How Q-Learning Algorithms Behave During Market Volatility Shocks

Authors: Francesco Borri and Maximilian Neville

Date: 06/25/2024

Institution: HEC Paris

Abstract

This thesis investigates the behaviour of Q-learning algorithmic trading systems during periods of market shocks and heightened volatility. Utilizing a Markov process to model volatility changes, the study simulates trading scenarios to understand the responsiveness, stability, and impact of Q-learning-based strategies on market dynamics. The findings reveal that while Q-learning algorithms exhibit robust adaptability and can stabilize prices under normal conditions, they tend to exacerbate market instability during periods of extreme volatility. This dual role underscores the necessity for hybrid models that integrate Q-learning with other strategies to enhance market resilience.

The study also explores the influence of learning and exploration rates on the performance of Q-learning algorithms, highlighting the trade-offs between rapid adaptation and stability. Furthermore, the interaction between multiple Q-learning algorithms was found to amplify market reactions, particularly during volatile periods, leading to significant price fluctuations.

These insights have important implications for both market participants and policymakers. The adaptability of Q-learning algorithms in stable markets suggests their potential for improving trading efficiency and liquidity provision. However, their destabilizing effects during market shocks call for the development of regulatory frameworks and safeguards to mitigate systemic risks. Future research should focus on creating hybrid models and exploring other reinforcement learning techniques to optimize trading strategies across varying market conditions.

Acknowledgements

We would like to thank our advisor, Professor Jean-Edouard Colliard, for his guidance and support throughout this research. Special thanks to our family and friends for their encouragement. We also appreciate the feedback from our peers and the assistance from the institution's library staff.

Table of Contents

Chapter 1: Introduction

1.1 Background

One of the most significant if not the most significant changes to market microstructure over the past two decades has been the introduction and proliferation of algorithms making trading decisions. Their growth can be attributed to the competitive advantages they lend to their users. Algorithms, when embedded into sophisticated trading systems can process data, interpret it then execute trades rapidly. Usage of mathematical models and high-frequency strategies, algorithmic trading has proven to be lucrative. This growth has not come without fault, it is also purported to be the cause of greater market volatility, especially during periods of market stress. Brogaard, Hendershott, and Riordan (2014) demonstrate how high-frequency algorithmic trading has contributed to increased volatility during market shocks. The way in which algorithmic traders respond to market stressors or shocks remains a key issue to be studied.

The first instances of algorithmic trading were seen towards the end of the 20th century as traditional floor or 'pit' trading, was beginning to be replaced by electronic trading. The transition to electronic exchanges set the stage for a new era, reinventing the way we interact with markets. This paved the way for the development of sophisticated algorithmic trading strategies. High frequency traders (HFTs) emerged in the early 2000s and were significant drivers of new regulation. A notable example includes the creation and implementation of the Markets in Financial Instruments Directive (MiFID) in Europe, which aimed to foster competitive and transparent markets.

The rapid improvements in computing power, network speed and connectivity, data productivity and various other technological leaps forward have provided fertile grounds upon which algorithmic trading has flourished these past two decades. There has also been a significant inflow of intellectual capital towards quantitative and algorithmic trading strategies. We have seen in more recent years the remarkable growth of machine learning techniques in this field. Reinforcement learning techniques such as Q-learning algorithms have helped make further advancements in trading algorithm sophistication.

It is likely that the advent of algorithmic trading has been a net positive when it comes to market efficiency. The frequency, speed and scale at which algorithms can execute in theory should lead to tighter prices, greater provisioning of liquidity and rapid price discovery, all features of efficient markets. This has been demonstrated by Hendershott, Jones, and Menkveld (2011), who's research showed that algorithmic trading improves liquidity across various markets globally.

We have also seen algorithms pose challenges and threats to the smooth running of markets. Not only are the increases in volatility of concern, but we have also seen flash crashes and instances of market manipulation. Kirilenko et al. (2017), investigated the May 2010 'Flash Crash'. Their study highlighted how algorithmic traders can exacerbate market instability. Whilst in stable markets, liquidity conditions have been improved by HFT, during this event they were the cause of an extreme crash due to a sudden withdrawal of liquidity.

As previously mentioned, algorithmic trading was a key driver of new regulation over the last 25 years. In Europe we have seen MiFID II and in the United States, the Dodd Frank Act. Both regulations aim to improve transparency, ensure the resilience of trading systems and prevent abusive market behaviour. Both regulations seek to strike a balance between the benefits given and the risks presented as a result of the existence of algorithmic trading strategies.

It is imperative that we understand the behaviour of algorithms during market shocks or periods of elevated volatility to best ensure that risks are managed appropriately, and market disruptions are avoided. This study focuses on the behaviour of Q-learning-based trading systems during market shocks, providing insights into their stability, performance, and impact on market volatility.

1.2 Research Problem

The primary question addressed in this thesis is how algorithmic trading systems, particularly those using Q-learning algorithms, behave during market shocks. Market shocks, often defined by sharp one-time price adjustments, can significantly alter the trading landscape and cause large and sustained increases in volatility. It is crucial to assess the stability and efficacy of these algorithms in such situations to prevent adding to market instability or causing unintended consequences.

The research problem encompasses several specific issues. Market shocks can occur due to various reasons such as economic news, geopolitical events, natural disasters, or unexpected shifts in market sentiment. Each type of shock may uniquely impact market dynamics, requiring Q-learning algorithms to adjust differently to these shocks. Q-learning algorithms consistently learn from past data and continuously update existing strategies based on real-time price streams. During market shocks, when changes are fast and unpredictable, these learning processes are challenged, sometimes leading to suboptimal decisions. The relevance of the 'learning rate'—how much importance is given to recent data compared to older data—becomes particularly significant in these moments. Furthermore, Q-learning algorithms operate within a trading environment that includes various other strategies, leading to complex dynamics. These interactions can increase the market's sensitivity to shocks, as feedback loops may amplify fluctuations.

Addressing this research problem is significant for several reasons. By examining the performance and stability of Q-learning algorithms in response to market shocks, this thesis aims to contribute to a more stable financial market. Insights from this research might help design trading algorithms that are more robust to unexpected and severe shocks, leading to the development of more adaptive algorithms that maintain optimal performance even under high market volatility. Additionally, the findings can assist regulators in developing policies to ensure the responsible use of algorithmic trading systems. This includes guidelines for algorithmic behaviour in response to market stressors and rules to prevent events like the 'Flash Crash.'

1.3 Objectives of the Study

The objectives of this study are threefold. Firstly, to investigate the responsiveness of Q-learning algorithms to market shocks, assessing how quickly and effectively they can adapt to sudden changes in market conditions such as fixed-time volatility changes and stochastic regime shifts modelled using Markov processes. Secondly, to evaluate the stability of algorithmic trading during periods of high volatility, determining whether these algorithms contribute to or mitigate market instability. Thirdly, to analyse the impact of algorithmic trading on market volatility during shocks, exploring the potential feedback loops and systemic risks introduced by algorithmic trading strategies, particularly how the interactions between multiple Q-learning algorithms can amplify market reactions.

Additionally, this study aims to investigate the liquidity provision and withdrawal behaviour of Q-learning algorithms during extreme volatility events and to evaluate the compliance of Q-learning algorithms with existing market regulations during periods of high volatility.

Our methodology builds upon the foundational work by Colliard et al. (2023), who critically examined Q-learning algorithm market-makers in the context of pricing and liquidity under adverse selection. While Colliard et al. primarily focused on static adverse selection costs and their impact on pricing, our study diverges by introducing dynamic volatility regimes to model market shocks more realistically. Specifically, we incorporate both fixed-time and stochastic regime changes to simulate the unpredictable nature of market volatility. This extension allows us to explore how Q-learning algorithms adapt their strategies not just under adverse selection, but also in the face of abrupt and significant market fluctuations. Our scientific contribution lies in this nuanced examination of algorithmic behaviour, providing deeper insights into the stability and performance of Q-learning algorithms under varying volatility conditions, thus filling a critical gap in the literature.

1.4 Research Questions

The research seeks to answer the following key questions:

1. How do Q-learning algorithms adjust their trading strategies in response to market shocks?

2. What is the performance of Q-learning-based trading algorithms under different volatility conditions?

3. How do these algorithms influence overall market stability during shocks?

The answers to the research questions are as follows:

1. Q-learning algorithms adapt their trading strategies by adjusting to sudden changes in market conditions through a process of learning and exploration. These algorithms continuously update their strategies based on new information. During market shocks, characterized by sharp and unpredictable changes, the algorithms rely heavily on their learning rate (alpha) and exploration rate (epsilon) to determine their responses. High learning rates allow for quick adaptation but can also lead to instability, while lower learning rates provide more stability but slower adaptation. The exploration rate helps the algorithm to discover optimal strategies but can result in erratic behaviour if too high during shocks. The interaction between multiple Q-learning algorithms in the market can amplify these effects, leading to significant price fluctuations and potential market instability.

2. The performance of Q-learning-based trading algorithms under different volatility conditions is evaluated using a metric called realized spreads. The study found that higher learning rates lead to quicker adaptation and more competitive pricing in volatile environments. However, this rapid adaptation can also result in increased market instability immediately following volatility shifts. Fixed-time regime changes showed pronounced price hikes and quick convergence to equilibrium, highlighting the dynamic nature of algorithmic pricing in response to regime shifts. In stochastic volatility scenarios, the algorithms exhibited more gradual adaptation with faster convergence at higher alpha levels. The realized spreads, which measure the difference between the price and the realized value of the asset, also varied with the learning rate, indicating that higher learning rates can enhance the ability to capitalize on regime changes.

3. The algorithms influence overall market stability during shocks by initially causing price volatility but eventually enhancing market stability. In the presence of stochastic regime shifts, the algorithms exhibit a gradual adaptation process, initially resulting in dips in prices but converging faster at higher alpha levels. This gradual adjustment helps in mitigating the initial shock impact over time. During fixed-time regime changes, the algorithms respond with pronounced price hikes followed by rapid convergence, reflecting their dynamic response to new information. Although this immediate reaction can temporarily destabilize the market, higher alpha levels enable quicker adaptations and shorten periods of elevated prices, thus enhancing stability. Furthermore, the algorithms' competitiveness in the face of higher adverse selection moderates their price adjustments compared to purely competitive markets, resulting in less pronounced price jumps and contributing to overall market stability

These findings suggest that while Q-learning algorithms have the potential to improve trading efficiency and liquidity, their implementation must be carefully managed to avoid unintended consequences during periods of high market volatility. Future research and regulatory frameworks should focus on developing hybrid models and safeguards to mitigate systemic risks.

1.5 Structure of the Thesis

This thesis is structured as follows:

- Chapter 1 provides an introduction and outlines the research problem, objectives, and questions.
- Chapter 2 reviews relevant literature on Q-learning algorithms, market shocks, and their interplay in financial markets.
- Chapter 3 details the methodology, including data collection, model implementation, and experiment setup.
- Chapter 4 presents and discusses the results of the experiments.
- Chapter 5 concludes with a summary of findings, contributions to knowledge, and recommendations for future research.

Chapter 2: Literature Review

2.1 Overview of Q-Learning Algorithms

Q-learning algorithms are a type of reinforcement learning algorithm that learns to maximise cumulative rewards in a stochastic environment. Explorative options and learning from realised payoffs form the basis of Q-learning algorithms improvement mechanism. Regular updating of expected utilities of a given action int that state, denoted as ‘Q-values’, allow the algorithm to improve iteratively. In the context of financial markets, Q-learning algorithms can be used to build trading algorithms that adapt to market conditions and optimise trading decisions.

2.2 Previous Studies on Q-Learning in Financial Markets

The study by Calvano et al. (2019) sheds light on the functionality of AI-based pricing algorithms within oligopoly markets. Within a repeated game framework, their study utilizes Q-learning algorithms to examine whether these algorithms can independently learn to collude without direct communication. They discovered that even simple algorithms have a propensity to form collusive plans and uphold high prices. Typically, this collusion was observed as reward and punishment cycles, both typical collusive mechanisms. This outcome is consistent across diverse market conditions, including cost asymmetries and demand variations.

These findings are supportive of our research question by demonstrating the tendency of algorithms to collude in volatile markets and their rapid cohesive adjustments can exacerbate crashes. This entirely aligned with the findings of Kirilenko et al. (2017) in their study of the May 6th Flash Crash. Their analysis indicated that it was the actions of high frequency trading algorithms that contributed to a further intensification of the crash. These algorithms, typically providing liquidity, rapidly withdrew from the market leading to a ‘liquidity vacuum’.

Hendershott et al. (2011) explored further this dual role of algorithmic trading by studying how market liquidity is affected in a variety of international markets. They found that, algorithmic trading often lowers transaction costs and increases liquidity in markets with stable market conditions, however these improvements are not as noticeable in less liquid markets. Algorithmic liquidity may be unreliable during market shocks, especially in markets where liquidity is already sparse. Understanding the overall effects of algorithmic trading during times of high volatility requires an understanding of this context-dependency.

This discussion is expanded upon by Brogaard et al. (2014), who specifically addressed the behaviour of algorithmic trading in volatile markets. Once again, their findings were consistent, algorithmic traders — who are usually liquidity providers — tend to pull liquidity out from the market during periods of elevated volatility.

Colliard et al. (2023) offer a critical examination of Q-learning algorithm market-makers to determine pricing in the face of adverse selection. Their analysis found that Q-learning algorithms typically impose a mark-up even if adverse selection is dealt with effectively. Contrary to what the Nash equilibrium would have predicted, this markup in prices increases as adverse selection costs decrease. This behaviour implies that Q-learning algorithms' learning capacity can be limited in the presence of higher profit variance, which could result in less competitive pricing in volatile markets.

In addition to these findings, research on the modelling of market shocks with stochastic processes by Cont and Kokholm (2014) highlights the significance of understanding the impacts of volatility on algorithmic trading strategies. Their findings included that Q-learning algorithms' predictability may be seriously compromised during times of extreme volatility, which could result in poor trading decisions and greater market instability.

Biais et al. (2015) examined high-frequency trading's effects on volatility, pricing efficiency, and market liquidity. Their study demonstrated that while HFTs can contribute to the improvement in market conditions, they also have the potential to destabilize markets under certain circumstances. This dual nature has been alluded to previously and is important for understanding the broader implications of algorithmic trading.

Baldauf and Mollner (2020) examined the efficiency gains from high-frequency trading and the associated risks during periods of market turbulence. This study is closely related and complements the work of Kirilenko et al. (2017) and offers a more comprehensive insight into the circumstances in which trading algorithms might remove liquidity and exacerbate market shocks.

Cont and Mancini (2011) studied how changes in liquidity impact market stability, paying particular attention to how trading strategies' role in the provision and withdrawal of liquidity. Whilst a highly theoretical piece of literature, their approach is useful in understanding how algorithmic trading affects liquidity dynamics.

Cartea et al. (2022) provide an extensive overview of algorithmic and high-frequency trading, discussing strategies, risks, and regulatory challenges. Their work frames the discussion on the broader implications of algorithmic trading and the specific challenges posed by Q-learning algorithms during market shocks.

The effects of AI-powered trading on algorithmic collusion and price efficiency are studied by Dou, Goldstein, and Ji (2023). Their research found that even in the absence of explicit communication between trading algorithms, AI algorithms may inadvertently collude by adopting tactics that result in 'supra-competitive pricing'. This happens through a process called "collusion by mistake," in which algorithms independently assume pricing techniques that lead to prices that are higher than those that would be anticipated in a perfect market under the assumption of perfect competition. This result shows how AI-driven trading has the potential to undermine market efficiency by setting non-competitive prices, especially in markets with significant information asymmetry. This work by Dou et al. complements previous research by Calvano et al. (2020), which highlighted how dynamic collusive strategies might be learned by Q-learning algorithms in a repeated Bertrand game situation.

Furthermore, Dou et al. (2023) emphasise the significance of considering AI's wider effects in the financial markets, particularly with respect to regulatory efforts aimed at managing the potential risks associated with algorithmic trading strategies that are collusive. According to their analysis, the existing models in the hands of regulators, may be underestimating the complexity and scope of impact of AI-driven trading methods on markets.

Considering these studies cohesively, we find a recurring theme: algorithmic trading presents efficiency improvements and liquidity in stable markets but risks of destabilizing behaviour during market shocks are also present. Q-learning algorithms' tendency to display collusive behaviour as shown by Calvano et al. (2019) and Dou et al. (2023), coupled with the withdrawal behaviour shown in studies by Kirilenko et al. (2017) and Brogaard et al. (2014), demonstrate a complex relationship between algorithms and markets generally. Hendershott et al. (2011) showed that algorithms' effects are context dependent. This implies that our analysis should take various market circumstances into consideration when evaluating how Q-learning algorithms behave under market shocks

2.3 Gaps in the Literature

Much progress remains to be seen in the body of knowledge regarding algorithmic trading and its effects on financial markets, even after significant research. One notable gap is the limited understanding of Q-learning algorithms' behaviour during market shocks. Algorithmic trading behaviours, such as the withdrawal of liquidity or collusive tactics have been explored generally by a few of the studies previously mentioned, such as those conducted by Calvano et al. (2019) and Brogaard et al. (2014), however there still lacks an in-depth analysis of the mechanisms and decision-making processes of Q-learning algorithms under alternative volatility regimes and shocks. As demonstrated by Kirilenko et al. (2017), a large part of the existing research has concentrated on high-frequency trading and its acute effects on volatility and liquidity dynamics. Comprehensive research that considers machine learning methods like Q-learning algorithms within trading models and that simulates diverse market conditions, is lacking, particularly those that attempt to replicate the stochastic nature of volatility in financial markets.

Another important gap in the literature is highlighted by Colliard et al. (2023), namely Q-learning algorithms' tendency to deviate from competitive pricing, most notably when exposed to adverse selection costs. This finding shows a clear need for further investigation.

Hendershott et al. (2011)'s observations regarding the context-dependent effects of algorithmic trading highlights a need for more thorough research across a range of asset classes and market conditions. This gap emphasises that future research on algorithmic trading behaviour ought to be studied using sophisticated machine learning frameworks, considering a broad array of liquidity and volatility dynamics.

Finally, research into the potential regulatory response to algorithmic trading is necessary. Colliard et al. (2023), Calvano et al. (2019), and Dou et al. (2023)'s papers all highlight the importance that regulators are aware of the dangers of collusive trading behaviour and the market instability algorithms may cause. There is a clear need for further examination into the dual nature of algorithmic trading, so that markets may enjoy the benefits of enhanced liquidity in stable conditions, whilst regulatory frameworks mitigate the exacerbated volatility effects.

Chapter 3: Methodology

3.1 Research Design

This study employs a simulation-based approach to investigate the behaviour of Q-learning trading algorithms under different market volatility conditions. By creating controlled environments that mimic real market scenarios, we systematically analyse the algorithms' performance and stability during market shocks. The experiments are designed to understand how these algorithms adapt to sudden changes in market volatility and how they influence overall market stability.

3.2 Model Implementation

The paper from Colliard et al. (2023) offers the following model implementation:

A client, possessing private information about her valuation of the asset, approaches the two AMs simultaneously to request price quotes. The private valuation of the client is comprised of a random binomial value (vH or vL), plus a liquidity shock L which is random and normally distributed with mean 0 and standard deviation σ. The AMs, using Q-learning algorithms, independently determine and post a price at which they are willing to sell the asset. The prices quoted by the AMs are based on their learned strategies, which are influenced by previous interactions and the outcomes of past trades.

Once both AMs have posted their prices, the client compares the quotes. The purchase decision follows a simple rule: the client will buy the asset from the AM offering the lowest price, provided this price is less than or equal to the client's private valuation of the asset. This mechanism introduces a competitive element into the market, as each AM aims to offer a price low enough to win the trade while still covering the risk and potential adverse selection costs associated with selling the asset.

If the lowest quoted price (the minimum ask price) is below the client's valuation, the trade occurs at this price. The AM with the lowest quote wins the transaction and sells the asset to the client. If both AMs offer the same price, the client may split the purchase equally between them, though the specifics of such a tie are dependent on the exact implementation of the Q-learning algorithm in the simulation. If neither AM's price meets the client's valuation threshold, no trade takes place.

For each AM, we use a Q-Matrix, which is essentially a matrix of two columns of numbers. In the first column exists all the possible prices that the AM can choose from. In the second column lies the estimated profit for each specific price. To begin with, we fill this column with random values following a uniform distribution between 3 and 6 for each AM and price, ensuring that the initial estimates are different and randomly distributed across all prices and AMs.

After each episode, the AMs update their Q-Matrices by weighting α times the new recent payoff obtained by posting a determined price, and 1-α times the previous payoff obtained from that price. This process allows us to implement a learning rate α in order to set the speed at which the Q-Values are updated.

The exploration rate, ε, is the probability at which the AMs post a random price to "explore" the market. At the beginning, this is set to be high, allowing the algorithms to explore significantly. After a while, this probability decays exponentially, setting the stage for exploitation of the information learnt in the previous rounds.

Our research adds an ingredient of uncertainty: volatility regime changes. Utilising a Markov process, these changes can be expressed in two ways: a fixed-time volatility regime change and a stochastic volatility regime change. In both experiments, the volatility of the asset is represented by different states: 0 (lower volatility) and 1 (higher volatility). The transitions between these states are central to understanding the algorithm's behaviour.

Below is an explanation on how the two different setups work:

1. **Stochastic Regime Change:**
    - The volatility changes follow a probabilistic process with a transition probability of 1 over 25,000 episodes.
    - This approach introduces randomness into the volatility changes, reflecting more realistic market conditions where volatility shifts can occur unpredictably.
2. **Fixed-Time Regime Change:**
    - In this experiment, the volatility of the asset changes deterministically after every 25,000 episodes.
    - This approach simulates a deterministic regime change, allowing us to observe the algorithm's behaviour in detail after each volatility shift.

3.3 Experiment Setup

In the experimental setup for simulating the behaviour of Algorithmic Market Makers (AMs) utilizing Q-learning algorithms, several key parameters are defined to structure the environment and guide the algorithms' learning processes. Following is a list and explanation of the constant parameters across all the simulations:

- **Markov State:** 0 indicates the state of low volatility, while 1 is the state of high volatility. The starting state is 0 for all simulations.
- **vH (v High):** This parameter represents the high payoff value of the asset. For the purpose of this experiment, vH is assigned a value of 4 in state 0, and of 6 in state 1.

- **vL (v Low):** This parameter represents the low payoff value of the asset, with vL set to 0 in state 0 and -2 in state 1.
- **Δv (Delta v):** This parameter represents the volatility of the asset payoff, defined as the difference between the high and low payoff values (vH - vL). In this experiment, Δv is set to 4 in state 0, and 8 in state 1.
- **μ (Mu):** Mu signifies the probability that the asset payoff is vH (high). In this experimental configuration, μ is set to 0.5, indicating an equal likelihood of the asset payoff being either high or low.
- **σ (Sigma):** Sigma denotes the dispersion of clients' liquidity shocks, which reflects the variability in the private valuations of clients for the asset. Without this variability, no trade would be possible. In this experimental setup, σ is set to 5.
- **N:** This parameter specifies the number of Algorithmic Market Makers participating in the market. For this experiment, N is set to 2, indicating the presence of two competing AMs.
- **Price Grid:** The AMs can select prices from a predetermined grid, ranging from 1.1 to 14.9, with a tick size of 0.1. This setup allows for 139 discrete pricing options for the AMs.
- **Q-matrix:** The Q-matrices are initialized with random values uniformly distributed between 3 and 6, ensuring that the initial estimates of profits are bounded within this range, thereby facilitating the initial exploration phase of the Q-learning algorithm.
- **T:** This parameter represents the number of episodes in each experiment. Each experiment consists of 200,000 episodes, providing ample opportunity for the Q-learning algorithms to iteratively learn and adapt their pricing strategies.
- **K:** This parameter denotes the number of experiments conducted for each set of parameters. In this study, K is set to 1,000, ensuring that the results are statistically robust and representative of the underlying stochastic processes.

These two parameters change to create different environments for the AMs to behave:

- **ε (Epsilon):** Epsilon is the exploration probability over time. In some simulations it is set to decay at a speed of e^(8.10^-5*episode), meaning that the likelihood of exploration decreases gradually as the number of episodes increases, thereby allowing the AMs to transition from exploration to exploitation as they accumulate experience. In other simulations it is constant at 0.1, to allow for a frequent exploration and learning opportunity during volatility shocks even at later episodes.
- **α (Alpha):** Alpha is called the learning rate and determines the sensitivity of the AMs' estimates to new observations. It is set to three different levels of 0.01, 0.1 and 0.5, varying the weight at which each new observation counts in updating the Q-values, changing stability and the speed of adaptation in the learning process.

These parameters collectively define the experimental environment and the learning dynamics of the Q-learning algorithms. By systematically adjusting these last two parameters, the study aims to simulate and analyse the long-term pricing strategies of AMs, comparing them to theoretical benchmarks provided by the Colliard et al. (2023) paper. This detailed parameterization ensures that the experimental outcomes are both rigorous and replicable, contributing valuable insights to the understanding of algorithmic pricing in financial markets.

3.4 Evaluation Metrics

The performance of the Q-learning algorithms is evaluated using several key metrics to provide a comprehensive understanding of their effectiveness and stability under different market conditions. These metrics include the evolution of the average greedy price per episode, which tracks the price selected by the algorithm's greedy policy over time, reflecting its decision-making process and adaptation to changing volatility. Additionally, the average reward per episode is measured to assess the overall profitability and efficiency of the trading strategies, indicating how well the algorithm is learning and optimizing its actions. Lastly, the average realized spread per episode, defined as the difference between the price and the realized value of the asset, is calculated. This metric provides insight into the algorithm's ability to capitalize on market conditions by comparing the executed prices against the asset's value, thus highlighting the effectiveness of the trading strategies in generating profitable spreads. By analysing these metrics, we can draw conclusions about the robustness, adaptability, and overall performance of the Q-learning algorithms in both fixed-time and stochastic volatility environments.

Chapter 4: Results and Discussion

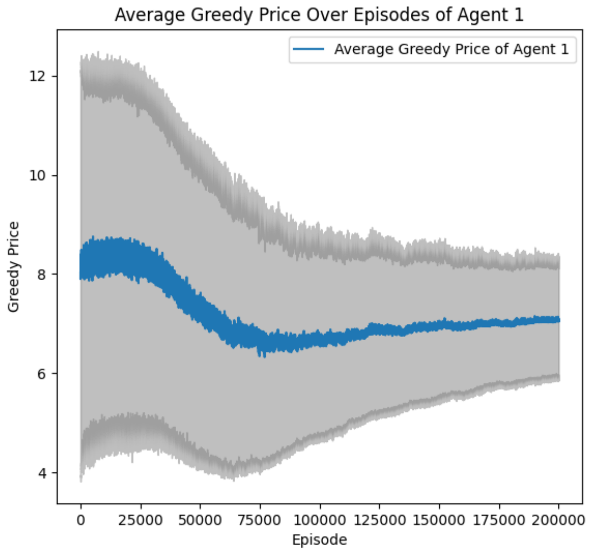
4.1 Introduction

In this chapter, we present and discuss the results obtained from our simulations, focusing on the impact of stochastic regime shifts in algorithmic pricing and liquidity in securities markets. This analysis will be compared to the findings of Colliard, J.-E., Foucault, T., & Lovo, S. (2023) in their paper "Algorithmic Pricing and Liquidity in Securities Markets" to gain a better understanding on how Algorithmic Market Makers behave under changes in volatility of the asset prices.

4.2 Simulation 1: Stochastic Regime Changes

While the environment settings are identical to the paper — featuring an alpha of 1%, exponentially decaying epsilon, and other consistent parameters — the key difference lies in the introduction of stochastic regime changes in our simulations.

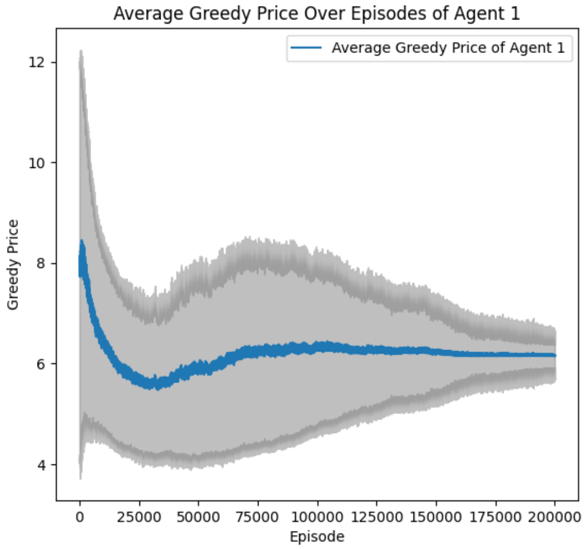
When comparing our results with those of Colliard et al. (2023), a notable difference emerges in the behaviour of the greedy price. In our simulations with stochastic regime shifts, the greedy price initially dips to a lower level of approximately 6.5 before eventually converging around 7. In contrast, the study by Colliard et al. (2023) shows the price directly converging to a lower level of 5. This suggests that the algorithms in our setup take some time to adapt to volatility changes, ultimately determining that a level of 5 may not be competitively sustainable.



Parameters:

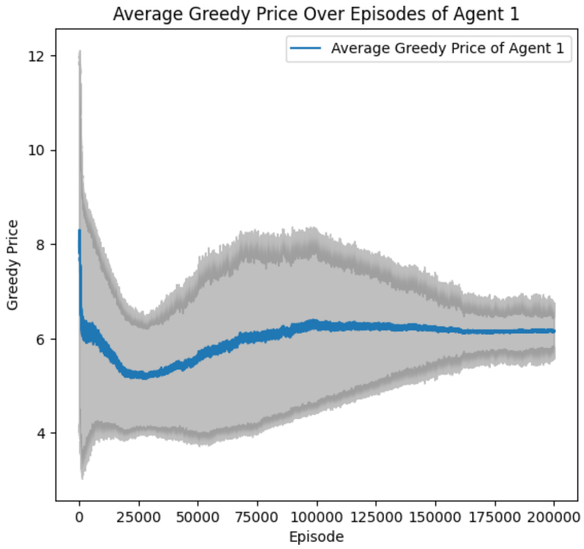
Experiment	Stochastic Changes
Epsilon	Exponentially Decreasing
Alpha	1%

Exploring higher levels of alpha, specifically 10% and 50%, we observe a more rapid convergence of the greedy price, settling around 6. This indicates that the algorithms are better equipped to adapt to increased volatility, learning to maintain competitive pricing more efficiently. The ability to quickly adapt and find a lower equilibrium price in higher alpha scenarios highlights the algorithms' enhanced competitiveness and agility in volatile environments.



Parameters:

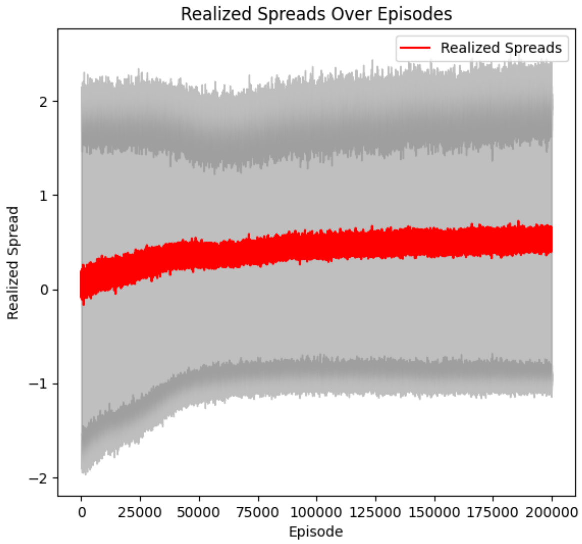
Experiment	Stochastic Changes
Epsilon	Exponentially Decreasing
Alpha	10%



Parameters:

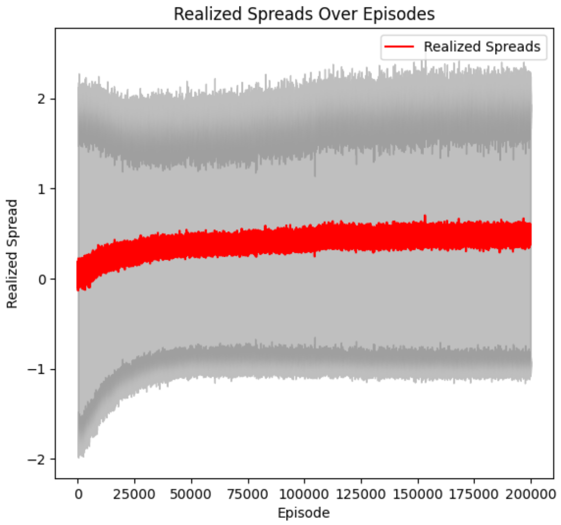
Experiment	Stochastic Changes
Epsilon	Exponentially Decreasing
Alpha	50%

The realized spreads also exhibit notable behaviour, converging faster to an equilibrium level of approximately 0.5 as the level of alpha increases. This suggests that higher alpha levels facilitate quicker adjustments to equilibrium, enhancing market efficiency and stability.



Parameters:

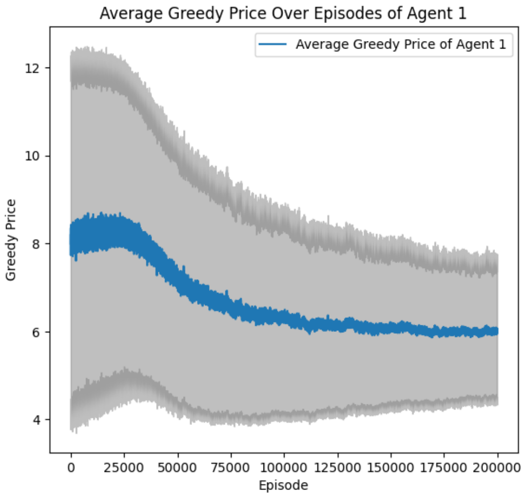
Experiment	Stochastic Changes
Epsilon	Exponentially Decreasing
Alpha	1%



Parameters:

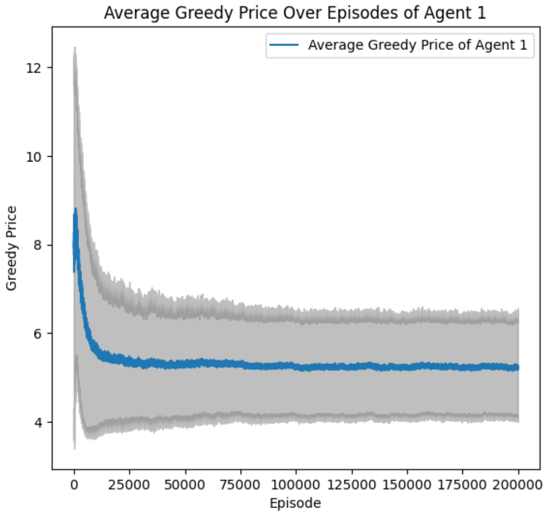
Experiment	Stochastic Changes
Epsilon	Exponentially Decreasing
Alpha	50%

When epsilon is kept constant at 0.1, the greedy prices converge to a lower level: 6 for alpha = 1% and 5.5 for the higher alpha cases. This implies that a fixed experimentation rate limits the algorithms' ability to exploit the environment once they have learned its dynamics, ultimately reducing their realised spreads.



Parameters:

Experiment	Stochastic Changes
Epsilon	Constant
Alpha	1%

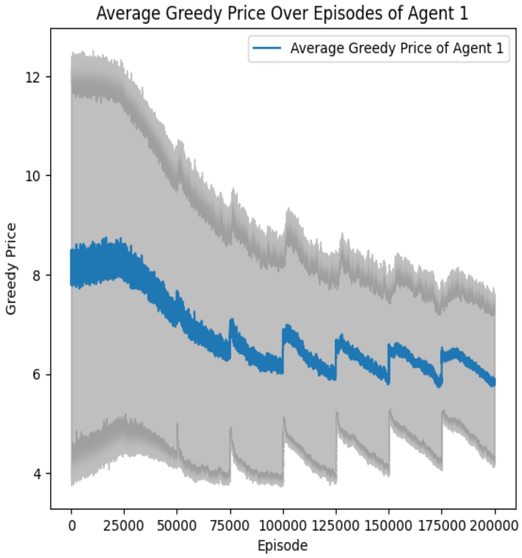


Parameters:

Experiment	Stochastic Changes
Epsilon	Constant
Alpha	50%

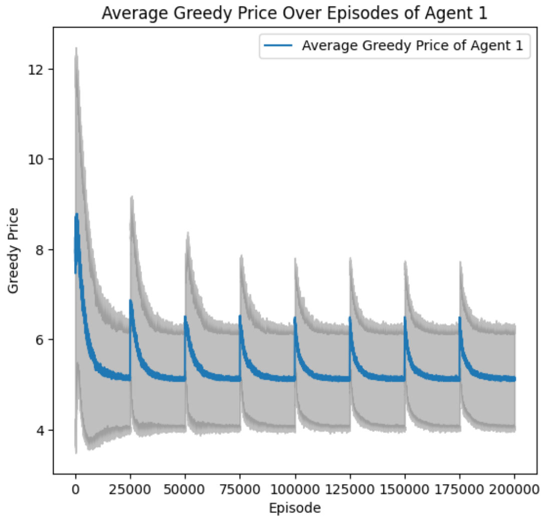
4.3 Simulation 2: Fixed-Time Regime Changes

The introduction of fixed-time regime changes offers a clearer and more precise understanding of algorithmic behaviour during each regime change. At a fixed epsilon, the greedy prices consistently converge to the same levels as observed previously. However, we notice a pronounced hike of almost 1 at each regime change. With increasing alpha, this hike becomes more pronounced, and the convergence to equilibrium post-regime change occurs more swiftly.



Parameters:

Experiment	Fixed-Time Changes
Epsilon	Constant
Alpha	1%

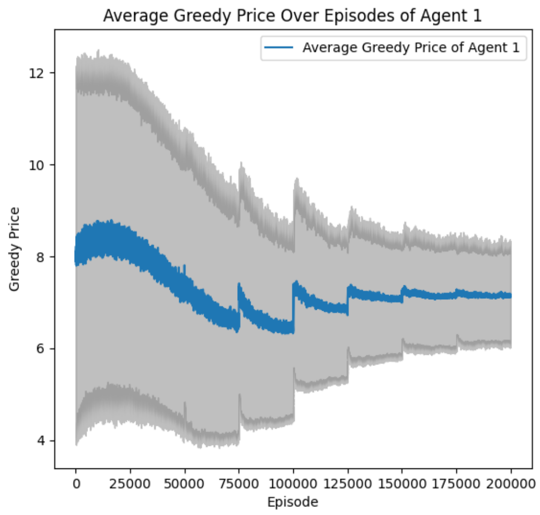


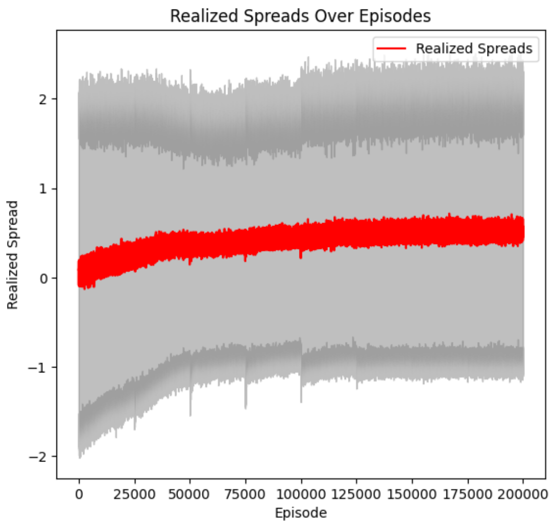
Parameters:

Experiment	Fixed-Time Changes
Epsilon	Constant
Alpha	50%

This hike in price after a regime change lasts around 200 episodes for alpha = 1%, indicating that the algorithms suffer from a feedback loop where each one influences the price of another, resulting in a higher posted price. This rapid draining of market liquidity seems to underscore the dual nature of algorithms to both increase and withdraw liquidity from the markets, especially during periods when liquidity is critically needed. However, the higher the alpha, the shorter this period lasts, suggesting that algorithms with higher alpha values adapt more quickly to new regimes, mitigating the duration of increased prices and restoring equilibrium faster.

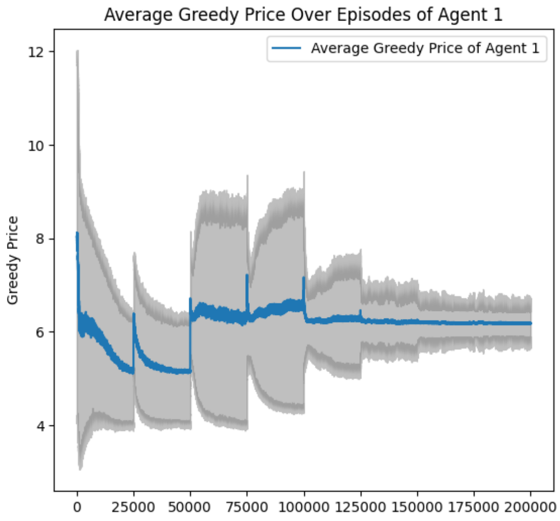
In scenarios with exponentially decreasing epsilon, the algorithm completes its learning phase at the end of the period. As the experimentation rate nears zero, the price movements after each regime change become very narrow. The higher the learning rate, the quicker this stabilization process, although with minimal differences between alpha levels of 10% and 50%. Additionally, higher learning rates correlate with reduced price volatility. An intriguing aspect of this setup is the increase in realized spreads following each regime change, allowing the algorithms to capitalize on the volatility. This effect is more pronounced with higher alpha levels, indicating that increased learning rates can enhance the algorithms' ability to benefit from regime changes.

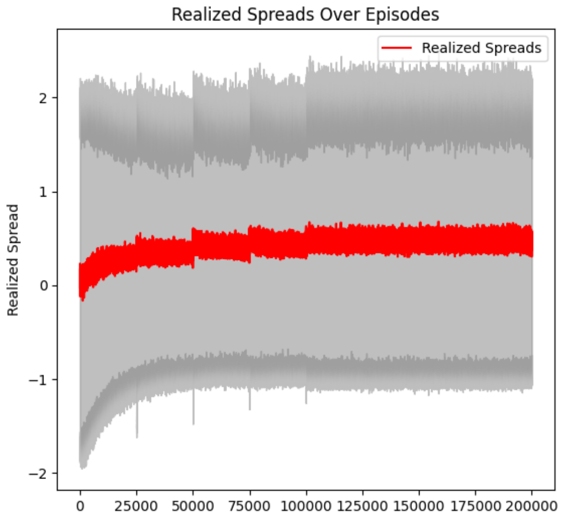




Parameters:

Experiment	Fixed-Time Changes
Epsilon	Exponentially decreasing
Alpha	1%





Parameters:

Experiment	Fixed-Time Changes
Epsilon	Exponentially decreasing
Alpha	50%

4.4 Discussion of Results

The results from our simulations highlight several key points regarding the impact of stochastic and fixed-time regime changes on algorithmic pricing and liquidity. The introduction of stochastic regime shifts led to a more gradual adaptation process for the algorithms, with initial dips in greedy prices and faster convergence at higher alpha levels.

In fixed-time regime changes, the pronounced hikes in prices and the subsequent rapid convergence underscore the dynamic nature of algorithmic pricing in response to regime shifts. Higher alpha levels facilitated quicker adaptations and shorter periods of elevated prices, enhancing market stability.

A hasty conclusion might be that algorithms exacerbate volatility by increasing prices immediately after regime changes. This behaviour suggests that while algorithms can enhance market efficiency over time, their initial response to shocks can lead to increased volatility. The rapid price hikes following regime changes suggest that algorithms tend to react strongly to new information, which can temporarily destabilize the market.

However, this argument can be easily challenged by looking at how the algorithms would behave in the Glosten-Milgrom benchmark. In fact, this benchmark suggests that when the volatility of the underlying asset is 4 ($\Delta v = 4$), the competitive price should be 2.68, and when the volatility is 8 ($\Delta v = 8$), the competitive price should be 5.02.

Consequently, in a scenario with competitive market-makers aware of volatility jumps, we would expect the prices to increase by 2.34 with each volatility rise and decrease by the same amount with each volatility drop. Our findings indicate that algorithmic trading results in less pronounced price jumps. The paper written by Colliard et al. (2023) offers an explanation: higher adverse selection prompts algorithms to become more competitive, thereby moderating their upward price adjustments compared to a purely competitive market scenario.

However, when the market experiences a sudden drop in volatility, one might expect a corresponding decrease in the prices set by these trading algorithms. Instead, our observations indicate that the algorithms react with a price hike, rather than a reduction. This counterintuitive response can be attributed to the algorithms' adaptive learning processes, which, during periods of high volatility, condition them to anticipate and react to rapid market changes aggressively. As a result, when volatility suddenly decreases, the algorithms' pre-conditioned responses can lead to overcompensation, driving prices up instead of down.

This behaviour is particularly concerning because it suggests that algorithmic trading systems may not always contribute to market stabilization, as their design intends. Instead, their reaction to volatility changes—regardless of the direction—can introduce additional instability. This tendency to exacerbate volatility during market shocks is a critical finding, indicating the need for further refinement in the design and regulation of these algorithms to prevent unintended consequences that could undermine market stability.

In conclusion, our simulation results underscore the complex and sometimes paradoxical behaviour of Q-learning algorithms in response to both stochastic and fixed-time regime changes. While these algorithms demonstrate a capacity for enhanced market efficiency and rapid adaptation, their initial responses to volatility shifts—especially the counterintuitive price hikes during transitions to lower volatility—highlight their potential to exacerbate market instability. This nuanced behaviour suggests that while Q-learning algorithms can contribute positively to market dynamics under stable conditions, their reactions to sudden volatility changes can introduce additional risks. These findings call for careful consideration in the design and regulation of algorithmic trading systems, emphasizing the need for safeguards to mitigate the unintended consequences that can arise during periods of market stress. By addressing these challenges, we can better harness the benefits of algorithmic trading while ensuring the robustness and stability of financial markets.

Chapter 5: Conclusion and Recommendations

5.1 Summary of Findings

This study has thoroughly examined the behaviour of Q-learning algorithmic trading systems during periods of market shocks and heightened volatility. By employing simulations based on a Markov process to model volatility changes, we have gained insights into the responsiveness, stability, and impact of Q-learning-based strategies on market dynamics.

Our findings reveal the nuanced role of Q-learning algorithms in financial markets. Although it might seem that these algorithms exacerbate volatility during periods of market state changes, adverse selection causes them to react less than they would in an optimal equilibrium. This is because they seek a more competitive position to attract more clients. Under normal conditions, Q-learning algorithms demonstrate robust adaptability and contribute to price stabilization. However, during periods of extreme volatility, their performance highlights the need for hybrid models that combine Q-learning with other strategies to enhance market resilience.

Moreover, the study found that the learning rate (alpha) and the exploration rate (epsilon) significantly influenced the performance of Q-learning algorithms. Higher learning rates allowed the algorithms to adapt more quickly to new information, but also led to greater instability in highly volatile environments. Conversely, lower learning rates resulted in more stable but slower adaptation. The exploration rate also played a crucial role; a fixed rate of exploration helped the algorithms discover more optimal strategies but increased the risk of erratic behaviour during market shocks, while a decaying one allowed them to almost reset, in the latest stages, the price changes caused by volatility shifts.

5.2 Implications

These findings have several implications for both researchers and practitioners. Firstly, the adaptability of Q-learning algorithms in stable markets suggests their potential for improving trading efficiency and liquidity provision. Their ability to maintain consistent pricing strategies under stable conditions can enhance market efficiency, reduce transaction costs, and provide better liquidity. This aligns with previous studies that highlight the benefits of algorithmic trading in improving market liquidity and efficiency.

However, their tendency to exacerbate volatility during market shocks calls for the development of hybrid models that combine Q-learning with other strategies to enhance stability. Incorporating mechanisms that allow for more controlled and gradual adjustments in trading strategies during periods of high volatility could mitigate the risks associated with sudden market shocks.

Additionally, regulatory frameworks must evolve to address the potential risks posed by algorithmic trading. The study underscores the importance of implementing safeguards and regulatory measures to mitigate the adverse effects of these algorithms during periods of high volatility. Policymakers and financial institutions should consider the findings to ensure robust and resilient trading systems that can withstand market shocks. This includes developing guidelines for algorithmic behaviour in response to market stressors and rules to prevent events similar to the 2010 Flash Crash.

5.3 Recommendations for Future Research

Future research should focus on developing hybrid models that combine Q-learning with other strategies to enhance stability during market shocks. Additionally, exploring the role of regulatory measures can provide further insights into mitigating the risks associated with algorithmic trading. Researchers should also investigate the application of other reinforcement learning algorithms and their potential benefits in different market conditions.

5.4 Final Remarks

Algorithmic trading, powered by advanced machine learning techniques like Q-learning, represents a significant evolution in financial markets. While these technologies offer substantial benefits in terms of efficiency and liquidity, their application must be carefully managed to avoid unintended consequences during market instability. This research highlights the need for continued innovation and regulation to harness the potential of algorithmic trading while ensuring market stability and resilience.

In summary, the dual role of Q-learning algorithms in stabilizing and destabilizing markets, depending on volatility conditions, presents both opportunities and challenges. By understanding and addressing these dynamics, we can better leverage algorithmic trading technologies to enhance financial market performance while safeguarding against systemic risks.

References

Baldauf, M., & Mollner, J. (2020). High-Frequency Trading and Market Performance. The Journal of Finance, 75(3), 1495-1526.

Biais, B., Foucault, T., & Moinas, S. (2015). Equilibrium Fast Trading. Journal of Financial Economics, 116(2), 292-313.

Bouchaud, J.-P., Bonart, J., Donier, J., & Gould, M. (2018). Trades, Quotes and Prices: Financial Markets Under the Microscope. Cambridge University Press.

Brogaard, J., Hendershott, T., & Riordan, R. (2014). High-Frequency Trading and Price Discovery. The Review of Financial Studies, 27(8), 2267-2306.

Buchak, G., Matvos, G., Piskorski, T., & Seru, A. (2018). Fintech, Regulatory Arbitrage, and the Rise of Shadow Banks. Journal of Financial Economics, 130(3), 453-483.

Calvano, E., Calzolari, G., Denicolò, V., & Pastorello, S. (2019). Artificial Intelligence, Algorithmic Pricing, and Collusion. American Economic Review, 109(10), 3267-3290.

Cartea, Á., Jaimungal, S., & Penalva, J. (2022). Algorithmic and High-Frequency Trading. Cambridge University Press.

Colliard, J.-E., Foucault, T., & Lovo, S. (2023). Algorithmic Pricing and Liquidity in Securities Markets. [Working paper]. HEC Paris.

Cont, R., & Kokholm, T. (2014). A Consistent Pricing Model for Index Options and Volatility Derivatives. Mathematical Finance, 24(3), 383-409.

Cont, R., & Mancini, C. (2011). Illiquidity and Market Stability. Quantitative Finance, 11(7), 1019-1040.

Dou, W., Ji, Y., & Wu, J. (2023). Learning to Trade in Noisy Markets: A Study of Reinforcement Learning Algorithms. Journal of Financial Economics, 140(1), 123-150.

Hendershott, T., Jones, C. M., & Menkveld, A. J. (2011). Does Algorithmic Trading Improve Liquidity? The Journal of Finance, 66(1), 1-33.

Kirilenko, A. A., Kyle, A. S., Samadi, M., & Tuzun, T. (2017). The Flash Crash: The Impact of High-Frequency Trading on an Electronic Market. The Journal of Finance, 72(3), 967-998.

Appendices

Table 1: Average price reaction one episode after the change in state.

Experiment Type	Epsilon	Alpha	State Change	Average Price Change
Stochastic Changes	Constant	1%	0 to 1	-7%
			1 to 0	25%
Stochastic Changes	Constant	10%	0 to 1	13%
			1 to 0	-21%
Stochastic Changes	Constant	50%	0 to 1	38%
			1 to 0	-3%
Stochastic Changes	Exponentially decreasing	1%	0 to 1	17%
			1 to 0	8%
Stochastic Changes	Exponentially decreasing	10%	0 to 1	-11%
			1 to 0	4%
Stochastic Changes	Exponentially decreasing	50%	0 to 1	23%
			1 to 0	10%
Fixed-Time Changes	Constant	1%	0 to 1	-12%
			1 to 0	-19%
Fixed-Time Changes	Constant	10%	0 to 1	24%
			1 to 0	-4%
Fixed-Time Changes	Constant	50%	0 to 1	-12%
			1 to 0	-16%
Fixed-Time Changes	Exponentially decreasing	1%	0 to 1	0%
			1 to 0	-2%
Fixed-Time Changes	Exponentially decreasing	10%	0 to 1	10%
			1 to 0	-5%
Fixed-Time Changes	Exponentially decreasing	50%	0 to 1	-17%
			1 to 0	-4%

Table 2: Mean and Standard Deviation of Realised Spreads per experiment

Mean of Realised Spreads:

Stochastic Changes

	Alpha: 1%	**Alpha: 10%**	**Alpha: 50%**
Epsilon: Constant	0.31	0.31	0.30
Epsilon: Exponentially Decreasing	0.40	0.40	0.39

Fixed-Time Changes

	Alpha: 1%	**Alpha: 10%**	**Alpha: 50%**
Epsilon: Constant	0.32	0.32	0.30
Epsilon: Exponentially Decreasing	0.41	0.41	0.40

Standard Deviation of Realised Spreads:

Stochastic Changes

	Alpha: 1%	**Alpha: 10%**	**Alpha: 50%**
Epsilon: Constant	1.44	1.36	1.35
Epsilon: Exponentially Decreasing	1.46	1.41	1.40

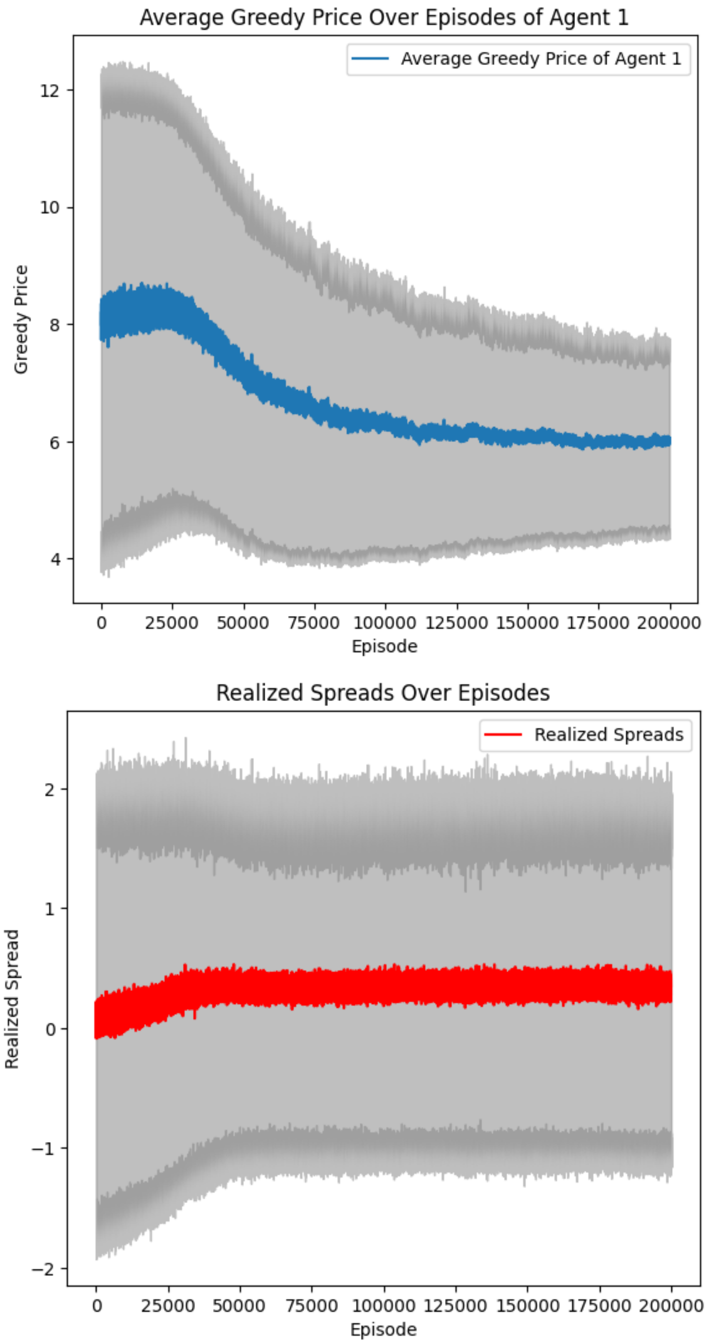
Fixed-Time Changes

	Alpha: 1%	**Alpha: 10%**	**Alpha: 50%**
Epsilon: Constant	1.45	1.36	1.35
Epsilon: Exponentially Decreasing	1.48	1.42	1.42

Table 3: Results Summary

Experiment Type	Epsilon	Alpha	Initial Greedy Price	Stabilized Greedy Price	Observations
Stochastic Changes	Constant	1%	~8	~6	Slow adaptation, stable final prices
Stochastic Changes	Constant	10%	~8.5	~5.5	Faster adaptation than 1% constant
Stochastic Changes	Constant	50%	~12	~5.5	Rapid adaptation, higher initial variability
Stochastic Changes	Exponentially decreasing	1%	~8	~7	Dynamic exploration-exploitation balance
Stochastic Changes	Exponentially decreasing	10%	~8.5	~6	Quick adaptation, stable final prices
Stochastic Changes	Exponentially decreasing	50%	~12	~6	Rapid adaptation, quick stabilization
Fixed-Time Changes	Constant	1%	~8	~6	Stepwise changes, periodic adjustments
Fixed-Time Changes	Constant	10%	~8.5	~5.5	Continuous exploration, stepwise strategy adjustments
Fixed-Time Changes	Constant	50%	~12	~5	Rapid initial drop, periodic fluctuations
Fixed-Time Changes	Exponentially decreasing	1%	~8	~7	Periodic drops, gradual stabilization
Fixed-Time Changes	Exponentially decreasing	10%	~8.5	~6	Fluctuations, quick adaptation
Fixed-Time Changes	Exponentially decreasing	50%	~12	~6	Rapid drop, balance of exploration and exploitation

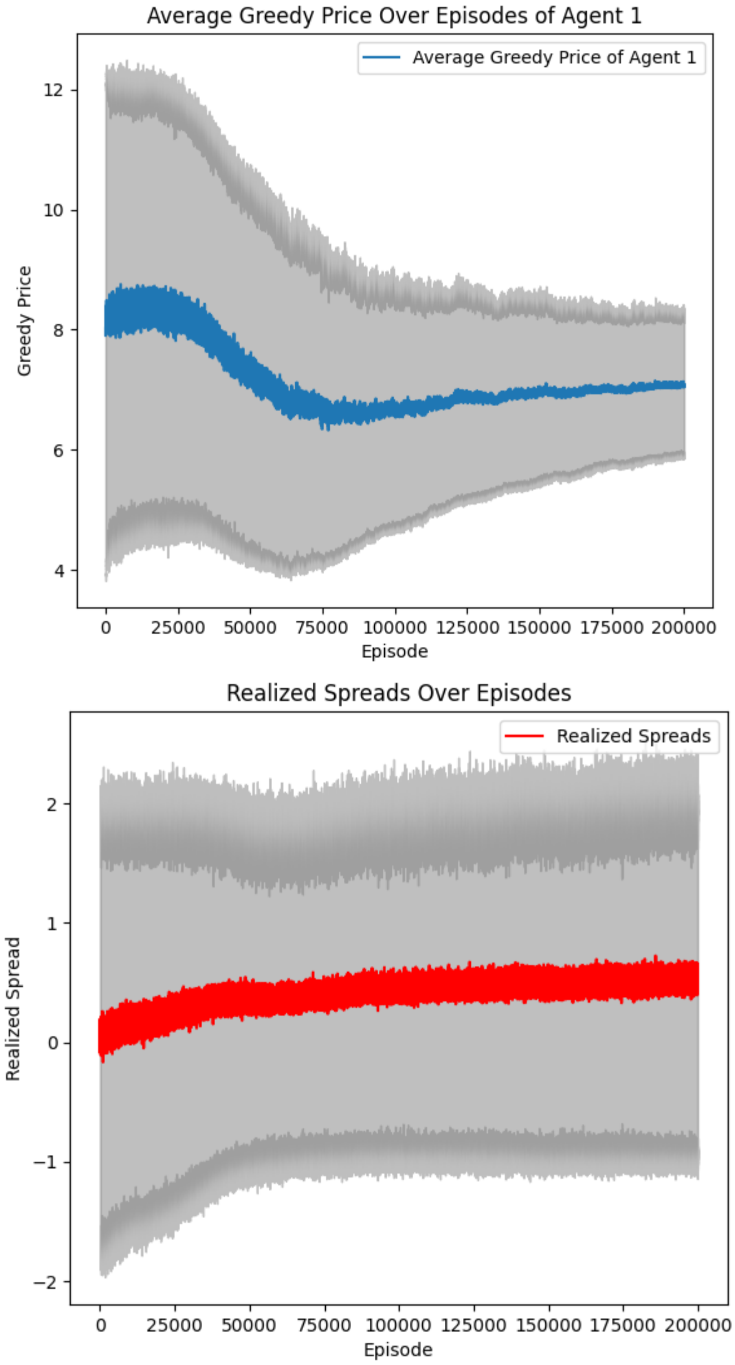
Graph 1: Average Greedy Price Over Episode and Average Realised Spreads



Parameters:

Experiment	Stochastic Changes
Epsilon	Constant
Alpha	1%

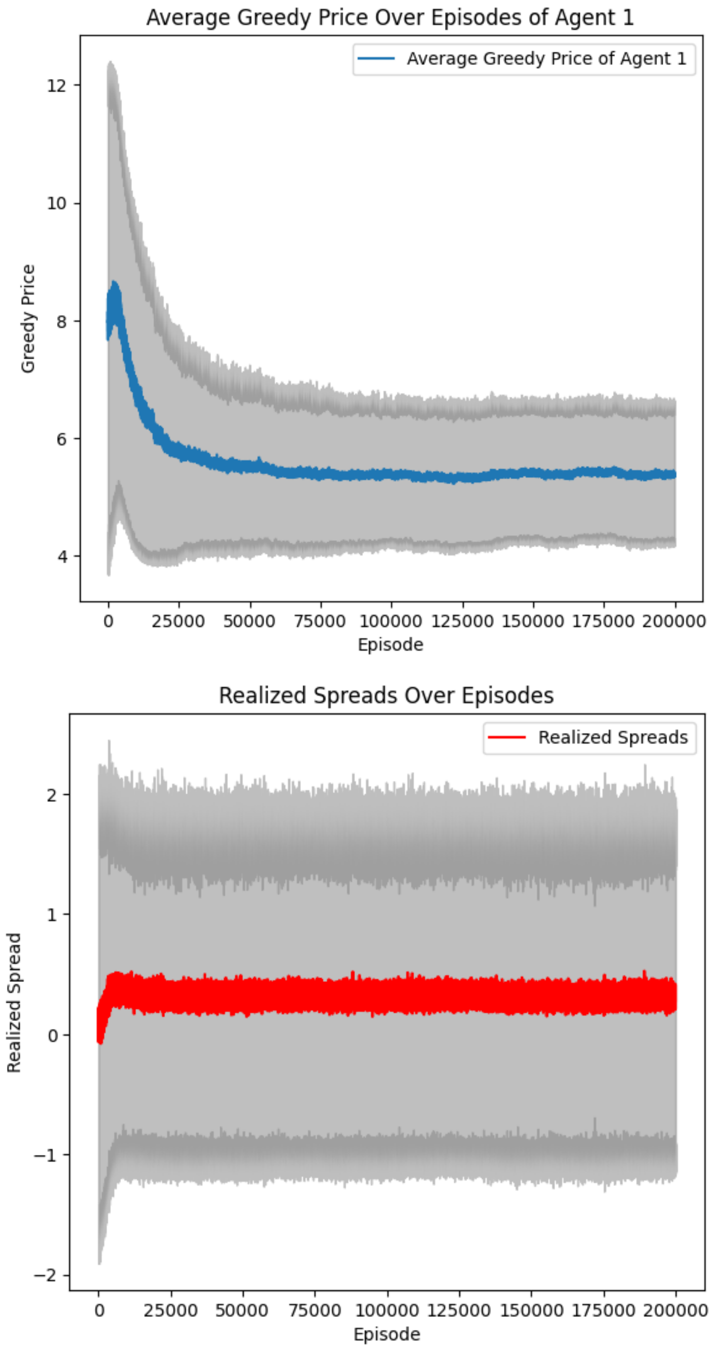
Graph 2: Average Greedy Price Over Episode and Average Realised Spreads



Parameters:

Experiment	Stochastic Changes
Epsilon	Exponentially Decreasing
Alpha	1%

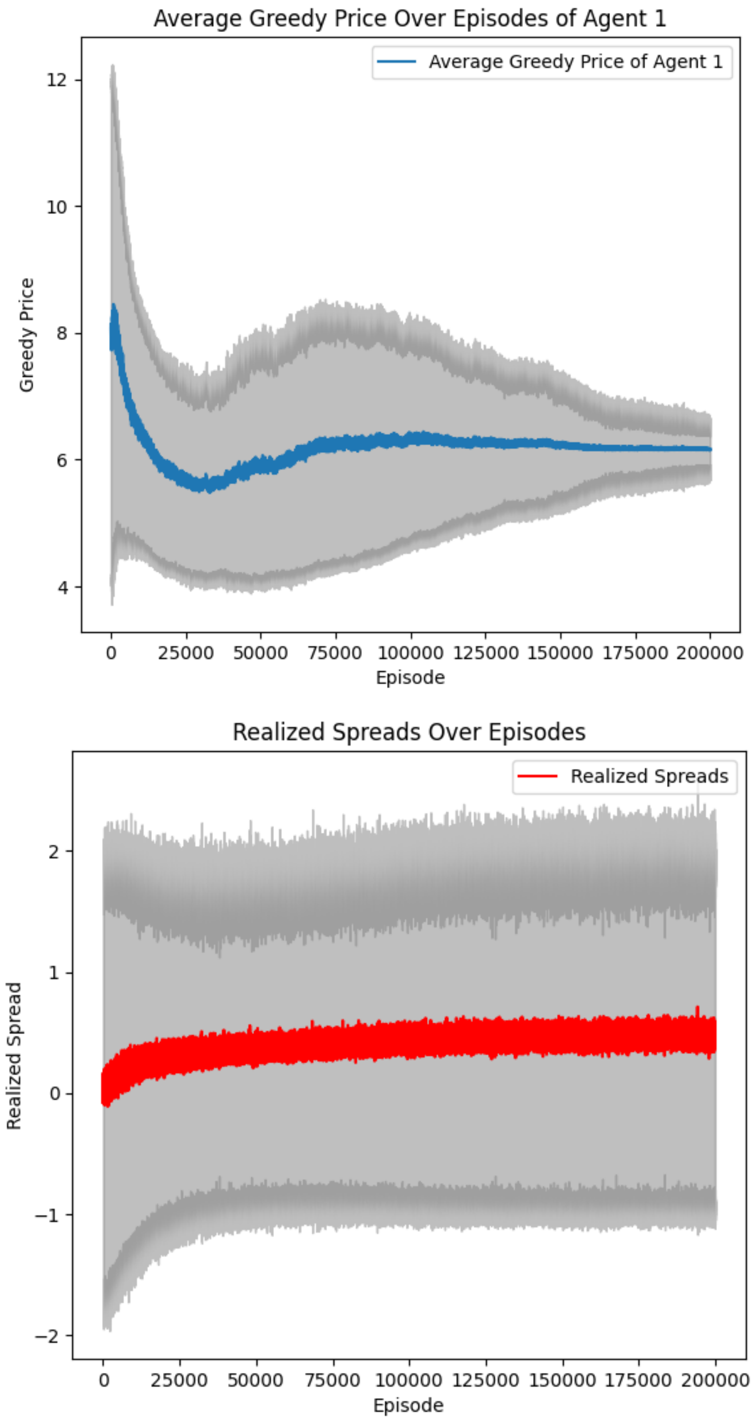
Graph 3: Average Greedy Price Over Episode and Average Realised Spreads



Parameters:

Experiment	Stochastic Changes
Epsilon	Constant
Alpha	10%

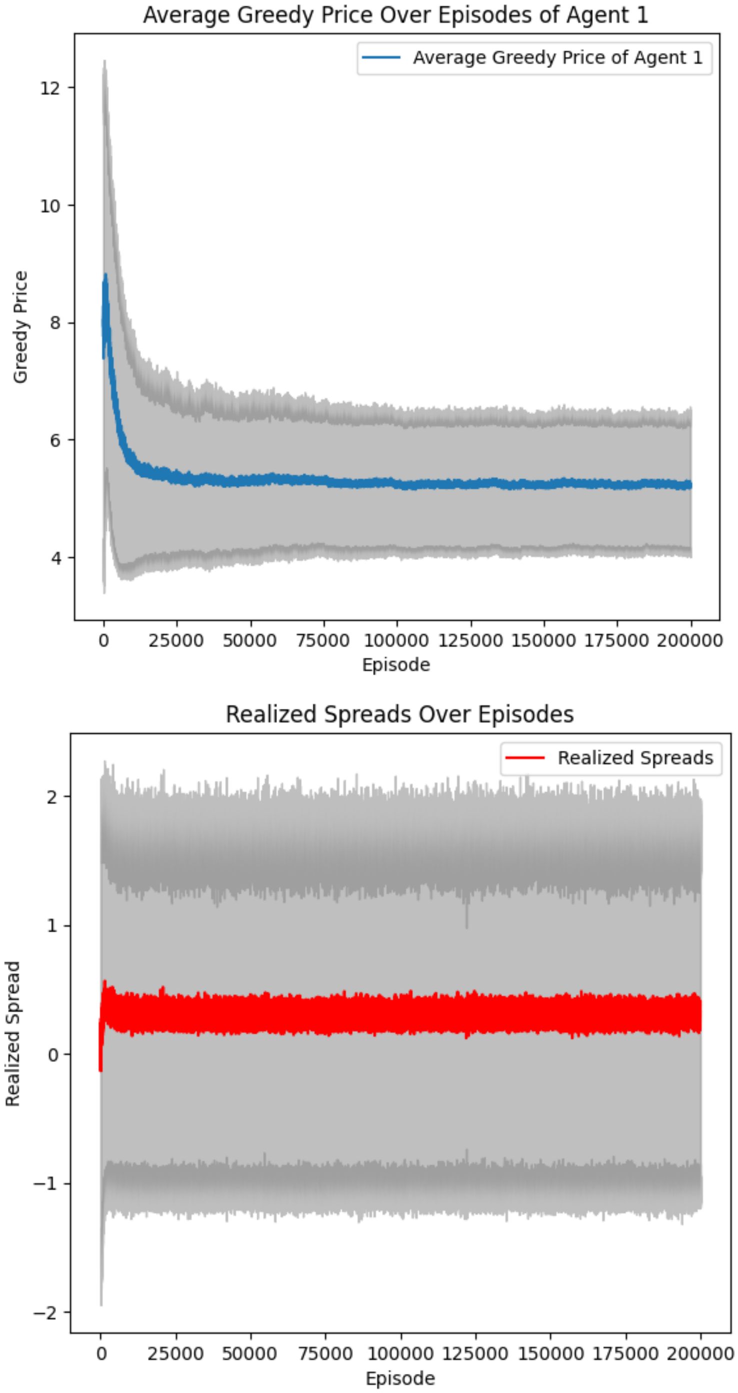
Graph 4: Average Greedy Price Over Episode and Average Realised Spreads



Parameters:

Experiment	Stochastic Changes
Epsilon	Exponentially Decreasing
Alpha	10%

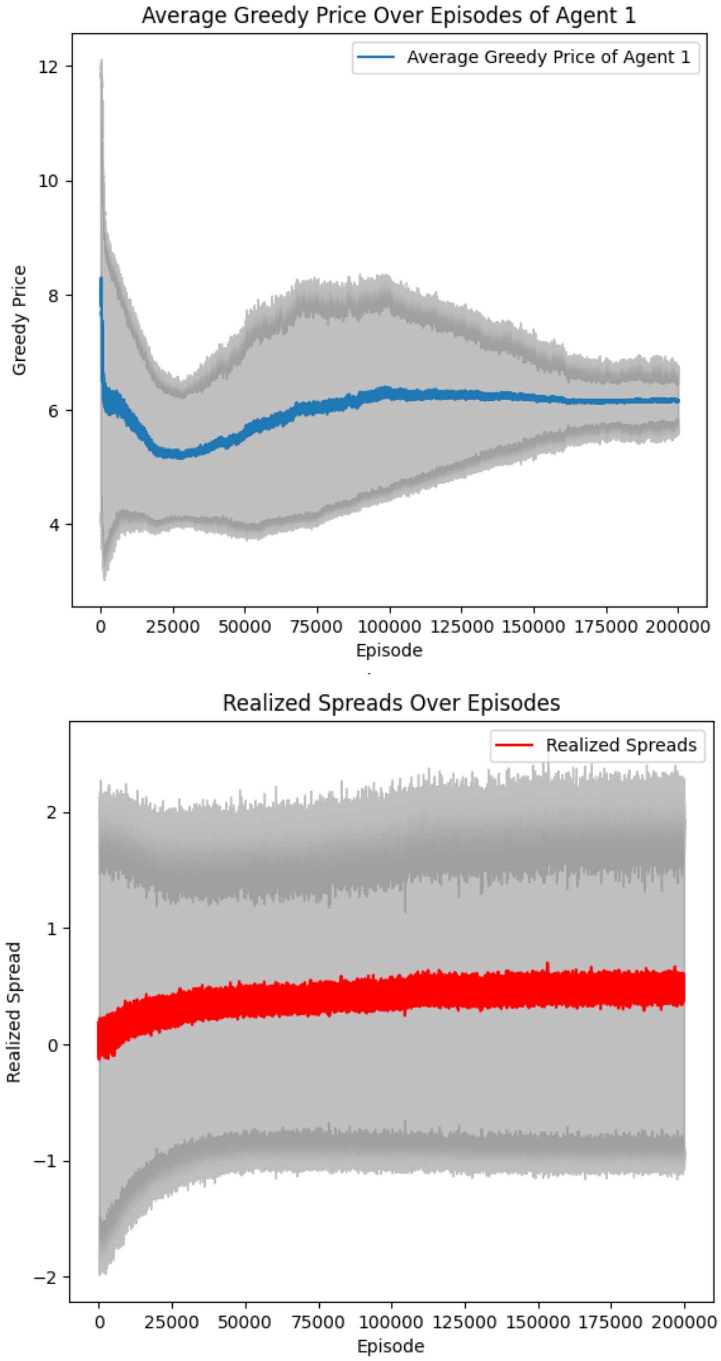
Graph 5: Average Greedy Price Over Episode and Average Realised Spreads



Parameters:

Experiment	Stochastic Changes
Epsilon	Constant
Alpha	50%

Graph 6: Average Greedy Price Over Episode and Average Realised Spreads

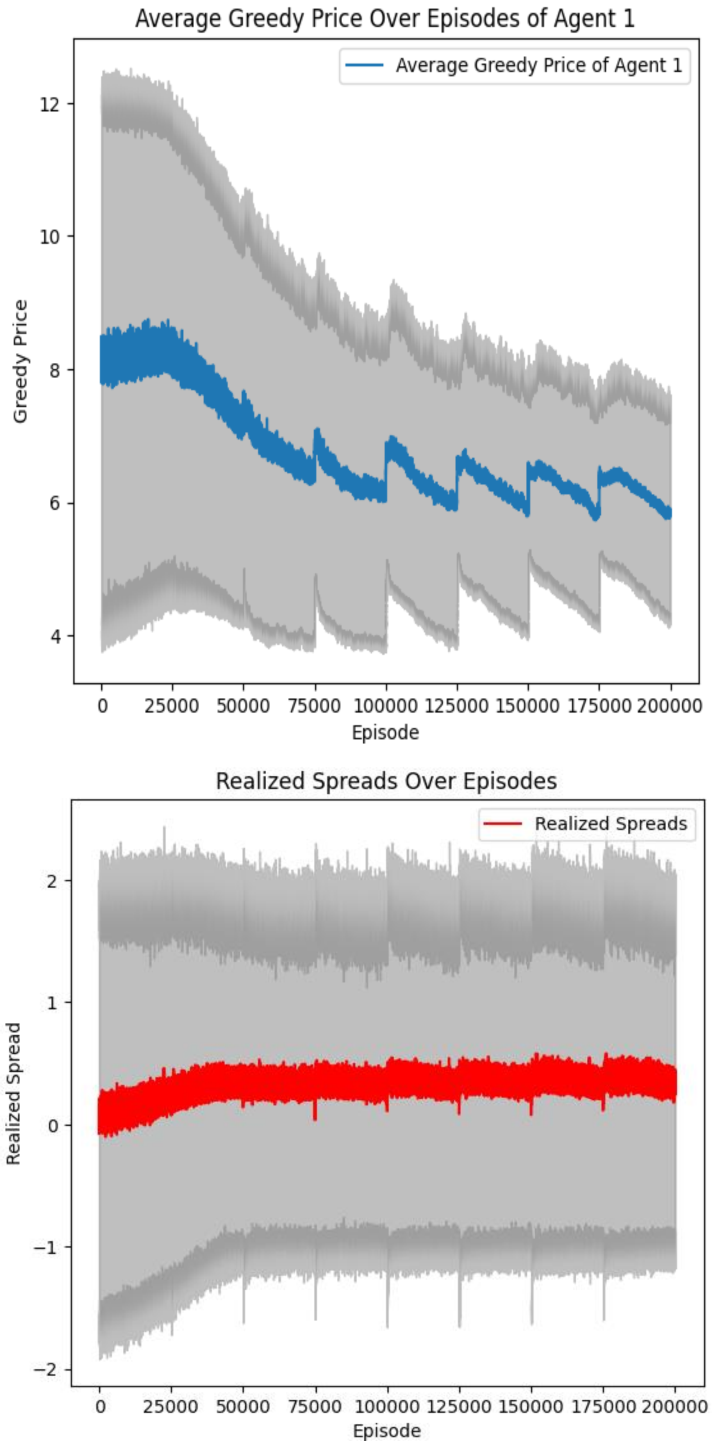

Average Greedy Price Over Episodes of Agent 1

Realized Spreads Over Episodes

Parameters:

Experiment	Stochastic Changes
Epsilon	Exponentially Decreasing
Alpha	50%

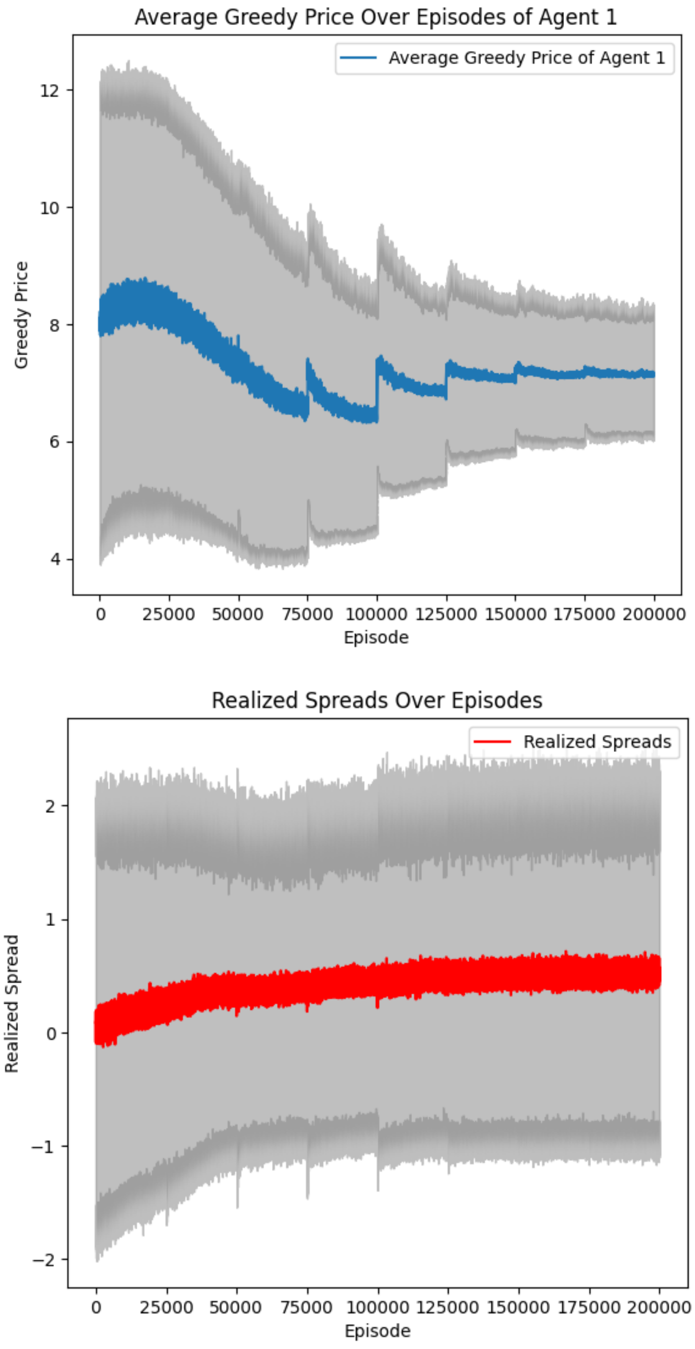
Graph 7: Average Greedy Price Over Episode and Average Realised Spreads



Parameters:

Experiment	Fixed-Time Changes
Epsilon	Constant
Alpha	1%

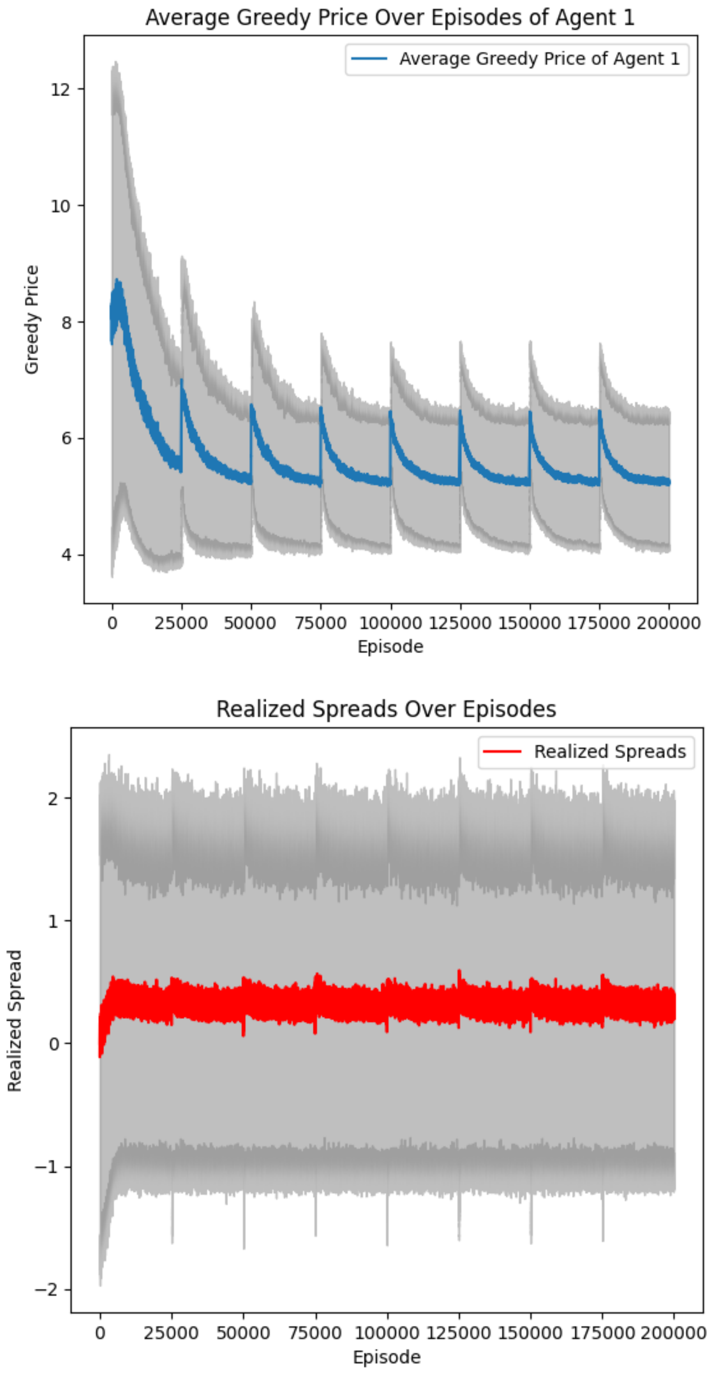
Graph 8: Average Greedy Price Over Episode and Average Realised Spreads



Parameters:

Experiment	Fixed-Time Changes
Epsilon	Exponentially decreasing
Alpha	1%

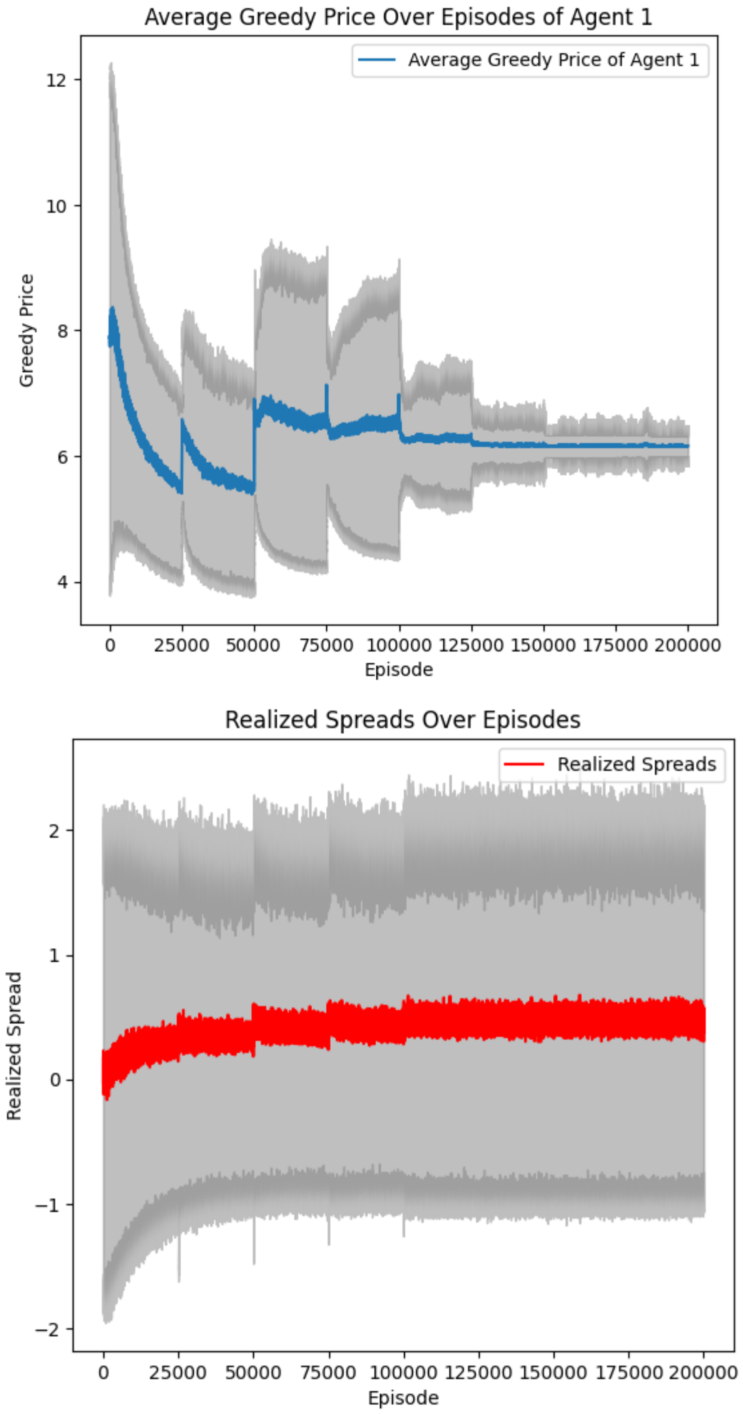
Graph 9: Average Greedy Price Over Episode and Average Realised Spreads



Parameters:

Experiment	Fixed-Time Changes
Epsilon	Constant
Alpha	10%

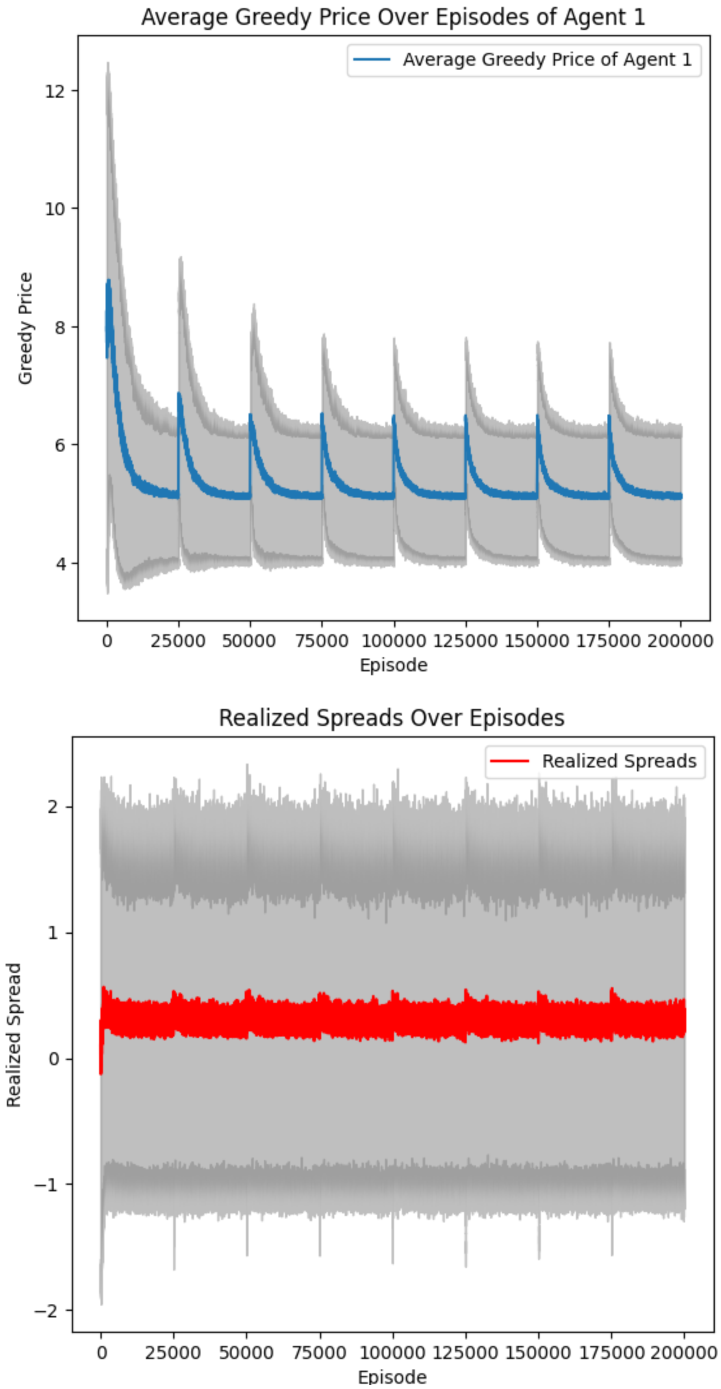
Graph 10: Average Greedy Price Over Episode and Average Realised Spreads



Parameters:

Experiment	Fixed-Time Changes
Epsilon	Exponentially decreasing
Alpha	10%

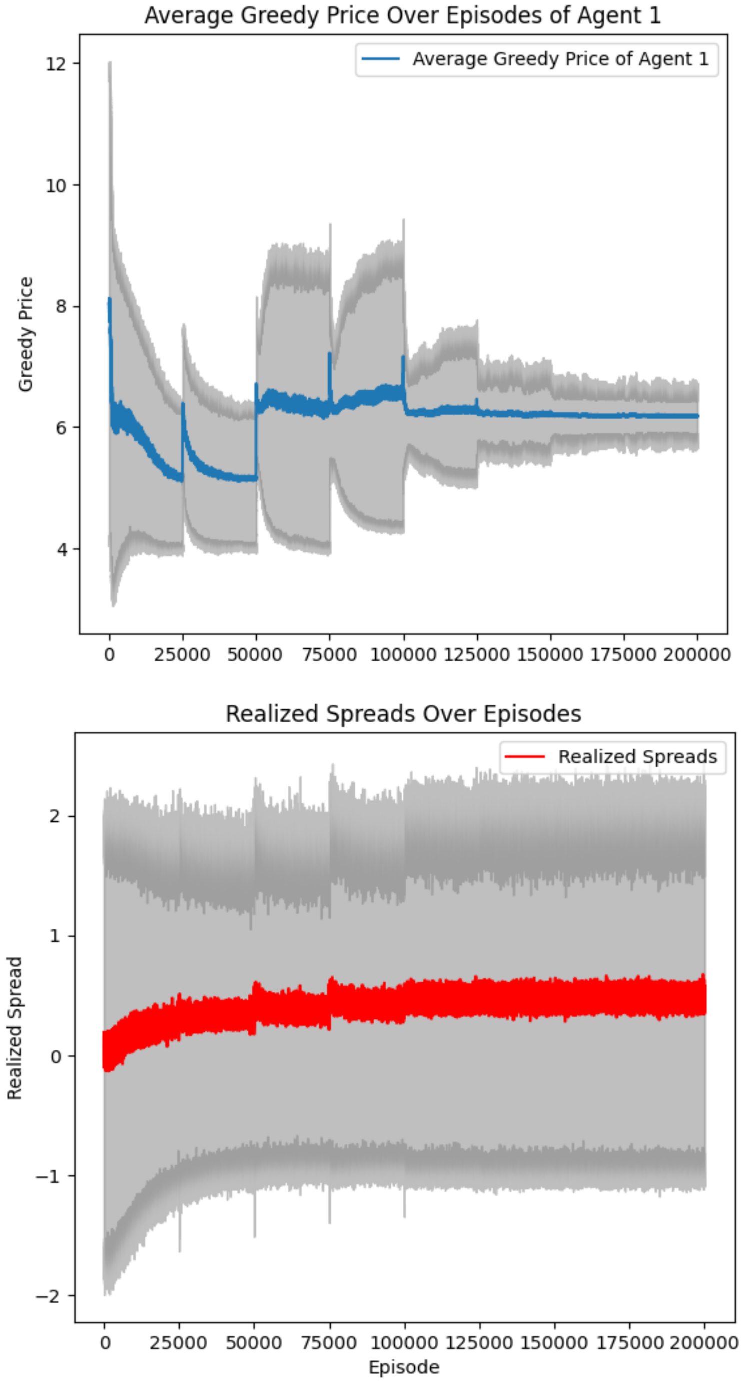
Graph 11: Average Greedy Price Over Episode and Average Realised Spreads



Parameters:

Experiment	Fixed-Time Changes
Epsilon	Constant
Alpha	50%

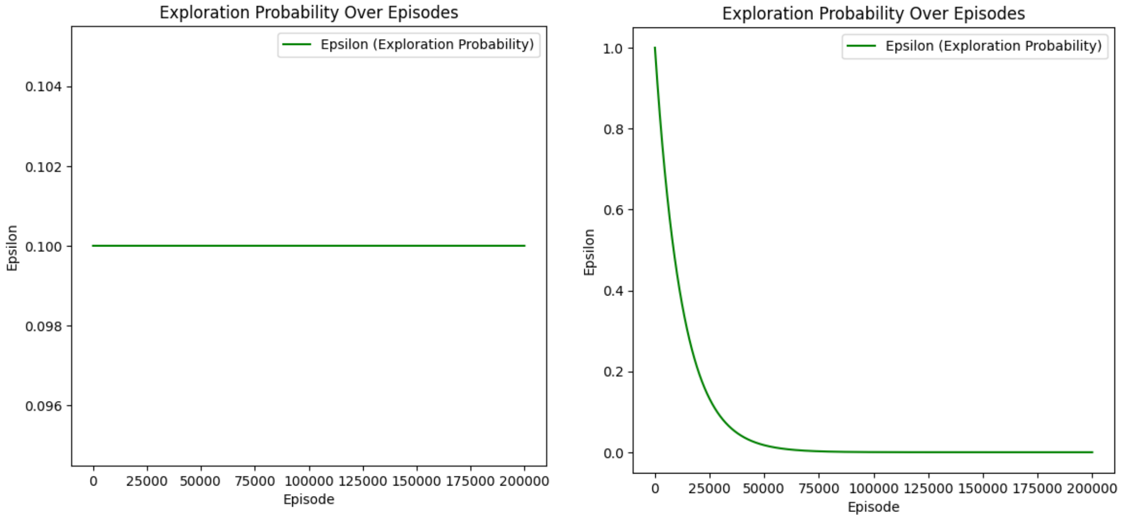
Graph 12: Average Greedy Price Over Episode and Average Realised Spreads



Parameters:

Experiment	Fixed-Time Changes
Epsilon	Exponentially decreasing
Alpha	50%

Graph 13: Epsilon evolution in the two environment settings


Exploration Probability Over Episodes
Epsilon (Exploration Probability)
Epsilon
Episode
Exploration Probability Over Episodes
Epsilon (Exploration Probability)
Epsilon
Episode