\ \bar{\mu}_M^{k-1} = \sum_{j=0}^{k-1} g_{k-1}(j) \mu_M^j. \end{cases} \tag{7}$$

Solving for the price function gives:

$$\mathcal{P}_A^{k-1} : p_t = \frac{\bar{\beta}_{M\eta}^{k-1}}{\bar{\beta}_{Mp}^{k-1}} \cdot v_t + \frac{1}{\psi \bar{\beta}_{Mp}^{k-1}} \cdot x_{At}^{k-1} + \frac{1}{\psi \bar{\beta}_{Mp}^{k-1}} \cdot z_t + \frac{1}{\psi \bar{\beta}_{Mp}^{k-1}} \cdot \widetilde{\mu}_M^{k-1},$$

The price function implies that the perceived price impact, λ_A^{k-1}, is:

$$\lambda_A^{k-1} = \frac{1}{\psi \bar{\beta}_{Mp}^{k-1}}.$$

We derive the posterior beliefs based on Bayes' law. Then, we substitute the posterior beliefs and the "perceived" price signal into the optimal demand function (6) and match the terms to solve for the coefficients of the demand function:

Proposition 2. *The strategy for level-(k-1) AI investor is characterized by:*

$$\begin{aligned}
\beta_{Ap}^{k-1} &= \left[\rho_A + \lambda_A^{k-1} + \frac{\psi(\hat{\sigma}_A^{k-1})^2 \bar{\beta}_{M\eta}^{k-1}}{\sigma_z^2}\right]^{-1} \cdot \left[1 - \frac{\psi^2(\hat{\sigma}_A^{k-1})^2 \bar{\beta}_{Mp}^{k-1} \bar{\beta}_{M\eta}^{k-1}}{\sigma_z^2}\right], \\
\mu_A^{k-1} &= \left[\rho_A + \lambda_A^{k-1} + \frac{\psi(\hat{\sigma}_A^{k-1})^2 \bar{\beta}_{M\eta}^{k-1}}{\sigma_z^2}\right]^{-1} \cdot \left[\frac{(\hat{\sigma}_A^{k-1})^2}{\sigma_v^2}\bar{v} - \frac{\psi(\hat{\sigma}_A^{k-1})^2 \bar{\beta}_{M\eta}^{k-1}}{\sigma_z^2} \cdot \widetilde{\mu}_M^{k-1}\right],
\end{aligned} \tag{8}$$

where $\left(\hat{\sigma}_A^{k-1}\right)^2 = \left(\frac{1}{\sigma_v^2} + \frac{1}{\left(\psi \bar{\beta}_{M\eta}^{k-1}\right)^{-2}\sigma_z^2}\right)^{-1}$*, and* $\widetilde{\mu}_M^{k-1} = \psi \bar{\mu}_M^{k-1} - S$*.*

The detailed proof is provided in Appendix A.2.

Solving the Strategies of Step-k Human Investors. The optimal demand for a step-k human investor is given by solving the first-order condition:

$$\begin{aligned}
x_{it}^k(\eta_{it}, p_t) &= \frac{\mathbb{E}\left[v_t \mid \eta_{it}, (\mathcal{P}_M^k)^{-1}(p_t)\right] - p_t}{\rho_M} \\
&\equiv \beta_{M\eta}^k \eta_{it} - \beta_{Mp}^k p_t + \mu_M^k
\end{aligned} \tag{9}$$

Based on her beliefs about the behavior of other human and AI investors, her perceived market-clearing condition is:

$$\sum_{j=0}^{k-1}\left(\beta_{M\eta}^j \int_0^{g_{k-1}(j)} \eta_{it} di - \psi g_{k-1}(j)\beta_{Mp}^j p_t + \psi g_{k-1}(j)\mu_M^j\right) + \left(\mu_A^{k-1} - \beta_{Ap}^{k-1} \cdot p_t\right) + z_t - S = 0$$

Solving for the price yields

$$\mathcal{P}_M^k : p_t = \left(\psi \bar{\beta}_{Mp}^{k-1} + \beta_{Ap}^{k-1}\right)^{-1} \left[\psi \bar{\beta}_{M\eta}^{k-1} v_t + z_t + \widetilde{\mu}^{k-1}\right],$$

where $\widetilde{\mu}^{k-1} \equiv \psi \bar{\mu}_M^{k-1} + \mu_A^{k-1} - S$.

Substituting the posterior beliefs and the expression for the price signal (A4) into the optimal demand function (9), and matching terms, we solve for the coefficients of the demand function:

Proposition 3. *The strategies for level-k human investors are*

$$\begin{aligned}
\beta_{M\eta}^k &= \rho_M^{-1} \frac{(\hat{\sigma}_M^k)^2}{\sigma_M^2}, \\
\beta_{Mp}^k &= \rho_M^{-1} - \rho_M^{-1} \frac{(\hat{\sigma}_M^k)^2}{(\sigma_{Mp}^k)^2} \cdot \frac{\left(\psi \bar{\beta}_{Mp}^{k-1} + \beta_{Ap}^{k-1}\right)}{\psi \bar{\beta}_{M\eta}^{k-1}}, \\
\mu_M^k &= \rho_M^{-1} \frac{(\hat{\sigma}_M^k)^2}{\sigma_v^2} \bar{v} - \rho_M^{-1} \frac{(\hat{\sigma}_M^k)^2}{(\sigma_{Mp}^k)^2} \cdot \frac{\widetilde{\mu}^{k-1}}{\psi \bar{\beta}_{M\eta}^{k-1}}.
\end{aligned} \tag{10}$$

where $(\hat{\sigma}_M^k)^2 = \left(\frac{1}{\sigma_v^2} + \frac{1}{\sigma_M^2} + \frac{1}{(\sigma_{Mp}^k)^2}\right)^{-1}$, *and* $(\sigma_{Mp}^k)^2 = \frac{\sigma_z^2}{\left(\psi \bar{\beta}_{M\eta}^{k-1}\right)^2}$.

The detailed proof is provided in Appendix A.3.

2.3 Properties of Step-k Strategies

The strategies of step-k human investors and the perceived strategies of the AI investor are determined by a system of nonlinear difference equations derived from Propositions 1 through 3. This system can be solved recursively. Due to the high degree of nonlinearity, closed-form solutions are intractable. However, given the model parameters, we can numerically compute the strategies for any finite k and analyze their limiting behavior as k approaches infinity.

Figure 1 illustrates the strategies of step-k human investors compared to those of the step-k AI investor. Step-0 human investors mistakenly believe that prices contain no information about the asset payoff. They rely solely on their private signals, so the coefficient $\beta_{M\eta}^0$ is high. They trade very conservatively in the sense that their price elasticity, β^0_{Mp}, is high. This means that if prices increase, they will greatly decrease their demand because they think price fluctuations come entirely from the demand of noise traders. The step-0 AI investor is trained in an environment against all step-0 human investors. The algorithm correctly understands the behavior of other market participants and accurately learns the price function. It exploits the biased beliefs of step-0 human investors by trading aggressively on the information contained in prices: β^0_{Ap} is low and below zero, meaning the AI investor is actually a momentum investor. It increases its asset positions when prices rise.

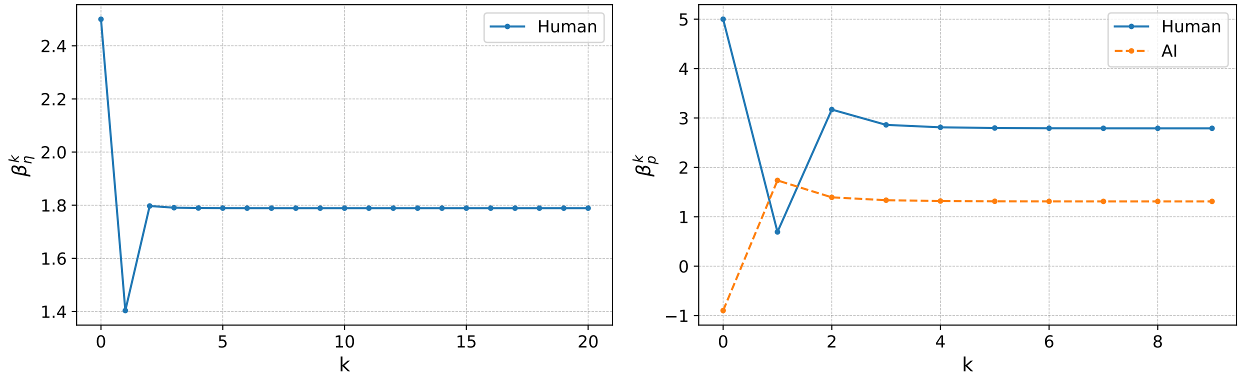
As k increases, the strategies of human investors become more sophisticated. A step-1 human investor believes she is the only step-1 investor in the environment, trading against all step-0 human investors and a step-0 AI investor. Similar to the step-0 AI investor, she profits by relying more on the information in prices and less on her private information, resulting in a low β^1_{Mp} and a low $\beta^1_{M\eta}$. The magnitude of this effect is less than that of the step-0 AI investor because she recognizes that the presence of the step-0 AI investor makes prices less informative about future payoffs than in an environment with only step-0 human investors. Due to the presence of step-1 human investors, prices become less informative and move less with the future asset payoffs, so the step-1 AI investor reacts by trading more conservatively: β^1_{Ap} is higher than β^0_{Ap}. Also, β^1_{Ap} becomes positive, meaning that the AI investor abandons the momentum strategy.

For even higher steps, the strategies of human investors converge. The proportion of step-k human investors in the market becomes smaller as k increases, so the impact of higher-step human investors on prices diminishes, and the change from step $k-1$ to step k shrinks. Since the changes in the training environment decrease, the AI investor's strategy also converges. We have the following proposition:

Proposition 4. *When $\bar{\beta}^k_{Mp} > 0, \forall k$, in the limit as $k \to \infty$, the strategies of step-k human investors and the step-k AI investor converge to limit strategies, respectively.*

The proof is in Appendix A.4. The condition for convergence is mild and holds for most

Figure 1: Strategies of Step-k Human Investors vs. Step-k AI Investor



(a) Demand Response to Private Information (b) Slope of the Demand Function w.r.t Price

Notes: This figure shows the strategies of step-k human investors vs. step-k AI investor. Blue solid lines plot the coefficients of step-k human investors' strategies, while orange dashed lines plot the coefficients of step-k AI investor's strategies. Panel (a) shows the demand response to private information, $\beta^k_{M\eta}$. Panel (b) shows the slope of the demand function with respect to price, β^k_{Mp} vs. β^k_{Ap}. The parameters are set as follows: $\rho_M = 0.2$, $\rho_A = 0.2$, $\sigma_v = 1$, $\sigma_M = 1$, $\sigma_z = 20$, $\psi = 10$, $S = 0$, and $\tau = 2$.

model parameterizations. It requires that the aggregate price elasticity of demand for all human investors is negative, meaning that, on average, human investors are not momentum traders. If human investors were, in aggregate, momentum traders, a positive feedback loop could emerge: higher demand from the AI would drive up prices, prompting human investors to buy even more. This could lead to explosive trading behavior and prevent the strategies from converging. A less restrictive condition for convergence is also discussed in Appendix A.4.

The limiting strategies are not necessarily equal to the strategies in the noisy rational expectations equilibrium because there are always lower-step human investors in the environment who hold incorrect beliefs about the distribution of other human investors' reasoning steps.

2.4 Market Outcomes

We analyze the market outcomes resulting from the competition between AI and human investors in different objective environments. We make the following definitions of the environments. A k_{Env} objective environment is populated by the k_{Env}-distribution of human investors and a k_{Env}-step AI investor. This k_{Env} objective environment is the same as the training environment of the k_{Env}-step AI investor. We define the objective environments in this way to ensure that the AI algorithm operates in the same environment in which it was trained.[10]

Equilibrium prices. The distribution of human investors' thinking steps is therefore $g_{k_{Env}}(h)$. The market-clearing condition in this objective environment is:

$$\sum_{j=0}^{k_{Env}} \int_0^{\psi g_{k_{Env}}(j)} x_{it}^j di + x_{A_t}^{k_{Env}} + z_t - S = 0.$$

Substituting the strategies of different steps of human investors, x_{it}^j, and the strategy of the AI investor, $x_{A_t}^{k_{Env}}$, into the market-clearing condition yields:

$$p_t = \theta v_t + \xi^{-1} z_t + \xi^{-1} \widetilde{\mu}^{k_{Env}},$$

where

$$\begin{cases} \theta^{k_{Env}} = \frac{\left(\psi \bar{\beta}_{M\eta} + \beta_{A\eta}^{k_{Env}}\right)}{\left(\psi \bar{\beta}_{MP} + \beta_{AP}^{k_{Env}}\right)} \\ \xi^{k_{Env}} = \psi \bar{\beta}_{MP} + \beta_{AP}^{k_{Env}} \end{cases}$$

and

$$\begin{cases} \bar{\beta}_{M\eta}^{k_{Env}} = \sum_{j=0}^{\infty} g_{k_{Env}}(j) \beta_{M\eta}^j \\ \bar{\beta}_{Mp}^{k_{Env}} = \sum_{j=0}^{\infty} g_{k_{Env}}(j) \beta_{Mp}^j \\ \bar{\mu}_M^{k_{Env}} = \sum_{j=0}^{\infty} g_{k_{Env}}(j) \mu_M^j \end{cases}$$

[10] This consistency abstracts from the performance loss due to a mismatch between the training and operating environments of AI algorithms. In practice, investors who deploy AI algorithms for trading will monitor the market and ensure their algorithms adapt to the market environments.

Price information sensitivity. The parameter $\theta^{k_{Env}}$ describes the price information sensitivity. It measures the information content of prices about future asset payoffs. A higher $\theta^{k_{Env}}$ indicates that prices are more sensitive to changes in the asset payoff, making prices more informative.

Market efficiency. Market efficiency is another measure that describes how informative the market prices are about future asset payoffs. It is defined as the precision of the posterior about the asset payoff conditional on the prices:

$$\begin{aligned}\text{Mkt Eff} &= \frac{1}{\text{Var}(v \mid p)} \\ &= \frac{1}{\sigma_v^2} + \frac{\left(\theta^{k_{Env}} \xi^{k_{Env}}\right)^2}{\sigma_z^2}\end{aligned}$$

Market liquidity. Following Goldstein and Yang (2017), we define market liquidity as the inverse of the price impact of noise trading. Market liquidity is measured by the parameter $\xi^{k_{Env}}$. A higher $\xi^{k_{Env}}$ indicates that prices are less sensitive to noise trader demand, making the market more liquid.

Price volatility. The unconditional volatility of prices, $\text{Var}(p_t)$, can be expressed as:

$$\text{Var}(p_t) = \left(\theta^{k_{Env}}\right)^2 \sigma_v^2 + \left(\xi^{k_{Env}}\right)^{-2} \sigma_z^2.$$

Price volatility depends on two components: the volatility due to changes in the asset payoff and the volatility due to noise trader demand. The first component is increasing in the price information sensitivity, $\theta^{k_{Env}}$, while the second component is decreasing in the market liquidity, $\xi^{k_{Env}}$. A higher $\theta^{k_{Env}}$ leads to more volatile prices because prices react more strongly to changes in the asset payoff. Conversely, a higher $\xi^{k_{Env}}$ leads to less volatile prices because noise trader demand has a smaller impact on prices.

2.5 Expected Profits

In order to have a better understanding of the competition between AI and human investors, we analyze the expected net profits of the AI investor and human investors in different objective environments. The expected profits are determined by the investors' strategies and the market conditions derived in the previous section. We have the following proposition for the expected profits of human investors in the k_{Env} objective environment:

Proposition 5. *The single-period expected profit for a step-k_M human investor in the objective environment k_{Env} is*

$$\begin{aligned}\mathbb{E}\left[\pi_{it}^{k_M}\right] =& \left[\beta_{M\eta} - \frac{1}{2}\rho_M\beta_{M\eta}^2 + \theta\left(\rho_M\beta_{M\eta}\beta_{Mp} - \beta_{Mp} - \beta_{M\eta}\right)\right] \cdot \left(\sigma_v^2 + \bar{v}^2\right) \\ &+ \left(\beta_{Mp} - \frac{1}{2}\rho_M\beta_{Mp}^2\right) \cdot \left[\theta^2\sigma_v^2 + \xi^{-2}\sigma_z^2 + \left(\theta\bar{v} + \xi^{-1}\widetilde{\mu}\right)^2\right] \\ &- \frac{1}{2}\rho_M\beta_{M\eta}^2\sigma_M^2 + \left(\rho_M\beta_{M\eta}\beta_{Mp} - \beta_{Mp} - \beta_{M\eta}\right)\bar{v}\xi^{-1}\widetilde{\mu} \\ &+ \mu_M\bar{v} + \left(\rho_M\beta_{Mp} - 1\right)\mu_M\left(\theta\bar{v} + \xi^{-1}\widetilde{\mu}\right) - \frac{1}{2}\rho_M\left(\mu_M^2 + 2\beta_{M\eta}\bar{v}\mu_M\right)\end{aligned}$$

where, for ease of notation, we drop the superscript k_M from the strategy coefficients, $\beta_{M\eta}^{k_M}$, $\beta_{Mp}^{k_M}$, and $\mu_M^{k_M}$, and the superscript k_{Env} from the market condition parameters, $\theta^{k_{Env}}$ and $\xi^{k_{Env}}$.

We also derive the expected profits of the AI investor in the objective environment:

Proposition 6. *For the step-k_A AI investors in the objective environment k_{Env}, the single-period expected profit equals*

$$\begin{aligned}\mathbb{E}\left[\pi_{At}^{k_A}\right] =& -\theta\beta_{Ap}\left(\sigma_v^2 + \bar{v}^2\right) + \left(\beta_{Ap} - \frac{1}{2}\rho_A\beta_{Ap}^2\right)\left[\theta^2\sigma_v^2 + \xi^{-2}\sigma_z^2 + \left(\theta\bar{v} + \xi^{-1}\widetilde{\mu}\right)^2\right] \\ &- \beta_{Ap}\bar{v}\xi^{-1}\widetilde{\mu} + \mu_A\bar{v} + \left(\rho_A\beta_{Ap} - 1\right)\mu_A\left(\theta\bar{v} + \xi^{-1}\widetilde{\mu}\right) - \frac{1}{2}\rho_A\mu_A^2\end{aligned}$$

where, for ease of notation, we drop the superscript k_A from the strategy coefficients, $\beta_{Ap}^{k_A}$ and $\mu_A^{k_A}$, and the superscript k_{Env} from the market condition parameters, $\theta^{k_{Env}}$ and $\xi^{k_{Env}}$.

The detailed proofs for these two propositions are provided in Appendix A.5.

3 Achieving Rational Expectations through Reinforcement Learning

In the benchmark model above, human investors believe that after extensive interaction with the environment, the AI algorithm can accurately understand the environment it faces, infer the correct price formation, and converge to the optimal strategy as if it has rational expectations. In this section, we provide both theoretical and simulation-based evidence to show the ability of AI algorithms to learn the environment and converge to "rational expectations" optimal strategies.

3.1 Details of AI Algorithms

Reinforcement learning (RL) is a machine learning paradigm for learning optimal sequential decisions. An agent learns by interacting directly with an environment, typically without a pre-specified model of its dynamics. The process is iterative. The agent takes an action, observes the resulting state and reward, and updates its strategy. The goal is to find a policy—a mapping from states to actions—that maximizes the expected cumulative discounted reward. Learning requires balancing the fundamental trade-off between exploration (gathering new information) and exploitation (using known information). RL encompasses a broad class of algorithms, including Policy Gradients, Actor-Critic methods, and Value Function methods like Q-learning (Sutton, Barto et al., 1998). Many of these algorithms are proven to converge to an optimal policy under standard assumptions (Bertsekas, 2019). We use Q-learning, a foundational RL algorithm, to show that it can converge to the rational expectations optimal strategy within our benchmark model.

Q-Learning algorithm. Following the definition in Watkins (1989), Q-learning is detailed as follows. The optimization problem the AI investor faces satisfies the definition of a Markov decision process (MDP). Let $\mathcal{A}$ and $\mathcal{S}$ be the action and state spaces, respectively. The problem is to choose a sequence of actions $a_t \in \mathcal{A}$ to maximize the expected total discounted rewards. The value function of a state s is defined as:

$$V(s) = \max_{\{a_t \in \mathcal{A}\}} \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \mid s_0 = s\right],$$

where $r(s_t, a_t)$ is the reward for period t and $\{s_t\}$ is the state trajectory evolving according to the transition probabilities. This expression can be rewritten as:

$$V(s) = \max_{a \in \mathcal{A}} \left\{r(s,a) + \gamma \mathbb{E}\left[V(s') \mid s, a\right]\right\}.$$

To find an algorithm that converges to the optimal policy, we define the Q-value as:

$$Q(s,a) = r(s,a) + \gamma \mathbb{E}\left[V(s') \mid s, a\right].$$

$Q(s,a)$ is the value of taking action a in state s and following the optimal policy thereafter. Let $\hat{Q}_t$ denote the estimated Q-function at time t. The algorithm recursively updates its Q-value estimates as follows:

$$\hat{Q}_{t+1}\left(s_t, a_t\right) = (1-\alpha)\hat{Q}_t\left(s_t, a_t\right) + \alpha \left[r(s_t, a_t) + \gamma \max_{a' \in \mathcal{A}} \hat{Q}_t\left(s_{t+1}, a'\right)\right],$$

where $\alpha \in [0,1]$ controls the learning rate of the Q-function. A larger α means that the algorithm puts more weight on the new observation and updates the Q-function more quickly. This temporal-difference update method bootstraps from current estimates and does not require a model of the environment's dynamics.

Necessary convergence conditions. The Q-learning algorithm converges to the optimal action-value function Q^* with probability 1 provided that (Bertsekas, 2019):

1. (Step Size) The learning rates satisfy $0 < \alpha_t \leq 1$, $\sum_{t=0}^{\infty} \alpha_t = \infty$, and $\sum_{t=0}^{\infty} \alpha_t^2 < \infty$.

2. (Sufficient Exploration) Every state-action pair is visited infinitely often. Let $N_T(s,a) := \sum_{t=0}^{T-1} \mathbf{1}\{S_t = s, A_t = a\}$ be the number of times the state-action pair (s,a) has been visited up to time T. Then $\forall (s,a) \in \mathcal{S} \times \mathcal{A}: \quad \lim_{T\to\infty} N_T(s,a) = \infty \quad$ almost surely .

Exploration method. We use a simple exploration method, the ε-greedy policy, which satisfies the above convergence conditions. Let $\pi_{\varepsilon_t}(a \mid s)$ be the policy, which gives the probability of choosing action a in state s at time t. Let $m = |\mathcal{A}|$ be the size of the action space. The ε-greedy policy is defined as:

$$\pi_{\varepsilon_t}(a \mid s) = \begin{cases} 1 - \varepsilon_t + \frac{\varepsilon_t}{m}, & \text{if } a \in \arg\max_{a \in \mathcal{A}} Q(s,a) \\ \frac{\varepsilon_t}{m}, & \text{if } a \notin \arg\max_{a \in \mathcal{A}} Q(s,a) \end{cases}$$

This method implies that, with probability $1 - \varepsilon_t$, the agent exploits by choosing among the greedy actions, and with probability ε_t, it explores by choosing uniformly at random over all actions. In practice, we let ε_t decay toward zero over time ($\varepsilon_t \downarrow 0$). A decaying ε_t is not required for the Q-values to converge to Q^*. However, it is necessary to ensure the policy itself converges to the optimal policy.

Specific features. Several additional adjustments are made to apply the Q-learning algorithm to the investment problem faced by the AI investor. First, the AI investor faces a simplified environment in our benchmark model. The Markov decision process is reduced to a multi-armed bandit problem because the evolution of the state does not depend on the investor's decisions, and there is no observable state (e.g., private signals) on which the investor can base its actions.[11] This simplification broadens the set of algorithms applicable to solving the problem. Second, the AI investor is assumed to submit a linear demand function, $x_{At} = \mu_A - \beta_{Ap} p_t$. The parameters for this function are chosen by maximizing the current Q-value estimate:

$$(\mu_A, \beta_{Ap}) = \arg \max_{\{\mu'_A, \beta'_{Ap}\}} \hat{Q}_t(\{\mu'_A, \beta'_{Ap}\})$$

Thus, the action space is a continuous, two-dimensional set, $\mathcal{A} = \mathbb{R} \times \mathbb{R}$, where each action $a = (\mu_A, \beta_{Ap})$ is the pair of parameters for the demand function.

3.2 Simulation Evidence

We set up the environment following the benchmark model and let the AI investor employ the Q-learning algorithm to make investment decisions. We then simulate the algorithm and evaluate the ability of this simple approach to learn the true Q-value function, Q^*, and recover the optimal strategy predicted by our benchmark equilibrium.

Hyperparameters. First, we choose the learning rate:

$$\alpha_t = \frac{c_1}{1 + c_2 t},$$

where $c_1 = 1$ and $c_2 = 0.001$. This learning rate decays at a moderate speed with respect to time t. It satisfies the convergence conditions: it is relatively large when t is small, allowing the Q-value function to be significantly updated during early-stage exploration, and it shrinks to zero as t increases, ensuring the convergence of the Q-value estimates.

[11] The general MDP framework and our discussion of RL algorithms preserve the flexibility to incorporate richer market settings. For example, in the benchmark model, we abstract from predictive AI algorithms designed primarily for information processing and assume that the AI investor does not observe any private signals about asset payoffs. If we instead allowed the AI investor to observe a private signal, this signal could be treated as one of the state variables on which the investor's actions depend. This fits naturally within the MDP structure. Another example is when the asset is long-lived and the market has overlapping generations of human investors. The AI investor's current actions affect the end-of-period price, which then influences the next period's decisions. The state evolution now depends on the agent's actions, which still satisfies the definition of an MDP.

Second, we choose the exploration rate as

$$\varepsilon_t = \frac{c_0}{c_0 + t},$$

where $c_0 = 1$. This ε-greedy policy satisfies the sufficient exploration condition. It guarantees that the agent continues to explore indefinitely while becoming asymptotically greedy.

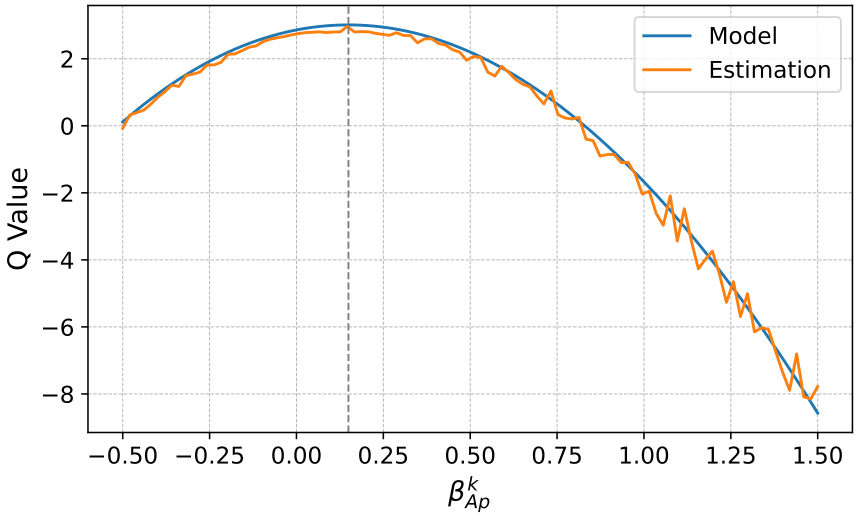
Third, we discretize the action space. Discrete grids are constructed based on the optimal strategy in the benchmark equilibrium. For the given objective environment, let μ_A^* and β_{Ap}^* denote the values of μ_A and β_{Ap} in the AI investor's optimal demand schedule. The action space is then specified by discretizing the intervals $[\mu_A^* - \delta_\mu \mu_A^*,\ \mu_A^* + \delta_\mu \mu_A^*]$ and $[\beta_{Ap}^* - \delta_\beta \beta_{Ap}^*,\ \beta_{Ap}^* + \delta_\beta \beta_{Ap}^*]$ into n_μ and n_β equally spaced grids, respectively. The parameters δ_μ and δ_β are chosen to cover a sufficiently wide range of deviations from the optimal strategy. The grid sizes n_μ and n_β are chosen to be large enough to provide a finer approximation of the Q-value function and produce strategies closer to the optimal strategy while balancing computational efficiency.

Results. In simulation, the simple Q-learning algorithm can effectively learn the optimal strategy. The algorithm converges to $\hat{\beta}_{Ap}^k = 0.1465$, which is very close to the benchmark model's optimal value of $\beta_{Ap}^k = 0.1496$. Not only does the Q-learning algorithm converge to the optimal strategy, but it also estimates the entire Q-value function curve reasonably well. Figure 2 compares the algorithm's estimated Q-value function with the benchmark model's Q-value function for the AI investor in equilibrium, assuming the AI investor has rational expectations. The estimated Q-value function is very close to the rational-expectations Q-value function predicted by the model.

3.3 Broad Generality Across RL Algorithms

The theoretical and simulation evidence of convergence does not mean that our analysis is limited to the Q-learning algorithm, but rather demonstrates that our theoretical framework

Figure 2: Estimated Q-Value Function by Algorithm



Notes: The blue solid line shows the Q-value function with respect to β^k_{Ap} for the AI investor in equilibrium when the AI investor has rational expectations. The orange solid line plots the algorithm's estimate of the Q-value function. The dashed vertical line shows the optimal value of β^k_{Ap} in the benchmark model.

can be applied to a wide range of RL algorithms. In the benchmark model environment, the action space is continuous. The above result shows that using a simple tabular Q-learning algorithm with a plain discretization of the action space can effectively learn the correct Q-value curve and converge to the optimal strategy predicted by the benchmark equilibrium. The setting of the problem fits the definition of problems for which many algorithms in the RL literature are designed. These algorithms have either been theoretically proven or empirically validated to converge effectively and find the (near) optimal strategy for the problem.[12] Therefore, we can assume that any of these algorithms could be employed by the AI investor in our model. Our subsequent theoretical results apply to all such algorithms.

[12] More advanced algorithms have been developed in recent years, e.g., algorithms with function approximation using deep neural networks. See Van Hasselt (2012) for a survey. Recent developments include Carden (2014); Lillicrap et al. (2016); Majzoubi et al. (2020); Kara, Saldi, and Yüksel (2023); Alaoui and Saoud (2024), etc.

4 Competition Between AI and Human Investors

This section details our main findings on the competitive dynamics between AI-powered and human investors. We first establish our central result: human traders can consistently outperform AI agents in different market environments. We then explore the conditions that enable this human advantage and analyze the underlying economic mechanism of human outperformance. First, we show that AI's performance is highly dependent on the precision of human investors' private signals. Then, we show that sophisticated human investors reduce the informativeness and volatility of market prices, which in turn limits the AI's ability to outperform humans. Finally, we demonstrate that the size of the AI sector itself constrains its performance due to the price impact of its trades.

4.1 Human Outperformance in Different Environments

Our primary analysis compares the profitability of human and AI investors. We find that human investors can consistently outperform AI-powered investors across various market environments. The average expected profit for human investors, including those employing naive strategies, is frequently higher than that of the sophisticated AI investor.

Figure 3 presents the expected profit per investor for human investors of varying sophistication levels competing against a step-k_{Env} AI investor in different k_{Env}-objective environments. The top two panels show the results for $k_{Env} = \infty$ environments, which are populated by a mix of human investors of all sophistication levels (step-0, step-1, step-2, etc.) and a step-∞ AI investor. Both panels show that human investors at all sophistication levels outperform the AI, even the most naive ones (step-0 and step-1). The parameter τ controls the distribution of human investors' reasoning steps. In the top-left panel, the reasoning steps of human investors follow a Poisson distribution with $\tau = 0.7$, meaning the majority of human investors have a low level of reasoning. The top-right panel shows the results for $\tau = 2.0$, where reasoning steps are more evenly distributed, with a larger proportion of higher-step human investors. Human outperformance is consistent across these different distributions. The graphs also show that more sophisticated human investors (those with more steps of thinking) earn greater expected profits.

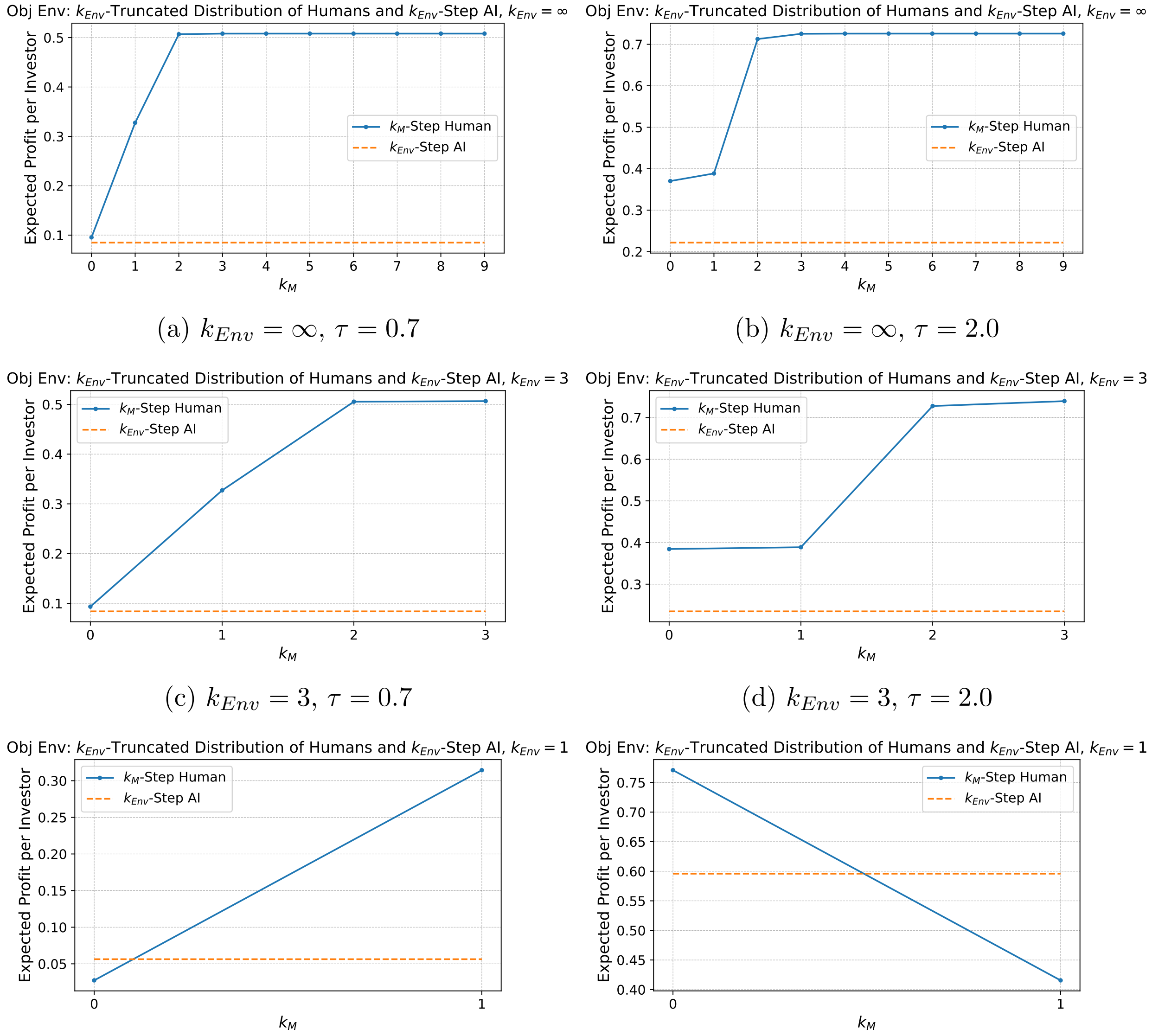
The middle panels show the results for $k_{Env} = 3$ environments. The results are similar to those in the top panels, with human investors consistently outperforming the AI investor. The bottom two panels show the results for $k_{Env} = 1$ environments, which contain a mix of step-0 and step-1 human investors and a step-1 AI investor. In these environments, the AI's performance relative to human investors depends on the distribution of human sophistication. When τ is low, the proportion of step-0 human investors is relatively large. Since all step-0 investors use the same strategy, their expected profits are crowded out. The step-1 AI investor's strategy is optimized against a mixed population of step-0 and step-1 human investors. When step-0 investors constitute a larger proportion of the environment, the AI's strategy becomes specialized to exploit their behavior. As a result, the step-1 AI investor outperforms step-0 humans but does not outperform step-1 human investors. Conversely, for a large τ, step-1 human investors become the dominant group. The AI's strategy consequently adapts to target these more sophisticated investors, enabling it to achieve superior performance against them. Under these conditions, however, the AI fails to outperform the less prevalent step-0 humans.

4.2 AI's Performance Depends on Human Signal Precision

The competitive balance between AI and human traders is critically dependent on the quality of the fundamental signals available to humans. The previous sections show that AI can outperform naive human investors in environments without sophisticated human investors. However, even in these environments, the AI's outperformance is highly influenced by the precision of human investors' signals.

The AI's ability to profit is non-monotonic with respect to human signal precision and is constrained at both extremes of human signal quality. Figure 4 shows the difference

Figure 3: Expected Profits of Human vs. AI Investors in Various Environments



(a) $k_{Env} = \infty$, $\tau = 0.7$

(b) $k_{Env} = \infty$, $\tau = 2.0$

(c) $k_{Env} = 3$, $\tau = 0.7$

(d) $k_{Env} = 3$, $\tau = 2.0$

(e) $k_{Env} = 1$, $\tau = 0.7$

(f) $k_{Env} = 1$, $\tau = 2.0$

Notes: This figure shows the expected profit per investor for human investors of varying sophistication levels against the step-k_{Env} AI investor in different k_{Env}-objective environments. The blue solid lines plot the expected profits of human investors. The orange dashed lines show the level of expected profits of the step-k_{Env} AI investor. The parameters are set as follows: $\rho_M = 0.2$, $\rho_A = 0.2$, $\sigma_v = 1$, $\sigma_M = 1$, $\sigma_z = 20$, $\psi = 10$, and $S = 0$.

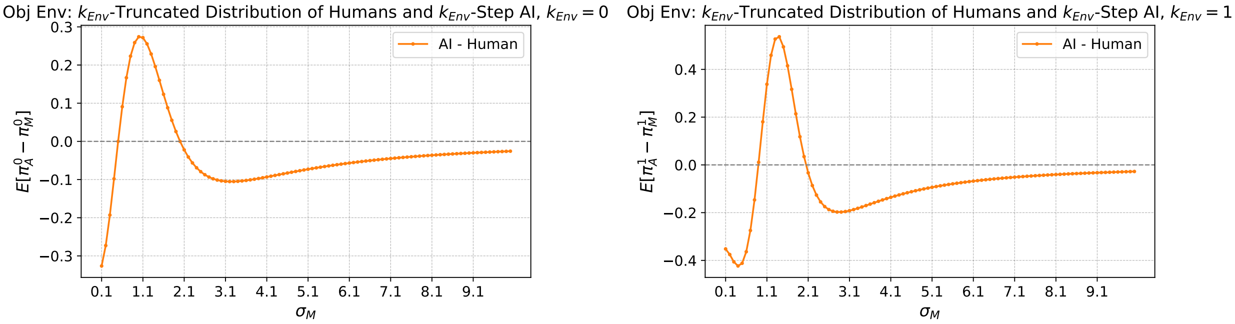
in expected profits between the AI investor and the highest-level human investors in k_{Env} objective environments against the standard deviation of human investors' private signals (σ_M). The results are presented for two objective environments: $k_{Env} = 0$ and $k_{Env} = 1$. When human investors have very precise signals (σ_M is low), they can outperform the AI investor due to this informational advantage. Even though the AI investor can correctly understand the information contained in equilibrium prices while naive humans misperceive it due to their limited reasoning capacity, the signals from prices are still noisier than the private signals available to human investors. As the precision of human signals drops (σ_M increases), this informational advantage decreases, and the AI begins to outperform humans. However, as signal precision drops further, the AI's competence against human investors deteriorates again. Market prices aggregate the private information from different human investors. Excessive noise in human fundamental signals makes prices less informative, limiting the AI's ability to profit from learning the price function. Although the AI investor can correctly learn the price function, the noise in prices makes it harder for the AI to extract useful information about future payoffs. Finally, as human signals become infinitely noisy, neither humans nor the AI can obtain useful information from either channel, and the difference in their expected profits approaches zero.

In summary, the key factor that limits the AI investor's profitability is the ability of human investors to shape the market environment through their trading. The AI investor's comparative advantage lies in its ability to correctly learn the price function and exploit the information contained in prices. However, prices are only informative because they aggregate private signals from human investors. If this source of information is cut off, there is not much information left in prices for the AI to learn from.

4.3 Sophisticated Humans Limit AI's Profitability

Another mechanism that constrains the AI investor's profitability is the presence of sophisticated human investors in the environment. Figure 5 shows how the expected profits of

Figure 4: Difference in Expected Profits against Human Signal Precision



(a) $k_{Env} = 0$

(b) $k_{Env} = 1$

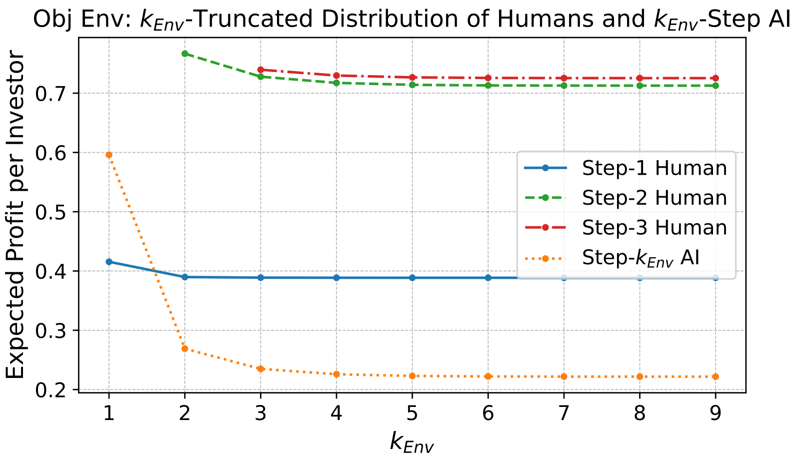
Notes: This figure plots the difference in expected profits between step-k_{Env} AI investor and step-k_{Env} human investors against the standard deviation of human investors' private signals (σ_M). A positive value indicates that the AI investor outperforms human investors. Panel (a) shows the results in a $k_{Env} = 0$ environment, where the AI competes against only step-0 human investors. Panel (b) shows the results in a $k_{Env} = 1$ environment. The parameters are set as follows: $\rho_M = 0.2$, $\rho_A = 0.2$, $\sigma_v = 1$, $\sigma_M = 1$, $\sigma_z = 20$, $\psi = 10$, $S = 0$, and $\tau = 2$.

human and AI investors change as more human investors become sophisticated. A higher k_{Env} environment contains a larger variety of human investors, with a larger proportion of more sophisticated investors. The figure shows that the presence of more sophisticated human investors reduces the expected profits of both humans and AI investors, but more significantly for the AI investor. The dashed orange line represents the AI investor's expected profits. It decreases more sharply than the profits of human investors as k_{Env} increases. For example, the decline is steeper for the AI than for step-1 to step-3 human investors (shown in blue, green, and red, respectively).[13]

The reason is that trading by sophisticated human investors stabilizes prices and limits the AI's ability to extract profits. As shown in Figure 1, the strategies of higher-step human investors converge to moderate strategies. Their demand sensitivity to private information, $\beta_{M\eta}$, and the price elasticity of demand, β_{Mp}, are not too high or too low. As the proportion of

[13] We present the results for the k_{Env}-step AI in each environment, which means that as more sophisticated human investors enter the market, the AI is retrained to adapt to the new environment. This ensures the consistency of the AI's operating and training environment. If we do not allow for retraining, the AI investor will earn even lower profits.

Figure 5: Expected Profit of Step-k_{Env} AI vs. Humans Across Different k_{Env}-Objective Environments



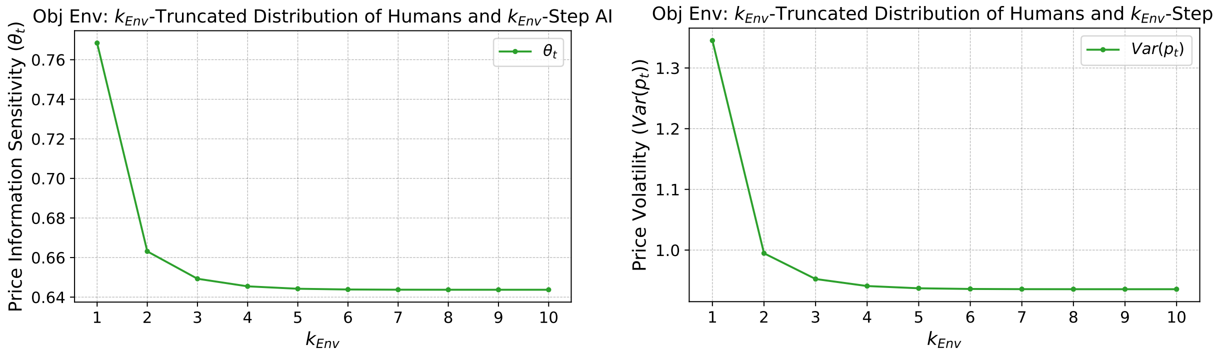
Notes: This figure plots the expected profit of step-k_{Env} AI investors against different steps of human investors across different k_{Env}-objective environments. The parameters are set as follows: $\rho_M = 0.2$, $\rho_A = 0.2$, $\sigma_v = 1$, $\sigma_M = 1$, $\sigma_z = 20$, $\psi = 10$, $S = 0$, and $\tau = 2$.

human investors who adopt moderate strategies increases, the information content of prices, $\theta^{k_{Env}}$, and price volatility, $\mathrm{Var}(p_t)$, decrease. Figure 6 shows that both the information content of prices and price volatility decrease in a higher-k_{Env} environment. This limits the ability of the AI investor to profit from learning the price function and exploiting price fluctuations. The analysis above shows that diversity in human investors' strategies stabilizes prices and limits the AI investor's exploitation of naive human investors.

4.4 Growth of AI Sector Size Limits Its Profitability

Another key force constraining the AI's profitability is the relative size of the AI sector itself. Figure 7 shows the difference in expected profits between the AI investor and the highest-level human investors in k_{Env} objective environments against the proportion of human investors. In all environments, increasing the proportion of AI investors (achieved by decreasing the proportion of human investors, ψ) monotonically decreases the AI's profitability compared

Figure 6: Price Information Sensitivity and Price Volatility Across Different k_{Env}-Objective Environments



(a) Price Information Sensitivity, $\theta^{k_{Env}}$

(b) Price Volatility, $\mathrm{Var}\,(p_t)$

Notes: This figure plots the price information sensitivity and price volatility across different k_{Env}-objective environments. The parameters are set as follows: $\rho_M = 0.2$, $\rho_A = 0.2$, $\sigma_v = 1$, $\sigma_M = 1$, $\sigma_z = 20$, $\psi = 10$, $S = 0$, and $\tau = 2$.

to human investors.[14] A large AI investor sector has a large price impact, which limits its profitability against human investors.

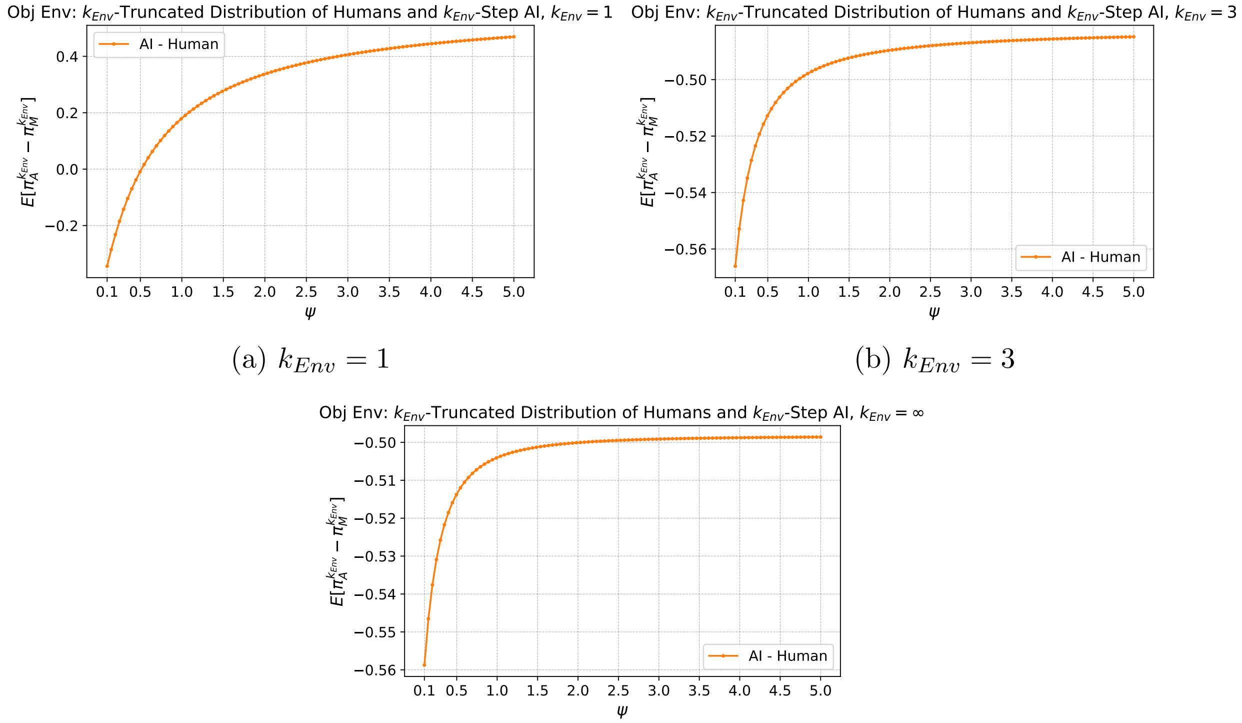
5 Conclusion

We develop a theoretical framework to study the competition between AI-powered and human investors in financial markets. We challenge the prevailing narrative that AI's superior processing capabilities will lead to its dominance in financial markets. Our model features human investors who possess informational advantages but are limited by bounded rationality, competing against an AI agent that learns its strategy through reinforcement learning. Our central finding is that human investors can consistently outperform AI agents across various market environments.

Our results highlight three key mechanisms that constrain the AI's performance. First, the AI's advantage depends critically on the quality of human private information. Its

[14] We adjust only the relative size of human and AI investors while holding constant the total measure of investors in the market relative to the size of noise trading and asset supply, and keeping the measure of the AI investor normalized to one.

Figure 7: Difference in Expected Profits against Human Proportion ψ



(a) $k_{Env} = 1$

(b) $k_{Env} = 3$

(c) $k_{Env} = \infty$

Notes: This figure plots the difference in expected profits between step-k_{Env} AI investor and step-k_{Env} human investors against the proportion of human investors (ψ). A positive value indicates that the AI investor outperforms human investors. The parameters are set as follows: $\rho_M = 0.2$, $\rho_A = 0.2$, $\sigma_v = 1$, $\sigma_M = 1$, $\sigma_z = 20$, $\psi = 10$, $S = 0$, and $\tau = 2$.

competitive edge gained from correctly learning the price function diminishes if those prices are not sufficiently informative due to very noisy human signals. Second, the presence of sophisticated human investors stabilizes prices, reducing the AI's profit opportunities from market fluctuations. Third, we show that as the AI sector grows, its own price impact limits its ability to outperform human investors.

These findings have important implications. They suggest that human-centric advantages, such as primary research and theory-guided fundamental analysis that generates unique private signals, will remain valuable in the age of AI. For regulators, the focus should perhaps be less on the absolute superiority of AI and more on the complex market ecology created by human-AI interaction. Future research could extend this framework by incorporating multiple, competing AI agents, allowing humans to learn about the AI's behavior, or exploring the dynamic co-evolution of human and algorithmic strategies.

References

Abada, I., and X. Lambin. 2023. Artificial intelligence: Can seemingly collusive outcomes be avoided? *Management Science* 69:5042–65.

Alaoui, S. B., and A. Saoud. 2024. How to discretize continuous state-action spaces in q-learning: A symbolic control approach. In *2024 IEEE 63rd Conference on Decision and Control (CDC)*, 8314–9. IEEE.

Allen, F., S. Morris, and H. S. Shin. 2006. Beauty contests and iterated expectations in asset markets. *The Review of Financial Studies* 19:719–52.

Angeletos, G.-M., and C. Lian. 2023. Dampening general equilibrium: incomplete information and bounded rationality. In *Handbook of Economic Expectations*, 613–45. Elsevier.

Banchio, M., and G. Mantegazza. 2024. Artificial intelligence and spontaneous collusion. *arXiv preprint arXiv:2202.05946* .

Banerjee, S., and M. Szydlowski. 2025. Trading against algorithms: Price dynamics and risk-sharing in a market with q-learners. *Available at SSRN 5380152* .

Bertsekas, D. 2019. *Reinforcement learning and optimal control*, vol. 1. Athena Scientific.

Calvano, E., G. Calzolari, V. Denicolo, and S. Pastorello. 2020. Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review* 110:3267–97.

Calvano, E., G. Calzolari, V. Denicoló, and S. Pastorello. 2021. Algorithmic collusion with imperfect monitoring. *International journal of industrial organization* 79:102712–.

Camerer, C. F., T.-H. Ho, and J.-K. Chong. 2004. A cognitive hierarchy model of games. *The Quarterly Journal of Economics* 119:861–98.

Carden, S. W. 2014. Convergence of a q-learning variant for continuous states and actions. *Journal of Artificial Intelligence Research* 49:705–31.

Cartea, Á., P. Chang, M. Mroczka, and R. Oomen. 2022. Ai-driven liquidity provision in otc financial markets. *Quantitative Finance* 22:2171–204.

Cartea, Á., P. Chang, and J. Penalva. 2022. Algorithmic collusion in electronic markets: The impact of tick size. *Available at SSRN 4105954* .

Colliard, J.-E., T. Foucault, and S. Lovo. 2022. Algorithmic pricing and liquidity in securities markets. *HEC Paris Research Paper No. FIN-2022-1459* .

Costa-Gomes, M., V. P. Crawford, and B. Broseta. 2001. Cognition and behavior in normal-form games: An experimental study. *Econometrica* 69:1193–235.

Costa-Gomes, M. A., and V. P. Crawford. 2006. Cognition and behavior in two-person guessing games: An experimental study. *American economic review* 96:1737–68.

Crawford, V. P., M. A. Costa-Gomes, and N. Iriberri. 2013. Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications. *Journal of Economic Literature* 51:5–62.

Dou, W. W., I. Goldstein, and Y. Ji. 2025. Ai-powered trading, algorithmic collusion, and price efficiency. *Jacobs Levy Equity Management Center for Quantitative Financial Research Paper, The Wharton School Research Paper* .

Eyster, E., M. Rabin, and D. Vayanos. 2019. Financial markets where traders neglect the informational content of prices. *The Journal of Finance* 74:371–99.

Farhi, E., and I. Werning. 2019. Monetary policy, bounded rationality, and incomplete markets. *American Economic Review* 109:3887–928.

García-Schmidt, M., and M. Woodford. 2019. Are low interest rates deflationary? a paradox of perfect-foresight analysis. *American Economic Review* 109:86–120.

Gârleanu, N., and L. H. Pedersen. 2013. Dynamic trading with predictable returns and transaction costs. *The Journal of Finance* 68:2309–40.

Goldstein, I., and L. Yang. 2017. Information disclosure in financial markets. *Annual Review of Financial Economics* 9:101–25.

Grossman, S. J., and J. E. Stiglitz. 1980. On the impossibility of informationally efficient markets. *American Economic Review* 70:393–408.

Han, J., and A. S. Kyle. 2018. Speculative equilibrium with differences in higher-order beliefs. *Management Science* 64:4317–32.

Hansen, K. T., K. Misra, and M. M. Pai. 2021. Frontiers: Algorithmic collusion: Supra-competitive prices via independent algorithms. *Marketing Science* 40:1–12.

Johnson, J. P., A. Rhodes, and M. Wildenbeest. 2023. Platform design when sellers use pricing algorithms. *Econometrica* 91:1841–79.

Kara, A., N. Saldi, and S. Yüksel. 2023. Q-learning for mdps with general spaces: Convergence and near optimality via quantization under weak continuity. *Journal of Machine Learning Research* 24:1–34.

Kyle, A. S. 1989. Informed speculation with imperfect competition. *The Review of Economic Studies* 56:317–55.

Lillicrap, T. P., J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra. 2016. Continuous control with deep reinforcement learning. *International Conference on Learning Representations* .

Majzoubi, M., C. Zhang, R. Chari, A. Krishnamurthy, J. Langford, and A. Slivkins. 2020. Efficient contextual bandits with continuous actions. *Advances in Neural Information Processing Systems* 33:349–60.

Marimon, R., E. McGrattan, and T. J. Sargent. 1990. Money as a medium of exchange in an economy with artificially intelligent agents. *Journal of Economic dynamics and control* 14:329–73.

Nagel, R. 1995. Unraveling in guessing games: An experimental study. *The American economic review* 85:1313–26.

Routledge, B. R. 1999. Adaptive learning in financial markets. *The Review of Financial Studies* 12:1165–202.

———. 2001. Genetic algorithm learning to choose and use information. *Macroeconomic dynamics* 5:303–25.

Stahl, D. O., and P. W. Wilson. 1994. Experimental evidence on players' models of other players. *Journal of Economic Behavior & Organization* 25:309–27.

———. 1995. On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior* 10:218–54.

Sutton, R. S., A. G. Barto, et al. 1998. *Reinforcement learning: An introduction*, vol. 1. MIT press Cambridge.

Van Hasselt, H. 2012. Reinforcement learning in continuous state and action spaces. In *Reinforcement Learning: State-of-the-Art*, 207–51. Springer.

Waltman, L., and U. Kaymak. 2008. Q-learning agents in a cournot oligopoly model. *Journal of Economic Dynamics and Control* 32:3275–93.

Watkins, C. J. C. H. 1989. Learning from delayed rewards. *PhD thesis, Cambridge University* .

Zhou, H. 2022. Informed speculation with k-level reasoning. *Journal of Economic Theory* 200:105384–.

Appendix

A Derivations and Proofs of Propositions

This section includes the detailed derivations and proofs of the propositions for the benchmark model.

A.1 Proof of Proposition 1

After incorporating the private signal, their posterior beliefs can be expressed as:

$$\begin{aligned} \mathbb{E}\left[v_t \mid \eta_{it}\right] &= \frac{\sigma_v^2 \eta_{it}}{\sigma_v^2+\sigma_M^2}+\frac{\sigma_M^2 \cdot \bar{v}}{\sigma_v^2+\sigma_M^2}, \\ \operatorname{Var}\left[v_t \mid \eta_{it}\right] &= \frac{\sigma_v^2 \sigma_M^2}{\sigma_v^2+\sigma_M^2}. \end{aligned} \tag{A1}$$

Substituting posterior (A1) into step-0 humans' optimal demand (4) yields:

$$\begin{aligned} x_{it}^0(\eta_{it}, p_t) &= \frac{1}{\rho_M}\left(\frac{\sigma_v^2 \cdot \eta_{it}}{\sigma_v^2+\sigma_M^2}+\frac{\sigma_M^2 \cdot \bar{v}}{\sigma_v^2+\sigma_M^2}-p_t\right), \\ &= \underbrace{\frac{\sigma_v^2}{\rho_M\left(\sigma_v^2+\sigma_M^2\right)}}_{\beta_{M\eta}^0} \cdot \eta_{it}-\underbrace{\frac{1}{\rho_M}}_{\beta_{Mp}^0} p_t+\underbrace{\frac{\sigma_M^2 \bar{v}}{\rho_M\left(\sigma_v^2+\sigma_M^2\right)}}_{\mu_M^0} \end{aligned}$$

Thus, we have:

$$\begin{aligned} \beta_{M\eta}^0 &= \frac{\sigma_v^2}{\rho_M\left(\sigma_v^2+\sigma_M^2\right)}, \\ \beta_{Mp}^0 &= \frac{1}{\rho_M}, \\ \mu_M^0 &= \frac{\sigma_m^2 \bar{v}}{\rho_M\left(\sigma_v^2+\sigma_m^2\right)}. \end{aligned}$$

A.2 Proof of Proposition 2

The market-clearing condition as perceived by the step-$(k-1)$ AI investor is:

$$\sum_{j=0}^{k-1} \int_0^{\psi g_{k-1}(j)} x_{it}^j di + x_{At}^{k-1} + z_t = S,$$

Substituting the optimal demand of step-0 to step-$(k-1)$ human investors, we can rewrite the perceived market-clearing condition as

$$\sum_{j=0}^{k-1} \left(\beta_{M\eta}^j \int_0^{g_{k-1}(j)} \eta_{it} di - \psi g_{k-1}(j) \beta_{Mp}^j p_t + \psi g_{k-1}(j) \mu_M^j \right) + x_{At}^{k-1} + z_t - S = 0,$$

where

$$\begin{cases} \bar{\beta}_{M\eta}^{k-1} = \sum_{j=0}^{k-1} g_{k-1}(j) \beta_{M\eta}^j \\ \bar{\beta}_{Mp}^{k-1} = \sum_{j=0}^{k-1} g_{k-1}(j) \beta_{Mp}^j \\ \bar{\mu}_M^{k-1} = \sum_{j=0}^{k-1} g_{k-1}(j) \mu_M^j \end{cases}$$

$\bar{\beta}_{M\eta}^{k-1}$, $\bar{\beta}_{Mp}^{k-1}$, and $\bar{\mu}_M^{k-1}$ denote the weighted averages of the demand coefficients for all lower-step human investors.

Solving for the price function gives:

$$\mathcal{P}_A^{k-1} : p_t = \frac{\bar{\beta}_{M\eta}^{k-1}}{\bar{\beta}_{Mp}^{k-1}} \cdot v_t + \frac{1}{\psi \bar{\beta}_{Mp}^{k-1}} \cdot x_{At}^{k-1} + \frac{1}{\psi \bar{\beta}_{Mp}^{k-1}} \cdot z_t + \frac{1}{\psi \bar{\beta}_{Mp}^{k-1}} \cdot \widetilde{\mu}_M^{k-1},$$

where $\widetilde{\mu}_M^{k-1} = \psi \bar{\mu}_M^{k-1} - S$. Define the residual supply $h_{A,t}^{k-1}$ as

$$h_{A,t}^{k-1} = p_t - \lambda_A^{k-1} x_{At}^{k-1} = \frac{\bar{\beta}_{M\eta}^{k-1}}{\bar{\beta}_{Mp}^{k-1}} \cdot v_t + \frac{1}{\psi \bar{\beta}_{Mp}^{k-1}} \cdot z_t + \frac{1}{\psi \bar{\beta}_{Mp}^{k-1}} \cdot \widetilde{\mu}_M^{k-1},$$

and thus, the signal from price $\widetilde{h}_{A,t}^{k-1}$ is

$$\widetilde{h}_{A,t}^{k-1} \equiv \frac{\bar{\beta}_{Mp}^{k-1}}{\bar{\beta}_{M\eta}^{k-1}} \left(h_{A,t}^{k-1} - \frac{\widetilde{\mu}_M^{k-1}}{\psi \bar{\beta}_{Mp}^{k-1}} \right) = v_t + \frac{z_t}{\psi \bar{\beta}_{M\eta}^{k-1}}. \tag{A2}$$

According to Bayes' law, the posterior distribution of v_t is given by:

$$\begin{aligned}
\mathbb{E}\left[v_t \mid (\mathcal{P}_A^{k-1})^{-1}(p_t)\right] &= \frac{(\hat{\sigma}_A^{k-1})^2}{\sigma_v^2}\bar{v} + \frac{(\hat{\sigma}_A^{k-1})^2}{\left(\psi\bar{\beta}_{M\eta}^{k-1}\right)^{-2}\sigma_z^2}\widetilde{h}_{A,t}^{k-1}, \\
\left(\hat{\sigma}_A^{k-1}\right)^2 \equiv \mathrm{Var}\left[v_t \mid (\mathcal{P}_A^{k-1})^{-1}(p_t)\right] &= \left(\frac{1}{\sigma_v^2} + \frac{1}{\left(\psi\bar{\beta}_{M\eta}^{k-1}\right)^{-2}\sigma_z^2}\right)^{-1}.
\end{aligned}$$

Substituting the posterior beliefs and the expression for the price signal (A2) into the optimal demand function (6) yields:

$$\begin{aligned}
x_{A,t}^{k-1}(\rho_A + \lambda_{At}^{k-1}) &= \frac{(\hat{\sigma}_A^{k-1})^2}{\sigma_v^2}\bar{v} + \frac{(\hat{\sigma}_A^{k-1})^2}{(\psi\bar{\beta}_{M\eta}^{k-1})^{-2}\sigma_z^2}\widetilde{h}_A^{k-1} - p_t^{k-1} \\
&= \frac{(\hat{\sigma}_A^{k-1})^2}{\sigma_v^2}\bar{v} + \frac{(\hat{\sigma}_A^{k-1})^2}{(\psi\bar{\beta}_{M\eta}^{k-1})^{-2}\sigma_z^2}\frac{\bar{\beta}_{Mp}^{k-1}}{\bar{\beta}_{M\eta}^{k-1}}\left(p_t - \frac{1}{\psi\bar{\beta}_{Mp}^{k-1}}x_{A,t}^{k-1} - \frac{\widetilde{\mu}_M^{k-1}}{\psi\bar{\beta}_{Mp}^{k-1}}\right) - p_t^{k-1} \\
&= \frac{(\hat{\sigma}_A^{k-1})^2}{\sigma_v^2}\bar{v} + \frac{(\hat{\sigma}_A^{k-1})^2\psi^2\bar{\beta}_{M\eta}^{k-1}\bar{\beta}_{Mp}^{k-1}}{\sigma_z^2}p_t - \frac{(\hat{\sigma}_A^{k-1})^2\psi\bar{\beta}_{M\eta}^{k-1}}{\sigma_z^2}x_{A,t}^{k-1} \\
&\quad - \frac{(\hat{\sigma}_A^{k-1})^2\psi\bar{\beta}_{M\eta}^{k-1}}{\sigma_z^2}\widetilde{\mu}_M^{k-1} - p_t^{k-1}
\end{aligned}$$

Moving and combining terms, we get:

$$\begin{aligned}
x_{At}^{k-1}\left[\rho_A + \lambda_{At}^{k-1} + \frac{\psi(\hat{\sigma}_A^{k-1})^2\bar{\beta}_{M\eta}^{k-1}}{\sigma_z^2}\right] &= \left(\frac{(\hat{\sigma}_A^{k-1})^2}{\sigma_v^2}\bar{v} - \frac{\psi(\hat{\sigma}_A^{k-1})^2\bar{\beta}_{M\eta}^{k-1}}{\sigma_z^2}\widetilde{\mu}_M^{k-1}\right) \\
&\quad + \left[\frac{\psi^2(\hat{\sigma}_A^{k-1})^2\bar{\beta}_{M\eta}^{k-1}\bar{\beta}_{Mp}^{k-1}}{\sigma_z^2} - 1\right]p_t
\end{aligned}$$

By matching terms, we solve for the coefficients of the demand function for the step-$(k-1)$ AI investor:

$$\begin{aligned}
\beta_{Ap}^{k-1} &= \left[\rho_A + \lambda_A^{k-1} + \frac{\psi(\hat{\sigma}_A^{k-1})^2\bar{\beta}_{M\eta}^{k-1}}{\sigma_z^2}\right]^{-1} \cdot \left[1 - \frac{\psi^2(\hat{\sigma}_A^{k-1})^2\bar{\beta}_{Mp}^{k-1}\bar{\beta}_{M\eta}^{k-1}}{\sigma_z^2}\right], \\
\mu_A^{k-1} &= \left[\rho_A + \lambda_A^{k-1} + \frac{\psi(\hat{\sigma}_A^{k-1})^2\bar{\beta}_{M\eta}^{k-1}}{\sigma_z^2}\right]^{-1} \cdot \left[\frac{(\hat{\sigma}_A^{k-1})^2}{\sigma_v^2}\bar{v} - \frac{\psi(\hat{\sigma}_A^{k-1})^2\bar{\beta}_{M\eta}^{k-1}}{\sigma_z^2} \cdot \widetilde{\mu}_M^{k-1}\right].
\end{aligned} \tag{A3}$$

A.3 Proof of Proposition 3

Based on their beliefs about the behavior of other human and AI investors, the market-clearing condition as perceived by step-k human investors is:

$$\sum_{j=0}^{k-1}\left(\beta_{M\eta}^{j}\int_{0}^{g_{k-1}(j)}\eta_{it}di-\psi g_{k-1}(j)\beta_{Mp}^{j}p_t+\psi g_{k-1}(j)\mu_M^{j}\right)+\left(\mu_A^{k-1}-\beta_{Ap}^{k-1}\cdot p_t\right)$$
$$+z_t-S=0$$

Expanding the equation yields:

$$\psi\sum_{j=0}^{k-1}g_k(j)(\beta_{M\eta}^{j}v_t-\beta_{Mp}^{j}p_t+\mu_M^{j})+(-\beta_{AP}^{k-1}p_t+\mu_A^{k-1})+z_t-S=0$$

Using the definition in (7), this equation can be rewritten as:

$$\psi(\bar{\beta}_{M\eta}^{k-1}v_t-\bar{\beta}_{Mp}^{k-1}p_t+\bar{\mu}_M^{k-1})+\left(-\beta_{Ap}^{k-1}p_t+\mu_A^{k-1}\right)+z_t-S=0$$
$$\Rightarrow\quad(\psi\bar{\beta}_{Mp}^{k-1}+\beta_{Ap}^{k-1})p_t=(\psi\bar{\beta}_{M\eta}^{k-1}+\beta_{A\eta}^{k-1})v_t+z_t+\underbrace{\psi\bar{\mu}_M^{k-1}+\mu_A^{k-1}-S}_{\widetilde{\mu}^{k-1}}$$

Solving for the price yields

$$\mathcal{P}_M^{k}:p_t=\left(\psi\bar{\beta}_{Mp}^{k-1}+\beta_{Ap}^{k-1}\right)^{-1}\left[\psi\bar{\beta}_{M\eta}^{k-1}v_t+z_t+\widetilde{\mu}^{k-1}\right],$$

where $\widetilde{\mu}^{k-1}\equiv\psi\bar{\mu}_M^{k-1}+\mu_A^{k-1}-S$. Define the signal from price $\widetilde{h}_{M,t}^{k}$ as

$$\widetilde{h}_{M,t}^{k}\equiv\frac{(\psi\bar{\beta}_{Mp}^{k-1}+\beta_{Ap}^{k-1})}{\psi\bar{\beta}_{M\eta}^{k-1}}\left(p_t-\frac{\widetilde{\mu}^{k-1}}{\psi\bar{\beta}_{Mp}^{k-1}+\beta_{Ap}^{k-1}}\right)=v_t+\frac{z_t}{\psi\bar{\beta}_{M\eta}^{k-1}}.\qquad\text{(A4)}$$

According to Bayes' rule, the posterior distribution of v_t is given by:

$$\mathbb{E}\left[v_t \mid \eta_{it}, (\mathcal{P}_M^k)^{-1}(p_t)\right] = \frac{(\hat{\sigma}_M^k)^2}{\sigma_v^2}\bar{v} + \frac{(\hat{\sigma}_M^k)^2}{\sigma_M^2}\eta_{i,t} + \frac{(\hat{\sigma}_M^k)^2}{(\sigma_{Mp}^k)^2}\widetilde{h}_{M,t}^k,$$

$$(\hat{\sigma}_M^k)^2 \equiv \operatorname{Var}\left[v_t \mid \eta_{it}, (\mathcal{P}_M^k)^{-1}(p_t)\right] = \left(\frac{1}{\sigma_v^2} + \frac{1}{\sigma_M^2} + \frac{1}{(\sigma_{Mp}^k)^2}\right)^{-1},$$

where

$$(\sigma_{Mp}^k)^2 = \frac{\sigma_z^2}{\left(\psi\bar{\beta}_{M\eta}^{k-1}\right)^2}$$

Substituting the posterior beliefs and the expression for the price signal (A4) into the optimal demand function (9), gives:

$$\begin{aligned}
x_{it}^k(\eta_{it}, p_t) =& \frac{\mathbb{E}\left[v_t \mid \eta_{it}, (\mathcal{P}_M^k)^{-1}(p_t)\right] - p_t}{\rho_M} \\
=& \rho_M^{-1}\left[\frac{\left(\hat{\sigma}_M^k\right)^2}{\sigma_v^2}\bar{v} + \frac{\left(\hat{\sigma}_M^k\right)^2}{\sigma_M^2}\eta_{it} + \frac{\left(\hat{\sigma}_M^k\right)^2}{\left(\sigma_{Mp}^k\right)^2}\frac{\left(\psi\bar{\beta}_{Mp}^{k-1} + \beta_{Ap}^{k-1}\right)}{\psi\bar{\beta}_{M\eta}^{k-1}}\left(p_t - \frac{\widetilde{\mu}^{k-1}}{\left(\psi\bar{\beta}_{Mp}^{k-1} + \beta_{Ap}^{k-1}\right)}\right) - p_t\right] \\
=& \rho_M^{-1}\frac{\left(\hat{\sigma}_M^k\right)^2}{\sigma_M^2}\eta_{it} + \rho_M^{-1}\left[\frac{\left(\hat{\sigma}_M^k\right)^2}{\left(\sigma_{Mp}^k\right)^2}\frac{\left(\psi\bar{\beta}_{Mp}^{k-1} + \beta_{Ap}^{k-1}\right)}{\psi\bar{\beta}_{M\eta}^{k-1}} - 1\right]p_t \\
&+ \rho_M^{-1}\frac{\left(\hat{\sigma}_M^k\right)^2}{\sigma_v^2}\bar{v} - \rho_M^{-1}\frac{\left(\hat{\sigma}_M^k\right)^2}{\left(\sigma_{Mp}^k\right)^2}\frac{\left(\psi\bar{\beta}_{Mp}^{k-1} + \beta_{Ap}^{k-1}\right)}{\psi\bar{\beta}_{M\eta}^{k-1}}\frac{\widetilde{\mu}^{k-1}}{\left(\psi\bar{\beta}_{Mp}^{k-1} + \beta_{Ap}^{k-1}\right)}
\end{aligned}$$

By matching terms, we solve for the coefficients of the demand function for step-k human investors:

$$\begin{aligned}
\beta_{M\eta}^k &= \rho_M^{-1}\frac{(\hat{\sigma}_M^k)^2}{\sigma_M^2}, \\
\beta_{Mp}^k &= \rho_M^{-1} - \rho_M^{-1}\frac{(\hat{\sigma}_M^k)^2}{(\sigma_{Mp}^k)^2}\cdot\frac{\left(\psi\bar{\beta}_{Mp}^{k-1} + \beta_{Ap}^{k-1}\right)}{\psi\bar{\beta}_{M\eta}^{k-1}}, \\
\mu_M^k &= \rho_M^{-1}\frac{(\hat{\sigma}_M^k)^2}{\sigma_v^2}\bar{v} - \rho_M^{-1}\frac{(\hat{\sigma}_M^k)^2}{(\sigma_{Mp}^k)^2}\cdot\frac{\widetilde{\mu}^{k-1}}{\psi\bar{\beta}_{M\eta}^{k-1}}.
\end{aligned} \tag{A5}$$

A.4 Proof of Proposition 4

We first restate Proposition 4 under a less restrictive condition:

Proposition 7 (Conditional Convergence)**.** *If the system's parameters* $(\rho_A, \rho_M, \psi, \sigma_v^2, \sigma_M^2, \sigma_z^2)$

are such that the state trajectory of the averaged coefficients $\bar{x}^k = (\bar{\beta}^k_{M\eta}, \bar{\beta}^k_{Mp})$ *remains within a domain* $D_S^+ \subset D_G$ *that is bounded away from the system's singularities, then the sequences of strategy coefficients* $\beta^k_{M\eta}$*,* β^k_{Mp}*, and* β^{k-1}_{Ap} *converge to unique limits* $\beta^\infty_{M\eta}$*,* β^∞_{Mp}*, and* β^∞_{Ap}*, respectively.*

Proof. The proof proceeds in several steps. We first formulate the system as a non-autonomous vector recurrence, analyze its stability, prove boundedness, and then prove convergence.

System formulation. Let the non-averaged state at step k be $x^k = (x^k_1, x^k_2)^T = (\beta^k_{M\eta}, \beta^k_{Mp})^T$, and the averaged state be $\bar{x}^k = (\bar{x}^k_1, \bar{x}^k_2)^T = (\bar{\beta}^k_{M\eta}, \bar{\beta}^k_{Mp})^T$. Let $y^k = \beta^k_{Ap}$.

The system definitions show that x^k and y^{k-1} are functions of $\bar{x}^{k-1}$.

$$y^{k-1} = F_A(\bar{x}^{k-1}) \tag{A6}$$

$$x^k = G(\bar{x}^{k-1}) \tag{A7}$$

We first define the component functions, which depend on $\bar{x} = (\bar{x}_1, \bar{x}_2)$.

Definition 1 (Component Functions)**.**

$$(\sigma_{Mp}(\bar{x}))^2 = \frac{\sigma_z^2}{(\psi \bar{x}_1)^2} \tag{A8}$$

$$(\hat{\sigma}_M(\bar{x}))^2 = \left(\frac{1}{\sigma_v^2} + \frac{1}{\sigma_M^2} + \frac{1}{(\sigma_{Mp}(\bar{x}))^2}\right)^{-1} = \left(\frac{1}{\sigma_v^2} + \frac{1}{\sigma_M^2} + \frac{(\psi \bar{x}_1)^2}{\sigma_z^2}\right)^{-1} \tag{A9}$$

$$(\hat{\sigma}_A(\bar{x}))^2 = \left(\frac{1}{\sigma_v^2} + \frac{1}{(\psi \bar{x}_1)^{-2}\sigma_z^2}\right)^{-1} = \left(\frac{1}{\sigma_v^2} + \frac{(\psi \bar{x}_1)^2}{\sigma_z^2}\right)^{-1} \tag{A10}$$

$$\lambda_A(\bar{x}) = \frac{1}{\psi \bar{x}_2} \tag{A11}$$

Definition 2 (Component Functions C_A and C_B)**.** *For brevity, we define two functions*

$C_A(\bar{x}_1)$ and $C_B(\bar{x}_1)$, which are continuous and bounded on I_1 (as $\bar{x}_1 \in I_1$):

$$C_A(\bar{x}_1) = \rho_A + \frac{\psi(\hat{\sigma}_A(\bar{x}))^2 \bar{x}_1}{\sigma_z^2} \tag{A12}$$

$$C_B(\bar{x}_1) = \frac{\psi^2(\hat{\sigma}_A(\bar{x}))^2 \bar{x}_1}{\sigma_z^2} \tag{A13}$$

Since $\bar{x}_1 > 0$ and all other terms are positive, $C_A(\bar{x}_1) > 0$ (assuming $\rho_A \geq 0$).

Definition 3 (System Maps F_A and G)**.** *The map $F_A : D_G \to \mathbb{R}$ from the problem definition is:*

$$F_A(\bar{x}) = \left[\rho_A + \lambda_A(\bar{x}) + \frac{\psi(\hat{\sigma}_A(\bar{x}))^2 \bar{x}_1}{\sigma_z^2}\right]^{-1} \cdot \left[1 - \frac{\psi^2(\hat{\sigma}_A(\bar{x}))^2 \bar{x}_1 \bar{x}_2}{\sigma_z^2}\right] \tag{A14}$$

Using (A11), (A12), and (A13), this can be written as:

$$F_A(\bar{x}) = \left[C_A(\bar{x}_1) + \frac{1}{\psi \bar{x}_2}\right]^{-1} \cdot [1 - C_B(\bar{x}_1)\bar{x}_2] \tag{A15}$$

The map $G : D_G \to \mathbb{R}^2$ has components $G(\bar{x}) = (G_1(\bar{x}), G_2(\bar{x}))^T$:

$$G_1(\bar{x}) = \rho_M^{-1} \frac{(\hat{\sigma}_M(\bar{x}))^2}{\sigma_M^2} \tag{A16}$$

$$G_2(\bar{x}) = \rho_M^{-1} - \rho_M^{-1} \frac{(\hat{\sigma}_M(\bar{x}))^2}{(\sigma_{Mp}(\bar{x}))^2} \cdot \frac{\psi \bar{x}_2 + F_A(\bar{x})}{\psi \bar{x}_1} \tag{A17}$$

Substituting (A8) into (A17), we get the final form for G_2:

$$G_2(\bar{x}) = \rho_M^{-1} - \rho_M^{-1} \frac{(\hat{\sigma}_M(\bar{x}))^2 \psi \bar{x}_1}{\sigma_z^2} (\psi \bar{x}_2 + F_A(\bar{x})) \tag{A18}$$

The domain D_G is the set of $\bar{x}$ where all denominators (in F_A and G) are non-zero.

The recurrence for the averaged state $\bar{x}^k$ is:

$$\bar{x}^k = \sum_{j=0}^{k} g_k(j) x^j = \frac{\sum_{j=0}^{k} f(j) x^j}{\sum_{l=0}^{k} f(l)} \tag{A19}$$

Let $S_k = \sum_{l=0}^{k} f(l)$. This recurrence can be written recursively:

$$\bar{x}^k = \frac{1}{S_k}\left(\sum_{j=0}^{k-1} f(j)x^j + f(k)x^k\right) = \frac{S_{k-1}}{S_k}\bar{x}^{k-1} + \frac{f(k)}{S_k}x^k$$

Let $\alpha_k = f(k)/S_k$. Since $S_{k-1} = S_k - f(k)$, we have $S_{k-1}/S_k = 1 - \alpha_k$. Substituting (A7) into this recurrence gives the final system dynamics:

$$\bar{x}^k = (1-\alpha_k)\bar{x}^{k-1} + \alpha_k G(\bar{x}^{k-1}) \tag{A20}$$

Boundedness of the $\bar{\beta}^k_{M\eta}$ component. $x^k_1 = \beta^k_{M\eta} = G_1(\bar{x}^{k-1})$. From (A16), $G_1(\bar{x}) = \rho_M^{-1}\frac{(\hat{\sigma}_M(\bar{x}))^2}{\sigma_M^2}$. From (A9), $(\hat{\sigma}_M(\bar{x}))^2 = \left(\frac{1}{\sigma_v^2} + \frac{1}{\sigma_M^2} + \frac{(\psi\bar{x}_1)^2}{\sigma_z^2}\right)^{-1}$. Since all terms in the parenthesis are positive, $(\hat{\sigma}_M)^2 > 0$. This implies $G_1(\bar{x}) > 0$. Furthermore, $(\hat{\sigma}_M)^2 \leq \left(\frac{1}{\sigma_v^2} + \frac{1}{\sigma_M^2}\right)^{-1} = \frac{\sigma_v^2\sigma_M^2}{\sigma_v^2+\sigma_M^2}$. Substituting this bound into the expression for $G_1(\bar{x})$:

$$G_1(\bar{x}) \leq \rho_M^{-1}\frac{1}{\sigma_M^2}\left(\frac{\sigma_v^2\sigma_M^2}{\sigma_v^2+\sigma_M^2}\right) = \frac{\sigma_v^2}{\rho_M(\sigma_v^2+\sigma_M^2)} = \beta^0_{M\eta}$$

Thus, $0 < x^k_1 \leq \beta^0_{M\eta}$ for all $k \geq 1$. Since $x^0_1 = \beta^0_{M\eta}$, the sequence x^k_1 is bounded in $(0, \beta^0_{M\eta}]$ for all $k \geq 0$. The averaged state $\bar{x}^k_1$ is a convex combination of $\{x^0_1, \ldots, x^k_1\}$ and therefore must also be bounded in $I_1 = (0, \beta^0_{M\eta}]$. This confines the trajectory $\bar{x}^k$ to a vertical strip $D_S = I_1 \times \mathbb{R}$.

Stability analysis and assumption. The convergence of $\bar{x}^k_2 = \bar{\beta}^k_{Mp}$ is not guaranteed. The function $G(\bar{x})$ is not continuous on all of $\mathbb{R}^2$. Its domain D_G excludes:

- The line $\bar{\beta}_{M\eta} = 0$. (Previous sections proved the trajectory avoids this).
- The curve where the denominator of F_A in (A15) is zero. This is where $C_A(\bar{x}_1) + 1/(\psi\bar{x}_2) = 0$, which simplifies to $\bar{x}_2 = -1/(\psi C_A(\bar{x}_1))$. Since $C_A > 0$, this is a curve in the negative half-plane $\bar{x}_2 < 0$.

The "safe" domain is $D_S^+ = I_1 \times (0, \infty)$. The initial state $\bar{x}^0 = (\beta_{M\eta}^0, 1/\rho_M)$ is in this domain, because $\rho_M > 0$.

However, G does not necessarily map the safe domain D_S^+ to itself. As shown by (A18), the term $F_A(\bar{x})$ grows linearly with $\bar{x}_2$ (specifically, $F_A(\bar{x}) \approx (-C_B/C_A)\bar{x}_2$). This means $\psi\bar{x}_2 + F_A(\bar{x})$ also grows linearly. This, in turn, implies that $G_2(\bar{x})$ grows linearly with $\bar{x}_2$. For example, when $\psi > C_B/C_A$, as $\bar{x}_2 \to \infty$, $G_2(\bar{x})$ will tend towards $-\infty$ (assuming relevant terms are positive).

This proves that $G_2(\bar{x})$ can be negative. The map G can take a state $\bar{x} \in D_S^+$ and produce an instantaneous state $x = G(\bar{x})$ that is outside D_S^+ (i.e., $x_2 < 0$). The next average state, $\bar{x}^k = (1-\alpha_k)\bar{x}^{k-1} + \alpha_k x^k$, is a convex combination of a positive $\bar{x}_2^{k-1}$ and a negative x_2^k. It is entirely possible for $\bar{x}_2^k$ to be negative. If the trajectory $\bar{x}^k$ ever hits one of the singularities, the system explodes.

A sufficient condition for stability (the condition in Proposition 4) is path-dependent and requires that:

$$\bar{x}_2^k = (1-\alpha_k)\bar{x}_2^{k-1} + \alpha_k G_2(\bar{x}^{k-1}) > 0 \quad \text{(and avoids the other singularity)}$$

for all k. If G_2 is negative, we need $\bar{x}_2^{k-1}$ to be large enough to absorb the negative pull:

$$\frac{\bar{x}_2^{k-1}}{-G_2(\bar{x}^{k-1})} > \frac{\alpha_k}{1-\alpha_k}$$

The key observation is that $\alpha_k = f(k)/S_k \to 0$ as $k \to \infty$ (since $f(k) \to 0$ and $S_k \to 1$). This means the ratio $\alpha_k/(1-\alpha_k) \to 0$. The system becomes more stable over time. The weight on the previous average, $(1-\alpha_k)$, dominates, and the contribution of the new term, $\alpha_k G_2$, vanishes. This condition is more difficult to satisfy for small k (e.g., $k = 1, 2, 3, \dots$) when α_k is large. The system must survive these first few steps.

This analysis shows that convergence is not universal. The proof must be conditional.

Assumption 1 (System Stability)**.** *The system parameters are such that the trajectory*

$\bar{x}^k$ defined by (A20), starting from $\bar{x}^0 \in D_S^+ = I_1 \times (0, \infty)$, remains within this domain $D_S = I_1 \times (0, \infty)$ for all k.

Lemma 8 (At-Most-Linear Growth of G)**.** *The function $G(\bar{x})$ has at-most-linear growth on the domain $D_S^+ = I_1 \times (0, \infty)$. That is, there exist constants $L, M > 0$ such that $\|G(\bar{x})\| \leq L\|\bar{x}\| + M$ for all $\bar{x} \in D_S^+$.*

Proof. We analyze the components $G = (G_1, G_2)$.

1. *Growth of $G_1(\bar{x})$:* As shown in Step 2, $G_1(\bar{x})$ depends only on $\bar{x}_1$, and it is uniformly bounded:

$$|G_1(\bar{x})| \leq \beta^0_{M\eta}$$

   for all $\bar{x} \in D_S$. A bounded function trivially satisfies linear growth.

2. *Growth of $G_2(\bar{x})$:* From (A17), $G_2(\bar{x}) = \rho_M^{-1}\left[1 - \frac{(\hat{\sigma}_M(\bar{x}))^2}{(\sigma_{Mp}(\bar{x}))^2} \cdot \frac{\psi\bar{x}_2 + F_A(\bar{x})}{\psi\bar{x}_1}\right]$. Let $C_1(\bar{x}) = \frac{(\hat{\sigma}_M(\bar{x}))^2}{(\sigma_{Mp}(\bar{x}))^2}$. Using (A8) and (A9), C_1 depends only on $\bar{x}_1$. Since $\bar{x}_1 \in I_1$ (a compact set), $C_1(\bar{x}_1)$ is bounded. Let $C_2(\bar{x}_1) = C_1(\bar{x}_1)/(\psi\bar{x}_1)$. C_2 is also continuous and positive on the compact set I_1, and thus bounded by some constant $K_C > 0$.

$$|G_2(\bar{x})| \leq |\rho_M^{-1}| \left(1 + |C_2(\bar{x}_1)| \cdot |\psi\bar{x}_2 + F_A(\bar{x})|\right) \leq |\rho_M^{-1}| \left(1 + K_C|\psi\bar{x}_2 + F_A(\bar{x})|\right)$$

   Now we analyze the growth of $F_A(\bar{x})$ with respect to $\bar{x}_2$. Using the form from (A15):

$$F_A(\bar{x}) = \left[\frac{C_A\psi\bar{x}_2 + 1}{\psi\bar{x}_2}\right]^{-1} \cdot [1 - C_B\bar{x}_2] = \frac{\psi\bar{x}_2}{C_A\psi\bar{x}_2 + 1} \cdot (1 - C_B\bar{x}_2)$$

$$F_A(\bar{x}) = \frac{\psi\bar{x}_2 - C_B\psi\bar{x}_2^2}{C_A\psi\bar{x}_2 + 1}$$

   where C_A and C_B are positive, bounded functions of $\bar{x}_1$ as defined in (A12) and (A13). This is a rational function of $\bar{x}_2$. By polynomial long division, we can write $F_A(\bar{x}) = (a\bar{x}_2 + b) + \frac{c}{C_A\psi\bar{x}_2 + 1}$, where a, b, c are coefficients that depend on C_A, C_B, ψ (and thus are bounded, since C_A, C_B are).

This linear growth bound is not global, as the hyperbolic term $\frac{c}{C_A\psi\bar{x}_2+1}$ has a singularity at $\bar{x}_2 = -1/(C_A\psi)$. However, our domain is $D_S = I_1 \times (0, \infty)$. Since $C_A > 0$, $\psi > 0$, and $\bar{x}_2 > 0$, the denominator $C_A\psi\bar{x}_2 + 1$ is strictly greater than 1. Because the denominator is bounded below by 1 (and thus bounded away from zero), the hyperbolic term $\frac{c}{C_A\psi\bar{x}_2+1}$ is uniformly bounded on our entire domain D_S. Therefore, $F_A(\bar{x}) =$ (linear part)+(bounded part). This proves that F_A has at-most-linear growth on D_S.

$$|F_A(\bar{x})| \leq |a||\bar{x}_2| + |b| + K_{hyp} \leq K_A|\bar{x}_2| + M_A$$

for some constants K_{hyp}, K_A, M_A.

Now substitute this back into the expression for $|G_2(\bar{x})|$:

$$|G_2(\bar{x})| \leq |\rho_M^{-1}| \left(1 + K_C(|\psi\bar{x}_2| + |F_A(\bar{x})|)\right)$$

$$|G_2(\bar{x})| \leq |\rho_M^{-1}| \left(1 + K_C(|\psi\bar{x}_2| + K_A|\bar{x}_2| + M_A)\right)$$

$$|G_2(\bar{x})| \leq |\rho_M^{-1}| \left(1 + K_C M_A + K_C(|\psi| + K_A)|\bar{x}_2|\right)$$

This proves $|G_2(\bar{x})| \leq L_2|\bar{x}_2| + M_2$ for some constants L_2, M_2.

3. *Norm of $G(\bar{x})$:* Finally, we bound the norm of $G(\bar{x})$:

$$\|G(\bar{x})\| = \sqrt{|G_1(\bar{x})|^2 + |G_2(\bar{x})|^2} \leq \sqrt{(\beta^0_{M\eta})^2 + (L_2|\bar{x}_2| + M_2)^2}$$

Since $\bar{x}_1$ is bounded, $\|\bar{x}\| = \sqrt{|\bar{x}_1|^2 + |\bar{x}_2|^2}$ is “equivalent” to $|\bar{x}_2|$ for large $|\bar{x}_2|$.

$$\|G(\bar{x})\| \leq \sqrt{(\beta^0_{M\eta})^2 + (L_2(\|\bar{x}\|) + M_2)^2}$$

For large $\|\bar{x}\|$, this is bounded by $L_2\|\bar{x}\| +$ constants. Thus, there exist constants L, M such that $\|G(\bar{x})\| \leq L\|\bar{x}\| + M$ for all $\bar{x} \in D_S$.

□

Theorem 9 (Boundedness of $\bar{x}^k$). *Given Assumption 1, the sequence $\bar{x}^k$ is bounded.*

Proof. From (A20) and Lemma 1, for $k \geq 1$:

$$\|\bar{x}^k\| \leq (1-\alpha_k)\|\bar{x}^{k-1}\| + \alpha_k\|G(\bar{x}^{k-1})\|$$

$$\|\bar{x}^k\| \leq (1-\alpha_k)\|\bar{x}^{k-1}\| + \alpha_k(L\|\bar{x}^{k-1}\| + M)$$

$$\|\bar{x}^k\| \leq (1+\alpha_k(L-1))\|\bar{x}^{k-1}\| + \alpha_k M$$

Let $C_k = \prod_{j=1}^{k}(1+\alpha_j(L-1))$. Then by unrolling the recurrence:

$$\|\bar{x}^k\| \leq C_k\|\bar{x}^0\| + M\sum_{j=1}^{k}\alpha_j C_k C_j^{-1}$$

Since $f(k)$ is a probability distribution, $\sum f(k) = 1$. The partial sum $S_k = \sum_{l=0}^{k} f(l)$ is bounded away from zero (by $S_0 = f(0) = e^{-\tau} > 0$) and $S_k \to 1$. Because $\sum f(k)$ converges and $S_k \geq f(0)$, the series $\sum \alpha_k = \sum(f(k)/S_k)$ also converges. Since $\sum \alpha_k < \infty$, the infinite product $\prod(1+\alpha_j(L-1))$ converges (if $L \geq 1$). This means C_k converges to a finite constant C_∞. The sequence $\|\bar{x}^k\|$ is bounded by $C_\infty\|\bar{x}^0\| + MC_\infty\sum\alpha_j$, which is a finite constant. Therefore, the sequence $\bar{x}^k$ is bounded. □

Theorem 10 (Convergence of $\bar{x}^k$). *Given Assumption 1, the sequence $\bar{x}^k$ converges.*

Proof. From Theorem 1, $\bar{x}^k$ is bounded. It lies in some compact set $\mathcal{K} \subset D_S$.

1. *x^k is bounded.* G is continuous on the compact set $\mathcal{K}$, so its image $G(\mathcal{K})$ is compact. Since $x^k = G(\bar{x}^{k-1})$, the sequence x^k is also bounded.

2. *Successive differences are summable.* Since x^k and $\bar{x}^{k-1}$ are bounded (from part 1 and Theorem 1), their difference is bounded by some constant $C > 0$. From (A20), we have

$\bar{x}^k = (1-\alpha_k)\bar{x}^{k-1} + \alpha_k x^k$. Rearranging gives the successive difference: $\bar{x}^k - \bar{x}^{k-1} = \alpha_k(x^k - \bar{x}^{k-1})$. Taking the norm:

$$\|\bar{x}^k - \bar{x}^{k-1}\| = \alpha_k \|x^k - \bar{x}^{k-1}\| \leq C \cdot \alpha_k$$

As $\sum \alpha_k$ converges, the series of successive differences $\sum \|\bar{x}^k - \bar{x}^{k-1}\|$ converges by the comparison test.

3. *$\bar{x}^k$ is a Cauchy sequence.* A sequence whose successive differences are absolutely summable is a Cauchy sequence. The sequence $\bar{x}^k$ exists in $\mathbb{R}^2$, which is a complete metric space. Therefore, the sequence must converge to a limit $\bar{x}^\infty \in \mathbb{R}^2$. By Assumption 1, the sequence $\bar{x}^k$ is in $D_S = I_1 \times (0,\infty)$, so its limit $\bar{x}^\infty$ must lie in the closure of this set, $\bar{D}_S = I_1 \times [0,\infty)$.

□

Convergence of all Sequences and the fixed point Given $\lim_{k\to\infty} \bar{x}^k = \bar{x}^\infty$:

- *Convergence of $x^k = (\beta^k_{M\eta}, \beta^k_{Mp})$.* By continuity of G:

$$\lim_{k\to\infty} x^k = \lim_{k\to\infty} G(\bar{x}^{k-1}) = G(\lim_{k\to\infty} \bar{x}^{k-1}) = G(\bar{x}^\infty)$$

- *Convergence of $y^k = \beta^k_{Ap}$.* By continuity of F_A:

$$\lim_{k\to\infty} y^k = \lim_{k\to\infty} F_A(\bar{x}^k) = F_A(\lim_{k\to\infty} \bar{x}^k) = F_A(\bar{x}^\infty)$$

This completes the proof. □

A.5 Proof of Propositions 5 and 6

Expected profit for human investor. The expected profit for a step-k_M human investor is derived as follows.

$$\begin{aligned}
\mathbb{E}\left[\pi_{it}^{k_M}\right] &= \mathbb{E}\left[\left(v_t - p_t^{k_{Env}}\right) x_{it}^{k_M}\right] - \frac{1}{2}\rho_M \mathbb{E}\left(x_{it}^{k_M}\right)^2, \\
&= \underbrace{\mathbb{E}\left[\left(v_t - p_t^{k_{Env}}\right)\left(\beta_{M\eta}^{k_M} \cdot (v_t + e_{it}) - \beta_{Mp}^{k_M} p_t^{k_{Env}} + \mu_M^{k_M}\right)\right]}_{\mathbb{E}(gross\ profit)} \\
&\quad \underbrace{-\frac{1}{2}\rho_M \mathbb{E}\left[\left(\beta_{M\eta}^{k_M}(v_t + e_{it}) - \beta_{Mp}^{k_M} p_t^{k_{Env}} + \mu_M^{k_M}\right)^2\right]}_{\mathbb{E}(transaction\ cost)}
\end{aligned}$$

For ease of notation, we drop the superscript k_M from the strategy coefficients, $\beta_{M\eta}^{k_M}$, $\beta_{Mp}^{k_M}$, and $\mu_M^{k_M}$, and the superscript k_{Env} from the price, $p_t^{k_{Env}}$, and market condition parameters, $\theta^{k_{Env}}$ and $\xi^{k_{Env}}$.

First, we calculate some expectations of terms related to v_t and e_{it}:

$$\begin{aligned}
\mathbb{E}\left[v_t e_{it}\right] &= \mathbb{E}\left[v_t\right] \cdot \mathbb{E}\left[e_{it}\right] = 0, \\
\mathbb{E}\left[v_t^2\right] &= \operatorname{Var}\left(v_t\right) + \mathbb{E}\left[v_t\right]^2 = \sigma_v^2 + \bar{v}^2, \\
\mathbb{E}\left[e_{it}^2\right] &= \operatorname{Var}\left(e_{it}\right) + \mathbb{E}\left(e_{it}\right)^2 = \operatorname{Var}\left(e_{it}\right) = \sigma_M^2
\end{aligned}$$

Some of the expectations of terms related to p_t can be derived as:

$$\begin{aligned}
\mathbb{E}\left[p_t e_{it}\right] &= \mathbb{E}\left[p_t\right] \cdot \mathbb{E}\left[e_{it}\right] = 0, \\
\mathbb{E}\left[v_t p_t\right] &= \mathbb{E}\left[v_t\left(\theta v_t + (\xi)^{-1} e_{p_t} + (\xi)^{-1}\widetilde{\mu}\right)\right], \\
&= \theta \mathbb{E}\left[v_t^2\right] + 0 + \bar{v}(\xi)^{-1}\widetilde{\mu}, \\
\mathbb{E}\left[p_t^2\right] &= \operatorname{Var}\left(p_t\right) + \mathbb{E}\left[p_t\right]^2 = \theta^2\sigma_v^2 + \xi^{-2}\sigma_z^2 + \left(\theta\bar{v} + \xi^{-1}\widetilde{\mu}\right)^2
\end{aligned}$$

The expected gross profit term can be calculated as:

$$\begin{aligned}\mathbb{E}(gross\ profit) =& \beta_{M\eta}\mathbb{E}\left[v_t^2\right] - \beta_{Mp}\mathbb{E}\left[v_t p_t\right] + \mathbb{E}\left[v_t \mu_M\right] \\ & - \beta_{M\eta}\mathbb{E}\left[p_t v_t\right] + \beta_{Mp}\mathbb{E}\left[p_t^2\right] - \mathbb{E}\left[p_t \mu_M\right] \\ =& \left[\beta_{M\eta} - \theta\left(\beta_{Mp} + \beta_{M\eta}\right)\right] \cdot \left(\sigma_v^2 + \bar{v}^2\right) \\ & + \beta_{Mp} \cdot \left[\theta^2\sigma_v^2 + \xi^{-2}\sigma_z^2 + \left(\theta\bar{v} + \xi^{-1}\widetilde{\mu}\right)^2\right] \\ & - \left(\beta_{Mp} + \beta_{M\eta}\right)\bar{v}\xi^{-1}\widetilde{\mu} + \mu_M\bar{v} - \mu_M\left(\theta\bar{v} + \xi^{-1}\widetilde{\mu}\right).\end{aligned}$$

Then we calculate the expected transaction cost term.

$$\begin{aligned}\mathbb{E}\left(x_{it}^{k_M}\right)^2 =& \mathbb{E}\left[\beta_{M\eta}^2\left(v_t + e_{it}\right)^2\right] + \mathbb{E}\left[\beta_{Mp}^2 p_t^2\right] + \mu_M^2 - 2\beta_{M\eta}\beta_{Mp}\mathbb{E}\left[v_t p_t\right] \\ & + 2\beta_{M\eta}\bar{v}\mu_M - 2\beta_{Mp}\mu_M\mathbb{E}\left[p_t\right] \\ =& \beta_{M\eta}^2\mathbb{E}\left(v_t^2\right) + \beta_{Mp}^2\mathbb{E}\left(p_t^2\right) - 2\beta_{M\eta}\beta_{Mp}\mathbb{E}\left(v_t p_t\right) + \mu_M^2 + \beta_{M\eta}^2\sigma_M^2 \\ & + 2\beta_{M\eta}\bar{v}\mu_M - 2\beta_{Mp}\mu_M\mathbb{E}[p_t].\end{aligned}$$

Then, we can derive:

$$\begin{aligned}\mathbb{E}(transaction\ cost) =& \left[-\frac{1}{2}\rho_M\beta_{M\eta}^2 + \theta\rho_M\beta_{M\eta}\beta_{Mp}\right] \cdot \left(\sigma_v^2 + \bar{v}^2\right) \\ & - \frac{1}{2}\rho_M\beta_{Mp}^2 \cdot \left[\theta^2\sigma_v^2 + \xi^{-2}\sigma_z^2 + \left(\theta\bar{v} + \xi^{-1}\widetilde{\mu}\right)^2\right] \\ & - \frac{1}{2}\rho_M\beta_{M\eta}^2\sigma_M^2 + \rho_M\beta_{M\eta}\beta_{Mp}\bar{v}\xi^{-1}\widetilde{\mu} \\ & + \rho_M\beta_{Mp}\mu_M\left(\theta\bar{v} + \xi^{-1}\widetilde{\mu}\right) - \frac{1}{2}\rho_M\left(\mu_M^2 + 2\beta_{M\eta}\bar{v}\mu_M\right)\end{aligned}$$

Combining the two terms, we obtain the net single-period expected profit for a step-k_M

human:

$$\begin{aligned}\mathbb{E}\left[\pi_{it}^{k_M}\right] =&\left[\beta_{M\eta}-\frac{1}{2}\rho_M\beta_{M\eta}^2+\theta\left(\rho_M\beta_{M\eta}\beta_{Mp}-\beta_{Mp}-\beta_{M\eta}\right)\right]\cdot\left(\sigma_v^2+\bar{v}^2\right)\\&+\left(\beta_{Mp}-\frac{1}{2}\rho_M\beta_{Mp}^2\right)\cdot\left[\theta^2\sigma_v^2+\xi^{-2}\sigma_z^2+\left(\theta\bar{v}+\xi^{-1}\widetilde{\mu}\right)^2\right]\\&-\frac{1}{2}\rho_M\beta_{M\eta}^2\sigma_M^2+\left(\rho_M\beta_{M\eta}\beta_{Mp}-\beta_{Mp}-\beta_{M\eta}\right)\bar{v}\xi^{-1}\widetilde{\mu}\\&+\mu_M\bar{v}+\left(\rho_M\beta_{Mp}-1\right)\mu_M\left(\theta\bar{v}+\xi^{-1}\widetilde{\mu}\right)-\frac{1}{2}\rho_M\left(\mu_M^2+2\beta_{M\eta}\bar{v}\mu_M\right)\end{aligned}$$

Expected profit for AI investor. The expected profit for the step-k_A AI investor is derived as follows.

$$\begin{aligned}\mathbb{E}\left[\pi_{it}^{k_A}\right]&=\mathbb{E}\left[\left(v_t-p_t^{k_{Env}}\right)x_{it}^{k_A}\right]-\frac{1}{2}\rho_A\mathbb{E}\left(x_{it}^{k_A}\right)^2,\\&=\underbrace{\mathbb{E}\left[\left(v_t-p_t^{k_{Env}}\right)\left(-\beta_{Ap}^{k_A}p_t^{k_{Env}}+\mu_A^{k_A}\right)\right]}_{\mathbb{E}(gross\ profit)}\\&\quad\underbrace{-\frac{1}{2}\rho_A\mathbb{E}\left[\left(-\beta_{Ap}^{k_A}p_t^{k_{Env}}+\mu_A^{k_A}\right)^2\right]}_{\mathbb{E}(transaction\ cost)}\end{aligned}$$

For ease of notation, we drop the superscript k_A from the strategy coefficients, $\beta_{Ap}^{k_A}$ and $\mu_A^{k_A}$, and the superscript k_{Env} from the price, $p_t^{k_{Env}}$, and market condition parameters, $\theta^{k_{Env}}$ and $\xi^{k_{Env}}$.

The expected gross profit term can be calculated as:

$$\begin{aligned}\mathbb{E}(gross\ profit)=&-\beta_{Ap}\mathbb{E}\left[v_tp_t\right]+\mathbb{E}\left[v_t\mu_A\right]+\beta_{Ap}\mathbb{E}\left[p_t^2\right]-\mathbb{E}\left[p_t\mu_A\right]\\=&-\theta\beta_{Ap}\left(\sigma_v^2+\bar{v}^2\right)+\beta_{Ap}\left[\theta^2\sigma_v^2+\xi^{-2}\sigma_z^2+\left(\theta\bar{v}+\xi^{-1}\widetilde{\mu}\right)^2\right]\\&-\beta_{Ap}\bar{v}\xi^{-1}\widetilde{\mu}+\mu_A\bar{v}-\mu_A\left(\theta\bar{v}+\xi^{-1}\widetilde{\mu}\right).\end{aligned}$$

Then we calculate the expected transaction cost term.

$$\mathbb{E}\left(x_{it}^{k_A}\right)^2=\mathbb{E}\left[\beta_{Ap}^2p_t^2\right]+\mu_A^2-2\beta_{Ap}\mu_A\mathbb{E}\left[p_t\right].$$

Then, we can derive:

$$\begin{aligned}\mathbb{E}(transaction\ cost) =& -\frac{1}{2}\rho_A\beta_{Ap}^2 \cdot \left[\theta^2\sigma_v^2 + \xi^{-2}\sigma_z^2 + \left(\theta\bar{v} + \xi^{-1}\widetilde{\mu}\right)^2\right] \\ &+ \rho_A\beta_{Ap}\mu_A\left(\theta\bar{v} + \xi^{-1}\widetilde{\mu}\right) - \frac{1}{2}\rho_A\mu_A^2\end{aligned}$$

Combining the two terms, we obtain the net single-period expected profit for the step-k_A AI:

$$\begin{aligned}\mathbb{E}\left[\pi_{At}^{k_A}\right] =& -\theta\beta_{Ap}\left(\sigma_v^2 + \bar{v}^2\right) + \left(\beta_{Ap} - \frac{1}{2}\rho_A\beta_{Ap}^2\right)\left[\theta^2\sigma_v^2 + \xi^{-2}\sigma_z^2 + \left(\theta\bar{v} + \xi^{-1}\widetilde{\mu}\right)^2\right] \\ &- \beta_{Ap}\bar{v}\xi^{-1}\widetilde{\mu} + \mu_A\bar{v} + \left(\rho_A\beta_{Ap} - 1\right)\mu_A\left(\theta\bar{v} + \xi^{-1}\widetilde{\mu}\right) - \frac{1}{2}\rho_A\mu_A^2\end{aligned}$$