A Q-Learning Approach to Algorithmic Market Making for Risky Securities

James Symons-Hicks*

September 20, 2024

Abstract

This paper explores algorithmic pricing in a multi-period market microstructure model using a Q-learning algorithm. We focus on how these algorithms respond to the risk posed by a stochastic asset value. In the baseline Q-learning model, the algorithms behave as if loss averse, leading to quotes only at loss-free prices. We consider a series of Q-learning algorithms to investigate the conditions under which these market makers will trade at competitive levels. Using 'Imperfect Counterfactual Updating', we show that Algorithmic Market Makers set prices close to competitive levels. In the presence of actions that are both profitable and risk-free for market makers, we find that counterfactual updating is sufficient for trade to occur at competitive levels; when these actions are removed, the algorithms require counterfactual updating and additional exploration to ensure trade. We find that this price-setting behaviour is caused by these algorithms wanting to avoid losses, rather than through profit-seeking collusion.

*I am grateful for the support and advice of my supervisor, Antonio Guarino, throughout this dissertation. I acknowledge the use of the UCL Myriad High Performance Computing Facility (Myriad@UCL), and associated support services, in the completion of this work.

1 Introduction

Owing to the ever-increasing demand for speed and efficiency, algorithmic trading has become commonplace in financial markets. These automated trading approaches undoubtedly increase the speed of trading, however, their impact on competition is still debated. Recent literature on algorithmic pricing, in both goods markets and financial markets, has focused on the collusive outcomes and supracompetitive prices of these algorithms. However, many of these studies assume a fixed asset value, in contrast with the stochastic value that is standard in market microstructure literature. As these studies on algorithmic pricing are often used to form expectations about algorithmic pricing in practice, it is crucial that we understand how they respond in different economic environments. In particular, there exists a need to better understand how these algorithms respond to risk.

In this paper, we consider a multi-period market microstructure model, based on the seminal paper by Glosten and Milgrom (1985), but allow Algorithmic Market Makers (AMMs) to set prices according to a Q-learning algorithm. Through updating the expected values attached to each action in a given state, these market makers learn which prices to set by trial-and-error over the course of one million trading "days". These algorithms balance "exploration" (learning through choosing random actions) with "exploitation" (choosing the action that has the highest expected value).

The introduction of informed traders, along with a stochastic asset value, creates risk for market makers. We find that this stochasticity significantly alters the behaviour of the AMMs; in the baseline Q-learning algorithm, the AMMs quotes only at the extremes of the asset values and behave as if loss averse. We consider a series of Q-learning algorithms and adjustments to the economic setup to better understand the root cause of this "loss aversion".

This paper contributes to the literature on algorithmic pricing in financial markets as follows: (1) we consider a traditional market microstructure model, based on the Glosten and Milgrom model, with a stochastic asset value; (2) we adapt the learning process of the Q-learning algorithms to incorporate additional information using counterfactual updating; (3) we give a greater consideration of the setup of intraday trading compared to that of the current literature, allowing for a model could be generalised to an arbitrary number of intraday periods.

We first consider the case of one intraday trading period. In the baseline Q-learning algorithm, AMMs act as if loss averse and trade only at loss-free prices. That is, they set the ask price at the high value of the asset and the bid price at the low value. Simply increasing the probability of exploration is not sufficient for trade to occur at risky, but theoretically prof-

itable, prices. The main adjustment to the baseline algorithm is the introduction of 'Imperfect Counterfactual Updating'. By providing the algorithms with a little extra information, the best market quotes, we can update the values attached to multiple actions simultaneously. We use the fact that, if a trader would buy at a given price, they would also buy at any price below this. Employing this technique leads to more competitive prices; by updating many values simultaneously, the algorithms learn better the true value of an action, even if this was not the action played, and avoid becoming "stuck" on the risk-free actions. This supports the findings of several papers (e.g., Abada et al. (2024), Asker et al. (2024), and Banchio and Skrzypacz (2022)) that increased algorithmic sophistication and counterfactual updating improve pricing competition.

We then describe the setup of the multi-period algorithm. Our interest is in studying how AMMs respond to order flow by allowing them to revise their quotes following a trade in the previous period. We present results for two intraday periods, finding that, in the baseline case, trade occurs only at loss-free prices in both time periods. With counterfactual updating, the prices are closer to the competitive levels and can respond to a trade in the first period as predicted by the theory; i.e., prices increase (decrease) after a buy (sell) in first period.

Overall, we find that, when facing a stochastic asset value, AMMs behave as if loss averse when using the baseline Q-learning algorithm. If there exist profitable risk-free actions, using Imperfect Counterfactual Updating is sufficient for trade to return to the competitive levels; when no such actions exist (i.e., when noise traders will not buy at the extremes of the asset value), trade breaks down, even with counterfactual updating. In the absence of profitable risk-free actions, the algorithms require a combination of counterfactual updating *and* increased exploration for trade to occur at competitive levels.

The remainder of this paper is organised as follows: Section 1.1 surveys the literature; Section 2 presents the economic model; Section 3 discusses the theoretical competitive prices; Section 4 explains the functioning of the algorithmic market makers; Section 5 explains how we simulate the market for one intraday trading period; Section 6 presents the findings from the one-period market; Section 7 explains how we generalise the algorithm to multiple periods; Section 8 presents the results from the multi-period market; Section 9 concludes.

1.1 Related Literature

The focus of this paper is on algorithmic pricing in financial markets; however, algorithmic pricing has also been studied in other contexts. We first discuss the wider economic applications,

before focusing on the application to financial markets. A common finding is that algorithmic pricing leads to collusive outcomes above competitive levels; the cause of this collusion, however, varies across studies.

A notable paper is that by Calvano et al. (2020), who consider a model of repeated Bertrand price competition in an oligopoly market. They find that "algorithms consistently learn to charge supracompetitive prices, without communicating". Using forward-looking algorithms, they find that collusion is sustained by punishment in the case of deviation. Furthermore, the algorithms frequently engage in price-cycles, orbiting around a target due to discretisation.

In a similar model, Asker et al. (2024) find that allowing algorithms to conduct counterfactuals pushes outcomes closer to competitive levels. Asker et al. attribute the supracompetitive pricing to limited information and learning, rather than punishment strategies. In first-price auctions, Banchio and Skrzypacz (2022) find that additional information on the winning bids improves competition. Using a Q-learning algorithm with continous time, Banchio and Mantegazza (2023) find that, without additional information, collusion arises due to "spontaneous coupling" (a situation in which learning and value estimates become correlated). In this paper, we also find that algorithms set more competitive prices when allowed to incorporate additional information.

Waltman and Kaymak (2008) consider Q-learning algorithms in Cournot oligopoly model; while the fully collusive outcome is not reached, profits are higher than the competitive levels. They find that outcomes depend on how the probability of exploration is determined, suggesting that it should depend on past experience. In a similar model, Abada et al. (2024) find that increased exploration mitigates collusion to some extent; when exploration rates decrease too quickly, the algorithms "burn-in a belief that cooperation outperforms competition". However, they find that increased exploration has limited effects compared to increased algorithmic sophistication in mitigating collusive outcomes.

Algorithmic pricing has also been studied in the context of financial markets. Cartea, Chang, Mroczka, and Oomen (2022) consider a model of market-maker spread-setting in Over-the-Counter markets using a multi-armed bandit setup, finding that *less* sophisticated algorithms can lead to collusive outcomes.[1] They consider a setup in which market makers cannot

[1] In particular, Cartea, Chang, Mroczka, and Oomen (2022) study three algorithms: the ϵ-greedy algorithm, the Upper Confidence Bound (UCB) algorithm, and the EXP3 algorithm. They find that, while the EXP3 algorithm converges to the Bertrand-Nash equilibrium, the same result cannot be guaranteed for the other algorithm types.

observe competitors' prices; in contrast, market makers can observe competing quotes in our exchange-based setup. Cartea, Chang, and Penalva (2022) analyse how tick size impacts competition in a setup with payoff uncertainty due to a "latency parameter" that randomly determines which trader is assigned the trade. They find that Q-learning leads to non-competitive outcomes and that a "larger tick size obstructs competition". Dou et al. (2024) consider a setup with a stochastic payoff but study how informed algorithmic speculators place orders when facing perfectly-competitive market makers, finding that these speculators use collusive trading strategies. Using a deep actor-critic algorithm, Xiong and Cont (2022) study how algorithmic spread-setting when market makers compete for market share, finding that collusive outcomes arise. In a similar paper, Cont and Xiong (2024) find that interactions between market makers simply through knowledge of market prices "may give rise to tacit collusion".

One important difference of this paper over much of the literature above is the use of a stochastic asset value. The closest paper in terms of setup is perhaps that of Colliard et al. (2022), which considers a binary asset value and sets (ask) prices using competing algorithmic market makers.[2] They find that the algorithms set prices considerably above the competitive levels, despite there existing profitable deviations. They suggest that these findings are due to the additional noise in the environment from the stochastic payoffs, and that, to compensate, these algorithms should increase experimentation. In their baseline model, the prices in the final trading day are distributed almost entirely above the maximum value of the asset. We find a similar result but show that this can be attributed to AMMs behaving as if loss averse. Our paper differs from the work of Colliard et al. (2022) in three ways: (1) we consider a more traditional market microstructure model that distinguishes between informed and noise traders; (2) we create a multi-period version of the model that can be generalised beyond two periods; (3) we consider the impact of counterfactual updating on pricing.

2 The Model

We consider a sequential trade model similar to that of Easley et al. (1997) and Cipriani and Guarino (2014), based on the seminal paper of Glosten and Milgrom (1985). A single risky asset is traded in a specialist market over multiple days by both informed and noise traders. The main novelty of this paper compared to these models is that the specialist (also known as

[2]Colliard et al. (2022) consider only the ask-side of the market. Thus, market makers set only ask prices and traders can either buy or not trade.

a market maker) sets ask and bid prices algorithmically, using a Q-learning algorithm.

2.1 The Asset

There is one risky asset which takes on a stochastic fundamental value, denoted by $\tilde{v}^d$, in day d. The value of the asset is realised for each new day and does not change during the trading day. With probability δ, $\tilde{v}^d$ is equal to v_H; with probability $1-\delta$, $\tilde{v}^d$ is equal to v_L, where $v_H > v_L$. At the end of the trading day, the value of the asset is made known to all market participants.

2.2 The Market

The asset is traded over D trading "days" by Algorithmic Market Makers (AMMs) who interact with a sequence of traders. We define a trading day, $d = 1, ..., D$, by the new realisation of the asset value. Each trading day consists of T trading "times", each of which is long enough to accommodate exactly one trade. A trading "time" may also be referred to as a trading "period". A trader is chosen at random to random to participate at each trading time and can choose whether to buy or sell one unit of the asset, or to not trade.

2.3 The Market Maker

In the existing literature, market makers operate in a perfectly competitive market, setting quotes such that they earn zero expected profit. Furthermore, they update expectations in a Bayesian manner. In our setup, market makers set prices using a Q-learning algorithm, which uses trial and error to learn the value from setting each possible price in a given state. A description of how AMMs choose prices is provided in Section 4.2.

We consider a model with N AMMs, indexed by $i = 1, ..., N$. At time t in day d, AMM_i submits a pair of quotes (ask and bid prices). Let us denote the ask price (the price at which a trader can buy) and bid price (the price at which a trader can sell) of AMM_i at time t of day d by $a^d_{i,t}$ and $b^d_{i,t}$, respectively. Based on the quotes submitted by the N market makers, the best ask and bid prices are defined as follows:

$$a^d_t = \min\{a^d_{1,t}, ..., a^d_{N,t}\} \quad \text{and} \quad b^d_t = \max\{b^d_{1,t}, ..., b^d_{N,t}\}. \tag{1}$$

If a trader does trade, they do so at the best price, i.e., at either a^d_t or b^d_t. Based on a^d_t and b^d_t, the trader will decide whether to trade or not. We denote the action of the trader at time t on day d by X^d_t, where $X^d_t \in \{1, 0, -1\}$, which corresponds to $\{\textit{Buy, No Trade, Sell}\}$, respectively.

In the case that there are multiple AMMs quoting the best price, one is randomly chosen.[3] We denote the number of AMMs offering the best ask and bid prices by Z_{ask} and Z_{bid}, respectively; the probability that a dealer offering the best price is assigned the trade is $\frac{1}{Z_{ask}}$ and $\frac{1}{Z_{bid}}$ for the ask and bid offers, respectively. We denote which AMMs have the prevailing ask and bid quotes as a result of the prices set (and potentially random selection for $Z > 1$) by $\mathcal{N}^*_{ask}$ and $\mathcal{N}^*_{bid}$. We then compute the volume of trade by AMM$_i$, $\Omega^d_{i,t}$, as follows:

$$\Omega^d_{i,t} = \begin{cases} 1 & \text{if } X^d_t = 1 \text{ and } i = \mathcal{N}^*_{ask}, \\ -1 & \text{if } X^d_t = -1 \text{ and } i = \mathcal{N}^*_{bid}, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The final profit on day d for AMM$_i$ is given by

$$\pi^d_i = \sum_{t=1}^{T} \left((a^d_{i,t} - \tilde{v}^d)\mathbf{1}_{\{\Omega^d_{i,t}=1\}} + (\tilde{v}^d - b^d_{i,t})\mathbf{1}_{\{\Omega^d_{i,t}=-1\}} \right), \tag{3}$$

where $\mathbf{1}_{\{\Omega^d_{i,t}=1\}}$ is an indicator function. The market makers aim to maximise their profit on a given day, without any regard for profits in the following days; that is, they are myopic.

2.4 The Traders

The asset is traded by both informed and noise traders. At each trading time, a trader can decide whether to buy, sell, or not to trade; each trader can trade one unit of the asset with the market maker. At each time t, a trader is informed with probability μ and has perfect information about the asset value; that is, they know the realisation of $\tilde{v}^d$. For the informed traders, they buy if $a^d_t \leq v_H$, they sell if $b^d_t \geq v_L$, and do not trade otherwise.[4]

With probability $1 - \mu$, a trader is a noise trader. These traders trade for liquidity reasons, rather than because they are informed of the fundamental value. In the baseline model, we consider the case of price-inelastic noise traders; they buy the asset with probability $\frac{\eta}{2}$, sell the asset with probability $\frac{\eta}{2}$, and do not trade with probability $1 - \eta$, irrespective of price.

[3] This can be thought of as (unmodeled) time competition where all market makers are equally fast and it is random as to which one is first.

[4] In the model, we constrain the trader to trading only one unit of the asset. However, as we do not necessarily constrain the algorithms to set $a^d_t > b^d_t$, a situation may arise in which it is profitable for the trader to both buy *and* sell the asset. When this is the case, the informed trader chooses the action that earns the highest profit; in the event of a tie between the profits from buying and selling, the trader will buy.

3 Theoretical Predictions for Competitive Market Makers

Before discussing the algorithmic price-setting, it is worth discussing the predictions of the theoretical model in the case of competitive market makers.

A market maker faces an adverse selection problem since a trader deciding to trade may be doing so because they know the value of the asset. In a perfectly competitive market, price competition between dealers should drive the prices to a point where they earn zero expected profits; that is, they set the quotes equal to the expected value of the asset *conditional* on their being trade. That is,

$$a_t^d = \mathbb{E}[\tilde{v}^d | h_t^d, X_t^d = 1], \tag{4}$$

$$b_t^d = \mathbb{E}[\tilde{v}^d | h_t^d, X_t^d = -1], \tag{5}$$

where $\mathbb{E}[\cdot]$ denotes the expectation, and h_t^d is the history, at time t, of trades up to and including time $t-1$.

The theoretical prices are

$$a_1^d = \frac{(1-\mu)\eta\mathbb{E}[\tilde{v}^d] + \mu v_H}{\mu + (1-\mu)\eta}, \tag{6}$$

$$b_1^d = \frac{(1-\mu)\eta\mathbb{E}[\tilde{v}^d] + \mu v_L}{\mu + (1-\mu)\eta}, \tag{7}$$

Appendix B.1 contains the derivations of these results. Therefore, as $v_H > v_L$, it follows that $a_1^d > b_1^d$, and the spread is equal to

$$a_1^d - b_1^d = \frac{\mu(v_H - v_L)}{\mu + (1-\mu)\eta} > 0. \tag{8}$$

In extending the quotes to time $t > 1$, one must account for the probability of observing the history of trades, h_t^d. For a given history, h_t^d, the ask quote is given by

$$a_t^d = \frac{v_L \Pr(h_t^d|v_L)\Pr(X_t^d = 1|v_L) + v_H \Pr(h_t^d|v_H)\Pr(X_t^d = 1|v_H)}{\Pr(h_t^d|v_L)\Pr(X_t^d = 1|v_L) + \Pr(h_t^d|v_H)\Pr(X_t^d = 1|v_H)}, \tag{9}$$

and the bid quote is given by

$$b_t^d = \frac{v_L \Pr(h_t^d|v_L)\Pr(X_t^d = -1|v_L) + v_H \Pr(h_t^d|v_H)\Pr(X_t^d = -1|v_H)}{\Pr(h_t^d|v_L)\Pr(X_t^d = -1|v_L) + \Pr(h_t^d|v_H)\Pr(X_t^d = -1|v_H)}. \tag{10}$$

These quotes are derived in Appendix B.1.2. In particular, the competitive quotes at $t = 2$

following a buy at time $t = 1$ are

$$a_2^d|(h_2^d = 1) = \frac{v_L[(1-\mu)\frac{\eta}{2}]^2 + v_H[\mu + (1-\mu)\frac{\eta}{2}]^2}{[(1-\mu)\frac{\eta}{2}]^2 + [\mu + (1-\mu)\frac{\eta}{2}]^2}, \tag{11}$$

$$b_2^d|(h_2^d = 1) = \frac{v_L + v_H}{2} = \mathbb{E}[\tilde{v}^d], \tag{12}$$

and following a sell at $t = 1$, the competitive quotes are

$$a_2^d|(h_2^d = -1) = \frac{v_L + v_H}{2} = \mathbb{E}[\tilde{v}^d], \tag{13}$$

$$b_2^d|(h_2^d = -1) = \frac{v_L[\mu + (1-\mu)\frac{\eta}{2}]^2 + v_H[(1-\mu)\frac{\eta}{2}]^2}{[\mu + (1-\mu)\frac{\eta}{2}]^2 + [(1-\mu)\frac{\eta}{2}]^2}. \tag{14}$$

If there are two opposite orders during the two periods (e.g., a buy followed by a sell), the expected value is equal to the unconditional expectation of the asset value. The two orders offset one another and are, therefore, uninformative to the market maker.

In the case of a no-trade at time $t = 1$, the quotes are not revised as this can only have come from a noise trader, who is uninformed of the fundamental value. Thus, the market maker does not revise expectations.

4 The Algorithm

In this section, we describe the functioning of the algorithm used by the AMMs. Section 4.1 provides a brief background on Q-learning; Section 4.2 describes the algorithm in our setup; Section 4.3 explains 'Imperfect Counterfactual Updating'.

4.1 A Description of Q-Learning

The AMMs set ask and bid prices according to a Q-learning algorithm (Watkins, 1989), an example of an off-policy reinforcement learning algorithm.[5] Other than the agents and environment, any reinforcement learning algorithm consists of at least three elements: (1) a value function which specifies the expected value of being in the current state, including immediate and future rewards; (2) a reward signal determined by the environment and the actions taken;

[5]The term "off-policy" refers to the fact that the target policy (the policy used as the *optimal* decision-making rule) differs from the behaviour policy (the policy that the algorithm uses to *interact* with its environment). The target policy is *greedy* (i.e., it selects the action with the highest Q-value), but the behaviour policy is ϵ-*greedy* (i.e., it explores with probability ϵ and exploits the greedy action with the complementary probability).

(3) a policy that defines how agents behave (Sutton and Barto, 2018).

At any point in time, agents observe a state variable, $s_t \in \mathcal{S}$, and choose an action $a_t \in \mathcal{A}$; both $\mathcal{S}$ and $\mathcal{A}$ are assumed to be finite. This constitutes a 'state-action' pair (s_t, a_t). A Q-function, $Q(s, a)$, is a mapping from a state-action pair (s, a) to the Q-value. A Q-value, denoted by $q(s_t, a_t)$, is an expected value of choosing action a_t in state s_t, accounting for both the expected immediate reward and the continuation value from being in that state. The value function can be obtained by choosing the action a that, for a given state s_t, corresponds to the maximum Q-value. That is,

$$V(s_t) = \max_{a \in \mathcal{A}} Q(s_t, a). \tag{15}$$

Upon choosing an action a_t in state s_t, the algorithm earns an immediate reward, π_t. The aim of the algorithm is to maximise its expected sum of discounted future rewards,

$$\mathbb{E}\left[\sum_{t=1}^{T} \gamma^{t-1} \pi_t\right], \tag{16}$$

where γ is the discount rate.[6]

An important feature of Q-learning is the trade-off between *exploration* and *exploitation*. At a given point in time, there is, for each state, an action corresponding to the highest Q-value. This is the "greedy" action. If the algorithm selects the greedy action, it is "exploiting [its] current knowledge of the values of the actions" (Sutton and Barto, 2018). Alternatively, the algorithm can choose to *explore*, that is, choose an action at random. Q-learning uses the ϵ-greedy behaviour policy; with probability ϵ, the algorithm explores, and with probability $1-\epsilon$, the algorithm exploits the greedy action.

The learning rule determines how the algorithm learns and updates its Q-value; in its most general form, the Q-learning rule is

$$Q_{n+1}(s, a) = (1 - \alpha) Q_n(s, a) + \alpha \left[\pi_t + \gamma \max_{a' \in \mathcal{A}} Q_n(s', a')\right], \tag{17}$$

where α is the learning rate, and s' corresponds to the state in the following period. The updated Q-value, $Q_{n+1}(s, a)$, is a weighted average of the current Q-value, $Q_n(s, a)$, and the reward for

[6] In our setup of the algorithm, the time dimension is slightly different. We assume that the AMM seeks to maximise the profit only in the current day, without any regard for profits in the following days. It is the intraday time periods that become the sequence of profits to be maximised; these are not discounted. Thus, for each day, in our setup, the algorithm seeks to maximise $\mathbb{E}\left[\sum_{t=1}^{T} \pi_t^d\right]$.

being in that state (the immediate reward plus a discounted future value). Q-learning uses a "Q-table" to store the value of each action-state pair; Table 1 shows an example of a Q-table where $\mathcal{S} = \{s^1, s^2, s^3\}$ and $\mathcal{A} = \{a^1, a^2, a^3\}$.

Table 1: Example 3×3 Q-table

State	**Action**		
	a^1	a^2	a^3
s^1	$q(s^1, a^1)$	$q(s^2, a^1)$	$q(s^3, a^1)$
s^2	$q(s^1, a^2)$	$q(s^2, a^2)$	$q(s^3, a^2)$
s^3	$q(s^1, a^3)$	$q(s^2, a^3)$	$q(s^3, a^3)$

4.2 Algorithmic Market Makers for T = 1

We now explain how we use Q-learning for algorithmic market making in the case of $T = 1$; we refer to this as the "one-period market".

4.2.1 Action Space

The action space is the set of prices that AMMs can charge. Given the tabular nature of the Q-learning algorithm, we discretise the action space into M discrete prices. It is not necessary to have the bid and ask prices sharing the same action space; we denote the action set for ask prices by $\mathcal{A}_{ask}$ and bid prices by $\mathcal{A}_{bid}$. The M prices are spaced evenly on the intervals $[a_{min}, a_{max}]$ and $[b_{min}, b_{max}]$ for the ask-side and bid-side, respectively. The price interval is set such that $\mathbb{E}[\tilde{v}^d]$, the theoretical competitive price, and either v_H or v_L (depending on ask or bid) are contained within the interior of the interval, allowing for quotes above and below these prices.

4.2.2 State Space

Similarly to Calvano et al. (2020), we define the states using the prices offered by market makers in the previous day. We use exclusively the *best* bid and ask prices to define the states.[7] Therefore, the state space is the set of all possible *best* prices that can occur. In the one-period market, the state spaces for the ask- and bid-side are equal to $\mathcal{A}_{ask}$ and $\mathcal{A}_{bid}$, respectively.

[7] Calvano et al. (2020) define their states as $s_t = \{\mathbf{p}_{t-1}, ..., \mathbf{p}_{t-k}\}$, where $\mathbf{p}$ is a vector of prices offered by all players and k is the length of the "bounded memory" of the algorithm. To avoid excessively large state spaces, they also set $k = 1$ in their baseline model.

In defining the states in this manner, the size of the state space is independent of N, yet retains its informativeness. As trade occurs only at the best quote, a market maker earns zero profit for any quote less competitive than this; therefore, it is the best quote that is informative of profit, rather than the entire list of quotes. We define the states for the one-period market as

$$S_1^{d,Ask} = a_1^{d-1}, \tag{18}$$

$$S_1^{d,Bid} = b_1^{d-1}. \tag{19}$$

That is, the ask-side (bid-side) state is the best ask (bid) quote in the previous day.

4.2.3 Actions, Feedback, and Learning

For each AMM, we generate two initial Q-tables: one for each the ask- and bid-side. Let us denote the ask and bid Q-tables for AMM_i at time $t = 1$ of day d by $\mathbf{Q}_{i,1}^{d,Ask} \in \mathbb{R}^{\mathbb{S}\times M}$ and $\mathbf{Q}_{i,1}^{d,Bid} \in \mathbb{R}^{\mathbb{S}\times M}$, where $\mathbb{S}$ is the size of the state space and M is the size of the action space. An element of these matrices, a single Q-value, is denoted by $q_{i,1}^{d,Ask}(S_1^{d,Ask}, a_m)$ for the ask-side and $q_{i,1}^{d,Bid}(S_1^{d,Bid}, b_m)$ for the bid-side. These are the Q-values that AMM_i attaches to price m in state S_1^d at time $t = 1$ of day d. The vector of Q-values for all M actions in state $S_1^{d,Ask}$ is denoted by

$$\mathbf{q}_{i,1}^{d,Ask}(S_1^{d,Ask}) \in \mathbb{R}^{1\times M}, \tag{20}$$

and for the bid-side state, $S_1^{d,Bid}$, the corresponding vector is

$$\mathbf{q}_{i,1}^{d,Bid}(S_1^{d,Bid}) \in \mathbb{R}^{1\times M}. \tag{21}$$

For the initial Q-tables, each Q-value is drawn from a uniform distribution over the interval $[\underline{q}, \bar{q}]$, where $\underline{q}$ and $\bar{q}$ are set such that they are above the maximum possible reward in each state. Setting the initial Q-values above the maximum reward induces exploration (Asker et al., 2024). The actions are perceived to have a higher expected reward and so will be tried more as the greedy action; as the true reward is learnt, values are revised downwards towards the final value. This additional exploration occurs, at some point, for all actions.

As explained in Section 4.1, the Q-learning algorithm uses the ϵ-greedy behaviour policy. When they exploit, they choose the ask and bid prices that correspond to the maximum values in $\boldsymbol{q}_{i,1}^{d,Ask}(S_1^{d,Ask})$ and $\boldsymbol{q}_{i,1}^{d,Bid}(S_1^{d,Bid})$, respectively. That is, given the ask state $S_1^{d,Ask}$ and the

bid state $S_1^{d,Bid}$, the greedy ask and bid prices at time $t = 1$ are given by

$$\bar{a}_{i,1}^d = \arg\max_m \boldsymbol{q}_{i,1}^{d,Ask}(S_1^{d,Ask}), \tag{22}$$

$$\bar{b}_{i,1}^d = \arg\max_m \boldsymbol{q}_{i,1}^{d,Bid}(S_1^{d,Bid}). \tag{23}$$

The other option is to explore. In this case, they randomly select one ask from $\mathcal{A}_{ask}$ and one bid from $\mathcal{A}_{bid}$; each price has an equal probability of being chosen. The AMMs explore with probability

$$\epsilon = \exp(-\beta d), \tag{24}$$

where $\beta > 0$ determines the rate at which this probability declines. With the complementary probability, the AMM will exploit their greedy action.

Once the AMMs submit the quotes, the trader decides whether to trade; the trading volume of each market maker, $\Omega_{i,1}^d$, is determined as discussed in Section 2.3. We compute the ask- and bid-side profits separately for each Q-table, rather than computing a joint profit; in doing so, we better reflect the profit on each side of the market, given that the ask-side and bid-side Q-values are separated. The profits are

$$\pi_i^{d,Ask} = (a_{i,1}^d - \tilde{v}^d)\mathbf{1}_{\{\Omega_{i,1}^d=1\}}, \tag{25}$$

$$\pi_i^{d,Bid} = (\tilde{v}^d - b_{i,1}^d)\mathbf{1}_{\{\Omega_{i,1}^d=-1\}}. \tag{26}$$

For example, if the trader bought from another AMM, the trader *sold* to AMM$_i$ ($\Omega_{i,1}^d = -1$), or there was no trade at all, then the ask-side profit would be zero.

We now define the updating rule for $T = 1$. For all actions $m = 1, ..., M$, we update the corresponding ask Q-value for AMM$_i$, $q_{i,1}^{d,Ask}(S_1^{d,Ask}, a_m)$, as

$$q_{i,1}^{d,Ask}(S_1^{d,Ask}, a_m) = \begin{cases} \alpha\pi_i^{d,Ask} + (1-\alpha)q_{i,1}^{d,Ask}(S_1^{d,Ask}, a_m) & \text{if } a_m = a_{i,1}^d, \\ q_{i,1}^{d,Ask}(S_1^{d,Ask}, a_m) & \text{if } a_m \neq a_{i,1}^d, \end{cases} \tag{27}$$

and the bid Q-value, $q_{i,1}^{d,Bid}(S_1^{d,Bid}, b_m)$, as

$$q_{i,1}^{d,Bid}(S_1^{d,Bid}, b_m) = \begin{cases} \alpha\pi_i^{d,Bid} + (1-\alpha)q_{i,1}^{d,Bid}(S_1^{d,Bid}, b_m) & \text{if } b_m = b_{i,1}^d, \\ q_{i,1}^{d,Bid}(S_1^{d,Bid}, b_m) & \text{if } b_m \neq b_{i,1}^d. \end{cases} \tag{28}$$

That is, if action m was played (either chosen randomly or as the greedy action), we compute

a weighted average of the profit earned and the current Q-value; if action m is not played, the Q-value is unchanged.

4.3 Imperfect Counterfactual Updating

We also consider an alternative to the standard updating method, which significantly increases the amount of learning with little additional information. If a trader was willing to buy (sell) the asset at a given price, they would also be willing to buy (sell) at any price below (above) this. So, a market maker that sells to a trader at price 100 knows that the trader would have still bought at price 99. For another market maker observing the trade, they know that they could have undercut their competitor at price 99 but would not have sold for any price above 100. We allow AMMs to use this information to conduct counterfactuals. Consider that a trader does not buy at the best price: for any price below the best, the market maker cannot be sure whether trade would have occurred; thus, these prices cannot be updated. We refer to this as 'Imperfect Counterfactual Updating' (ICU).[8]

Let us consider the case for the ask prices.[9] We denote a counterfactual ask price corresponding to the mth action in $\mathcal{A}_{ask}$ by $\hat{a}_m$. There are three potential levels of ICU, depending on the outcome in a period:

1. **No Trade**. This is the case with the least counterfactual updating. The trader did not buy at a_t^d, and so, they would also not have bought at any price above this. Thus, we update all counterfactual asks $\hat{a}_m \geq a_t^d$ using a counterfactual profit of zero. We do not know, however, whether they would have bought at a price below this; hence, we cannot update these actions.

2. **AMM$_\mathbf{i}$ sells the asset**. We compute the *realised* profit from AMM$_i$ as before. They do not know if the trader would have bought at a price *above* a_t^d, and so cannot update these. However, for any price *below* a_t^d, the trader would have still bought the asset; we can compute a counterfactual profit for all $\hat{a}_m < a_t^d$.

3. **AMM$_\mathbf{j}$ sells the asset**. If AMM$_j$ ($j \neq i$) sells, we can update all actions. For any price strictly *above* a_t^d, they would earn zero profit. For any price strictly *below* a_t^d, AMM$_i$ would

[8] Asker et al. (2024) employ a similar method using the fact that the demand curve slopes downwards, calling this "synchronous updating using downward demand". This paper extends this to both supply and demand, given that the market makers submit bid and ask prices.

[9] The logic is the same for the bid prices but is instead reversed, using the fact that, if a trader would sell at price b, they would also sell at any price *above* b.

have undercut AMM$_j$ and won the trade, earning a profit of $\hat{a}_m - \tilde{v}^d$. Setting $\hat{a}_m = a_t^d$, means that, due to competition, they would have earned $\frac{1}{Z+1}(a_t^d - \tilde{v}^d)$ in expectation, where Z is the number of market makers that already offered the best price. Thus, we can update all actions in this case.

Models of learning have also been studied as part of the behavioural game-theoretic literature. Camerer and Ho (1999) propose a model of learning called '*Experience-Weighted Attraction*' learning, in which players update experienced actions with a weight of one and counterfactual computations with a weight of $\rho < 1$. We impose a similar mechanism for algorithmic updating. The weighted payoff for a given ask quote is given by

$$(\rho + (1-\rho)\mathbf{1}_{\{a_m = a_{i,t}^d\}})\pi_i^d(a_m, \mathbf{a}_{-i}, \tilde{v}^d). \tag{29}$$

That is, the payoff for each a_m, given the quotes set by the other AMMs $\mathbf{a}_{-i}$ is discounted by $\rho < 1$ if this is a counterfactual action. $\mathbf{1}_{\{a_{m,i} = a_{d,t}\}}$ is an indicator function that is equal to one if ask m is the quote that was chosen by the AMM at time t of day d. Thus, in this case, the payoff, $\pi_i(a_m, \mathbf{a}_{-i}, \tilde{v}^d)$, has a weight of one.

5 Simulating the One-Period Market

This section describes the simulation of the one-period market, including parameterisation. Algorithm 1 provides a pseudo-code for the simulation; Figure C1 (Appendix C) provides a graphical representation.

Algorithm 1 One-Period Market Simulation

```
Initialise parameters and generate action and state spaces.
for k = 1, ..., K do
    Randomly initialise bid and ask Q-tables for all market makers.
    Set random initial state for bid and ask.
    for d = 1, ..., D do
        Update states as {b_t^{d-1}, a_t^{d-1}}.
        Realise ṽ^d and trader type.
        Determine exploit/explore and bid and ask prices.
        Traders decide whether to trade using {b_t^d, a_t^d}.
        Compute profits for each market maker.
        Update bid and ask Q-tables for each market maker.
    end for
end for
```

The changing actions of other AMMs, together with the stochastic asset value, mean that a stationary environment is not achieved; thus, there is no theoretical guarantee of convergence

with the Q-learning algorithm. We instead run many independent iterations of the the market simulation to generate a clear representation of the market. We run K separate iterations; for each, we re-initialise the Q-values and repeat steps 2-6 below. We set $K = 200$ to allow for a sufficiently large sample.

Step 1 - Setting Parameters

We set the parameters of the baseline simulation as in Table 2.

Table 2: Values for Baseline Parameterisation

Parameter	Value	Description
α	0.10	Learning rate
β	0.00004	Exploration rate
ρ	0.50	Weight on counterfactual profits
μ	0.30	Probability of informed traders
η	1	Probability of a noise trader buying/selling
K	200	Number of experiments
D	1,000,000	Number of trading days
N	2	Number of AMMs
v_H	102	High value of the asset
v_L	98	Low value of the asset
δ	0.5	$\Pr(\tilde{v}^d = v_H)$

We set α such that it is in line with the Q-learning computer science literature (Calvano et al., 2020); β is set to balance greater exploration with the extra noise that it induces and the "costs of experimentation", i.e., the opportunity cost of not choosing the greedy action.[10] One can choose the initial β value by considering the number of times that the price is expected to be tried from random exploration.[11] With $\beta = 0.00004$, we expect to explore 25,000 times, which, given $M = 71$, is 352 times per action.

Typical market microstructure models consider $\eta < 1$ so that the probability of a no-trade by a noise trader, $1 - \eta$, is strictly positive to account for days in which trade volume is lower.

[10] In particular, "if one algorithm experiments more extensively, this creates noise in the environment, which makes it harder for other to learn" (Calvano et al., 2020). Additionally, experimentation is typically viewed as costly when the algorithm is rolled out in an "online" setting, where it can learn from the live environment, potentially leading to losses from exploring.

[11] Given the setup of the exploration probability in Equation 24, the number of times the algorithm is expected to explore is finite. As $D \to \infty$, the number of times that the algorithm is expected to explore is

$$\sum_{d=1}^{\infty} e^{-\beta d} = \frac{e^{-\beta}}{1 - e^{-\beta}}.$$

In our simulations, the no-trade option is not of particular interest; thus, we set $\eta = 1$ so that the probability that a noise trader buys or sells is $\frac{1}{2}$.

We set the number of trading days, D, equal to one million to allow for sufficient time for learning. Both asset values are equally likely ($\delta = \frac{1}{2}$). We set $v_H = 102$ and $v_L = 98$; the unconditional expectation is $\mathbb{E}[\tilde{v}^d] = 100$.

Using the parameterisation in Table 2 and the expressions for the theoretical prices in Equations 6 and 7, we find that the competitive prices are 100.6 and 99.4 for the ask and bid prices, respectively. Thus, for the baseline model, we set $\{a_{min}, a_{max}\} = \{99.5, 103.0\}$ and $\{b_{min}, b_{max}\} = \{97.0, 100.5\}$. For each interval, there are $M = 71$ evenly-spaced discrete prices, corresponding to a tick size of 0.05.

Step 2 - Generating Q-Tables

With $M = 71$, there are 71 actions and states; we generate an initial Q-table of size 71×71 for both asks and bids, with Q-values drawn uniformly from [5,8].[12] For $d = 1$, the initial states are drawn randomly from the set of states.

Step 3 - Nature Results

'Nature' determines whether the value of the asset is high or low, and whether the trader is informed or noise. If they are a noise trader, it also determines what action they take.

Step 4 - Choosing Prices

We determine the greedy ask and bid quotes for each AMM using the Q-values in the current state. Based on the probability of exploration, each AMM will either explore or exploit the greedy actions. If exploring, we randomise uniformly over the action set.

Step 5 - Computing Trading Volume and Profit

The trader will decide whether to trade; if they trade, they do so at the best quotes. The AMM that submitted this quote (or, in the case that multiple AMMs submit the same quote, as determined by randomisation) is assigned the trade; all others have a trading volume of zero for that day. We compute the profit for each market maker as in Equation 3.

Step 6 - Updating Q-Tables

Using the profits computed in step 5, we update the Q-table following the updating rule in Equations 27 and 28 . This is the end of the day; for all $d < D$, we repeat steps 2-6.

[12] These values are above the maximum possible payoff given the values of v_L,v_H, and the ranges of the action sets. The maximum payoff would occur when the value is low (high) and the ask (bid) is 103 (97); this would yield a reward of 5.

6 Results for One-period Market (T = 1)

In this section, we present a series of results from the one-period market simulation. We aim to uncover the causes of the prices set by the AMMs by considering how different algorithmic specifications and adjustments to the economic setup impact pricing outcomes.

The results show that the baseline Q-learning algorithm behaves as if loss averse and requires counterfactual learning to price competitively. In the presence of a profitable risk-free action (i.e., a price at which trade occurs and the market maker cannot lose), counterfactual updating is sufficient; when this is not available, continued exploration is needed to augment the counterfactual updating.

With discrete prices, multiple Nash equilibria exist. We find that ask prices of 100.6, 100.65, and 100.7 and bid prices of 99.3, 99.35, and 99.4 are Bertrand-Nash equilibria. These prices are derived in Section B.2.

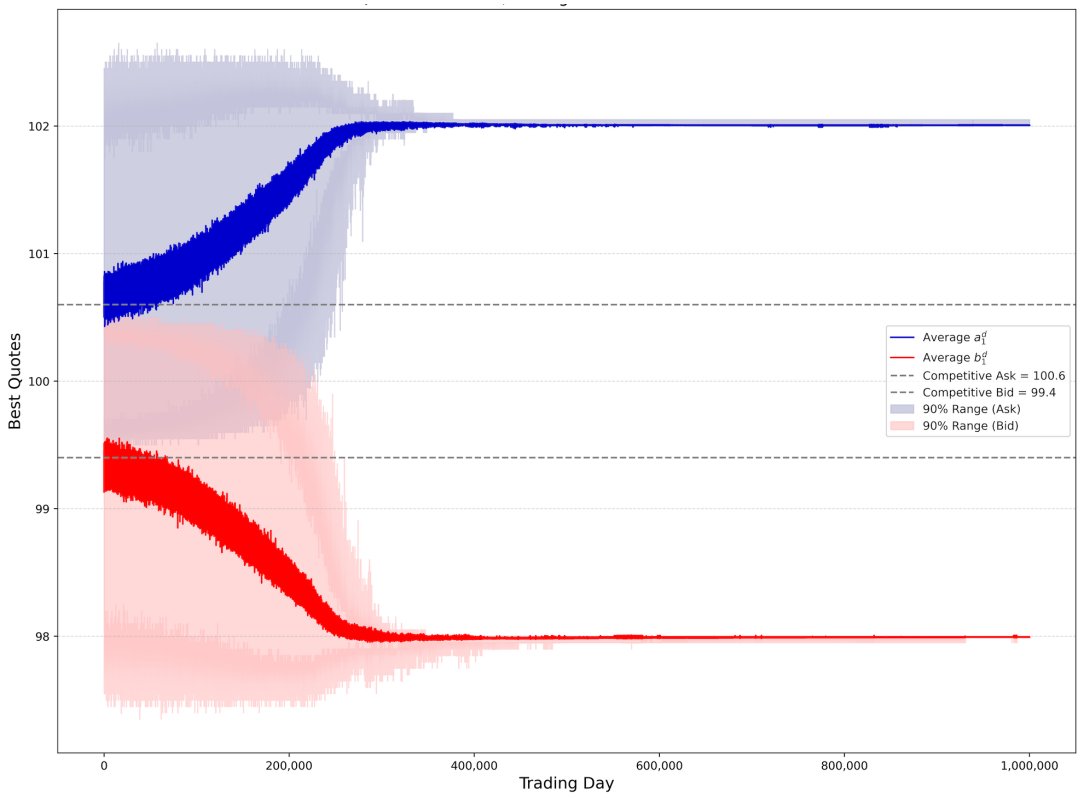
6.1 Baseline Setup

We start with the baseline setup without Imperfect Counterfactual Updating (ICU) and use the parameterisation in Table 2.

RESULT 1 *In the baseline setup, the prices converge to actions that are loss-free for market makers. That is, the ask price settles at v_H and the bid price settles at v_L.*

Figure 6.1 shows how the mean of the *best* ask and bid quotes across iterations change over the one million days. The average spread begins to widen and the ask (bid) tends towards v_H (v_L); this does not change significantly after approximately 300,000 days. At these prices, the market makers cannot lose from trade; an ask price of 102 earns zero profit if trading with an informed trader and an expected profit of 2 if trading with a noise trader. The 90% interquartile ranges also shrink over time; initially, it covers most of the action set, however, after around 300,000 days, the range is near zero. In the final trading day, we find that, in 189 of the 200 iterations, AMMs submit a best ask of 102 and, in 190 iterations, they submit a best bid of 98; the remaining asks are above 102 (v_H) and the remaining bids are below 98 (v_L). This implies that each iteration converged to loss-free actions.

The distributions of the best quotes change over the one million days (Figure 6.2). During the early stages, the distribution is near uniform due to random exploration. However, AMMs quickly reach quotes at the extreme values (102 for asks and 98 for bids); once these values are



Parameters: $\alpha = 0.1, \beta = 4e-05, \mu = 0.3$. Final Mean Ask: 102.01. Final Mean Bid: 97.99.

Figure 6.1: Baseline Setup - Average Best Ask and Bid Quotes in Each Day

reached and the probability of exploration is low, the algorithms rarely quote in the interval $(98, 102)$.

RESULT 2 *By the final day, AMMs attach, on average, a negative Q-value to any ask price below v_H and any bid price above v_L.*

Despite action-state pairs having theoretically positive expected values, many of the Q-values are instead negative. These values are *near* zero, but are just negative (Figure 6.3). In State 102.0, on average, an AMM attaches a Q-value of -0.003 to price 101.95; in contrast, assuming that the other AMM chooses an ask of 102, the expected profit is 1.35.

To better understand how AMMs respond to being in each state, we compute the greedy quotes in the final day.Figure 6.4 shows the median greedy ask and bid prices in a subset of states. In States 101.8-101.95, the median greedy price is strictly greater than the state value; i.e., in State 101.85, the median greedy price is 102. Assuming that AMM_2 chooses the price corresponding to the state, then AMM_1 is not compete as they charge a strictly higher price. On the bid-side, in States 98.05-98.2, the median greedy prices are either 98 or 97.95, which also mean they are not competing.

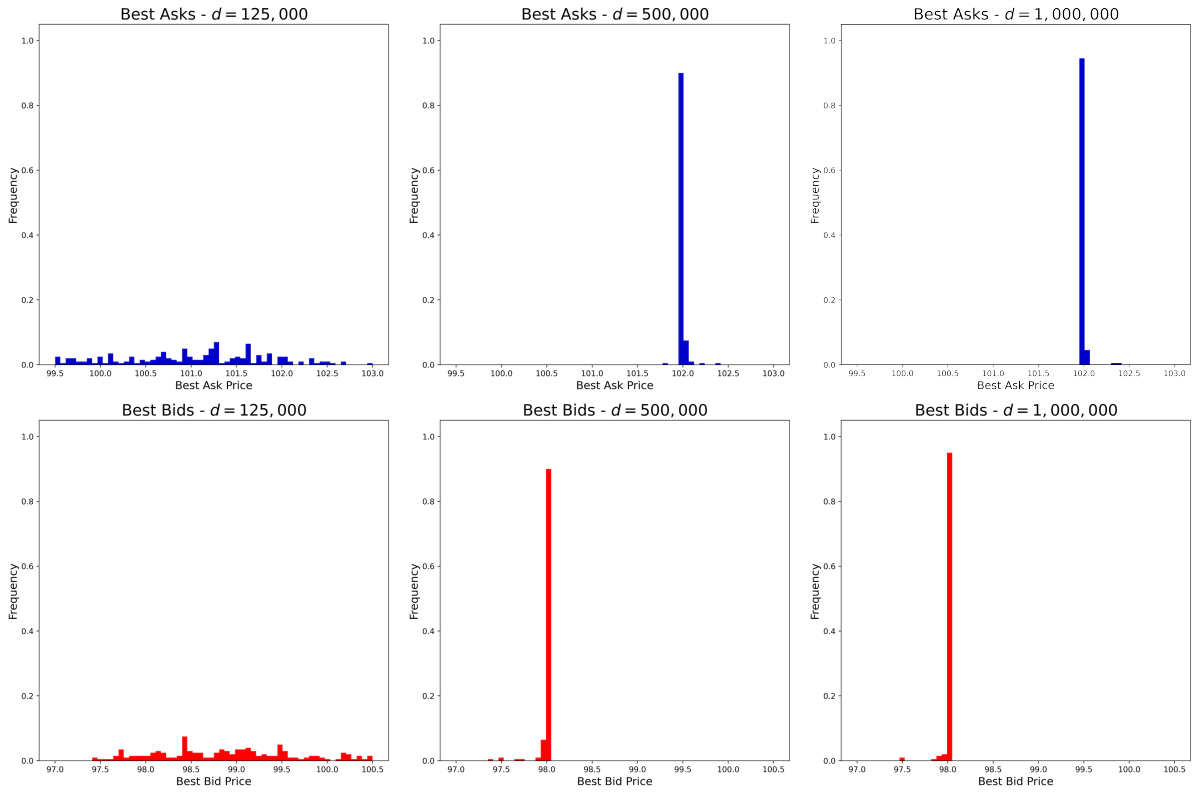
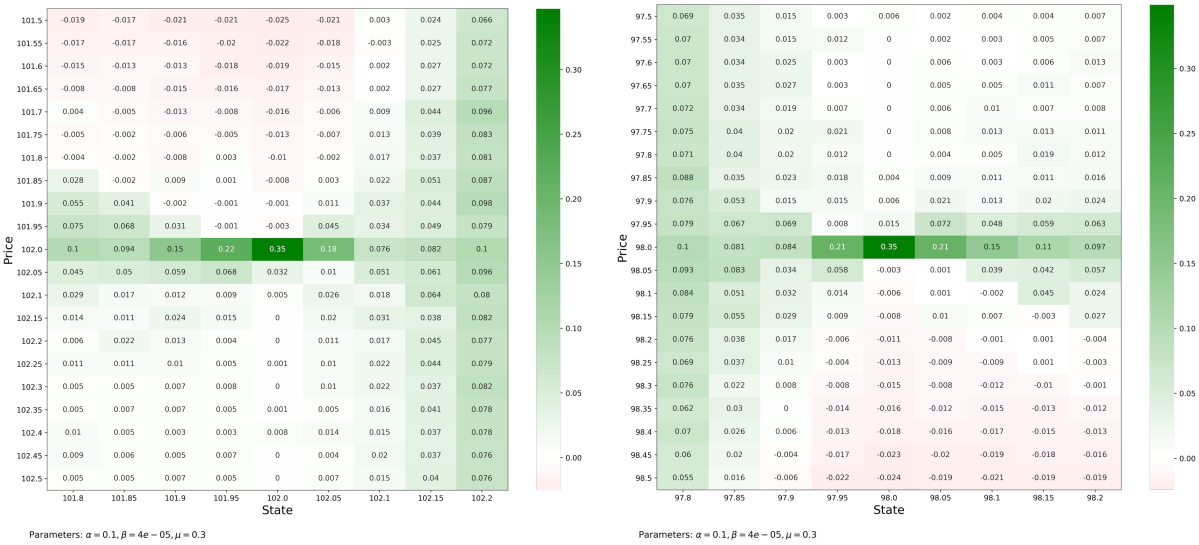


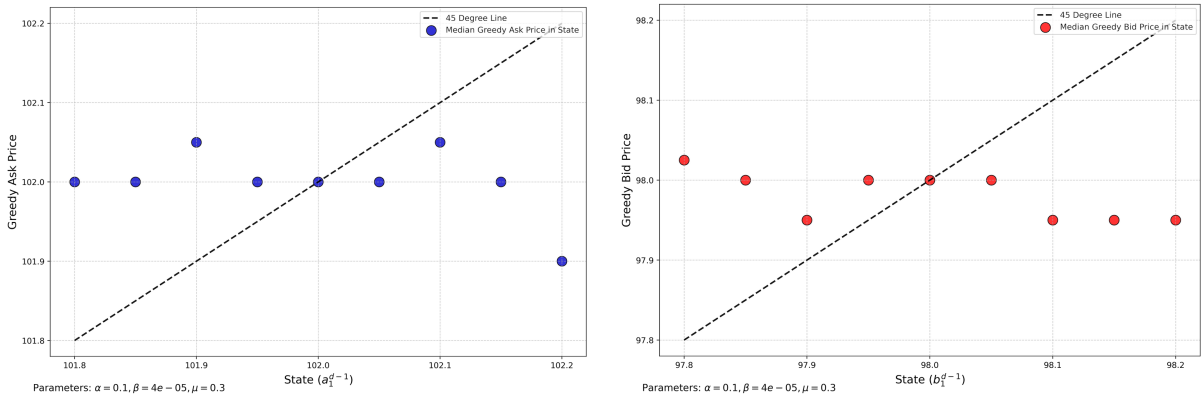
Figure 6.2: Baseline Setup - Distributions of Best Ask and Bid Quotes



(a) Ask (b) Bid

Figure 6.3: Baseline Setup - Averaged Final Q-Tables

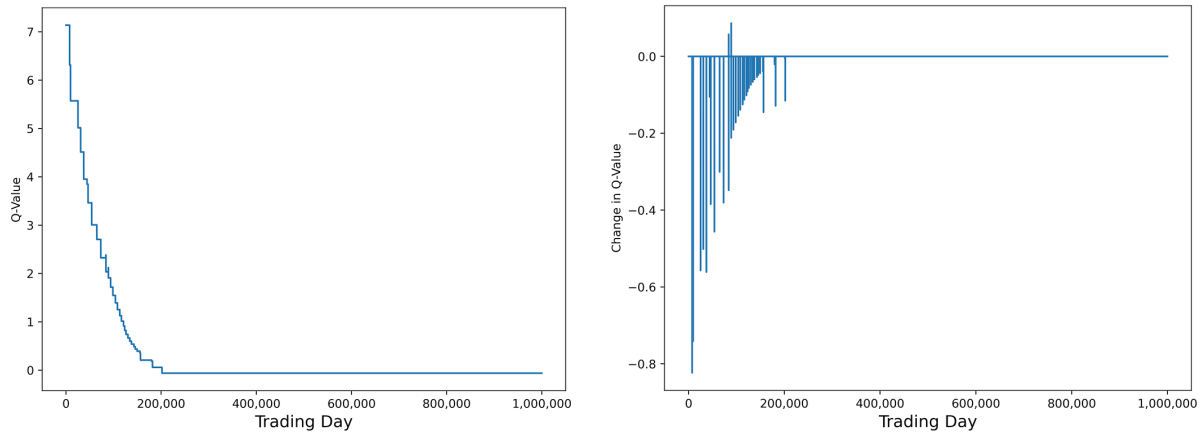
These results suggest the following: if the AMM is "unlucky" and experiences are series of negative profits due to the stochastic asset value, the Q-value may drop below zero. If this action-state pair is not revisited by random exploration, which occurs with a probability near zero in the later stages, this action is never tried again. This is because there exist actions that always yield a weakly positive profit (e.g., asks greater than or equal to 102) and so the



(a) Greedy Ask Quotes (b) Greedy Bid Quotes

Figure 6.4: Baseline Setup - Median Final Greedy Quotes by State

Q-values associated with these are always positive. Thus, when choosing the greedy action, the AMM would choose the action with a zero Q-value over one that has a negative Q-value. Figure 6.5 plots an example of how the Q-value changes in one iteration for ask price 100.9 in State 102.[13] Around day $d = 210,000$, the Q-value drops to -0.06 and does not change again after this, meaning that this action in State 102 was never retried. This process repeats for all risky actions over the one million days and the AMMs move to risk-free actions.



(a) Q-value (b) Change in Q-value

Figure 6.5: Baseline Setup - Example Q-value for $a_m = 100.9$ in State 102

6.2 Introducing ICU

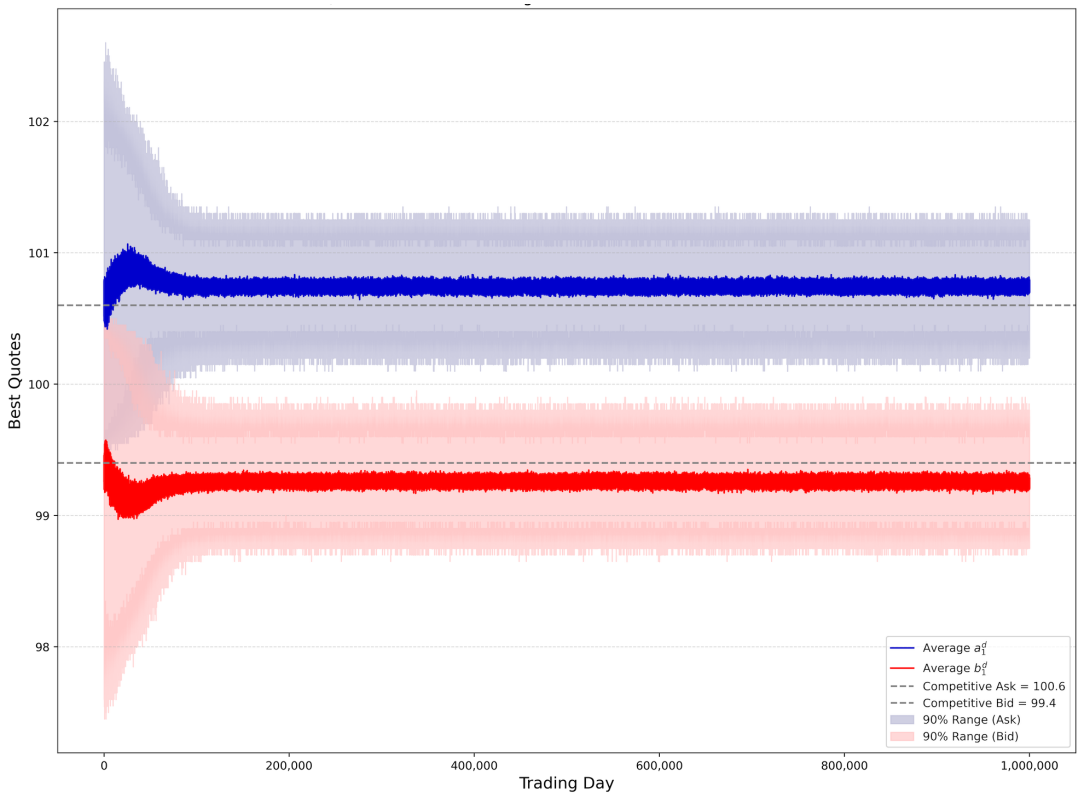
In this section, we consider how the results change when introducing Imperfect Counterfactual Updating (ICU). Whilst the final quotes are noisier, we find that they are distributed over a

[13] While we plot the example here for a single iteration, we find that this plot is representative of the pattern more generally.

more competitive range of prices.

RESULT 3 *When there exist profitable risk-free actions, ICU is sufficient to reach prices near the competitive levels.*

With ICU, the best quotes quickly reach a point near the competitive prices (Figure 6.6); the mean final ask (bid) is 100.71 (99.25) compared to the competitive price of 100.6 (99.4). However, there is significant dispersion in the best quotes across iterations; this can the seen by the 90% interquartile range, as well as the distributions in Figure 6.7. We find that the standard deviation of the best quotes across iterations is approximately 0.26 in the final day.



Parameters: $\alpha = 0.1, \beta = 4e-05, \delta = 0.5, \mu = 0.3$. Final Mean Ask: 100.71. Final Mean Bid: 99.25.

Figure 6.6: Introducing ICU - Average Best Ask and Bid Quotes in Each Day

Figure 6.8 shows how the averaged final Q-tables differ from the baseline setup. On the ask-side, AMMs attach positive Q-values to actions far below 102. In State 100.6, the Q-values for actions above 100.65 are strictly positive; the Q-values only become negative for prices 100.6 and below. When we introduce ICU, the Q-values for each action are updated more frequently compared to the case where we update only the experienced action; therefore, we expect the Q-values to be closer to the true expected value.

One may wonder why we do not experience the same convergence to risk-free actions with

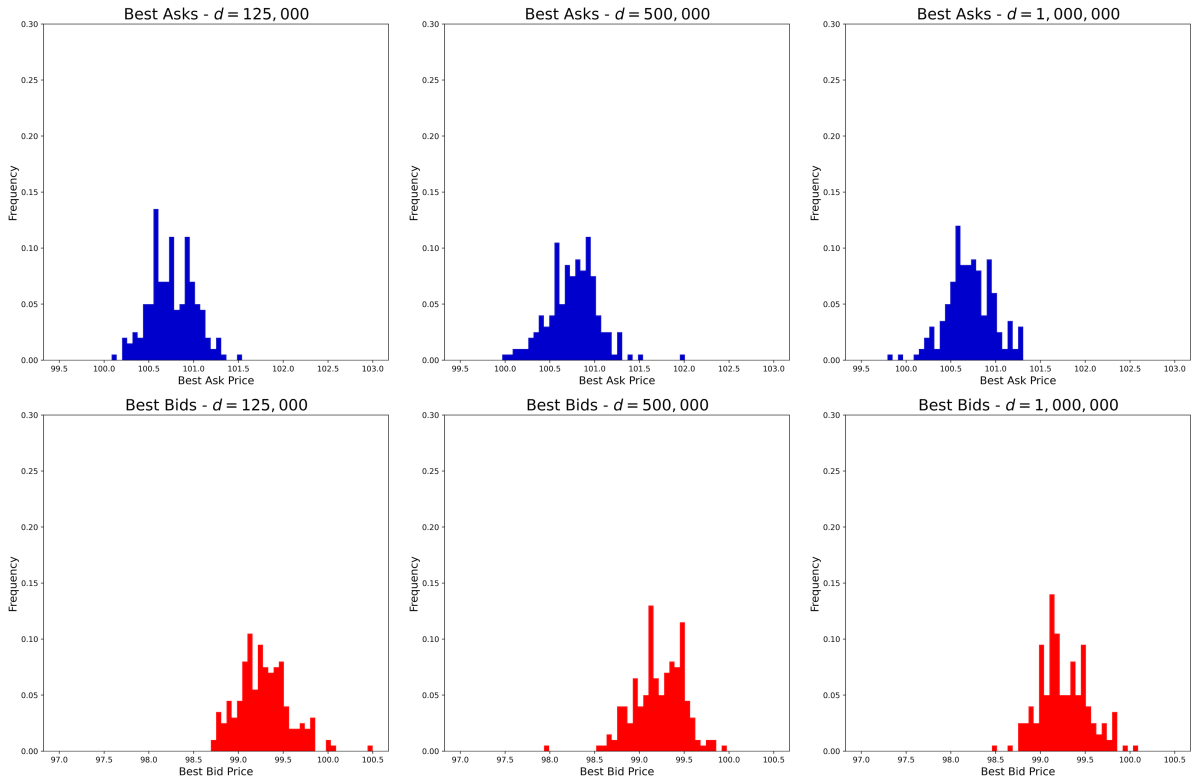
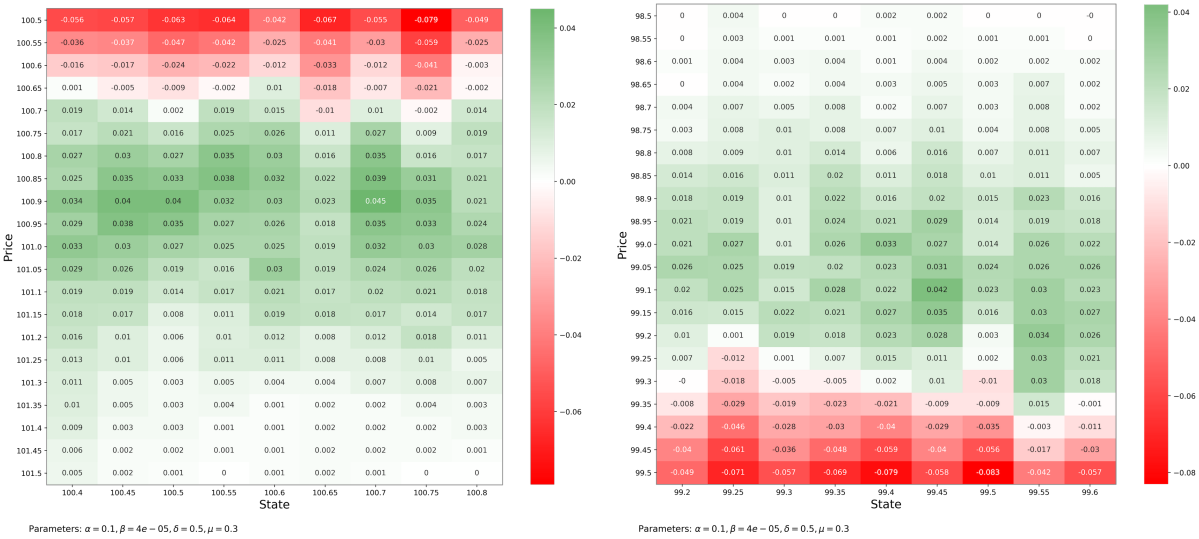


Figure 6.7: Introducing ICU - Distributions of Best Ask and Bid Quotes

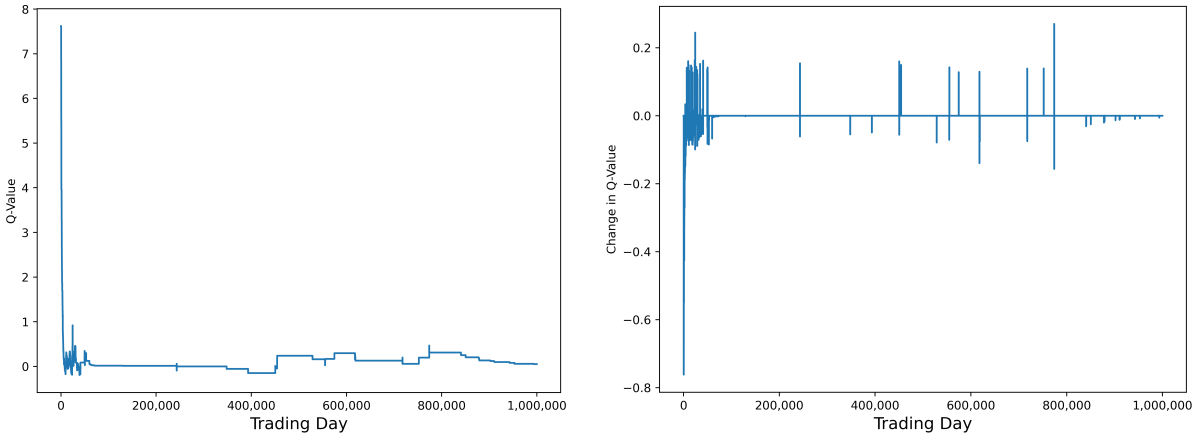


(a) Ask (b) Bid

Figure 6.8: Introducing ICU - Averaged Final Q-Tables

ICU. Indeed, as can be seen in the example for action 100.9 in State 102 (Figure 6.9), there are periods in which the Q-values drop below zero.. As before, these actions cannot be chosen as the greedy action but, when using ICU, the Q-values continue to be updated, even if that action was not chosen. For example, if trade occurs at an ask price above the one associated with a negative Q-value, then this Q-value is still updated by ICU. At some point during the simulation, the once negative Q-value can become positive again; therefore, AMMs do not become "stuck" on

risk-free prices.



(a) Q-value

(b) Change in Q-value

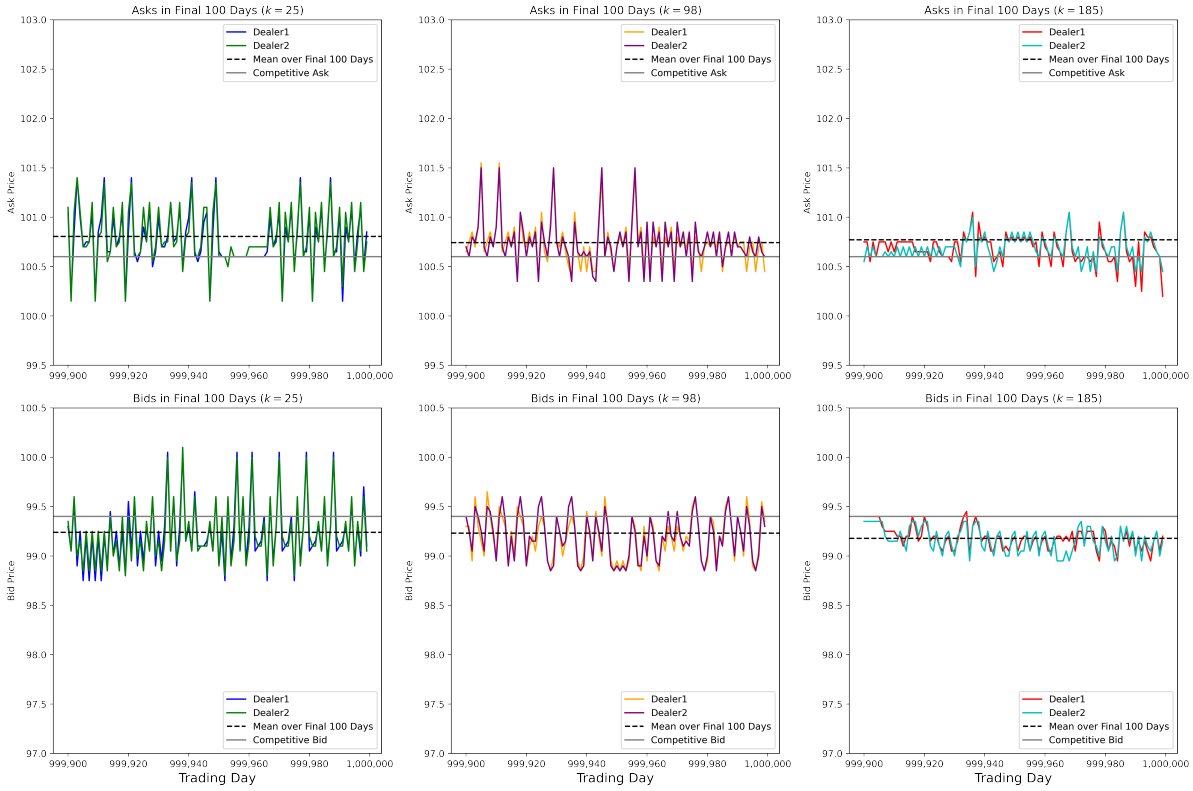
Figure 6.9: Introducing ICU - Example Q-value for $a_m = 100.9$ in State 102

We find that, in the final day, in only 24% of the iterations do AMMs set prices equal to one of the Nash equilibria on the ask-side; this is 18% on the bid-side. Interestingly, in 25% of the iterations, AMMs set the best ask prices *below* and 30% set the best bid prices *above* the competitive price in the final day. To see why this occurs, notice that many Q-values are close to zero and to one another. Thus, the Q-values (and greedy actions) are sensitive to the stochastic payoff. We see, in Figure 6.10, that prices fluctuate around the theoretical price. The ask prices fluctuate around 100.6 but, on average, are just above. The stochastic payoff, coupled with ICU, is the mechanism driving this; when profit is high, the AMMs update all Q-values below the chosen one and think that it is still profitable to charge less. When the profit is inevitably negative at the lower price, the prices increase to *above* the competitive price. As the Q-values are close to zero, this process repeats frequently.

The highly correlated prices in Figure 6.10 may not necessarily be from collusion. After a while, the Q-values become very close to the expected values, and thus, the Q-values for AMM_1 become close to those for AMM_2. Using ICU, the AMMs update in an almost identical manner; the only difference comes from the experienced action, which varies for the AMM that traded. Thus, the Q-values become correlated as they are updated with (almost) identical values.

6.3 Naive Adjustment of Exploration

Several papers (e.g., Abada and Lambin, 2023; Abada et al., 2024; Colliard et al., 2022) suggest that insufficient exploration is driving supracompetitve prices. We show that simply increasing the exploration probability is insufficient to reach competitive levels. We find similar, but noisier, results compared to the sections above when adjusting the probability of exploration



Note: The iterations chosen for the above figures were selected randomly.

Figure 6.10: Introducing ICU - Example Quotes in the Final 100 Days

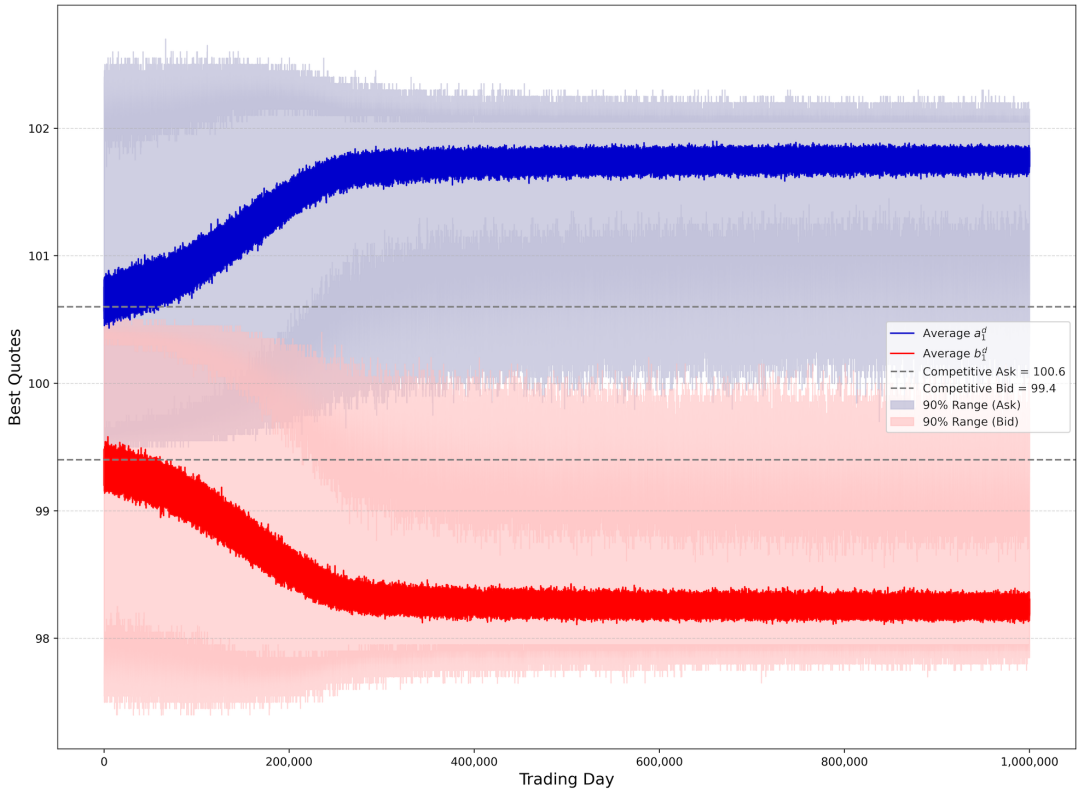
such that there is always *at least* a five percent probability of exploration (as Colliard et al. (2022) consider). The exploration probability becomes

$$\epsilon = 0.05 + 0.95 \exp(-\beta d). \tag{30}$$

RESULT 4 *Naive exploration alone is not sufficient to prevent AMMs from quoting at the extreme values in the presence of risk.*

Figure 6.11 shows that, without ICU, the increased probability of exploration has a limited effect on prices. Whilst the mean ask (bid) price in the final day is 101.76 (98.25), this is likely due to the additional noise induced by exploration. 44% of AMMs still quote asks of 102 or higher and 40% that quote bids of 98 or lower (Figure 6.12).[14] The median ask (bid) quote is 101.95 (98.05).

[14]One should also note that the lower mean ask and higher mean bid may also be affected by the fact that these quotes are the *best* prices offered. They are, therefore, biased (weakly) downwards for the asks and upwards for the bids compared to the mean of all quotes submitted by AMMs. We find 67% of market makers quote 102 or higher in the final day; this is 64% for bids of 98 or lower.



Parameters: $\alpha = 0.1, \beta = 4e-05, \mu = 0.3$. Final Mean Ask: 101.76. Final Mean Bid: 98.25.

Figure 6.11: Adjusted Probability - Average Best Ask and Bid Quotes in Each Day

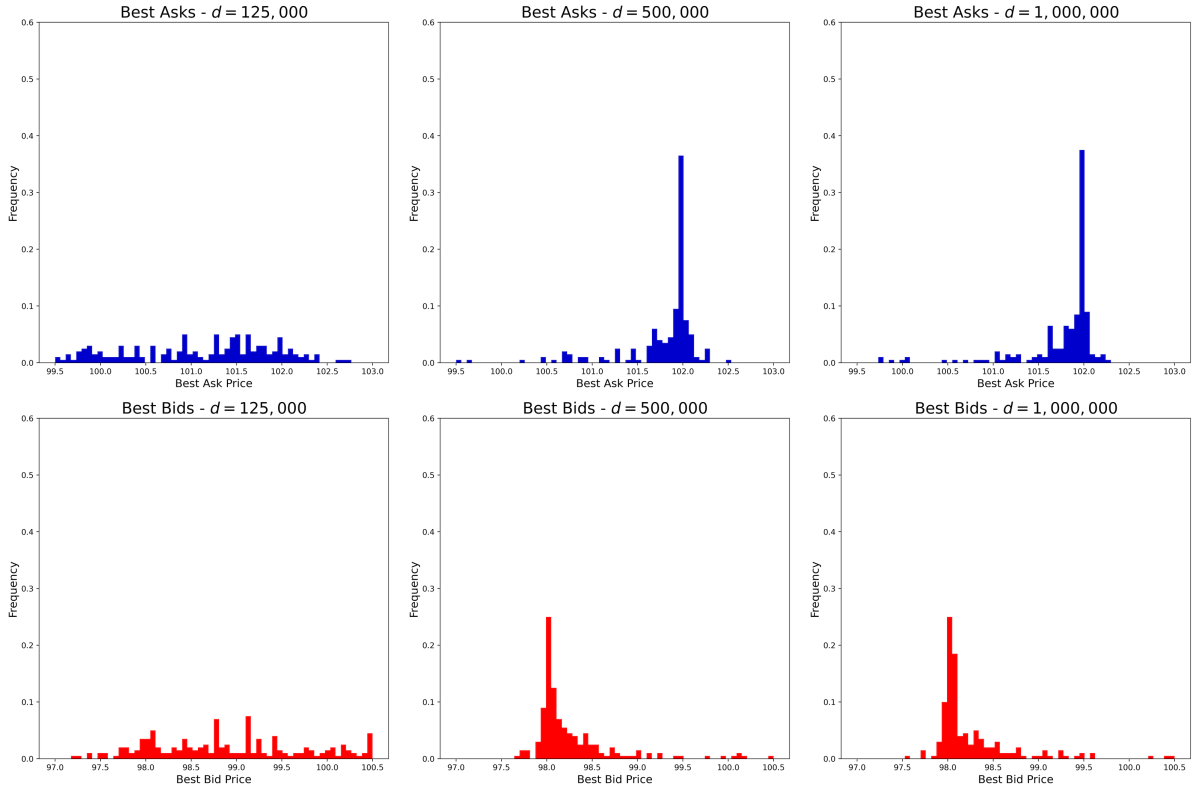
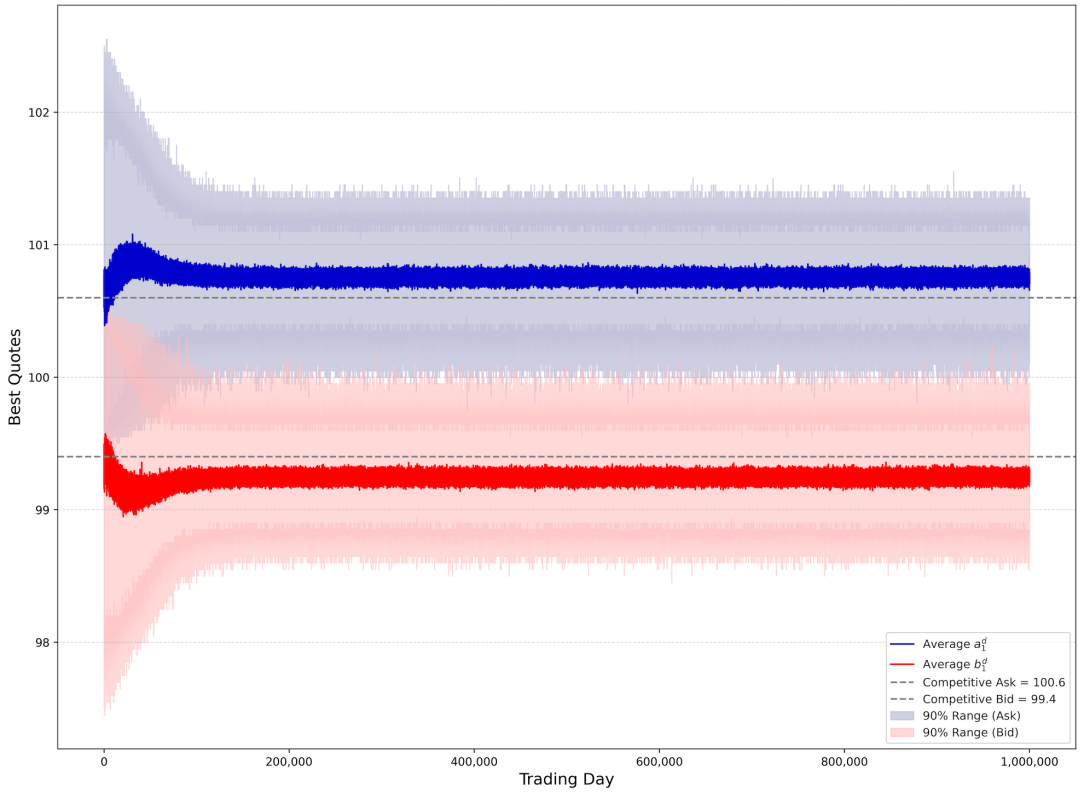


Figure 6.12: Adjusted Probability - Distributions of Best Ask and Bid Quotes

We also combine this adjusted probability with ICU. The final mean prices (Figure 6.13) are almost identical to those in the previous section; the main difference is that the quotes are noisier; the standard deviation of the final best quotes is 0.31 with the changed probability compared to 0.26 previously.

These results suggest that increasing the exploration probability in this way is not sufficient to reach competitive levels. There are minor differences with respect to prices but the main difference is the extra noise from greater exploration.



Parameters: $\alpha = 0.1, \beta = 4e-05, \delta = 0.5, \mu = 0.3$. Final Mean Ask: 100.77. Final Mean Bid: 99.23.

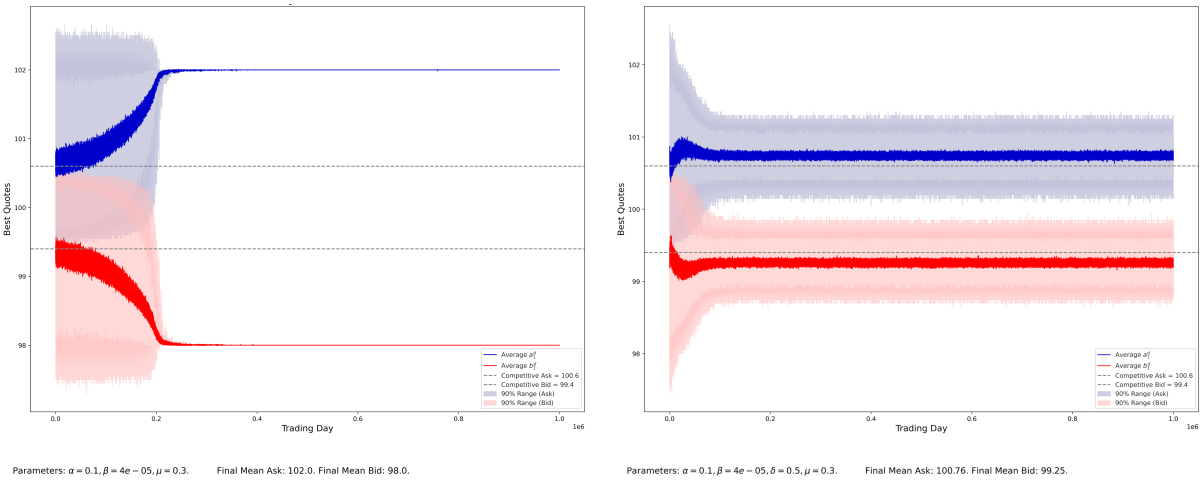
Figure 6.13: Adjusted Probability & ICU - Average Best Ask and Bid Quotes in Each Day

6.4 Introducing Price-Elastic Noise Traders

In this section, we adjust the behaviour of noise traders to make them price-elastic. We assume that they will only trade with ask quotes that are (weakly) below v_H and bid quotes that are (weakly) above v_L. When both $a_t^d \leq v_H$ *and* $b_t^d \geq v_L$, noise traders will buy and sell the asset with equal probability; when $a_t^d \leq v_H$ but $b_t^d < v_L$, then noise traders buy or no-trade with equal probability; finally, when $b_t^d \geq v_L$ but $a_t^d > v_H$, noise traders sell or no-trade with equal probability. If $a_t^d > v_H$ *and* $b_t^d < v_L$, then they will not trade.

Figure 6.14 presents the results when introducing price-elastic noise traders. Both with and

without ICU, there is little difference to the price-inelastic case. We notice that AMMs quickly learn not to charge ask prices above 102 and bids below 98, otherwise they earn zero profit; instead, they charge exactly these prices.



(a) Without ICU (b) With ICU

Figure 6.14: Elastic Noise Traders - Average Best Ask and Bid Quotes in Each Day

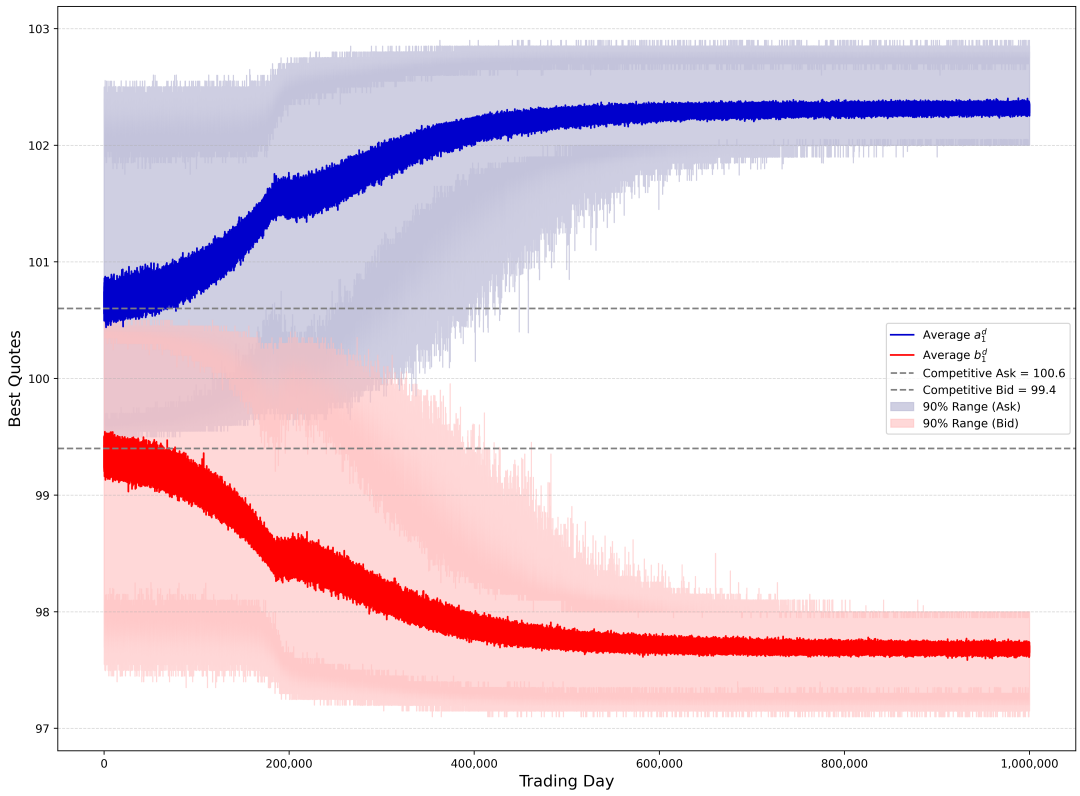
6.5 Removing Profitable Risk-Free Actions

We now consider the same price-elastic setup but constrain noise traders to only trade in the interval $[98.1, 101.9]$, which is strictly inside the interval $[v_L, v_H]$. In doing so, we remove the profitable risk-free actions, meaning AMMs cannot profit from noise traders at risk-free actions, i.e., at ask (bid) price 102 (98).[15]

With this narrowed range, trade with noise traders stops (Figure 6.15). The mean ask and bid prices in the final day are 102.32 and 97.7; the interquartile range shows that the quotes are distributed entirely *outside* the interval for which noise traders will trade. Recall our assumption that informed traders will still trade if indifferent; therefore, the only trade occurs with informed traders at an ask (bid) price of 102 (98), where both the AMMs and traders earn zero.

The rate at which the prices reach the extremes is slower than in the previous section. Previously, 102 and 98 were still profitable actions, and so AMMs realised that these are the best quotes when wanting to avoid losses. Now, any ask in $[102, 103]$ and bid in $[97, 98]$ will return a profit of zero, making AMMs indifferent between any of these actions. We see that the

[15] At these extreme values, the only trade would occur with informed traders and so the AMMs would earn zero profit.



Parameters: $\alpha = 0.1, \beta = 4e-05, \mu = 0.3$. Final Mean Ask: 102.32. Final Mean Bid: 97.7.

Figure 6.15: Removing Risk-Free Actions - Average Best Ask and Bid Quotes in Each Day

best quotes fluctuate in these ranges (Figure 6.16). As the best quotes cover this interval more completely, there are many more action-state pairs to update, extending the duration of the learning process. By the final day, almost the entire distribution of quotes are in the intervals where there is no trade with noise traders.

6.5.1 ICU in the Absence of Profitable Risk-Free Actions

We now consider ICU when there are no profitable risk-free actions. Even with ICU, we notice the beginning of a breakdown of trade (Figure 6.17), although considerably slower than in the case without ICU. As the exploration probability approaches zero, we notice that the *mean* quotes begin to diverge from the competitive levels. This change in the mean prices is caused by the dispersion *across* iterations, as AMMs gradually become "stuck" on quotes outside (v_L, v_H). The distributions gradually shift so that more AMMS have best quotes outside this interval (Figure 6.18). Figure 6.19 supports this hypothesis; after around day 150,000, the standard deviation across iterations increases and continues to do so over the course of the market simulation.

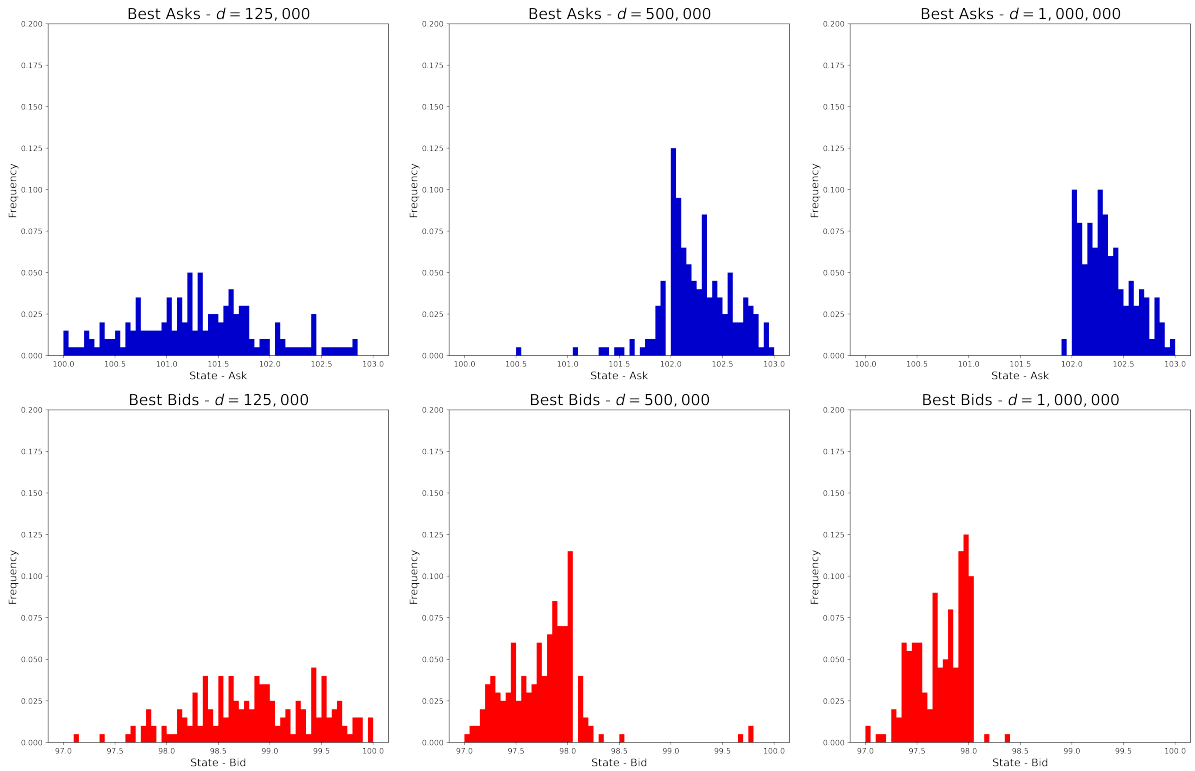


Figure 6.16: Removing Risk-Free Actions - Distributions of Best Ask and Bid Quotes

RESULT 5 *In the absence of profitable risk-free actions, the trade begins to break down, even when using ICU.*

Even with ICU, AMMs to move towards the risk-free actions despite them not being profitable. At some point, the Q-values for all risky actions in a state may become negative; when this happens and the exploration probability is near zero, they will no longer be updated by ICU. Previously, a noise trader would buy at price 102, meaning that the AMM could positively update all Q-values below 102. When we remove the risk-free actions, AMMs will not trade with a noise trader at price 102, so cannot positively update any actions below.[16] At some point during the simulation, it is possible for all Q-values below 102 to become negative due to the stochastic payoff. Previously, trade at the extremes would allow further updating. Now, there is a non-zero probability of quotes becoming "stuck" in the interval where there is no noise trade.

Appendix A.1 shows that this result is robust to changes μ. Even when $\mu = 0$ (i.e., no adverse selection), we notice the start of a trade breakdown; changing μ simply changes the rate at which this occurs. This suggests that the asset value stochasticity drives this result.

[16] In fact, at price 102, the only trade would occur with an informed price (so the value would be 102), meaning that the AMMs see any price below 102 as *less* attractive.

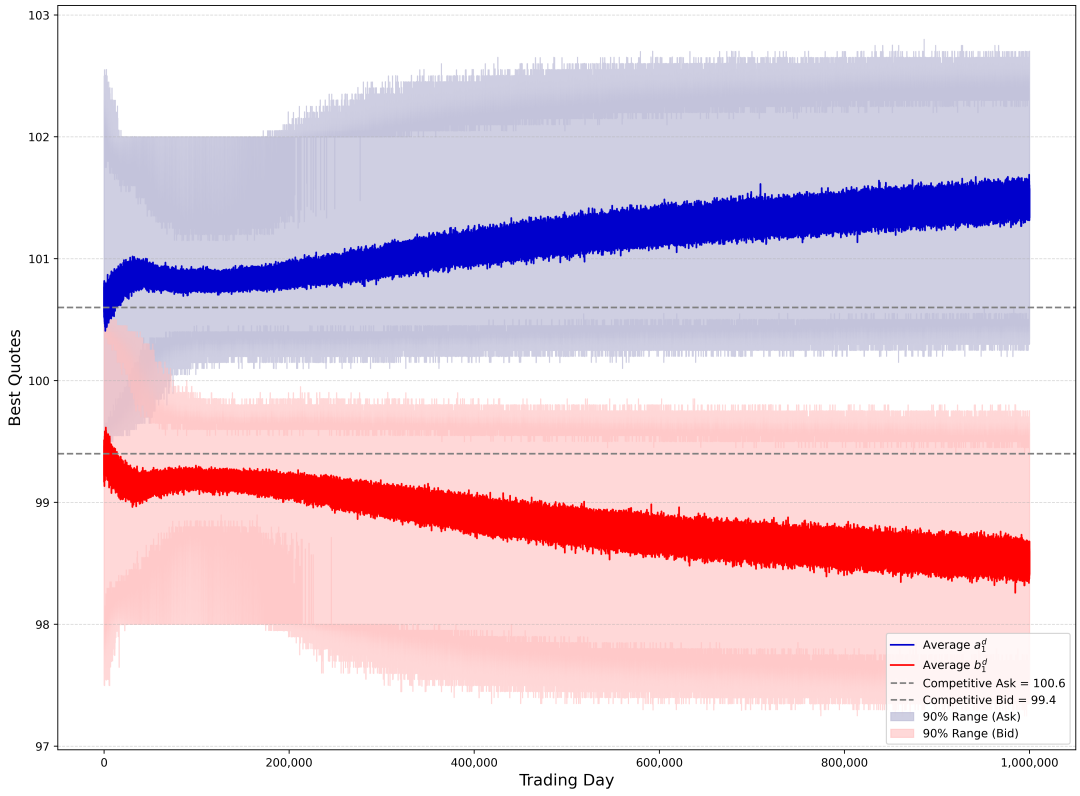


Figure 6.17: ICU & Removing Risk-Free Actions - Average Best Ask and Bid Quotes in Each Day

Overall, these findings confirm that AMMs do not charge extreme prices in the baseline case to take advantage of the profits from noise traders, but due to their desire to avoid losses. When we remove these safe actions, even with ICU, we notice a breakdown in trade, suggesting that AMMs indeed behave as if loss averse.

6.5.2 Combining ICU with Additional Exploration

Combining ICU with a minimum exploration probability avoids a breakdown in trade when there are no profitable risk-free actions. A minimum exploration probability reduces the likelihood that AMMs become "stuck" on price at which there is no noise trade.

RESULT 6 *In the absence of profitable risk-free actions, the AMMs require a combination of ICU and additional exploration to avoid a trade breakdown with noise traders.*

Figure 6.20 shows the mean prices when we combine ICU with the increased probability; there is no breakdown in this case. We find mean ask and bid prices of 100.81 and 99.24, respectively, although the additional exploration probability introduces more noise in the distribution

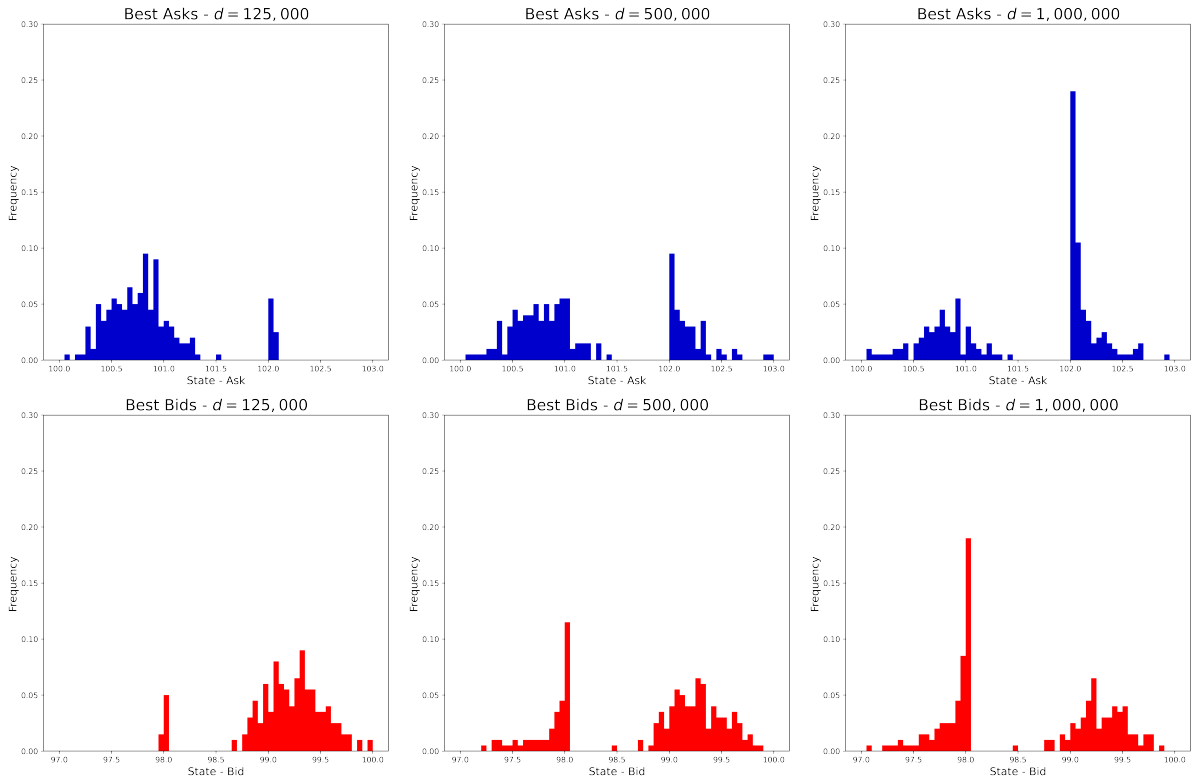
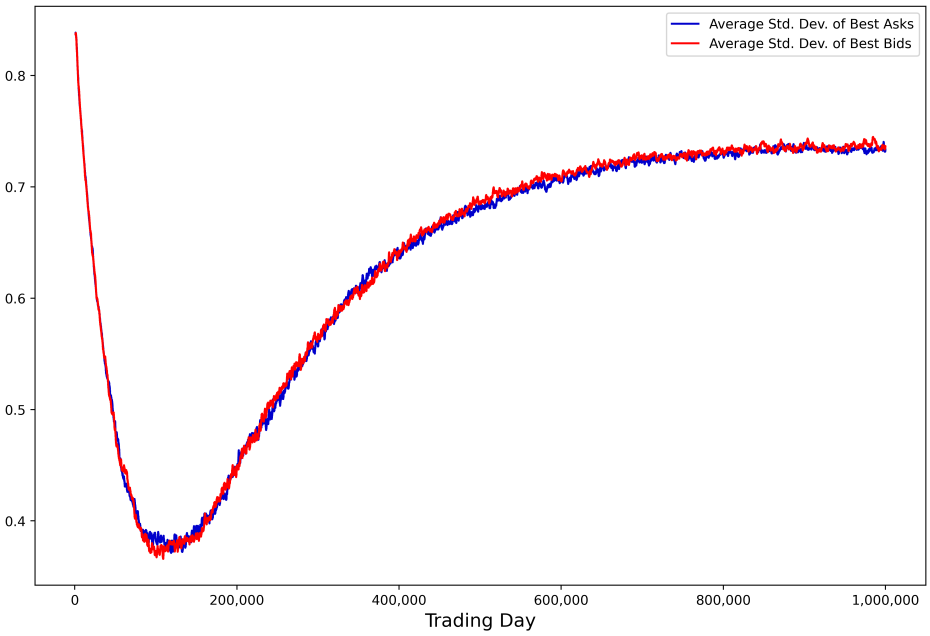


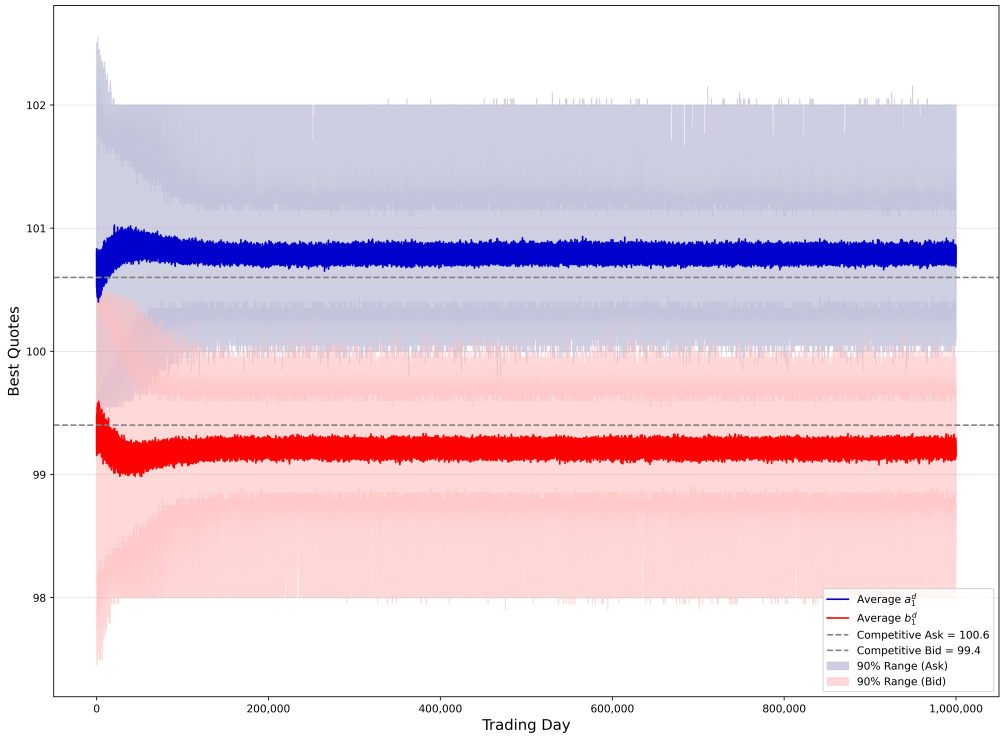
Figure 6.18: ICU & Removing Risk-Free Actions - Distributions of Best Ask and Bid Quotes

of the best quotes. The additional probability of exploration means that, even if a Q-value is negative, it will likely be updated again, either because it was played or through ICU. Hence, the probability of becoming "stuck" shrinks to zero.



Parameters: $\alpha = 0.1, \beta = 4e-05, \delta = 0.5, \mu = 0.3$. Final Mean StdDev of Asks: 0.73. Final Mean StdDev of Bids: 0.74.

Figure 6.19: ICU & Removing Risk-Free Actions - Standard Deviation of Best Quotes



Parameters: $\alpha = 0.1, \beta = 4e-05, \delta = 0.5, \mu = 0.3$. Final Mean Ask: 100.81. Final Mean Bid: 99.24.

Figure 6.20: ICU and & Adjusted Probability when Removing Risk-Free Actions - Average Best Ask and Bid Quotes in Each Day

7 The Multi-period Market

In financial markets, there are typically many trades within a day and, consequently, order flow is informative for a market maker setting prices. Extending to multiple intraday period allows us to understand how AMMs respond to trades. In Section 8, we analyse the results for $T = 2$.

7.1 Algorithmic Market Makers for T > 1

When generalising to $T > 1$ intraday periods, there are two key differences: one is the state space and the other is the updating of the Q-values. The action space is identical to that discussed for $T = 1$ (Section 4.2.1).

7.1.1 State Space

The primary challenge of multiple intraday periods is the 'curse of dimensionality' in the state space. When accounting for order flow through the state space, we must consider the effect on the number of potential states. For example, considering the entire sequence of trades causes the state space to grow exponentially with T. Thus, one needs to summarise the information in a low-dimensional way; we opt to use the *net* inventory of AMM$_i$ at time t as a measure of the order flow.

For $t = 1$, we define the states as for $T = 1$; that is, as the best quotes at time $t = 1$ in day $d-1$. For $t > 1$, the state is a triple consisting of: (1) the best quotes at time t in day $d-1$, (2) the net inventory of AMM$_i$ before the trade at time t, and (3) an indicator of the time. That is, for $t > 1$, the states are

$$S_{i,t}^{d,Ask} = \{a_t^{d-1}, \mathcal{I}_{i,t}^d, t\}, \tag{31}$$

$$S_{i,t}^{d,Bid} = \{b_t^{d-1}, \mathcal{I}_{i,t}^d, t\}, \tag{32}$$

where $\mathcal{I}_{i,t}^d$ is the net inventory at time t for AMM$_i$:

$$\mathcal{I}_{i,t}^d = \sum_{\tau=1}^{t-1} \Omega_{i,\tau}^d. \tag{33}$$

The state at time $t > 1$ differs by AMM as the net inventory is specific to each AMM. We use the period indicator to distinguish between periods.

For each $t = 1, ..., T$, there are $M(2t-1)$ states, which, even for a relatively small M, can

become large as T increases.[17] For a very fine set of actions, this would result in a particularly large state space, many of which may be visited infrequently. Therefore, we group the prices into intervals and, instead, use these in the state space. We define the intervals for the ask-side state space as

$$[a_{min}, a_{min}+\kappa), [a_{min}+\kappa, a_{min}+2\kappa), ..., [a_{max}-2\kappa, a_{max}-\kappa), [a_{max}-\kappa, a_{max}], \tag{34}$$

and as

$$[b_{min}, b_{min}+\kappa], (b_{min}+\kappa, b_{min}+2\kappa], ..., (b_{max}-2\kappa, b_{max}-\kappa], (b_{max}-\kappa, b_{max}], \tag{35}$$

for the bid-side. κ is a parameter used to set the width of the intervals; in our simulations, we set $\kappa = 0.25$. For computational efficiency, we will also group the $t = 1$ states into intervals.

For a large T, the state space is potentially very large; states corresponding to high values of net inventory would be visited infrequently. As a solution, one could truncate the net inventory levels in the state space and map any net inventory level above the truncation level to this value, thus ensuring that the state is visited with a sufficient probability. In Section 8, we consider $T = 2$; therefore, we do not need to truncate.

7.1.2 Updating Q-Values

Given our setup of the state space, states are not revisited during a trading day. Therfore, it seems natural that we update the Q-values only at the end of the day, once the value is revealed.

Recall that profits from a trade at time t, evaluated at time T, are given by

$$\pi_{i,t}^{d,Ask} = (a_{i,t}^d - \tilde{v}^d)\mathbf{1}_{\{\Omega_{i,t}^d=1\}}, \tag{36}$$

$$\pi_{i,t}^{d,Bid} = (\tilde{v}^d - b_{i,t}^d)\mathbf{1}_{\{\Omega_{i,t}^d=-1\}}. \tag{37}$$

For each time $t = 1, ..., T$ and all actions $m = 1, ..., M$, we update, at time T, the corresponding

[17] There are M potential values of the state corresponding to a_t^{d-1} for each time t. Also, there are $2t-1$ potential values that the net inventory can take. Considering all combinations of price components and net inventory components yields $M(2t-1)$ possible states.

ask Q-value given the state of each AMM at time t, $q_{i,T}^{d,Ask}(S_{i,t}^{d,Ask}, a_m)$, as

$$q_{i,T}^{d,Ask}(S_{i,t}^{d,Ask}, a_m) = \begin{cases} \alpha\pi_{i,t}^{d,Ask} + (1-\alpha)q_{i,T}^{d,Ask}(S_{i,t}^{d,Ask}, a_m) & \text{if } a_m = a_{i,t}^d, \\ q_{i,T}^{d,Ask}(S_{i,t}^{d,Ask}, a_m) & \text{if } a_m \neq a_{i,t}^d, \end{cases} \tag{38}$$

and the bid Q-value, $q_{i,T}^{d,Bid}(S_{i,t}^{d,Bid}, b_m)$, as

$$q_{i,T}^{d,Bid}(S_{i,t}^{d,Bid}, b_m) = \begin{cases} \alpha\pi_{i,t}^{d,Bid} + (1-\alpha)q_{i,T}^{d,Bid}(S_{i,t}^{d,Bid}, b_m) & \text{if } b_m = b_{i,t}^d, \\ q_{i,T}^{d,Bid}(S_{i,t}^{d,Bid}, b_m) & \text{if } b_m \neq b_{i,t}^d. \end{cases} \tag{39}$$

The Q-values are evaluated only at time T and are updated given the state of AMM_i at each time t. They are a weighted average of the profit and the current Q-value if action m corresponds to the experienced action; if not, the Q-value is unchanged.

7.2 Simulating the Multi-period Market

There are only minor changes in the order of the market simulation between the $T = 1$ case and the generalised case. Algorithm 2 provides a pseudo-code for the market simulation; Figure C2 (Appendix C) provides a visual representation. We use the baseline parameterisation from Table 2.

Algorithm 2 T-Period Market Simulation

```
Initialise parameters and generate action and state spaces.
for k = 1, ..., K do
    Randomly initialise bid and ask Q-Tables for market makers.
    for d = 1, ..., D do
        Realise ṽ^d.
        for t = 1, ..., T do
            Update states as {b^d_{t-1}, I^d_{i,t}, t}, {a^d_{t-1}, I^d_{i,t}, t}.
            Realise trader type for each time period.
            Choose exploit/explore and the bid and ask prices.
            Traders decide whether to trade using {b^d_t, a^d_t}.
        end for
        Update Q-Tables using ṽ^d.
    end for
end for
```

The main difference comes from the timing of the trader arrival. Whether a trader is an informed or noise trader is determined for *each* time t; the probability of encountering a noise trader at time t is independent of time $t-1$. Additionally, an AMM chooses to exploit or explore at *each* time t, rather than choosing one of these for the day; thus, this decision is also

independent of the decision at time $t-1$.

8 Results for the Multi-period Market (T = 2)

We now present the results from the multi-period market when $T = 2$. Given the baseline parameterisation with price-inelastic noise traders, the theoretical competitive prices are: $a_1^d = 100.6$ and $b_1^d = 99.4$ at time $t = 1$; $a_2^d = 101.1$ and $b_2^d = 100$ after a trader buys at $t = 1$; $a_2^d = 100$ and $b_2^d = 98.9$ after a sell at $t = 1$; after no trade, the prices are unchanged.

The results of this section echo the results of the one-period market: the baseline algorithm behaves as if loss averse, with AMMs only willing to trade at risk-free levels; introducing ICU allows for improved learning, resulting in quotes near to competitive levels; additionally, removing the risk-free actions leads to a breakdown of trade with noise traders, even with ICU. In responding to order flow, AMMs mirror the rationale of the theoretical prices: the quotes (both ask and bid) increase following a buy from a trader at $t = 1$ and decrease following a sell.

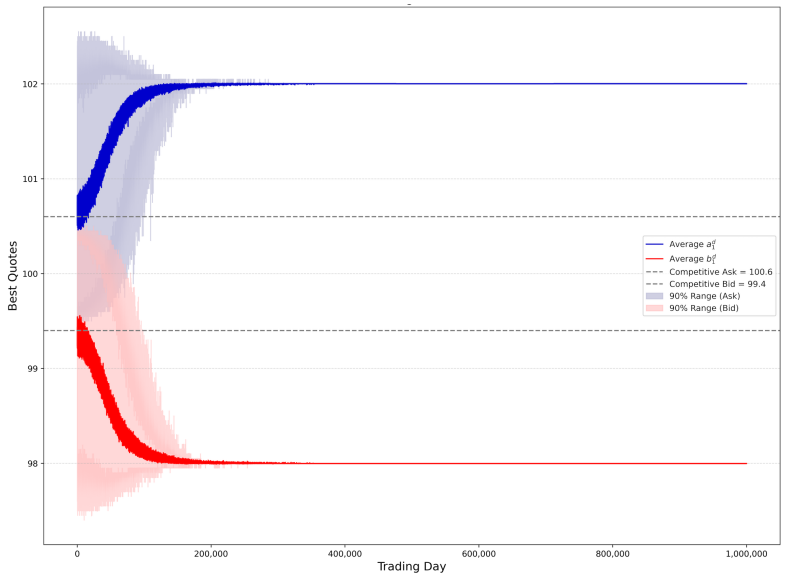
In our defined state space, it is the net inventory of AMM_i, rather than the market imbalance, that we use as our trade measure. When $T = 2$, the net inventory of each dealer coincides with the market imbalance when they are responsible for trade; for the case of zero net inventory, it could be either no trade at $t = 1$ or that the other AMM was responsible for trade. We will consider how AMMs respond after observing a net inventory of zero.

8.1 Baseline Case for T = 2

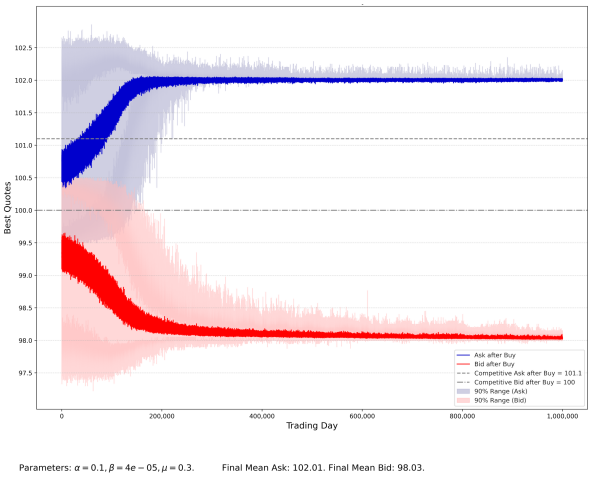
Panel (a) of Figure 8.1 shows that the $t = 1$ results mimic the results from the one-period market.[18] Prices move to the extremes in the baseline case, the mechanism for which is the same as the one-period market: when there are losses on risky actions, the Q-values drop below zero and are not chosen as the greedy action; as the exploration probability approaches zero, they become "stuck" on the risk-free actions.

RESULT 7 *In the baseline setup, the prices tend towards the extremes of the asset values in both time periods.*

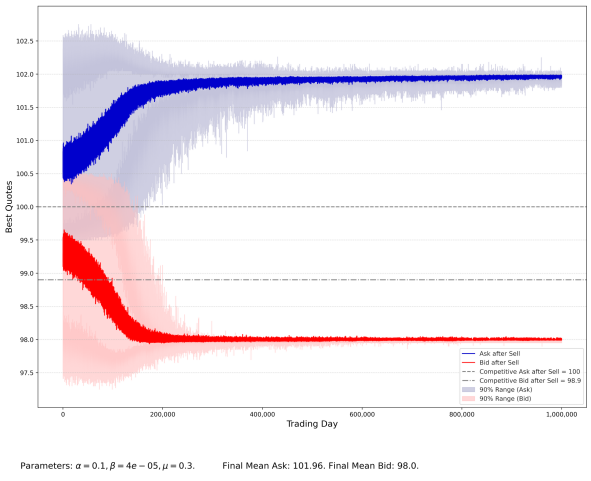
[18]The learning problem for the AMM in the first period mimics that of the one-period market, except that we group the states into intervals. This suggests that, for the first period, this extra coarseness has minimal effects on the prices.



(a) Best Ask and Bid Quotes at time $t = 1$



(b) Best quotes after buy at $t = 1$

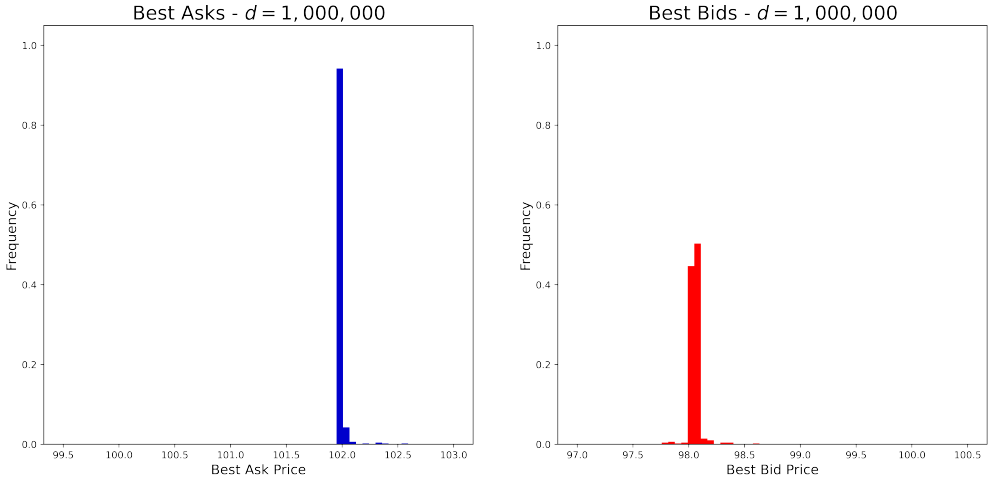


(c) Best quotes after sell at $t = 1$

Figure 8.1: $T = 2$ Baseline Setup - Best Ask and Bid Quotes in Each Day

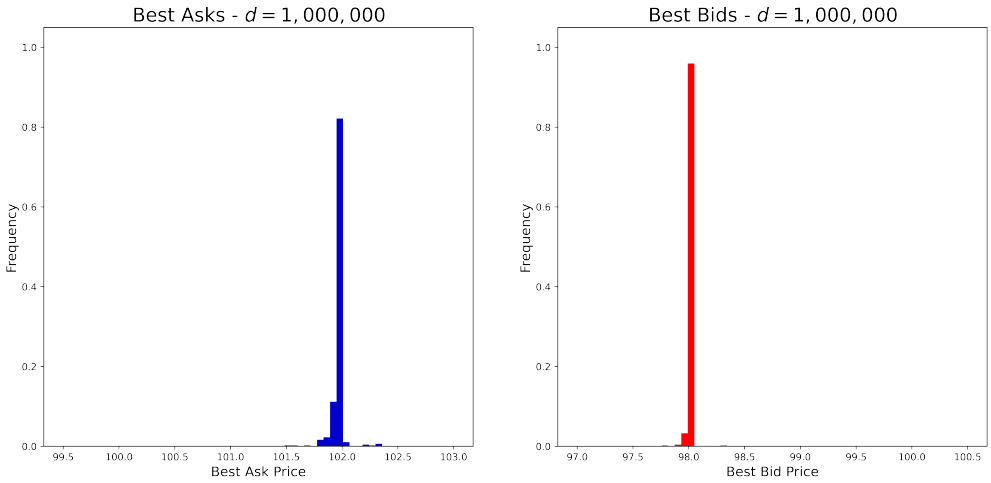
At time $t = 2$, prices also tend towards the extreme values after both a buy and sell at time $t = 1$. However, this is a noisier process than for $t = 1$; even in the final trading day, there is a non-zero range of quotes across iterations. Following a buy at $t = 1$, the ask prices are distributed above 102; following a sell, the bid prices are distributed below 98 (Figure 8.2). In some iterations, the best bid (ask) following a buy (sell) is inside the interval $[v_L, v_H]$. There are two likely causes for this additional noise at time $t = 2$. Firstly, when conditioning on whether there was a buy or sell at $t = 1$, there are less days for which each of these occur compared to $t = 1$, where the same problem is faced each day. As each is experienced less, this naturally creates a slower learning process. Furthermore, the state space for the dealer who has a net inventory of zero creates noise; they are facing the problem without knowing whether there was

no trade in the first round, or whether there was a buy *or* sell.



(a) Ask After Buy

(b) Bid After Buy



(c) Ask After Sell

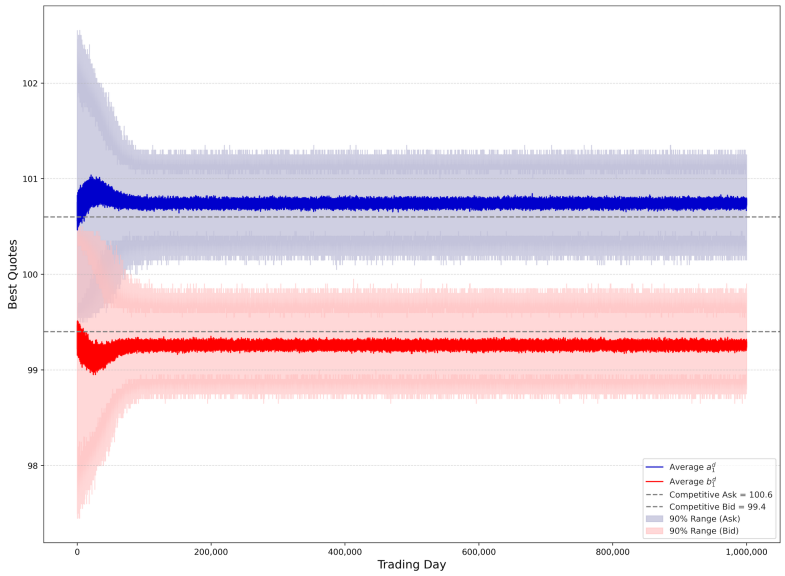
(d) Bid After Sell

Figure 8.2: $T = 2$ Baseline Setup - Distributions of Final Ask and Bid Quotes at $t = 2$

8.2 Introducing ICU for T = 2

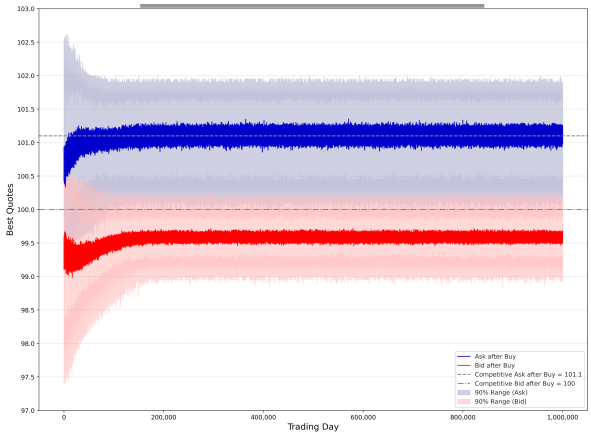
RESULT 8 *With ICU, the quotes increase following a buy from a trader at $t = 1$ and decrease following a sell, in line with the theoretical predictions.*

Introducing ICU creates almost identical results to the one-period market for $t = 1$ (Figure 8.3, panel (a)), with a mean final ask (bid) quote of 100.73 (99.24) compared to 100.71 (99.25) in the one-period market. For $t = 2$, panels (b) and (c) show that both ask and bid prices increase after a trader buys and decrease after a trader sells at $t = 1$. This is in line with the theoretical predictions, suggesting that AMMs can understand, to some extent, the change in the expected value of the asset.



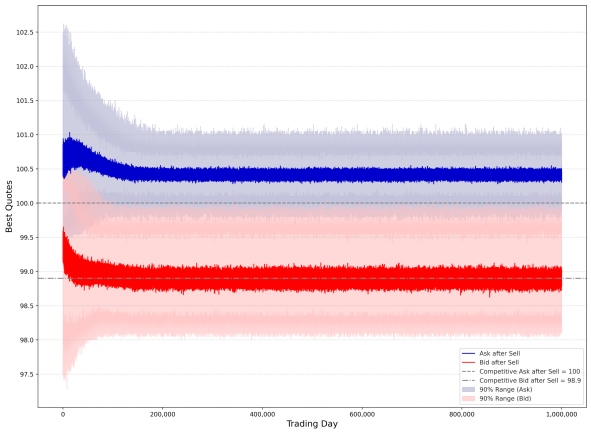
Parameters: $\alpha = 0.1, \beta = 4e-05, \delta = 0.5, \mu = 0.3$. Final Mean Ask: 100.73. Final Mean Bid: 99.24.

(a) Best Ask and Bid Quotes at time $t = 1$



Parameters: $\alpha = 0.1, \beta = 4e-05, \delta = 0.5, \mu = 0.3$. Final Mean Ask: 101.15. Final Mean Bid: 99.58.

(b) Best quotes after buy at $t = 1$



Parameters: $\alpha = 0.1, \beta = 4e-05, \delta = 0.5, \mu = 0.3$. Final Mean Ask: 100.45. Final Mean Bid: 98.84.

(c) Best quotes after sell at $t = 1$

Figure 8.3: $T = 2$ Introducing ICU - Best Ask and Bid Quotes

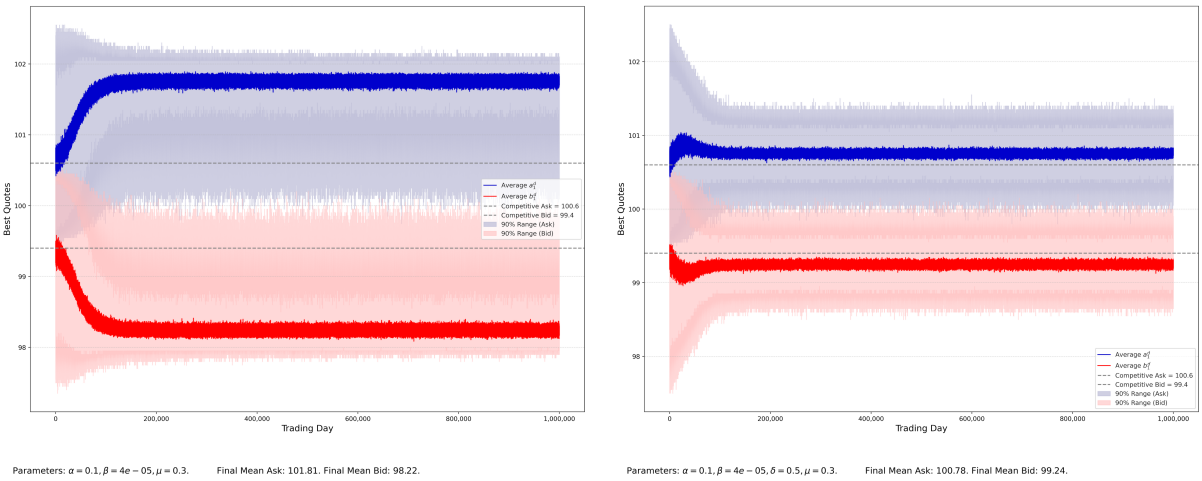
On the ask-side, the median response for a trader with net inventory of 1 is 101.3; when a trader has a net inventory of 0, they quote asks of 101.15. On the bid-side, when faced with a net inventory of -1, AMMs submit a median bid of 98.6 compared to 98.85 with a net inventory of zero. Thus, we find that those who trade in the first round set quotes *further* from the competitive prices than those who do not trade at $t = 1$. AMMs who do not trade at $t = 1$ still set wider spreads compared to $t = 1$. It is not unreasonable to expect this widening of spreads over the first period levels to account for the probability of trade by the other AMM.

8.3 Adjusting Exploration Probability for T = 2

We consider how the prices change when we increase the probability such that there is always at least a five percent probability of exploration. As before, the probability of exploration is

$$\epsilon = 0.05 + 0.95 \exp(-\beta d). \tag{40}$$

Without ICU, adjusting the probability of exploration creates considerable noise with only a minor change in the mean of the best quotes (Figure 8.4, panel (a)); we find mean ask and bid quotes of 101.81 and 98.22, respectively. In the final day, 47% of best ask quotes are 102 or higher; similarly, 47% of the best bid quotes are 98 or below. These results mimic those of the one-period market. At $t = 2$, we notice a similar movement towards the extremes of the asset values (Figure 8.5). The mean prices are slightly higher after a buy at $t = 1$ and lower after a sell, although they remain considerably above competitive levels.

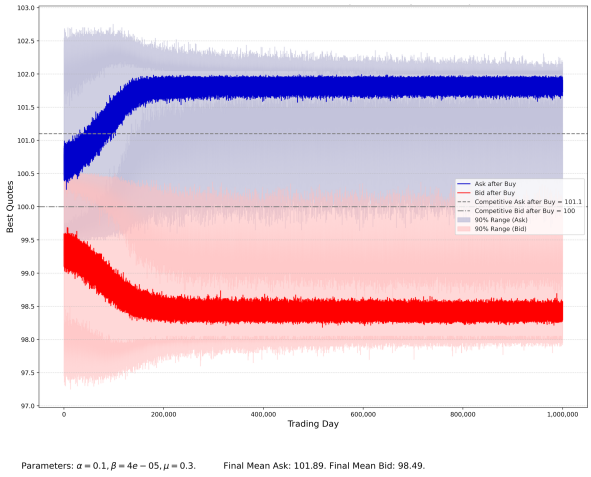


(a) Without ICU (b) With ICU

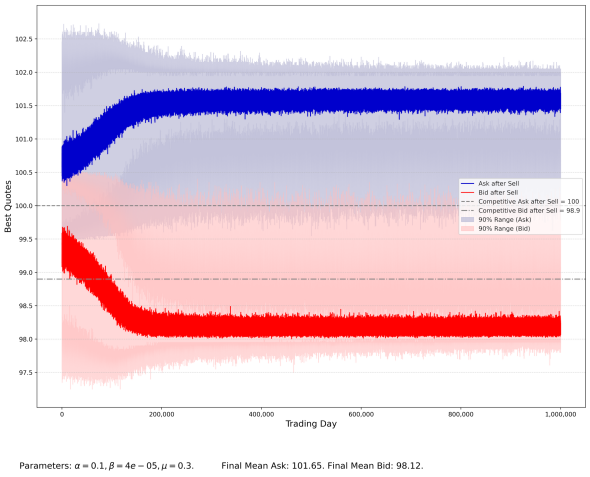
Figure 8.4: $T = 2$ Adjusted Probability - Best Ask and Bid Quotes at $t = 1$

Following a buy at $t = 1$, 70% of the asks, but only 12% of bids, are outside the interval $[v_L, v_H]$; when there is a sell at $t = 1$, then only 12% of the asks, but 70% of bids, are outside this interval. This suggests that, for the trade that is more likely to occur (e.g., a buy after a buy at $t = 1$), more AMMs switch to the risk-free actions; when this is probability is lower, there are more quotes within the interval. This suggests that increasing the exploration probability has some small effects on the $t = 2$ prices, but less than when using ICU.

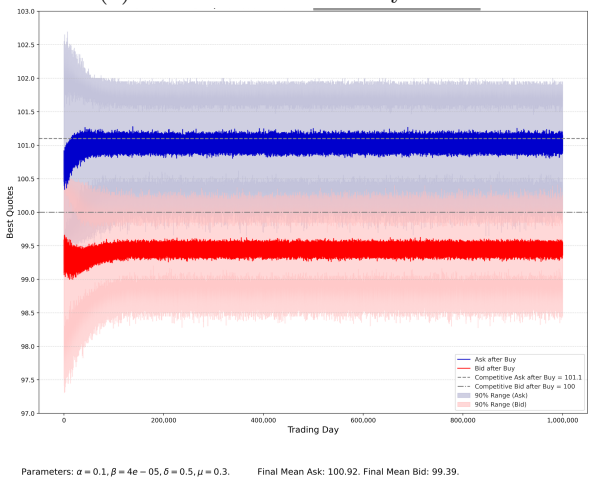
When combining ICU with the adjusted probability, prices are more competitive at both $t = 1$ and $t = 2$. At $t = 2$, prices increase following a buy and decrease following a sell at $t = 1$, as predicted by the theory. The prices deviate further from the theoretical prices compared



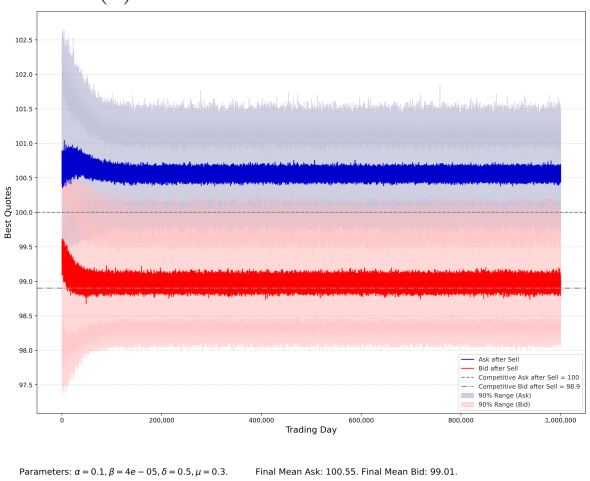
(a) Without ICU: After buy at $t = 1$



(b) Without ICU: After sell at $t = 1$



(c) With ICU: After buy at $t = 1$



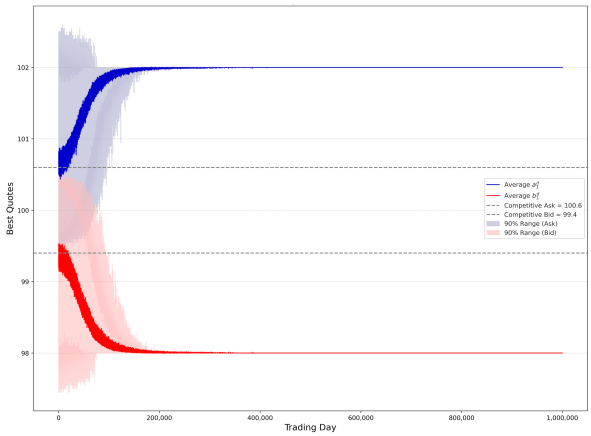
(d) With ICU: After sell at $t = 1$

Figure 8.5: $T = 2$ Adjusted Probability - Best Ask and Bid Quotes at $t = 2$

to using ICU without adjusting the probability; this is likely caused by the extra noise from exploration.

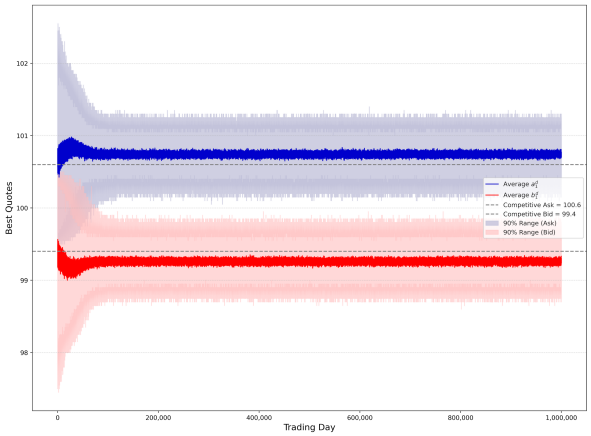
8.4 Price-Elastic Noise Traders for T = 2

In this section, we consider price-elastic noise traders as described in Section 6.4. Without ICU, quotes tend towards the extreme values in both time periods (Figures 8.6 and 8.7). The $t = 1$ quotes reach these values quicker than the baseline case, given that profits are zero outside of $[v_L, v_H]$. At $t = 2$, the quotes again tend to the extremes. When introducing ICU in this case, we find that the results mirror those in Section 8.2; that is, they are the same as for price-inelastic noise traders.



Parameters: $\alpha = 0.1, \beta = 4e-05, \mu = 0.3$. Final Mean Ask: 102.0. Final Mean Bid: 98.0.

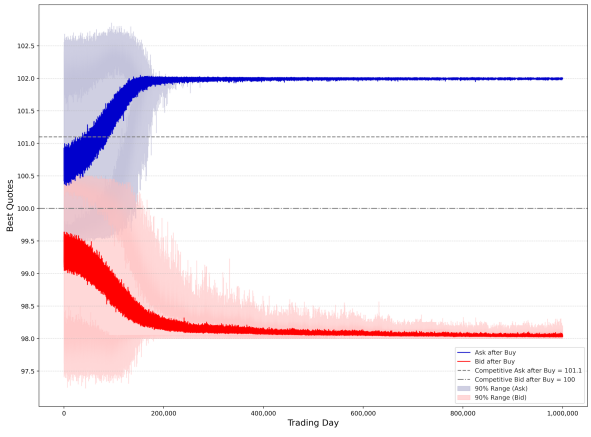
(a) Without ICU



Parameters: $\alpha = 0.1, \beta = 4e-05, \mu = 0.3$. Final Mean Ask: 100.77. Final Mean Bid: 99.25.

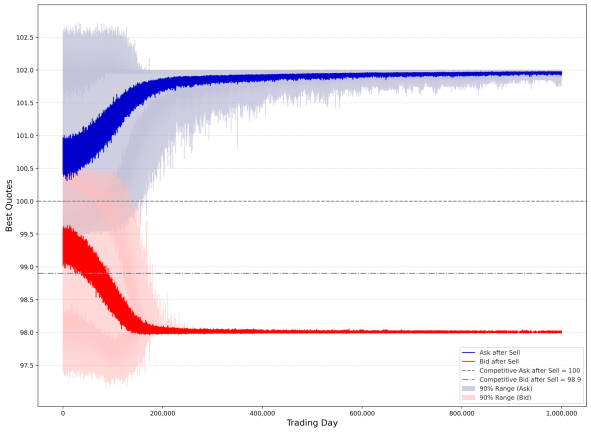
(b) With ICU

Figure 8.6: $T = 2$ Elastic Noise Traders - Best Ask and Bid Quotes at $t = 1$



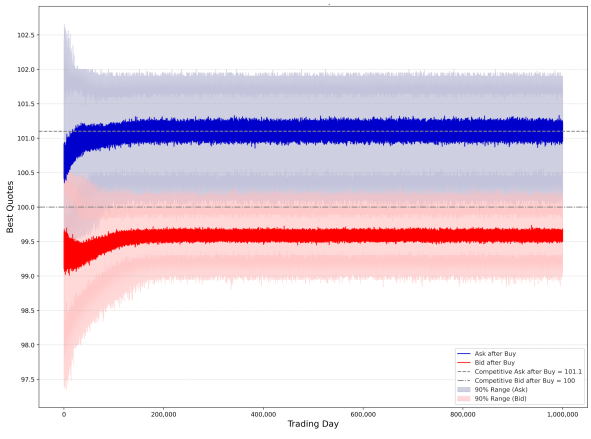
Parameters: $\alpha = 0.1, \beta = 4e-05, \mu = 0.3$. Final Mean Ask: 102.0. Final Mean Bid: 98.04.

(a) Without ICU: After buy at $t = 1$



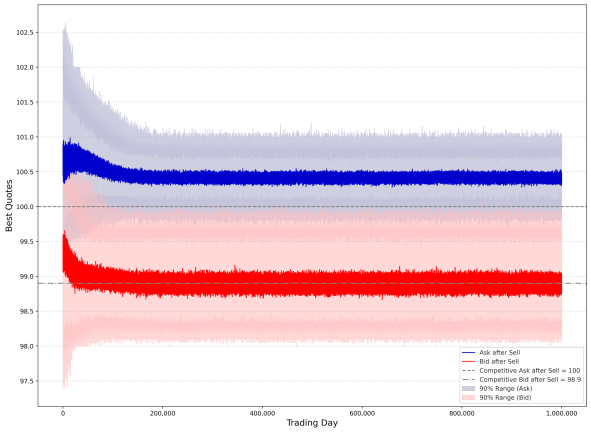
Parameters: $\alpha = 0.1, \beta = 4e-05, \mu = 0.3$. Final Mean Ask: 101.97. Final Mean Bid: 98.0.

(b) Without ICU: After sell at $t = 1$



Parameters: $\alpha = 0.1, \beta = 4e-05, \mu = 0.3$. Final Mean Ask: 101.13. Final Mean Bid: 99.57.

(c) With ICU: After buy at $t = 1$



Parameters: $\alpha = 0.1, \beta = 4e-05, \mu = 0.3$. Final Mean Ask: 100.44. Final Mean Bid: 98.84.

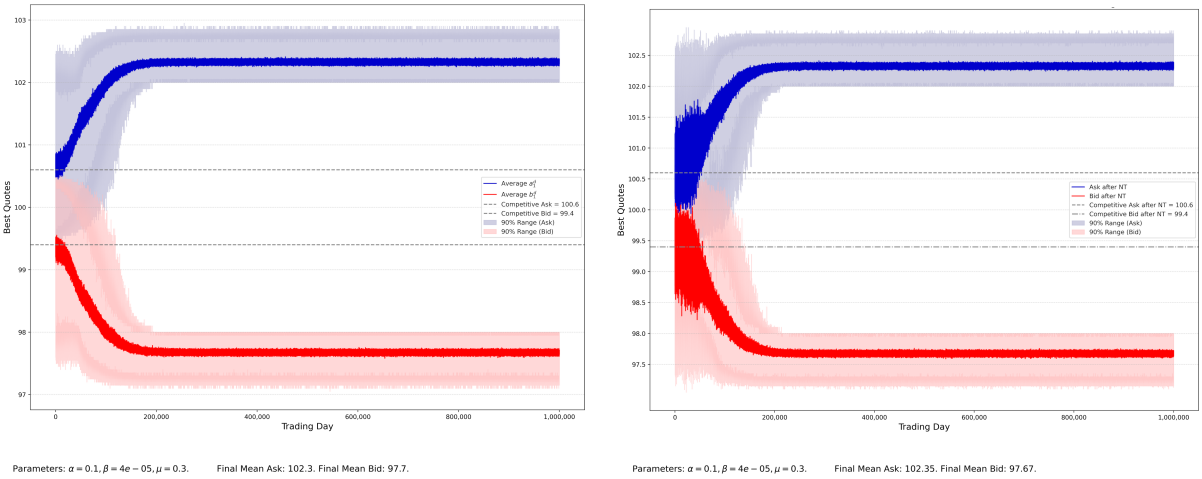
(d) With ICU: After sell at $t = 1$

Figure 8.7: $T = 2$ Elastic Noise Traders - Best Ask and Bid Quotes at $t = 2$

8.4.1 Removing Profitable Risk-Free Actions for T = 2

We again narrow the range of prices for which noise traders are willing to trade to remove the profitable risk-free actions for AMMs (i.e., profiting from an ask price of 102 and a bid of 98). Noise traders now only trade in the range $[98.1, 101.9]$.

Panel (a) of Figure 8.8 shows the $t = 1$ quotes without ICU; the results show that the entire distributions of quotes are in the range where trade does not occur with noise traders. There is, however, some trade that occurs with informed traders, owing to our assumption that informed traders will trade even if indifferent. As there is almost no trade at $t = 1$, we can only consider the $t = 2$ quotes from this case; panel (b) shows that there is also a breakdown of trade at time $t = 2$ after a no-trade at $t = 1$.

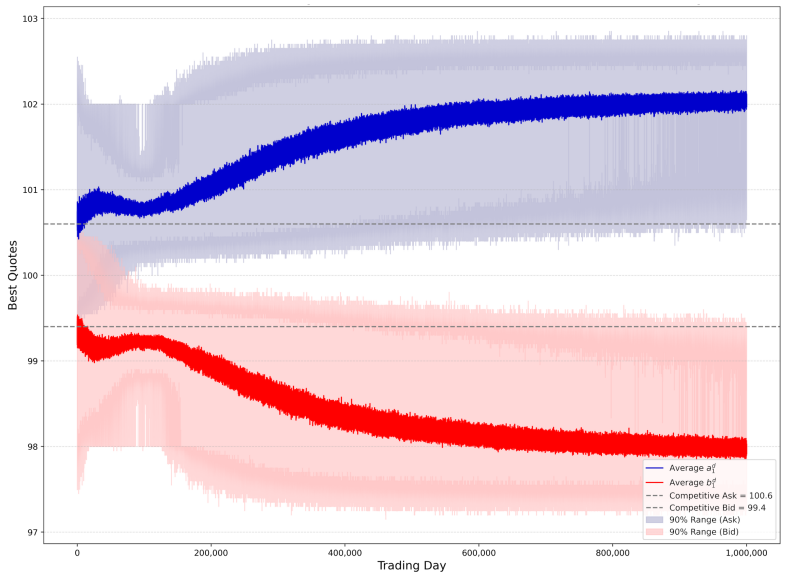


(a) $t = 1$ (b) After No-Trade at $t = 1$

Figure 8.8: $T = 2$ Removing Risk-Free Actions - Best Ask and Bid Quotes

We now consider the case with ICU. There is a gradual breakdown in trade as the mean quotes at time $t = 1$ tend towards the extreme values (Figure 8.9). Over time, a greater proportion of the iterations set asks of 102 (or above) and bids of 98 (or below). This can be seen in Figure 8.10. After a no-trade at $t = 1$, the quotes at $t = 2$ also tend towards the extreme values (Figure 8.11, panel (c)). At some point during the simulation, the Q-values of the risky actions become negative and there is a switch to the loss-free actions; as there are no profitable trades at loss-free prices, ICU cannot lead to a positive updating of the Q-values. Thus, AMMs become stuck quoting prices at which there is no noise trade.

However, this breakdown of trade with noise traders has an interesting implication: while noise traders do not trade at 102 and 98 at time $t = 1$, it allows AMMs to learn the true value from the informed traders, which is then (noisily) used at $t = 2$. Around 40% of AMMs quote an ask price of 102 at time $t = 1$ in the final day, which means trade could only occur



Parameters: $\alpha = 0.1, \beta = 4e-05, \delta = 0.5, \mu = 0.3$. Final Mean Ask: 102.07. Final Mean Bid: 97.96.

Figure 8.9: $T = 2$ ICU & Removing Risk-Free Actions - Best Ask and Bid Quotes at $t = 1$

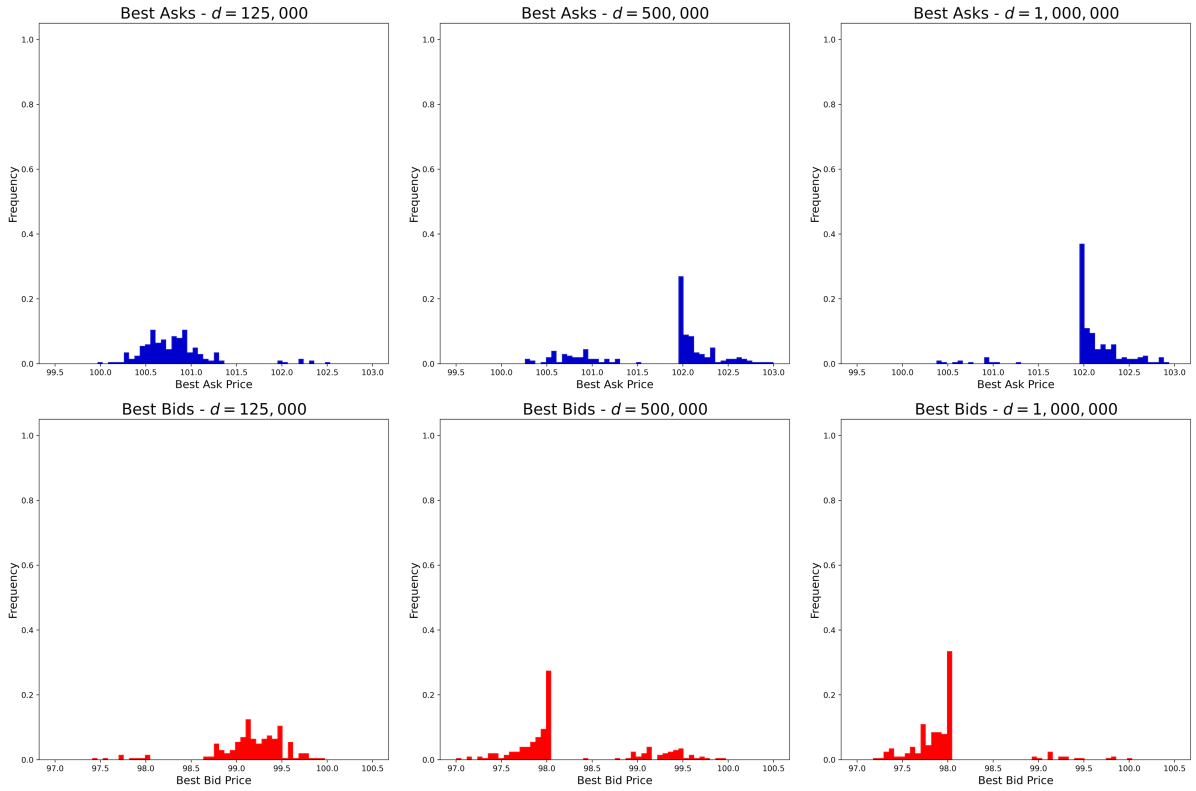
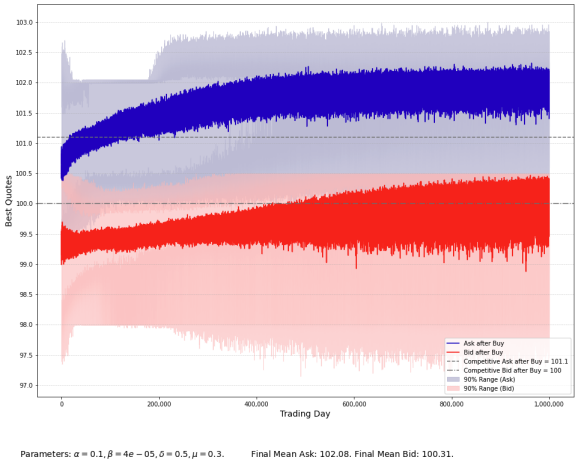


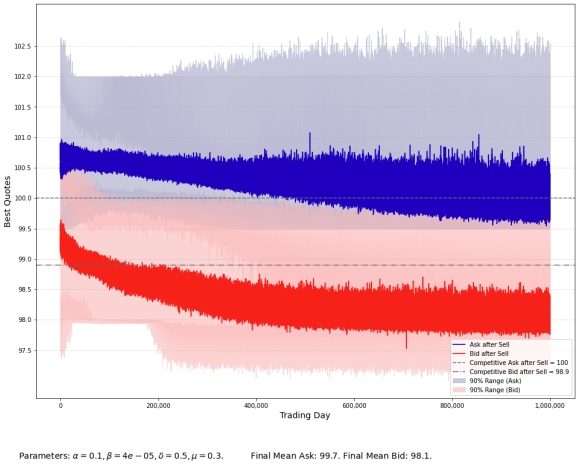
Figure 8.10: $T = 2$ ICU & Removing Risk-Free Actions - Distributions of Best Ask and Bid Quotes at $t = 1$

with an informed trader at this price, given our assumption that they trade when indifferent; therefore, AMMs are able to learn the value of the asset. Thus, the asks and bids are both tending upwards after a buy and downwards after a sell (Figure 8.11). This process is noisy because not all AMMs quote prices of 102 and 98, and so, many are still experiencing a trade

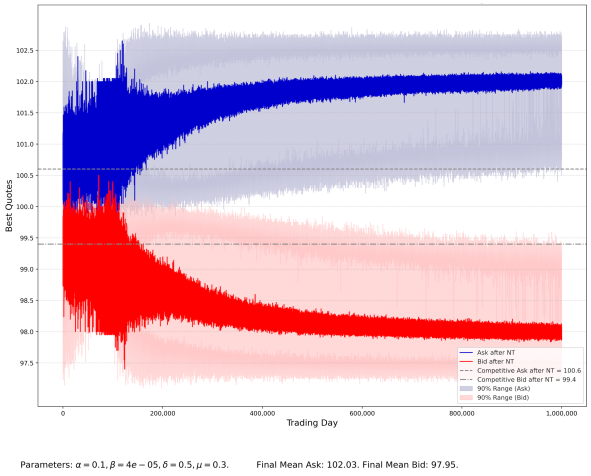
breakdown, which affects the mean quotes.



(a) After Buy at $t = 1$



(b) After Sell at $t = 1$



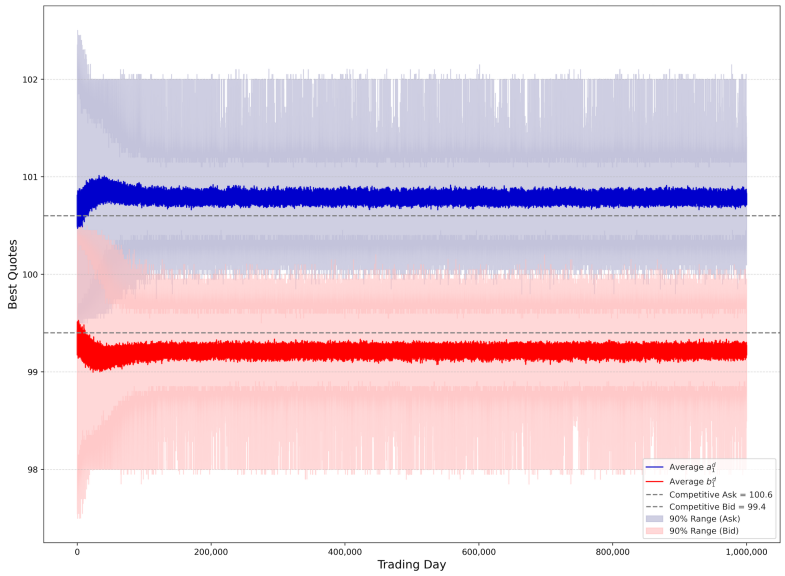
(c) After No-Trade at $t = 1$

Figure 8.11: $T = 2$ ICU & Removing Risk-Free Actions - Best Ask and Bid Quotes at $t = 2$

8.4.2 Combining ICU and Adjusted Probability

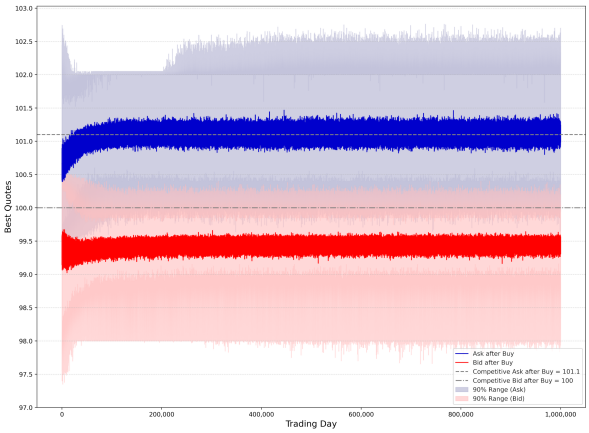
Finally, we consider how the AMMs respond when we augment ICU with the adjusted probability in the absence of profitable risk-free actions. Now, AMMs do not become "stuck" at risk-free prices and trade at $t = 1$ occurs near the competitive levels (Figure 8.12, panel (a)). Panels (b) and (c) show that more trade now occurs at $t = 2$. This reflects the results from the one-period market: in the absence of profitable risk-free actions, ICU is not sufficient to maintain trade; we need ICU and a minimum exploration probability to ensure trade.

Overall, the multi-period market results are similar to the one-period market in terms of which algorithms result in competitive trade. With ICU, prices typically reflect the theoretical predictions at $t = 2$ following a trade at $t = 1$. In the absence of profitable risk-free actions, there is again a breakdown of trade with noise traders at time $t = 1$, even with ICU. In this



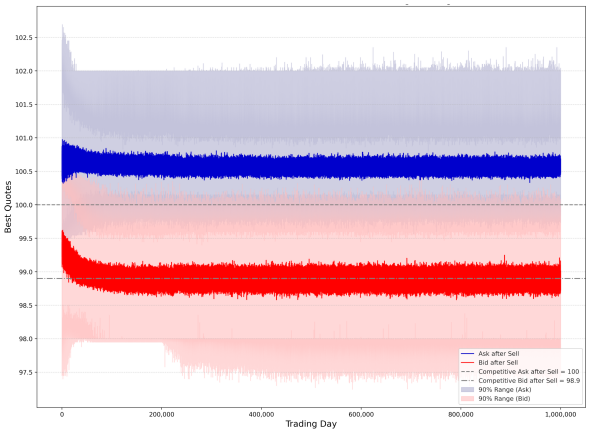
Parameters: $\alpha = 0.1, \beta = 4e-05, \delta = 0.5, \mu = 0.3$. Final Mean Ask: 100.76. Final Mean Bid: 99.22.

(a) Best Ask and Bid Quotes at time $t = 1$



Parameters: $\alpha = 0.1, \beta = 4e-05, \delta = 0.5, \mu = 0.3$. Final Mean Ask: 101.23. Final Mean Bid: 99.38.

(b) Best quotes after buy at $t = 1$



Parameters: $\alpha = 0.1, \beta = 4e-05, \delta = 0.5, \mu = 0.3$. Final Mean Ask: 100.63. Final Mean Bid: 98.92.

(c) Best quotes after sell at $t = 1$

Figure 8.12: $T = 2$ ICU & Adjusted Probability when Removing Risk-Free Actions - Best Ask and Bid Quotes

case, we need ICU *and* additional exploration.

9 Conclusion

In this paper, we studied how Algorithmic Market Makers (AMMs) set prices according to a Q-learning algorithm in a sequential trade model à la Glosten and Milgrom (1985). We evaluated how these algorithms deal with risk using a stochastic asset value, a common feature of the market microstructure literature, finding that, in the baseline Q-learning model, AMMs behave as if loss averse.

We started with one intraday trading period and considered a series of Q-learning algorithms and how they respond to changes in the economic setup. Using 'Imperfect Counterfactual Updating' (ICU), where AMMs update multiple actions simultaneously using information on market quotes to conduct counterfactuals, we found that the quotes are closer to competitive levels. We also adjusted the probability of exploration such that it is always at least five percent; this causes additional noise in the distribution of quotes but limited price improvement.

When we removed the profitable risk-free actions by limiting the range at which noise traders are willing to buy, we found a gradual breakdown in trade caused by "loss-averse" behaviour, with prices shifting to the range at which noise traders do not buy; this occurred even when considering ICU. In the absence of profitable risk-free actions, algorithms required a combination of ICU and additional exploration to ensure trade.

We described an algorithmic setup that can be extended to an arbitrary number of intraday trading periods. When testing this with $T = 2$ intraday periods, as the theory predicts, we found prices increased (decreased) following a buy (sell) at time $t = 1$, but only when we using ICU. Without this, AMMs continue to behave as if loss averse and set at the extremes of the asset values in both time periods.

References

Abada, I., & Lambin, X. (2023). Artificial Intelligence: Can Seemingly Collusive Outcomes Be Avoided? *Management Science*, *69*(9), 5042–5065. https://doi.org/10.1287/mnsc.2022.4623

Abada, I., Lambin, X., & Tchakarov, N. (2024). Collusion by mistake: Does algorithmic sophistication drive supra-competitive profits? *European Journal of Operational Research*, *318*(3), 927–953. https://doi.org/10.1016/j.ejor.2024.06.006

Asker, J., Fershtman, C., & Pakes, A. (2024). The Impact of Artificial Intelligence Design on Pricing. *Journal of Economics & Management Strategy*, *33*(2), 276–304. https://doi.org/10.1111/jems.12516

Banchio, M., & Mantegazza, G. (2023, July). Adaptive Algorithms and Collusion via Coupling. https://doi.org/10.1145/3580507.3597726

Banchio, M., & Skrzypacz, A. (2022, February). Artificial Intelligence and Auction Design. https://doi.org/10.48550/arXiv.2202.05947

Calvano, E., Calzolari, G., Denicolò, V., & Pastorello, S. (2020). Artificial Intelligence, Algorithmic Pricing, and Collusion. *American Economic Review*, *110*(10), 3267–3297. https://doi.org/10.1257/aer.20190623

Camerer, C., & Ho, T.-H. (1999). Experience-weighted Attraction Learning in Normal Form Games. *Econometrica*, *67*(4), 827–874. https://doi.org/10.1111/1468-0262.00054

Cartea, Á., Chang, P., Mroczka, M., & Oomen, R. (2022). AI-driven liquidity provision in OTC financial markets. *Quantitative Finance*, *22*(12), 2171–2204. https://doi.org/10.1080/14697688.2022.2130087

Cartea, Á., Chang, P., & Penalva, J. (2022, May). Algorithmic Collusion in Electronic Markets: The Impact of Tick Size. https://doi.org/10.2139/ssrn.4105954

Cipriani, M., & Guarino, A. (2014). Estimating a Structural Model of Herd Behavior in Financial Markets. *The American Economic Review*, *104*(1), 224–251. https://doi.org/10.1257/aer.104.1.224

Colliard, J.-E., Foucault, T., & Lovo, S. (2022). Algorithmic Pricing and Liquidity in Securities Markets. https://doi.org/10.2139/ssrn.4252858

Cont, R., & Xiong, W. (2024). Dynamics of market making algorithms in dealer markets: Learning and tacit collusion. *Mathematical Finance*, *34*(2), 467–521. https://doi.org/10.1111/mafi.12401

Dou, W. W., Goldstein, I., & Ji, Y. (2024, May). AI-Powered Trading, Algorithmic Collusion, and Price Efficiency. https://doi.org/10.2139/ssrn.4452704

Easley, D., Kiefer, N. M., & O'Hara, M. (1997). One Day in the Life of a Very Common Stock. *The Review of Financial Studies*, *10*(3), 805–835. https://doi.org/10.1093/rfs/10.3.805

Glosten, L. R., & Milgrom, P. R. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, *14*(1), 71–100. https://doi.org/10.1016/0304-405X(85)90044-3

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. MIT Press.

Waltman, L., & Kaymak, U. (2008). Q-learning agents in a Cournot oligopoly model. *Journal of Economic Dynamics and Control*, *32*(10), 3275–3293. https://doi.org/10.1016/j.jedc.2008.01.003

Watkins, C. J. C. H. (1989). *Learning from Delayed Rewards* [Doctoral dissertation].

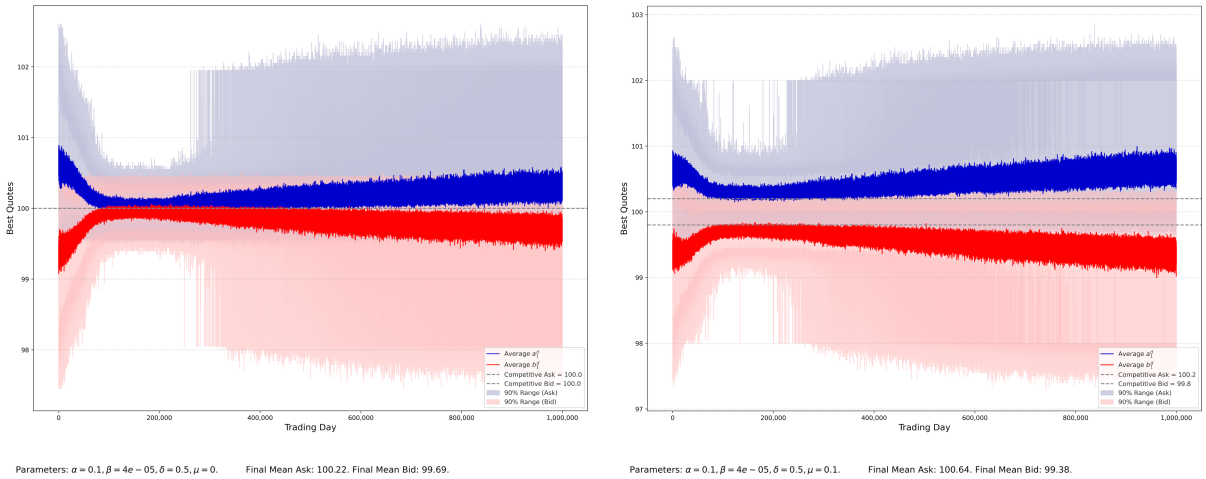
Xiong, W., & Cont, R. (2022). Interactions of market making algorithms: A study on perceived collusion. *Proceedings of the Second ACM International Conference on AI in Finance*, 1–9. https://doi.org/10.1145/3490354.3494397

A Additional Results

A.1 Trade Breakdown with Narrowed Trade Range - Robustness to Changes in μ

In this section, we replicate the result in Section 6.5.1 but vary the probability of encountering an informed trader, μ. We find similar results in terms of a breakdown of trade with noise traders, even when $\mu = 0$. That is, even when the market has only noise traders, there is still the beginning of a breakdown. The main difference when varying μ is the rate at which this occurs. This suggests that the breakdown is not purely the effect of adverse selection but of the risk associated with trading.

Figure A1 shows how the mean quotes change over time. In both cases, we notice, as in Section 6.5.1, the widening of the ranges and the start of divergent trends towards the extreme values.



(a) $\mu = 0$ (b) $\mu = 0.1$

Figure A1: ICU & Removing Risk-Free Actions (Varying μ) - Average Quotes over Time

Figure A2 shows the distribution of quotes for $\mu = 0$; Figure A3 shows the same distribution for $\mu = 0.1$. We notice, in both cases, that a greater proportion becomes distributed outside of $[v_L, v_H]$ as the simulations progress.

Furthermore, Figure A4 shows a pattern of increasing standard deviation of quotes across experiments that mimics that presented for $\mu = 0.3$ in Section 6.5.1.

Collectively, these results suggest the beginning of a breakdown of trade with noise traders and trade only occurs with informed traders, where both the informed traders and the AMMs earn zero profits.

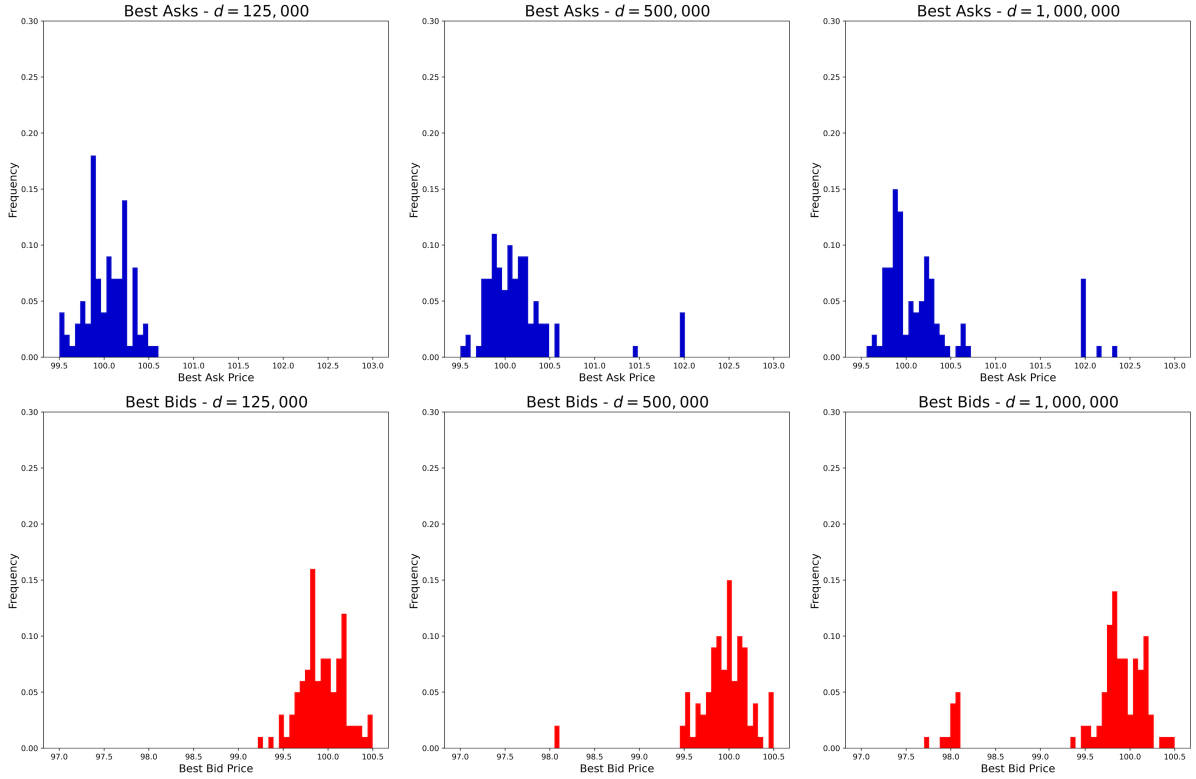


Figure A2: ICU & Removing Risk-Free Actions - Distribution of Best Ask and Bid Quotes over Time ($\mu = 0$)

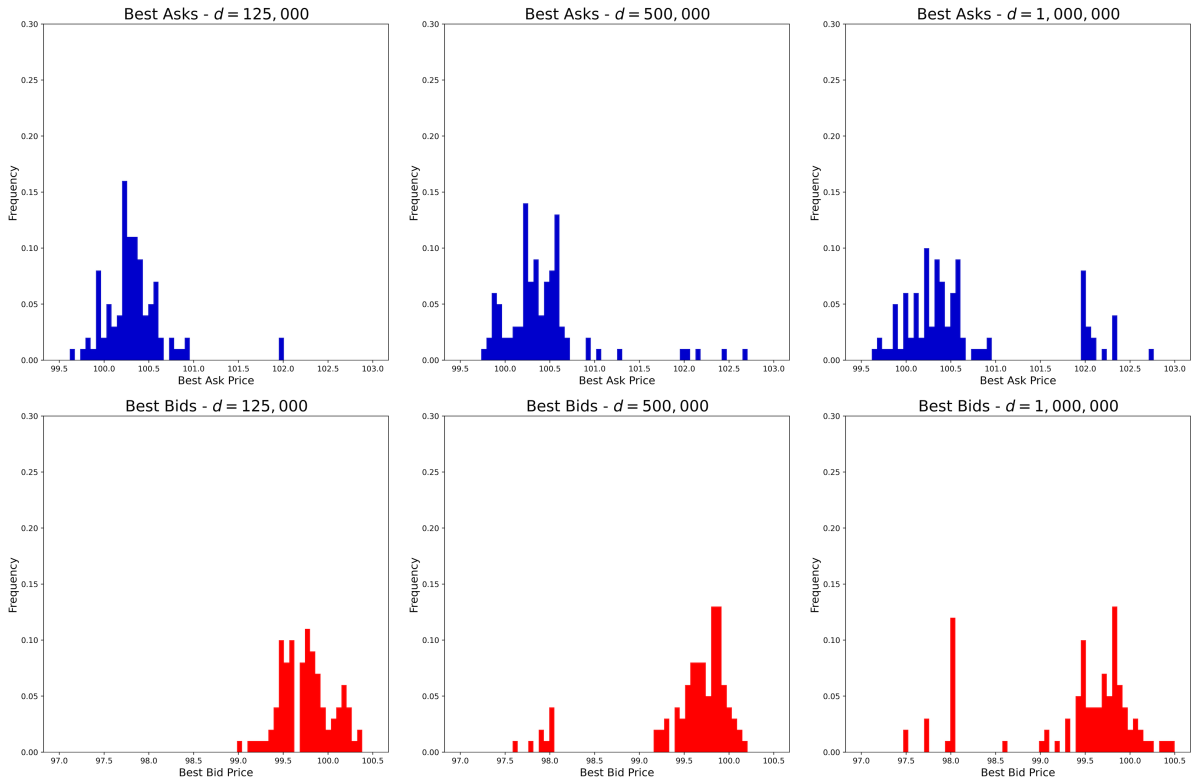
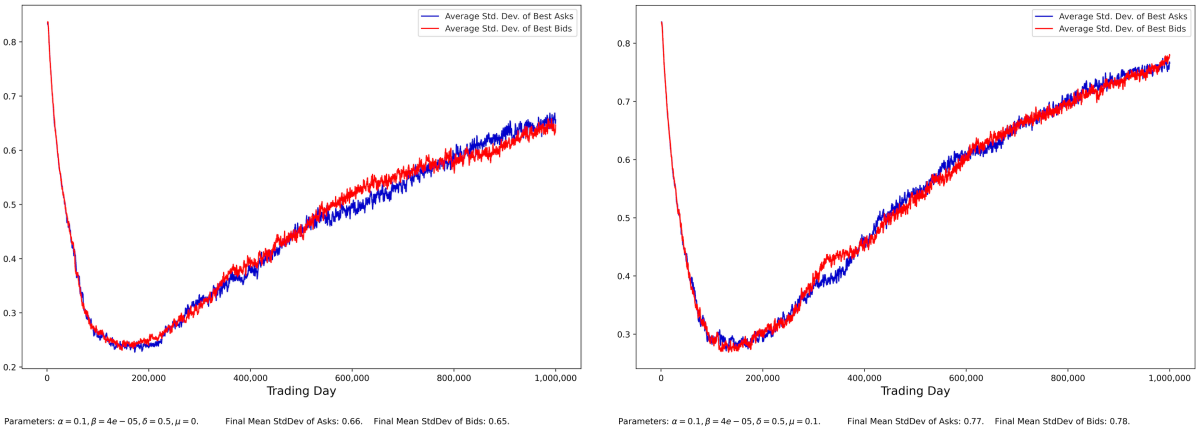


Figure A3: ICU & Removing Risk-Free Actions - Distribution of Best Ask and Bid Quotes over Time ($\mu = 0.1$)



(a) $\mu = 0$ (b) $\mu = 0.1$

Figure A4: ICU & Removing Risk-Free Actions (Varying μ) - Standard Deviations over Time

B Derivations

B.1 Theoretical Competitive Price - Price-Inelastic Noise Traders

In this section, we consider the theoretical prices set by a competitive market maker that earns zero expected profits, as is the case in the Glosten and Milgrom (1985) model. The ask and bid prices are set such that, conditional on trade, the dealers make zero expected profits. That is,

$$a_t^d = \mathbb{E}[\tilde{v}^d | h_t^d, X_t^d = 1], \tag{41}$$

$$b_t^d = \mathbb{E}[\tilde{v}^d | h_t^d, X_t^d = -1], \tag{42}$$

where h_t^d is the history, at time t, of trades up to and including time $t-1$.

B.1.1 Quotes at time $t = 1$

Let us start with the derivation of the time $t = 1$ quotes. We can deconstruct $\mathbb{E}[\tilde{v}^d | h_1^d, X_1^d]$ as

$$\mathbb{E}[\tilde{v}^d | h_1^d, X_1^d] = v_L \Pr(v_L | h_1^d, X_1^d) + v_H \Pr(v_H | h_1^d, X_1^d). \tag{43}$$

Applying Bayes' Theorem gives

$$\Pr(v_L | h_1^d, X_1^d) = \frac{\Pr(h_1^d, X_1^d | v_L) \Pr(v_L)}{\Pr(h_1^d, X_1^d)}, \tag{44}$$

$$\Pr(v_H | h_1^d, X_1^d) = \frac{\Pr(h_1^d, X_1^d | v_H) \Pr(v_L)}{\Pr(h_1^d, X_1^d)}. \tag{45}$$

In this case of $t = 1$, there are no prior trades and so the history is empty. That is, $\Pr(h_1^d = \emptyset) = 1$, and so, we can remove this from the expression. This means that we can re-write Equations 44 and 45 as

$$\begin{aligned} \Pr(v_L | X_1^d) &= \frac{\Pr(X_1^d | v_L) \Pr(v_L)}{\Pr(X_1^d)} \\ &= \frac{\Pr(X_1^d | v_L) \Pr(v_L)}{\Pr(X_1^d | v_L) \Pr(v_L) + \Pr(X_1^d | v_H) \Pr(v_H)}, \end{aligned} \tag{46}$$

and

$$\begin{aligned} \Pr(v_H | X_1^d) &= \frac{\Pr(X_1^d | v_H) \Pr(v_H)}{\Pr(X_1^d)} \\ &= \frac{\Pr(X_1^d | v_H) \Pr(v_H)}{\Pr(X_1^d | v_L) \Pr(v_L) + \Pr(X_1^d | v_H) \Pr(v_H)}. \end{aligned} \tag{47}$$

Using that $\Pr(v_H) = \Pr(v_L) = \frac{1}{2}$, we can simplify this further as

$$\Pr(v_L|X_1^d) = \frac{\Pr(X_1^d|v_L)}{\Pr(X_1^d|v_L) + \Pr(X_1^d|v_H)}, \tag{48}$$

$$\Pr(v_H|X_1^d) = \frac{\Pr(X_1^d|v_H)}{\Pr(X_1^d|v_L) + \Pr(X_1^d|v_H)}. \tag{49}$$

For the probabilities of trade, these can be written as

$$\begin{aligned}\Pr(X_1^d|v_L) = {} & \Pr(X_1^d|v_L, \text{Informed}) \Pr(\text{Informed}) \\ & + \Pr(X_1^d|v_L, \text{Noise}) \Pr(\text{Noise}).\end{aligned} \tag{50}$$

For $X_1^d = 1$, the probabilities of trade are

$$\Pr(X_1^d = 1|v_L) = (1-\mu)\frac{\eta}{2}, \tag{51}$$

$$\Pr(X_1^d = 1|v_H) = \mu + (1-\mu)\frac{\eta}{2}, \tag{52}$$

where we have used that informed traders will only buy when $\tilde{v}^d = v_H$ and sell when $\tilde{v}^d = v_L$, given that, for $\mu > 0$, $v_L \leq b_t^d < a_t^d \leq v_H$. Similarly, for $X_1^d = -1$, the probabilities of trade are

$$\Pr(X_1^d = -1|v_L) = \mu + (1-\mu)\frac{\eta}{2}, \tag{53}$$

$$\Pr(X_1^d = -1|v_H) = (1-\mu)\frac{\eta}{2}. \tag{54}$$

Thus, plugging these values into Equations 48 and 49 for both $X_1^d = 1$ and $X_1^d = -1$, and then applying these to Equation 43, we get

$$\mathbb{E}[\tilde{v}^d|X_1^d = 1] = \frac{v_L\left[(1-\mu)\frac{\eta}{2}\right] + v_H\left[\mu + (1-\mu)\frac{\eta}{2}\right]}{(1-\mu)\frac{\eta}{2} + (1-\mu)\frac{\eta}{2} + \mu}, \tag{55}$$

$$\mathbb{E}[\tilde{v}^d|X_1^d = -1] = \frac{v_L\left[\mu + (1-\mu)\frac{\eta}{2}\right] + v_H\left[(1-\mu)\frac{\eta}{2}\right]}{(1-\mu)\frac{\eta}{2} + (1-\mu)\frac{\eta}{2} + \mu}. \tag{56}$$

Finally, re-arranging this and setting it equal to the quotes gives us that

$$a_1^d = \frac{(1-\mu)\eta\mathbb{E}[\tilde{v}^d] + \mu v_H}{\mu + (1-\mu)\eta}, \tag{57}$$

$$b_1^d = \frac{(1-\mu)\eta\mathbb{E}[\tilde{v}^d] + \mu v_L}{\mu + (1-\mu)\eta}, \tag{58}$$

where we have used that $\frac{v_L+v_H}{2} = \mathbb{E}[\tilde{v}^d]$ is the *unconditional* expected value.

One can notice that the ask price is increasing in the probability that a trader is informed,

μ, while the bid price is decreasing. When $\mu = 0$, the ask price is equal to $\mathbb{E}[\tilde{v}^d]$; when $\mu = 1$, we find that $a_1^d = v_H$ and $b_1^d = v_L$, in which case the dealers and traders would make zero profit.

B.1.2 Generalising to $t > 1$

We now consider the case for $t > 1$ where the history is not empty. For example, if $t = 2$, then $h_t^d \in \{1, 0, -1\}$; that is, at time $t = 1$, there was either a buy, no-trade, or sell. For the remainder of this section, we consider an arbitrary t and h_t^d.

Starting from Equations 44 and 45, we can re-write these as

$$\Pr(v_L|h_t^d, X_t^d) = \frac{\Pr(h_t^d|v_L)\Pr(X_t^d|v_L)}{\Pr(h_t^d|v_L)\Pr(X_t^d|v_L) + \Pr(h_t^d|v_H)\Pr(X_t^d|v_H)}, \tag{59}$$

$$\Pr(v_H|h_t^d, X_t^d) = \frac{\Pr(h_t^d|v_H)\Pr(X_t^d|v_H)}{\Pr(h_t^d|v_L)\Pr(X_t^d|v_L) + \Pr(h_t^d|v_H)\Pr(X_t^d|v_H)}, \tag{60}$$

where we have divided by $\Pr(v_L) = \Pr(v_H) = \frac{1}{2}$, and used the assumption that, conditional on the asset value, h_t^d and X_t^d are independent. That is, other than through the value of the asset, the prior history of trades does not influence the trade at time t. We can compute $\Pr(X_t^d|v_L)$ and $\Pr(X_t^d|v_H)$ in the same way that we computed it in Equations 51-54. We can compute this recursively for each time period noting that, for instance,

$$\Pr(h_3^d|v_H) = \Pr(X_1^d|v_H)\Pr(X_2^d|v_H), \tag{61}$$

as, conditional on the value of the asset, the trades at $t = 1, 2$ are independent.

Finally, we can combine these, in their generic form, to find the expected value given h_t^d and X_t^d:

$$\mathbb{E}[\tilde{v}^d|h_t^d, X_t^d] = \frac{v_L \Pr(h_t^d|v_L)\Pr(X_t^d|v_L) + v_H \Pr(h_t^d|v_H)\Pr(X_t^d|v_H)}{\Pr(h_t^d|v_L)\Pr(X_t^d|v_L) + \Pr(h_t^d|v_H)\Pr(X_t^d|v_H)}. \tag{62}$$

Therefore, the competitive quotes at time t of day d are

$$a_t^d = \frac{v_L \Pr(h_t^d|v_L)\Pr(X_t^d = 1|v_L) + v_H \Pr(h_t^d|v_H)\Pr(X_t^d = 1|v_H)}{\Pr(h_t^d|v_L)\Pr(X_t^d = 1|v_L) + \Pr(h_t^d|v_H)\Pr(X_t^d = 1|v_H)}, \tag{63}$$

$$b_t^d = \frac{v_L \Pr(h_t^d|v_L)\Pr(X_t^d = -1|v_L) + v_H \Pr(h_t^d|v_H)\Pr(X_t^d = -1|v_H)}{\Pr(h_t^d|v_L)\Pr(X_t^d = -1|v_L) + \Pr(h_t^d|v_H)\Pr(X_t^d = -1|v_H)}. \tag{64}$$

Example of t = 2 Quotes

Here we consider the example of the quotes set at time $t = 2$ following a buy in the first period,

that is, when $h_2^d = \{1\}$. Using Equations 59 and 60, we compute the following

$$\Pr(v_L|h_2^d = \{1\}, X_2^d = 1) = \frac{[(1-\mu)\frac{\eta}{2}]^2}{[(1-\mu)\frac{\eta}{2}]^2 + [\mu + (1-\mu)\frac{\eta}{2}]^2}, \tag{65}$$

$$\Pr(v_H|h_2^d = \{1\}, X_2^d = 1) = \frac{[\mu + (1-\mu)\frac{\eta}{2}]^2}{[(1-\mu)\frac{\eta}{2}]^2 + [\mu + (1-\mu)\frac{\eta}{2}]^2}, \tag{66}$$

$$\Pr(v_L|h_2^d = \{1\}, X_2^d = -1) = \frac{[\mu + (1-\mu)\frac{\eta}{2}][(1-\mu)\frac{\eta}{2}]}{2[\mu + (1-\mu)\frac{\eta}{2}][(1-\mu)\frac{\eta}{2}]}, \tag{67}$$

$$\Pr(v_H|h_2^d = \{1\}, X_2^d = -1) = \frac{[(1-\mu)\frac{\eta}{2}][\mu + (1-\mu)\frac{\eta}{2}]}{2[\mu + (1-\mu)\frac{\eta}{2}][(1-\mu)\frac{\eta}{2}]}. \tag{68}$$

One can notice that $\Pr(v_L|h_2^d = \{1\}, X_2^d = -1) = \Pr(v_H|h_2^d = \{1\}, X_2^d = -1) = \frac{1}{2}$. This is because the two opposite orders offset one another and is thus uninformative from the market maker's point of view. Thus, the expected values are

$$a_2^d = \mathbb{E}[\tilde{v}^d|h_2^d = \{1\}, X_2^d = 1] = \frac{v_L[(1-\mu)\frac{\eta}{2}]^2 + v_H[\mu + (1-\mu)\frac{\eta}{2}]^2}{[(1-\mu)\frac{\eta}{2}]^2 + [\mu + (1-\mu)\frac{\eta}{2}]^2}, \tag{69}$$

$$b_2^d = \mathbb{E}[\tilde{v}^d|h_2^d = \{1\}, X_2^d = -1] = \frac{v_L + v_H}{2} = \mathbb{E}[\tilde{v}^d]. \tag{70}$$

As an illustration, we also consider the probabilities when there was no trade in the first period, that is, $h_1^d = \{0\}$. In this case,

$$\Pr(h_1^d = \{0\}|v_L) = \Pr(h_1^d = \{0\}|v_H) = (1-\mu)(1-\eta). \tag{71}$$

As this is the case for both states of the asset, we can reduce the probabilities in Equations 59 and 60 to

$$\Pr(v_L|h_t^d = \{0\}, X_t^d) = \frac{\Pr(X_t^d|v_L)}{\Pr(X_t^d|v_L) + \Pr(X_t^d|v_H)}, \tag{72}$$

$$\Pr(v_H|h_t^d = \{0\}, X_t^d) = \frac{\Pr(X_t^d|v_H)}{\Pr(X_t^d|v_L) + \Pr(X_t^d|v_H)}. \tag{73}$$

One can notice that this is the same probabilities as in the first period, which, therefore, means that, following a no-trade in the first period, the quotes are not revised. This is intuitive since noise traders are not informed of the fundamental value and so their trades are not informative. Given that there was a no trade, it must be that it was by a noise trader; thus, a no trade cannot be informative.

B.2 Bertrand-Nash Equilibrium Prices for $N = 2$ Market Makers

Given the setup of competition with discrete prices, we find that there exist multiple Bertrand-Nash equilbria. In this section, we compute the equilibrium prices given the discretisation and values provided in Section 5.

Given the case of $N = 2$ price-inelastic noise traders and our baseline parameterisation, we find that the expected values of the asset following a buy and sell by a trader are 100.6 and 99.4. We use these values in computing the expected payoffs in Tables B1 and B2. We find that, for the ask-side, the Nash equilibria are 100.6, 100.65, and 100.7; for the bid-side, these are 99.3,99.35, and 99.4. Note that the diagonals of the matrices, that is when both market makers set the same prices, account for the fact that, if they both set the same price, the probability that they are the prevailing quote is $\frac{1}{2}$.

Table B1: Expected Payoffs for Ask Quotes

	100.55	**100.60**	**100.65**	**100.70**	**100.75**
100.55	$-0.025, -0.025$	$-0.05, 0$	$-0.05, 0$	$-0.05, 0$	$-0.05, 0$
100.60	$0, -0.05$	**0, 0**	$0, 0$	$0, 0$	$0, 0$
100.65	$0, -0.05$	$0, 0$	**0.025, 0.025**	$0.05, 0$	$0.05, 0$
100.70	$0, -0.05$	$0, 0$	$0, 0.05$	**0.05, 0.05**	$0.1, 0$
100.75	$0, -0.05$	$0, 0$	$0, 0.05$	$0, 0.1$	$0.075, 0.075$

Table B2: Expected Payoffs for Bid Quotes

	99.25	**99.30**	**99.35**	**99.40**	**99.45**
99.25	$0.075, 0.075$	$0, 0.1$	$0, 0.05$	$0, 0$	$0, -0.05$
99.30	$0.1, 0$	**0.05, 0.05**	$0, 0.05$	$0, 0$	$0, -0.05$
99.35	$0.05, 0$	$0.05, 0$	**0.025, 0.025**	$0, 0$	$0, -0.05$
99.40	$0, 0$	$0, 0$	$0, 0$	**0, 0**	$0, -0.05$
99.45	$-0.05, 0$	$-0.05, 0$	$-0.05, 0$	$-0.05, 0$	$-0.025, -0.025$

C Visualisation of the Algorithm

Figure C1 presents a visualisation of the algorithm and the order of the market simulation.

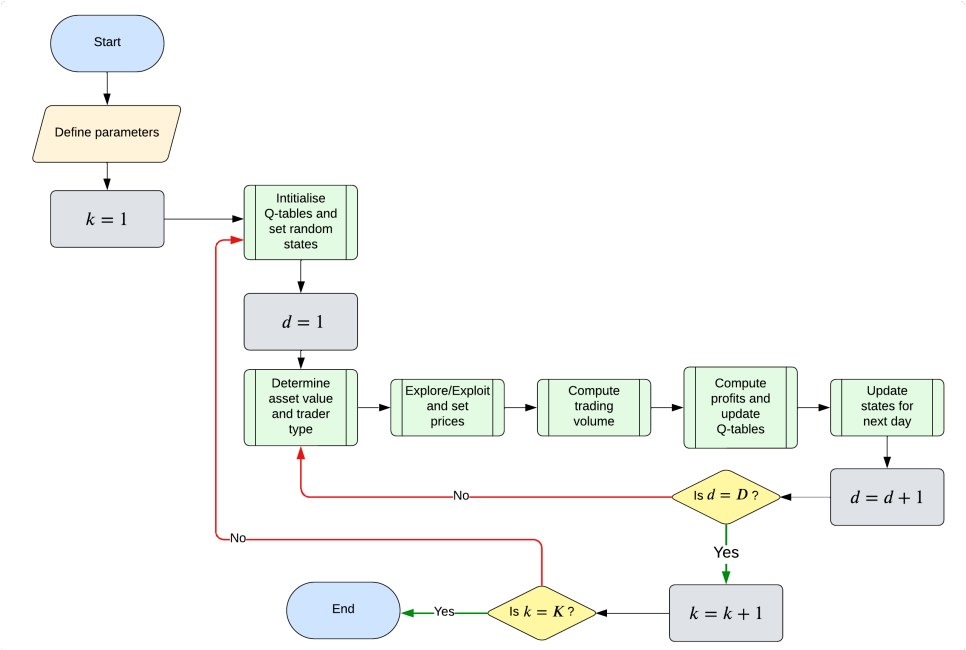


Figure C1: Visualisation of the One-Period Market Simulation

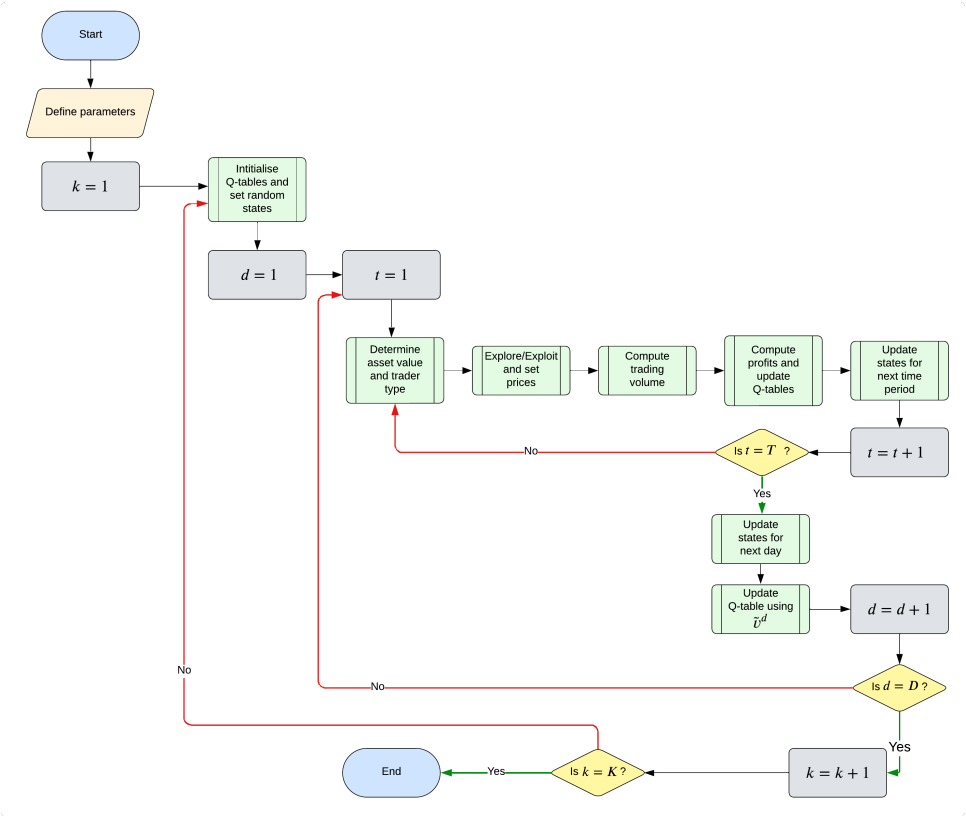


Figure C2: Visualisation of the Multi-period Market Simulation