Reinforcement learning in a dynamic limit order market*

Amy Kwan† **Richard Philip**‡

March 27, 2025

Abstract. What drives the value of limit orders? We use a novel machine learning approach to investigate optimal limit order management and the factors affecting the expected value of an order. A limit order is more valuable if positioned towards the front of the queue and when there is a large queue resting behind the order. When trading is constrained by minimum tick size requirements, volatility decreases the value of the order, but increases its value when trading is unconstrained. Further, the option to cancel an order is economically meaningful, contributing approximately 19% of the order's total expected value. This study uncovers pervasive market dynamics, advancing our understanding of financial markets.

Key words: Limit order markets, machine learning, big data, queue size, optimal limit order

JEL: G10; G20

*This paper has benefited from the comments of Michael Brolley, James Brugler, David Cimon, Vincent van Kervel, Pete Kyle, Tom McInish, Albert Menkveld, Ryan Riordan, Andriy Shkilko, Ester Felez Vinas, Gideon Saar, Wing Wah Tham, Yajun Wang, Ying Wu, Chen Yao, Marius Zoican, and the audiences at the SFS Cavalcade Asia Pacific, FIRN annual meeting, and the Microstructure Exchange. Thank you to seminar participants at The University of Sydney, The University of New South Wales, University of Wollongong, Wilfrid Laurier University, and the team at Vivienne Court Trading for their insightful comments.

†University of New South Wales, Australia, e-mail: amy.kwan@unsw.edu.au

‡University of Sydney, Australia, e-mail: richard.philip@sydney.edu.au

1 Introduction

Limit order submissions and cancellations account for an overwhelming 95% of trading activity in modern markets.[1] In response, exchanges and regulators have proposed measures to curb message traffic — such as limiting the order-to-trade ratio, enforcing minimum resting times, and introducing message taxes or cancellation fees. Despite these efforts, our understanding of how liquidity providers should optimally manage limit orders remains limited. What is the value of a limit order? At what price should it be submitted? Which market features most influence a liquidity provider's order decisions? When should an order be canceled, and how valuable is that cancellation option? Answering these questions is non-trivial; the dimensionality of the problem is extremely large and decisions are path dependent.

In this study, we propose a novel framework to quantify the expected value of a limit order and investigate the optimal order management behavior of liquidity providers. Using high-frequency, order level data, we apply a reinforcement learning (RL) approach to estimate the expected value of a resting limit order across more than 18,000 distinct market states. These estimated values enable us to address key questions related to optimal liquidity provision. Specifically, we identify conditions under which limit orders have negative expected value and should therefore be canceled, as well as conditions under which they yield high expected value.[2] Further, by analyzing the cross-sectional variation across market states, we uncover the most important features influencing a liquidity providers' limit order decisions. Last, we quantify the value of the option to cancel a limit order and identify the specific market conditions in which this option is most valuable.

We consider limit order management as a sequential Markov decision process within a reinforcement learning (RL) framework. RL is a form of machine learning that enables an agent to learn optimal actions through feedback from its own experience, given the current state of the environ-

[1] See Brogaard et al. (2019). Market orders, which have been the primary focus of much of the existing literature, represent less than 5% of total activity.

[2] These estimates also allow us to identify the optimal price level at which to submit a limit order, and to assess when it is advantageous to cancel a resting order and resubmit at an alternative price level with a higher expected value.

ment.[3] In our setup, a risk-neutral liquidity provider evaluates their order at short, discrete time intervals, repeatedly choosing whether to leave the order in the book or cancel it. This decision-making process continues until the order is either executed or withdrawn. At each decision point, the trader acts to maximize expected profit, keeping the order if its conditional expected value is positive, and canceling it otherwise. This expected value is based on (1) current market conditions and (2) the anticipated future management of the order. Our framework therefore captures the endogenous value of the cancellation option, as it depends on the expected evolution of the limit order book and the trader's future responses. The resulting conditional expected value at time t is inherently recursive, depending on all future conditional expected values and their associated transition probabilities. To overcome the recursive nature of the problem, we empirically estimate the conditional expected value via an iterative update function, known as Q-learning.

We show that the average expected value of a limit order resting at the best bid is approximately one quarter of a tick. The value of the limit order drops off substantially as we move away from the best quotes: the expected value of limit orders resting at one and two levels behind the best bid are 0.10 ticks and 0.03 ticks, respectively. There is also large variation in the expected value of a limit order. For example, our findings reveal that for all price levels, a resting limit order loses almost half its expected value when it transitions from the front to the back of the queue, which quantifies the theoretical predictions of Yueshen (2021), Li et al. (2020) and Yao and Ye (2018), who propose that queue priority is advantageous. Second, we extend the findings of Parlour (1998) by quantifying the importance of queue size. Specifically, we show that the expected value of an order increases with the queue size resting behind the order, and decreases with the queue size in front of the order. Third, we document that the expected value of a limit order at the best price declines as the size of the opposing queue increases.

Last, we show that volatility is important for the limit order's value, with its effect contingent on whether the stock is tick constrained, consistent with Li et al. (2020). Foucault (1999) predicts that volatility has two opposing forces on a limit order's profitability. The first force suggests an increase in volatility *decreases* the expected value of a limit order via an increase in the risk

[3] Unlike conventional theoretical models that rely on equilibrium assumptions and stylized trader behavior, our RL framework imposes structure on rich empirical data to identify which theoretical features most influence traders' order submission decisions.

of adverse selection. However, the second force suggests an increase in volatility *increases* the expected profit of a limit order as liquidity providers counteract losses from an increase in adverse selection risk by widening the bid ask spread. Further, Li et al. (2020) show that the minimum tick size also plays an important role. If the breakeven bid ask spread is always less than the one tick-mandated spread, liquidity providers do not widen the bid ask spread to compensate for the increased picking off risk even in times of high volatility. Thus, volatility decreases the expected value of the limit order as the compensation for providing liquidity (i.e., the difference between the quoted and breakeven bid ask spread) falls but remains positive.

Conversely, if volatility increases such that the breakeven bid ask spread widens beyond the one tick-mandated spread, liquidity providers react by widening their quoted bid ask spread to compensate themselves for the additional risk. In doing so, an increase in volatility increases the value of the limit order. Consistent with these predictions, we find that volatility has mixed effects depending on whether the stock is tick constrained. For stocks that are most tick constrained, an increase in volatility decreases the expected value of a limit order at the best price. On the other hand, we show that the value of a limit order increases with volatility for stocks that are most tick constrained.

Our RL approach also enables a comparative analysis of these market features while simultaneously accounting for the intricate interdependencies among them. Our analysis ranks the price level at which a limit order is placed as the most critical variable for traders to consider. This finding suggests that it is important for theory models to consider discrete price levels and order submissions at or away from the quote as in Goettler et al. (2005). Following the price level, the subsequent factors in order of importance are queue sizes at different price levels, market volatility, and the queue position of the order.

How valuable is the option to cancel a limit order? We find the option to cancel represents 19% of a limit order's total expected value, on average. This option becomes even more valuable during periods of high *ex-ante* adverse selection risk. In the most extreme case, we demonstrate that a limit order, which would otherwise have a negative expected value, can have a positive expected value purely because of the option to cancel the order at a later time.

Despite the complexity of limit order management, theory has shed some light on the way traders manage their orders. Among others, Parlour (1998), Foucault (1999), Goettler et al. (2005), Foucault et al. (2005), Goettler et al. (2009), Rosu (2009), Ricco et al. (2020), Rosu (2020), and Bhattacharya and Saar (2022) propose multi-period equilibrium models, which represent limit order markets as sequential games. In these models, traders arrive sequentially and submit, or update, the optimal order that maximizes their gains from trade. However, these models differ in the features that are modeled. For example, some models highlight the importance of volatility (Foucault (1999)) while others demonstrate the importance of queue size (Parlour (1998)). In some models, traders can only submit to one price level (Parlour (1998)) while in other models, traders can submit to prices beyond the best quotes (e.g., Goettler et al. (2005)). In Goettler et al. (2005), Foucault et al. (2005), and Ricco et al. (2020) traders can enter the market once, while in Goettler et al. (2009), Rosu (2009) and Bhattacharya and Saar (2022) traders can reenter the market. Many of the features of these models have not been jointly tested due to the lack of technologies available to researchers.

Our study complements traditional theory models in that our approach is completely driven by data, enabling us to empirically assess existing theories. In the spirit of Sandås (2015), who tests Glosten (1994) via a structural model, we use a structural RL model, which removes the need for assumptions about trader behavior or market dynamics. This allows us to identify optimal order management strategies under real-world market conditions, where traders may act in ways that deviate from rational or stylized behavior.

To address concerns about the 'black-box' nature of machine learning methods, which can obscure economic interpretation (see Chinco et al. (2019)), we ground our model in variables motivated by the theoretical literature. In this study, our estimation is based on an order of unit size, conditional on order submission. Nevertheless, the flexibility of our data-driven framework allows for the inclusion of alternative or additional features in the state space. For example, the framework can be adapted to investigate a trader's choice between a market or limit order, determine the optimal order size given current market conditions, or consider factors such as a trader's inventory, risk tolerance, and private information. We further explore these capabilities in Section 5.

We contribute to the literature in three ways. First, our approach is the first attempt to quantify the value of a limit order and to systematically assess the factors influencing its value. In doing so, we can explore previously untested theoretical predictions and uncover new interactions in financial markets. While existing theoretical models highlight the importance of queue size and queue priority (Parlour (1998), Yueshen (2021), Li et al. (2020) and Yao and Ye (2018)), we quantify the value of queue position to a liquidity provider across various market conditions. We empirically show that volatility has mixed effects on the value of a limit order depending on the degree of tick constraint, which is consistent with the predictions in Foucault (1999) and Li et al. (2020). Importantly, because ML techniques are well suited to complex environments characterized by multiple variables, we can evaluate these relations while holding all other market conditions constant.

Second, our research contributes to the literature on order cancellations. In Copeland and Galai (1983), when a market maker posts a bid or offer, they effectively write an option. However, this order also grants the market maker an option to cancel the order at some future point in time. Despite the substantial increase in order cancellations, constituting 47% of all messages (Brogaard et al. (2019)), understanding the implications and value associated with the option to cancel an order remains limited. Our study complements Dahlström et al. (2023), who investigate the determinants of order cancellations by liquidity providers, by highlighting the economic significance of the option to cancel.

Finally, we contribute to the growing literature applying learning algorithms to financial markets. Similar to this study, Bhattacharya and Saar (2022) use a recursive procedure to solve their model of dynamic limit order markets and Ait-Sahalia and Saglam (2023) model the high frequency trader's optimization problem as a Markov Decision Process. Dou et al. (2024) explore how AI-powered trading algorithms, specifically those combining algorithmic trading with reinforcement learning, impact price efficiency in a theoretical framework. In Colliard et al. (2022), algorithmic market makers set quotes using Q-learning algorithms and their trading outcomes are compared to the outcomes predicted by theory. We also apply a Q-learning approach to evaluate the decision making process of liquidity providers but in contrast to Colliard et al. (2022) and Dou et al. (2024), our analysis tests existing theory using empirical data. O'Hara (2015) highlights that in the modern era, markets and trading have changed, with limit orders now playing a more crucial role. Similarly,

Easley et al. (2021) issue a call to update the learning models and empirical methods used. Our paper answers this call by proposing a novel technique that provides a deeper understanding of limit order management than traditional learning models or empirical methods allow.

2 Method

2.1 Intuition

Consider a liquidity provider, or trader, who wants to optimally manage their limit orders to ensure that only limit orders with a positive expected value execute.[4] The dynamics of the limit order book make this task non-trivial, as the trader must constantly monitor their resting limit orders and cancel an order if it is expected to lose money. To achieve this task, the trader must estimate the expected value of a limit order conditional on the current state of the market *and* the future optimal management of the order over its life cycle.

Estimating the expected value of a limit order, conditional on its future optimal management, requires the trader to consider the evolution and likelihood of various market conditions. The trader must evaluate these conditions until one of two events occurs: 1) the order is executed, or 2) the order is canceled. The decision to cancel an order is endogenous and should be made when the limit order has a negative expected value.

Figure 1 illustrates the trader's problem. Initially, the limit order book is in a certain state at time t_0. The gray rectangles represent the volume available at the ask prices, and the white rectangles represent the volume available at the bid prices. The best bid and ask prices are 13 and 14, respectively, resulting in a bid ask spread of 1. In Figure 1, we assume a trader submits a limit buy order at t_0 at a price of 12 (one tick behind the best available bid) and depict this order as a black rectangle.

[Insert Figure 1]

[4] Similarly, automated market makers use a learning algorithm to pick the price that generates the largest expected profit in Colliard et al. (2022).

The trader then monitors the limit order book until the volume on the current best bid is removed, which occurs at t_1. For illustrative purposes, we assume the market evolves into one of only two possible states at t_1: State A or State B. In State A, since t_0, other market participants have submitted buy limit orders at 12, causing our trader's order to move up the queue at 12. Further, market participants have added buy limit orders at 11, and some of the sell limit orders at 14 have been removed, either due to cancellations or executions. In contrast, in State B, no new market participants have submitted additional buy limit orders. Instead, a large sell limit order at 13 has been submitted, removing the bid at 13 that existed at t_0.

If the volume available on the bid side of the order book is significantly larger (or smaller) than the volume available on the ask side of the order book, the midprice is more likely to increase (or decrease) in the near future (see Cao et al. (2009)). Therefore, the order in State A has a positive expected value, as the volume on the bid side is much larger than the volume on the ask side, suggesting a future price rise. In contrast, the order in State B has a negative expected value, as the volume available on the ask side is much larger than the volume available on the bid side, indicating the price is likely to decline in the future and the order would be adversely selected.

The expected value of the limit order submitted at t_0, if left unmonitored, is the sum of the expected values in State A and B, each weighted by their respective probabilities. Therefore, if the probability of transitioning to State B is much higher than the probability of transitioning to State A, the expected value of the unmonitored limit order at t_0 could be negative. However, if monitoring and the option to cancel the limit order are allowed, the expected value of the order becomes positive. This is because the trader will cancel the order if the market transitions to State B, resulting in a profit of 0, and leave the order if the market transitions to State A, where it has a positive expected value. This oversimplified example demonstrates that the option to cancel can transform an order from having a negative expected value to a positive one.

In this illustrative example, we make two tenuous assumptions. First, we assume the market can only transition to two possible states once the trader's order is submitted. In reality, the market can transition to an almost infinite number of states. Second, we arbitrarily assert that the limit order has a positive expected value in State A and a negative expected value in State B. Instead of

relying on these arbitrary assertions, we can achieve a precise estimate of the true expected values for State A and State B at time t_1 by calculating the expected value of the limit order within both states, conditional on the order's optimal management throughout its life cycle. This problem presents the same challenge we are attempting to address at t_0.

To overcome these limitations, we use a recursive state space technique known as reinforcement learning (RL). This technique allows us to accommodate numerous states and capture the inherently recursive nature of the problem.

2.2 Reinforcement Learning

Typically in an RL framework, an agent has knowledge of the current state, s, and then makes an action, a. Jointly, we refer to this state-action pair as an experience tuple defined as $\langle s, a \rangle$. If there are S states and A actions, then the agent has the choice of making A possible actions in S different states, which implies there are $S \times A$ unique experience tuples. We assume that each experience tuple can transition the agent to a new state, s', with probability $T(\langle s, a \rangle, s')$. For each action in a given state, the agent receives an immediate reward, $R(s, a)$. The agent's objective function is to maximize the total future reward by choosing the appropriate actions for each state that maximize the long-run discounted sum of all the immediate rewards received for each action in the future.

More formally, if we define the rules or policy an agent must follow as π, the optimal value of a state is computed as follows:

$$V^*(s) = \max_{\pi} E\big(\sum_{t=0}^{\infty} \gamma^t E[R(s_t, a_t)]\big), \tag{1}$$

where $E[R(s_t, a_t)]$ is the expected immediate reward at time t and γ is a discount factor bound between 0 and 1. $V^*(s)$ is the expected infinite discounted sum of reward the agent receives if they start in state s and execute the optimal policy defined by π^* moving forward. In our setup, the optimal policy, π^*, defines how the trader should optimally manage their limit order moving

forward (i.e., the action the trader should take given current market conditions and current order positioning). Similarly, the reward is the profit generated from earning the spread or favorable price movements after the order executes.

For every experience tuple, there is an associated Q-value, $Q^*(s,a)$, which is the expected infinite discounted sum of reward the agent gains if the agent takes action a while in state s, then subsequently follows the optimal policy path. Using (1), we note that $Q^*(s,a)$ can be expressed recursively as:[5]

$$\underbrace{Q^*(s,a)}_{\substack{\text{long run expected value}\\ \text{from taking action } a}} = \underbrace{E[R(s,a)]}_{\substack{\text{expected immediate}\\ \text{value from}\\ \text{taking action a}}} + \gamma \underbrace{\sum_{s' \in S} \overbrace{T(\langle s,a\rangle, s')}^{\substack{\text{probability of}\\ \text{transitioning to}\\ \text{future state } s'\\ \text{by taking}\\ \text{action } a}} \overbrace{\max_{a'}(Q^*(s',a'))}^{\substack{\text{expected long}\\ \text{run value from}\\ \text{taking optimal}\\ \text{action } a' \text{ when}\\ \text{in state s'}}}}_{\substack{\text{expected future value}\\ \text{from taking future optimal}\\ \text{actions, } a', \text{ while in future states, } s'}}, \tag{2}$$

where s' and a' define future states and actions, respectively. Equation (2) is the basis of our framework. In our setup, $Q^*(s,a)$ is the expected long run value of the limit order if the trader takes action a while in state s and in all future states s' takes the optimal action a'. We observe that this expected long run value equals any immediate value for taking action a plus the expected long run value the trader receives in future state s' if they make optimal future action a'. Recognizing that the future state s' is not known with certainty, our RL model assigns different transition probabilities, $T(\langle s,a\rangle, s')$, for all possible future states. Equation (2) is recursive because both the right hand side and the left hand contain a $Q^*(s,a)$ term. Thus, for estimation we use an iterative learning rule known as Q-learning.[6]

Estimating (2) requires us to first define a state-action space that reflects the problem of optimal limit order management. Specifically, the states should capture current market conditions and information about the order, while the actions should reflect the decisions available to the trader. Next, estimation requires two key input variables: the immediate reward and the transition probabilities. In the following sections, we describe how we cast the optimal limit order management

[5] See Watkins and Dayan (1992) for a full derivation.

[6] We provide a detailed illustrative example of the learning rule in Appendix C

problem within the RL framework. We explain the basic timing of our trader's decision process, define our state and action space, and describe how we empirically estimate the input variables: the immediate reward and the transition probabilities.

2.2.1 Timing

Figure 2 depicts the timing of our trader's decisions. In essence, the trader follows a recursive Markovian decision making system. At the start of each interval, the trader makes a decision based on observations of the current market conditions, for example, the existing shape of the order book and their own private information about their limit order's status. The trader decides whether to leave or cancel their existing limit order. At the start of the subsequent interval, the trader repeats the same decision making process. This decision-making process repeats continuously until the limit order executes or the order is canceled. If the limit order executes, the trader continues to monitor market conditions to observe the long-term value of the executed order.

[Insert Figure 2]

This recursive decision-making system allows the trader to keep the same limit order active for multiple consecutive intervals. During this time, the trader can monitor the order's queue position and market conditions. If at any point the order appears to have a high chance of adverse selection, indicated by a negative expected value, the trader cancels the order.

In our empirical section, we select a short time interval of 100ms. Choosing a short time interval offers three advantages. First, a short interval more closely reflects a trader who continuously monitors their orders. Second, a shorter interval provides more data points for model estimation. Third, it allows us to produce more accurate estimates of the likelihood of transitioning to future market conditions, as dramatic changes are less likely to occur over short intervals.

2.2.2 Actions

The A actions available define all possible decisions or individual actions, a, a trader can make given the current state. In our setup, the trader can make two possible actions. The trader can either cancel their resting limit order, which we define as C, or the trader can leave their existing limit order in the queue by taking no action, which we define as NA. Taken together, the trader's action space is defined by

$$a \in \{C, NA\}. \tag{3}$$

Figure 2 depicts the timing of the actions. Specifically, the trader decides on an action at the beginning of the interval. To ensure that the trader's limit order remains at price levels within our defined state space, we make the following adjustment: when the market transitions to a state in which limit order lies outside the state space, then the action C supersedes action NA. This implies that the resting limit order is canceled. This adjustment forces the trader to cancel resting limit orders if the best bid and offer has diverged away from the trader's resting limit order.

2.2.3 States

The state, s_t, reflects information available to the trader about the environment at time t. We decompose the environment into two sets of variables that reflect the current state: private and public. The public variables represent current market conditions available to all market participants. Parlour (1998) suggests that queue sizes in the limit order book influence the strategic behavior of traders. For this reason, we include the size of the queue at the best bid, one tick below the best bid and two ticks below the best bid, which we define as q^{B_0}, q^{B_1} and q^{B_2}, in our state space. Similarly, we include the size of the queue on the opposing side of the book (the best ask), which we define as q^{A_0}. Given queue sizes are essentially continuous, for tractability, we reduce the dimensionality of the state space by discretizing queue sizes. Specifically, we categorize queue lengths into five quintiles; extremely long (ELo), long (Lo), normal (No), short (Sh) and extremely short (ESh).[7]

[7]To further reduce dimensionality, we discretize the queue size at q^{B_2} to only three terciles.

Moreover, Foucault (1999) finds that volatility is a main determinant for limit order management. For this reason, we also include volatility, V, as a public variable, which we discretize into terciles; low (Low), medium (Med) and high (Hi).[8]

The private variables we use to define our state space capture information that is unique to the trader. Specifically, we capture the trader's current inventory position, I, which in our model is either 0 (no position) or 1 (long). We also include a variable, L, which captures the price level of the trader's limit order. We let L take on the value of $i \in 0, 1, 2$ if the trader has a resting limit order submitted at level i of the order book. Finally, because there is an advantage to being at the top of queue as the order has time priority (see Yueshen (2021), Li et al. (2020) and Yao and Ye (2018)), we include the queue position of any resting limit orders in our state space, which we define by Q. Similar to our previous variables, for tractability, we reduce the dimensionality of our queue position to five quintiles, which we define as *top*, *top-middle*, *middle*, *middle-back* and *back*.

Last, to ensure we estimate the expected value of a single limit order in isolation, we include a state that captures when the trader cancels their order. This state is a terminal absorbing state where the trader remains once they cancel their order. We define this terminal state by setting $Q = X$ and $L = X$. Taken together, these definitions let us express the current market state, s, as a vector

$$s = [I, L, Q, q^{B_0}, q^{B_1}, q^{B_2}, q^{A_0}, V] \tag{4}$$

[8]To proxy for volatility, we compute the difference between the log of the highest and the log of the lowest traded price over the last 100 trades in the stock.

where

$$I \in \{0, 1\}$$

$$L \in \{0, 1, 2, X\}$$

$$Q \in \{top, top\text{-}middle, middle, middle\text{-}back, back, X\}$$

$$q^j \in \{ELo, Lo, No, Sh, ESh\}, \forall j \in \{B_0, B_1, B_2, A_0\}$$

$$V \in \{Low, Med, Hi\}$$

In our setup, we restrict the trader to executing only one limit order. We achieve this restriction by ensuring no additional orders exist once a long position is achieved. As a result, the states when the trader is long are only defined by the four public limit order book information variables and volatility $(q^{B_0}, q^{B_1}, q^{B_2}, q^{A_0}, V)$. This restriction implies there are m possible states when the trader is long.[9] In contrast, when the trader has no inventory and is working their limit order, the state is defined by the public variables (i.e., queue sizes and volatility) plus the private variables (i.e., queue position and the price level of the resting limit order). The additional private variables results in n possible states when the trader has no inventory.[10] Collectively, in this setup, we have n states when the trader has no inventory, m states when the trader is long and one absorbing state for when the trader cancels their order, thereby resulting in $m + n + 1$ total possible states, where $n > m$ and $m + n + 1 = S$.

2.2.4 Transition matrix

With the states and action defined, we require estimates of the transition probabilities. Recall that if the limit order is currently in state s and the trader makes action a, the order transitions to states s' with probability $T(\langle s, a\rangle, s')$. Since the transition probabilities from state i to all other

[9] In our setup $m = 1,125$ as we have three public limit order book information variables $(q^{B_0}, q^{B_1}, q^{A_0})$, each with five possible values, one order book information variable with three possible values (q^{B_2}) and one volatility variable (V) with three possible values. Thereby resulting in $5^3 \times 3 \times 3$ possible combinations.

[10] In our setup $n = 16,875$. We have 1,125 possible public states, plus the private price level and queue position variables, which have three and five possible values respectively. Collectively, these variables result in $1,125 \times 5 \times 3$ possible combinations.

states must sum to 1 for a given action, for all i and a, $\sum_{j=1}^{S} T(\langle s_i, a\rangle, s_j) = 1$.

Because our framework has S unique market states, each action has an $S \times S$ transition probability matrix. When the trader makes no action (i.e., action NA), which leaves their resting limit order, the future state the limit order transitions to is not known with certainty. Thus, we empirically estimate the $S \times S$ transition probabilities for action NA. To estimate $T(\langle s_i, NA\rangle, s_j)$ we determine the number of times we observe a limit order in state s_i, followed by the limit order being in state s_j in the subsequent interval, and express this number as a fraction of all observations of limit orders in state s_i. More formally, if we define $N_{i,j}|NA$ as the number of times a limit order in state i transitions to state j, it is straightforward to show that the MLE estimate of $T(\langle s_i, NA\rangle, s_j)$ is

$$T(\langle s_i, NA\rangle, s_j) = \frac{N_{i,j}|NA}{\sum_{j=1}^{S} N_{i,j}|NA}. \tag{5}$$

In contrast, when a trader cancels their limit order, they transition to the absorbing cancel state with certainty. For this reason, we do not require empirical estimates for the $S \times S$ transition probabilities for action C, as the probability of transitioning to the absorbing cancel state is always 1. To ensure the trader only has one resting limit order, we restrict any state where the trader has an inventory position, or has already canceled their order, to not having a resting order. Because of this restriction, the action to cancel is prohibited and has a zero probability for all states where the trader has a long inventory position, or has canceled their order.[11]

To generate the full transition matrix, T, that captures all state actions, we vertically stack the $S \times S$ transition matrix for action NA on top of the $S \times S$ transition matrix for action C.

2.2.5 Immediate reward

An action from a given state can transition the trader to a new state and produce an immediate reward in the process. In our setup, the immediate reward captures any value, or profit, generated

[11] In Appendix A, we provide a detailed description of the structure and design of the transition matrices for each action.

during the transition from the current state to the next. This profit is derived from two sources: 1) price movements while carrying inventory, and 2) earning the spread through limit order execution. If the trader holds inventory when transitioning from state s to s', the immediate reward for this transition is the observed change in the midpoint during the transition (first component of equation (6)). Alternatively, if the trader's limit order executes during the transition from s to s', they profit by earning the spread. In this case, the immediate reward is the difference between the midpoint price observed in state s' and the execution price of the limit order (second component of equation (6)). Finally, if the trader has no inventory in state s and no order executes during the transition from state s to s', then the immediate reward must be zero. Formally, defining the midpoint price in state i as mid_i, the immediate reward from making action a while in state s that results in a transition to state s' is

$$R(\langle s, a\rangle, s') = \underbrace{(mid_{s'} - mid_s) \times I_s}_{\substack{\text{Profit from carrying} \\ \text{inventory}}} + \underbrace{(mid_{s'} - execPrice_s) \times Exec_{s,s'}}_{\text{Profit from execution}}, \tag{6}$$

where I_s equals 1 if the the trader has a long inventory position when in state s and 0 otherwise and $execPrice_s$ equals the price of the limit order and $Exec_{s,s'}$ equals 1 if the limit order executes during the transition from state s to s' and zero otherwise.

To compute the immediate reward for each transition, we require empirical estimation when the trader leaves their order (action NA). To obtain these estimates, we first compute the immediate reward using equation (6) for every observation in the data. Then, for each state-action transition, we compute the average immediate reward across all observations that belong to that state-action transition.[12] In contrast, when the trader cancels their order, the immediate reward must be zero as they have no limit orders executed and no inventory position. Therefore, for action C, the $S \times S$ immediate reward matrix contains only zeros.

Similar to the transition matrix, we create the immediate reward matrix for all experience

[12] In Appendix B, we provide a detailed description of the structure and design of the immediate reward matrices for action NA.

tuples by vertically stacking the immediate reward matrix for action NA and the immediate reward matrix for action C, resulting in a matrix of dimension $2S \times S$. We compute the expected immediate reward for taking action a while in state s as

$$E[R(s,a)] = \sum_{s' \in S} T(\langle s,a \rangle, s') \times R(\langle s,a \rangle, s'). \tag{7}$$

3 Data

We use ITCH data for the Australian Securities Exchange (ASX) extracted from the SIRCA database for the period July 3, 2017 to September 29, 2017. Table 1 contains summary statistics for the 20 sample stocks analyzed, ranging from the lowest price stock of Santos (STO), with a price of approximately 3.50 over the sample period to CSL Ltd. (CSL) with an average price of 129.86. The sample stocks also cover a wide range of average bid ask spreads, from 1.00 tick to 2.59 ticks.

[Insert Table 1]

The ITCH data provides comprehensive order book information with nanosecond-level timestamps, allowing us to fully reconstruct the order book at all price levels. We extract detailed information for each resting limit order, including its queue position. To facilitate our analysis, we process the data according to the following steps. First, we reconstruct the limit order book, enabling us to replay the market activity throughout a trading day. Second, for each trading day, we create 210,000 consecutive intervals, each 100 milliseconds long. The first interval begins at 10:10, coinciding with the start of continuous trading, and the last interval concludes at 16:00 when continuous trading ends.

At the beginning of each interval, we assume there is a series of hypothetical limit orders positioned at various price levels and queue positions. To align with our RL model, we consider hypothetical bids for one share placed at the prevailing best bid, one tick behind the best bid,

and two ticks behind the best bid. Additionally, at each of these price levels, we assume there is a hypothetical order positioned at the top of the queue, three-quarters of the way up the queue, halfway up the queue, one-quarter up the queue, and at the very back of the queue.[13]

Next, using the granularity of the data, we track these hypothetical limit orders over the next 100ms and determine if any of the orders execute.[14] In the event that the hypothetical limit order remains unexecuted within the 100ms timeframe, we monitor its progress by tracking the execution of real orders that precede it, as well as the cancellation and submission of real orders during the interval. This approach enables us to determine the position of the hypothetical order within the order book.

For each hypothetical order, we record information on the state space at both the beginning and the end of the interval. Specifically, at the start of the interval, we capture the volatility, initial queue position, and the total volume available at the first three best bid and best ask prices. At the end of the interval, we note whether the order executes. If the order does not execute, we report the order's new queue position. Further, regardless of execution, we record the volatility and total volume available at the first three best bid and best ask prices at the end of the interval.

With the extracted information, we can identify each order's initial starting state and its state at the end of the interval. This information allows us to estimate the transition matrix and immediate reward matrix using the process outlined in Section 2.

4 Results

In this section, we estimate our model using four public state variables based on the limit order book queue sizes (q^{B0}, q^{B1}, q^{B2} and q^{A0}), each with five possible values (except q^{B2}, which only has three states for tractability reasons).[15] Additionally, we include a public volatility state variable

[13] We assume each order consists of only one share to ensure it does not have a significant economic impact. Additionally, the price and queue position of the hypothetical orders are selected to ensure our observations cover the state space defined by our RL framework.

[14] We assume a hypothetical order executes if, during the 100ms interval, a real order positioned behind it in the queue executes, or if a trade occurs at a price worse than the hypothetical order's price.

[15] Queue size quantiles are formed for each stock at each price level.

with three possible values, where volatility is defined as the difference between the log of the highest traded price and the log of the lowest traded price over the last 100 trades. We also consider two private state variables related to the trader's resting limit order: L and Q, which have 3 and 5 possibles values, respectively.

This state space results in 16,875 different states when the trader has no inventory and is executing a limit order, 1,125 unique market states when the trader's order has executed and they hold an inventory position, and 1 absorbing cancel state. In total, this gives us $m = 16,875$, $n = 1,125$ and $o = 1$, resulting in 18,001 unique states.[16]

In Section 4.2 , we investigate the effect of each market feature on the expected value of a limit order. In Section 4.3, we assess the relative contribution of each market feature to the order's expected value. Finally, in Section 4.4, we evaluate the value of the option to cancel a limit order.

4.1 The expected value of a limit order

The results presented in Table 2, Panel A show that an average limit order submitted within two ticks of the best bid and ask price has an expected value of 0.146 ticks, conditional on the order's optimal management over its life.

Next, we investigate the relation between a limit order's price level and its expected value. The relation between a resting limit order's price and expected value is not immediately clear due to two opposing forces. First, the further a limit order is from the best bid or offer, the more favorable the execution price. However, this price improvement comes at the cost of lower execution probabilities (see Handa and Schwartz (1996)).

Figure 3 presents a boxplot of expected value for all markets states at each of the three price levels defined in our state space (best bid, one tick behind and two ticks behind the best bid). In Figure 3, we observe that the expected value of a limit order is positive, on average. This result is consistent with the empirical findings of Handa and Schwartz (1996), who report that a randomly submitted limit order is profitable and supports the hypothesis that liquidity providers

[16] For further clarity, we demonstrate the full estimation process via a detailed illustrative example in Appendix C.

who accommodate purchases (sales) should be compensated with a higher (lower) price than the fundamental value (see Scholes (1972)).

[Insert Figure 3]

Table 2, Panel B reports the summary statistics for our expected value estimates. The first row reports summary statistics for all market states, whereas rows 2 to 4 report summary statistics for limit orders conditional on their price level. Consistent with Figure 3, when the order is resting at the best bid, its mean expected value is highest at 0.262 ticks. When an optimally managed limit order moves away from the best bid, its expected value drops to 0.105 ticks when it is one tick behind the best bid, and drops further to 0.031 ticks, when it is two ticks behind the best best bid. Similarly, the variance in a limit order's conditional expected value decreases as the order moves away from the best price. The expected value of an order located at the best bid has a standard deviation of 0.134 ticks, but this value drops to only 0.01 ticks when the order is two behind the best bid.

[Insert Table 2]

The observation that the mean expected value and variance of expected value decreases as the order moves further away from the best bid or offer may explain why the majority of order cancellations occur at the best bid or ask (see Fong and Liu (2010)). Intuitively, a trader has little incentive to cancel a limit order resting far from the best price. Such an order likely has a small positive expected value because it carries minimal execution risk and could gain favorable queue priority in the future. However, if the market moves toward the resting limit order, increasing its probability of execution, the expected value could turn negative. At that point, the trader should consider canceling the order.

4.2 Features influencing a limit order's expected value

4.2.1 Queue position

In this section, we investigate the effect of a limit order's queue position on the order's expected value. Some argue that there is an advantage to being at the top of the queue, due to the time priority rule (see Yueshen (2021), Li et al. (2020) and Yao and Ye (2018)). In contrast, some literature suggests that small incoming market orders are more informed (see Brogaard et al. (2014)). Thus, orders at the top on the queue execute against these small informed orders, whereas orders further back in the queue can only execute against larger, less informed orders. To determine the effect of queue position on the expected value of a limit order, we estimate the following regression:

$$Q_s = \beta_1 QueuePos_s + State\ Fixed\ Effects + \epsilon_s, \tag{8}$$

where Q_s is the expected value of a limit order in state s and $QueuePos_s$ is the order's queue position (0 being the top and 1 being the back) in state s. We use fixed effects for all other variables that define our state space to isolate the effect of queue position.

Table 3 presents the mean coefficient for orders resting at three different positions: the best bid (column 1), one level behind the best bid (column 2), and two levels behind the best bid (column 3), across all 20 sample stocks. Additionally, Table 3 reports the number of stocks with significantly positive or negative coefficients at the 5% level, as in Engle and Patton (2004).[17]

Our results strongly indicate that queue priority benefits the liquidity provider. The coefficient for queue position is negative and significant across all 20 sample stocks and all price levels. This implies that the further back a limit order is in the queue, the lower its expected value. The magnitude of these coefficients suggests that queue position has a substantial economic impact on the expected value of a limit order. For instance, an order's expected value at the best bid decreases

[17]Each regression has 5,625 observations, encompassing the following states: five queue position states, five queue size states at the best bid, five queue size states one tick behind the best bid, three queue size states two ticks behind the best bid, five queue size states at the best ask, and three volatility states. This results in a total of $5 \times 5 \times 5 \times 3 \times 5 \times 3 = 5,625$ observations.

by 0.12 ticks when it moves from the top of the queue ($QueuePos_s = 0$) to the back of the queue ($QueuePos_s = 1$). This decrease is economically significant, representing almost half of the average value of a limit order resting at the best bid, which is 0.262 ticks.

We also observe that the magnitude of the mean coefficient decreases as the order moves further from the best bid. Specifically, for orders at the best bid, the mean coefficient is -0.12. For orders one level behind the best bid, the mean coefficient is -0.05, and for orders two levels behind the best bid, the mean coefficient further decreases to -0.01. This pattern indicates that queue priority becomes more critical as the order moves closer to the best price, where execution is most likely.

[Insert Table 3]

Our results highlight the advantages of having orders positioned at the front of the queue, which is consistent with Yueshen (2021), Li et al. (2020) and Yao and Ye (2018). Orders at the front of the queue have higher execution probabilities and lower adverse selection risk compared to those at the back. This lower risk occurs because an order at the back of the queue can only execute against large incoming market orders, which have a significant adverse price impact when all resting limit orders at the current price level are removed. In contrast, orders at the front of the queue can execute against small incoming market orders, which do not exhaust the liquidity at the current price level. The combination of higher execution likelihood and lower adverse selection risk results in a higher expected value for orders at the front of the queue.

Our findings also support the argument by Lo et al. (2002) that the simulated profits from placing a network of buy and sell limit orders, as reported by Handa and Schwartz (1996), may be overstated. This is because their assumption that the orders are placed at the top of the queue does not fully account for the critical importance of queue priority.

4.2.2 Queue size

Existing theoretical literature suggests that queue sizes affect the value of a limit order (see Parlour (1998), Goettler et al. (2005), Goettler et al. (2009)). However, there are few empirical tests. In this

section, we empirically investigate how queue size affects the expected value of a limit order. To investigate the relation between queue sizes and expected value, we estimate the following regression for orders at different price levels:[18]

$$Q_s = \beta_1 q_s^{B_0} + \beta_2 q_s^{B_1} + \beta_3 q_s^{B_2} + \beta_4 q_s^{A_0} + \textit{State Fixed Effects} + \epsilon, \tag{9}$$

where Q_s is the expected value of a limit order in state s, $q_s^{B_i}$ is the size of the bid queue at level i in state s and $q_s^{A_0}$ is the size of the queue on the ask. To isolate the effect of queue sizes, we use fixed effects for all other variables that define our state space.

Table 4 presents the results for orders resting at the three different prices levels defined in our state space (best bid, one tick behind the best bid, two ticks behind the best bid). For each variable, we report the mean coefficient across all 20 sample stocks. To ensure the mean coefficients are not driven by one stock, we also report the number of stocks with statistically positive or negative coefficients.

Overall, our results suggest that the larger the queue size *behind* a resting limit order, the higher the expected value of the order. Conversely, the larger the queue size *in front* of a resting limit order, the lower the expected value of the order. Observing the results for orders resting at the best bid, we find that the mean coefficients for q^{B_0}, q^{B_1}, q^{B_2} are all positive at 0.06, 0.05 and 0.04 respectively, suggesting that an increase in queue lengths at or behind the resting order's price level increases the limit order's expected value. This relation weakens at price levels further away from the best bid. Not only do the average coefficients drop monotonically from 0.06 to 0.04 as we transition from q^{B_0} to q^{B_2}, but we also see the number of stocks with positive and significant coefficients drop from 19 to 18 to 14 as we transition from q^{B_0} to q^{B_1} to q^{B_2}.

In contrast to our findings for orders resting at the best bid, for orders behind the best bid (columns 2 and 3), an increase in queue sizes at price levels ahead of the resting limit order

[18]There is no existing theory suggesting that price levels have a linear affect on limit order value. Thus, we estimate a regression for all orders at each price level individually.

decreases the order's expected value. For example, the coefficient for queue sizes at the best bid (q^{B_0}) is negative and significant for all 20 sample stocks for orders resting one level behind the best bid (column 2). Similarly, for orders resting two levels behind the best bid in column 3, the coefficients for queue lengths at the best bid (q^{B_0}) and one level behind the best bid (q^{B_0}) are negative and significant for all 20 sample stocks.

[Insert Table 4]

Our findings manifest in two ways. First, when the volume in front of the limit order increases, the order's probability of execution worsens. This is because the volume in front of the order exerts order book pressure that can drive the price away from the order. Cao et al. (2009) demonstrate that if the volume on the bid side of the order book is significantly larger (smaller) than the volume available on the ask side of the order book, then the midpoint price is more likely to increase (decrease) in the near future. Thus, if a resting limit order has a large volume ahead of it, that limit order is more likely to be on the thick side of the book. As a result, the price is likely to shift away from the order, resulting in non-execution.

Second, a limit order with more volume in front of the order faces higher adverse selection risk. This is because the volume in front of the order must first execute before the limit order can execute. For example, a limit order behind a large block of volume can only immediately execute when a larger incoming market order enters to first remove the large block of volume. These large market orders cause the largest adverse selection (see Hasbrouck (1991)). In contrast, an order with no volume in front of it can execute against the next incoming market order, regardless of how small it is.

The relation between a limit order's expected value and the volume on the opposite side of the book also depends on the order's price level. In Table 4, the coefficient for the volume on the opposite side of the book (q_{A_0}) is negative and significant for all sample stocks when the order is on the best bid. However, the sign becomes positive and significant for orders behind the best bid. This finding suggests an increase in volume on the opposite side of the order book decreases (increases) the expected value of the limit order if it is at (behind) the best price.

This difference in effect is due to a trade off between adverse selection and execution probability. Cao et al. (2009) document that a large volume on the opposite side of the book creates book pressure that causes shifts in the midpoint towards the limit order, which increases both the likelihood of adverse selection and the probability of execution. When a resting limit order is on the best bid, an increase in ask volume increases the likelihood of a downtick, thereby increasing the expected losses from adverse selection. This negative effect is stronger than the potential gains resulting from a higher probability of order execution. In contrast, when the order is behind the best bid, the expected value from increased execution probability outweighs the expected losses from adverse selection risk. Adverse selection risk remains low for orders behind the best bid as the order can be subsequently canceled if market conditions worsen.

Taken together, our results provide strong support for Parlour (1998); we find that the larger the queue size *behind* a resting limit order, the *higher* the expected value of the order, and the larger the queue size *in front* of a resting limit order, the *lower* the expected value of the order. We also document that the queue size on the opposite side of the book has mixed effects due to a trade off between adverse selection and execution probability. As the queue size on the other side of the book increases, the risk of adverse selection and execution probability both increase. For orders resting at the best price, the losses from adverse selection outweigh the gains from higher execution probability. In contrast, for orders resting behind the best price, the gains from higher execution probability outweigh the loses from adverse selection. Overall, our findings provide support for the predictions of Parlour (1998), Goettler et al. (2005) and Goettler et al. (2009) that strategic traders should consider queue sizes at multiple price levels and demonstrate pervasive features that exist for orders at different price levels.

4.2.3 Volatility

In this section, we explore the impact of volatility on the expected value of a limit order. While existing theoretical models provide some insights into the relation between volatility and the expected value of a limit order, there is no clear consensus due to two opposing forces identified in the literature.

The first force suggests that an increase in volatility decreases the expected value of a limit order. Specifically, Foucault (1999) predicts that when volatility increases, the picking off risk for a limit order increases, and the losses that ensue are larger, decreasing the expected value of the limit order. The second force acts as a response to the first; to compensate for the higher likelihood of adverse selection and the corresponding reduction in the expected value of a limit order, liquidity providers widen the bid ask spread when volatility increases (as discussed in Copeland and Galai (1983) and Foucault (1999)). In a continuous order book where limit orders can be placed at any price level, liquidity providers can adjust the bid ask spread precisely to offset the anticipated losses due to the increased picking-off risk.

However, in practice, price levels are discrete, and liquidity providers may not always be able to set a breakeven bid ask spread that perfectly offsets the increase in volatility, as described in Li et al. (2020). In the first scenario, the breakeven bid ask spread is below one tick, as shown in Figure 4, Panel A. Here, the liquidity provider receives the difference between the one tick mandated bid ask spread and the breakeven bid ask spread as compensation for providing liquidity. The bottom of Figure 4, Panel A illustrates the effect of an increase in volatility. As volatility increases, the breakeven bid ask spread widens but still remains within the one-tick mandated spread, resulting in a decrease in the liquidity provider's compensation. This reduction in compensation lowers the expected value of the limit order. This scenario is particularly pronounced in stocks that are constrained by the minimum tick size.

[Insert Figure 4]

In the second scenario, shown in Figure 4, Panel B, an increase in volatility leads to a breakeven bid ask spread that exceeds the mandated minimum tick size. In response, liquidity providers widen their quoted bid ask spread to a level that is at least as wide as the breakeven spread. As a result, an increase in volatility raises the expected value of a limit order resting at the best bid price. This scenario is most common in stocks that are least constrained by the minimum tick size.

In summary, due to price discretization, volatility can either increase or decrease the expected value of a limit order, depending on whether the stock's trading is constrained by minimum tick

size requirements. Our initial analysis examines the impact of market volatility on the expected value of a limit order across all stocks. We then further explore the effects of volatility on individual stocks, focusing on whether their trading is constrained by minimum tick size requirements.

For our initial investigation using the full sample of stocks, we conduct the following regression analysis:

$$Q_s = \beta_1 Volatility_s + State\ Fixed\ Effects + \epsilon_s, \tag{10}$$

where Q_s is the expected value of a limit order in state s, and $Volatility_s$ is the discretized volatility in state s.[19] To isolate the effect of volatility, we use fixed effects for all other variables that define our state space. Since our primary interest is in the effect of volatility on orders at the best bid or offer, and this effect may vary across price levels, we estimate (10) on the subset of limit orders at the first price level of our defined state space.

Table 5, Column 1 reports an average coefficient of 0.38 across all sample stocks. While this average coefficient suggests that volatility increases the expected value of a limit order, a closer inspection of the individual coefficients for each stock reveals a more complex picture. Specifically, we find positive coefficients for 9 out of the 20 sample stocks, while 11 out of 20 stocks show negative coefficients. Thus, the effect of volatility on the expected value of a limit order at the best bid is not consistent across all stocks.

[Insert Table 5]

Because liquidity providers can widen the bid ask spread to offset the increase in picking off risk, volatility could have mixed effects on the expected value of a limit order (Foucault (1999)). According to Li et al. (2020), the impact of volatility should vary depending on whether a stock's trading is tick constrained. For stocks that are typically tick constrained, we propose that an

[19] Volatility is measured as the highest traded price minus the lowest traded price over the last 100 trades and discretized into terciles.

increase in volatility negatively affects the expected value of a limit order, as the breakeven bid ask spread is narrower than the one tick mandated spread. Conversely, for stocks that are less tick constrained, we predict that an increase in volatility will raise the expected value of a limit order. In these cases, liquidity providers can widen their quoted bid ask spread to compensate for the increased picking off risk during volatile periods.

To test whether volatility has different effects on the expected value of a limit order for tick constrained and unconstrained stocks, we split the sample stocks into two subsamples. Table 5, column 2 reports results for the quartile of the most tick constrained stocks, while column 3 reports results for the quartile of the least tick constrained stocks. Consistent with our hypothesis, we find that volatility decreases the expected value of a limit order for tick constrained stocks. Specifically, *Volatility* is negative and significant for all stocks within the tick constrained subset. Conversely, for all stocks not constrained by the tick size in Column 3, we find that volatility increases the expected value of a limit order.

Overall, our results support the findings of Foucault (1999) and the tick size channel proposed by Li et al. (2020). For stocks that are most tick constrained, the breakeven spread lies within the one-tick mandated spread. In these cases, liquidity providers are unable to widen the quoted bid ask spread, leading to a decrease in the expected value of a limit order as volatility increases. Conversely, for stocks that are least tick constrained, an increase in volatility can push the breakeven spread beyond the one-tick minimum. In response, liquidity providers widen the quoted bid ask spread to compensate for the additional picking-off risk. As a result, volatility increases the expected value of a limit order for these less constrained stocks.

4.3 How important are the variables?

The results so far indicate that price levels, queue sizes, queue position, and volatility all influence the expected value of a limit order. In this section, we assess the importance of these variables using a technique from the machine learning literature known as Mean Decreased Accuracy (MDA). This method has been applied to the finance literature by Easley et al. (2021) and Kwan et al. (2024). In our context, MDA measures the decrease in accuracy of the forecasted expected value of a limit

order when one of the variables defining our states is intentionally measured with error.

Estimating the MDA requires two parameters. The first parameter is the true expected value of a resting limit order, $Q(s, NA)$, which we estimate using the RL model described in Section 2. Specifically, we have 18,001 $Q(s, NA)$ estimates corresponding to 16,875 different states when the trader has no inventory and is executing a limit order, 1,125 unique market states when the trader's order has executed and they hold an inventory position, and 1 absorbing cancel state.

The second parameter is the randomized expected value of a resting limit order, $Q(s_R^k, NA)$, which we estimate by randomizing one of the seven variables that define the state space while keeping all other variables constant.[20] $Q(s_R^k, NA)$ represents the expected value associated with the randomly altered state, s_R^k, created by randomizing variable k. This randomization helps isolate the effect of variable k on the $Q(s, NA)$ estimate.

Using these two parameters, we estimate the MDA for variable k as follows:

$$MDA^k = \sum_{s=1}^{S} \left(\frac{|(Q(s, NA) - Q(s_R^k, NA))|}{Q(s, NA)} \right) / S. \tag{11}$$

The MDA measures the error in expected value estimates that occurs when a variable is measured with error. Therefore, the larger a variable's MDA, the more important that variable is in determining the expected value of a limit order. For each variable, k, we estimate the MDA and repeat this process 100 times.

Table 6 presents the mean and standard deviation of the MDA for each variable. Our findings indicate that the most influential factor affecting the expected value of a limit order is the price level at which the order rests. The next most important variable is the queue size on the same side of the order book as the limit order. However, the importance of queue size decreases as the queues move further from the best bid. Specifically, the queue size at the best bid (MDA = 1.22) is the most important, followed by the queue size one tick behind the best bid (MDA = 1.17), and then the queue size two ticks behind the best bid (MDA = 0.68).

[20] The seven variables include the price level, queue position, queue sizes at different price levels (best bid, one tick behind the best bid, two ticks behind the best bid, best ask), and volatility.

After considering queue sizes on the same side of the book as the order, the next most important variable is the queue size on the opposite side of the order book (MDA = 0.68), followed by volatility (MDA = 0.56), and lastly, queue position (MDA = 0.2).

[Insert Table 6]

4.4 The option to cancel

Despite the prevalence of order cancellations, the option to cancel has received little attention in the literature. In this section, we investigate the value of the option to cancel and identify the market conditions under which this option is most valuable. For our analysis, we estimate a constrained version of our RL model, which restricts the trader to only one action, NA, meaning the trader is unable to cancel their order.[21] Thus, the estimated Q values from this restricted model represent the expected value of a limit order that is not optimally managed. To determine the value of the option to cancel, we calculate the difference between the Q value of the unrestricted model, which includes the option to cancel, and the Q value of the restricted model, which lacks this option.

Table 7 reports the summary statistics on the value of the option to cancel. The first row reports summary statistics for limit orders across all market states, while rows 2 to 4 focus on limit orders conditional on their price level. On average, the value of the option to cancel a limit order on the best bid is 0.049 ticks. This means that, on average, the option to cancel a limit order from the best level is worth approximately 19% of the total value of an optimally managed limit order.[22] This finding suggests that the endogenous option to cancel a limit order contributes an economically meaningful amount towards the overall expected value of an optimally managed limit order.

[Insert Table 7]

Table 7 also suggests that the value of the option to cancel varies significantly depending on

[21] Recall in the full model, the trader can cancel the order, C, or take no action, NA.
[22] Table 2 reports a mean value of an optimally managed limit order at the best bid of 0.258 ticks.

prevailing market conditions. Regardless of the price level at which orders are resting, the data shows a substantial disparity between the means and medians, indicating a pronounced right skew. Over the full sample, the mean value of the option to cancel is 0.024 ticks, whereas the median is only 0.008 ticks.

Theoretical considerations suggest that the option to cancel is most valuable when the limit order is most likely to be adversely selected. However, it is difficult for a trader to know when adverse selection risk is high *ex-ante*. To proxy for an *ex-ante* measure of adverse selection risk, we draw on Cao et al. (2009), who show that order book pressure can be used to predict short term price movements, and Goldstein et al. (2023) who show that limit orders with high ex-ante adverse selection risk are more likely to rest on the thin side of the book. A higher volume of buy orders (bids) in the market creates buying pressure in the stock, and the price is more likely to rise. As a result, the adverse selection risk of a resting buy limit order decreases and the value of having the option to cancel the buy order also decreases. On the other hand, when volume on the ask side increases, the selling pressure is more likely to result in a price fall. The adverse selection risk of a resting buy order rises and the option to cancel the order becomes more valuable.

Drawing from these studies, we use order book pressure as a proxy for *ex-ante* adverse selection and estimate the following regression for the subset of limit orders at the front of the queue at each price level:

$$value\ of\ option\ to\ cancel = \beta_0 + \beta_1 q^{B_0} + \beta_2 q^{B_1} + \beta_3 q^{B_2} + \beta_4 q^{A_0} + \epsilon. \tag{12}$$

If the value of the option to cancel increases when adverse selection increases, we expect an increase in volume on the *same* side of the order (i.e., $q^{B_0}, q^{B_1}, q^{B_2}$) to reduce the value of the option to cancel. Similarly, we expect an increase in volume on the *opposite* side of the order (i.e., q^{A_0}) to increase the value of the option to cancel. Table 8 confirms this hypothesis and demonstrates the option to cancel is most valuable when book pressure is going against the order (i.e., adverse selection is high). Specifically, for all regressions, Table 8 reports a negative relation between the

queue size at *any* price level on the bid side and the value of the option to cancel. Similarly, for all regressions, the value of the option to cancel has a positive relation with the queue size on the opposing ask (q^{A_0}). In other words, when there is greater trading volume on the same side as the resting limit order, the option to cancel holds a lower value. Conversely, if there is more volume present on the opposite side of the limit order, the option to cancel carries a higher value.

[Insert Table 8]

5 Extensions

In this section, we describe possible extensions to the basic RL framework. Such enhancements include the exploration of strategic decisions related to selecting between limit orders and market orders, incorporating considerations such as the liquidity provider's risk tolerance or existing inventory levels, and determining order sizes.

5.1 Market vs. limit orders

A substantial body of theoretical work explores the dynamics behind traders' preferences for market versus limit orders. For instance, Kaniel and Liu (2006) reveals that, contrary to previous assumptions, informed traders are more inclined to use limit orders rather than market orders. In a fully dynamic framework, Bhattacharya and Saar (2022) show that the liquidity of the market significantly influences informed traders' decision to place limit orders in less liquid markets and marketable orders in more liquid markets. From an empirical standpoint, Ranaldo (2004) demonstrates that a trader's decision on order aggressiveness is dependent on the market's depth, spread, and volatility. The flexibility of our proposed RL framework allows researchers to study a trader's order choice while also taking into account factors such as their existing inventory levels, risk aversion, private information, as well as the prevailing market conditions.

The RL specification specified in section 2 can be modified to investigate the decision between limit and market order submissions. Specifically, we can extend the action space to include an

additional action that simultaneously cancels the resting limit order, and executes a market order by crossing the spread, M. This augmentation results in three possible actions 1) to leave the existing limit order (NA), 2) to cancel the existing limit order (C) or 3) to cancel the existing limit order and immediately execute a market order (M). As before, we solve Equation (2) via the Q-learning rule for all $Q(s, a)$ where $a \in \{C, NA, M\}$. This modification allows us to estimate the expected profit for each action, enabling us to compare the potential outcomes of leaving a resting limit order or sending a market order in a given state.

Moreover, we can modify the RL framework to examine the order submission strategies of an impatient trader who penalizes wait time until execution. Specifically, we can adjust the discount factor γ in Equation 2. Recall that γ is a discount factor between 0 and 1. Values close to 1 apply minimal discounting to future payoffs or rewards, while values close to 0 place little weight on future payoffs (i.e., future payoffs are heavily discounted). Therefore, using a value close to zero places little emphasis on rewards from limit order executions that occur in the distant future. This setup effectively represents an impatient trader who underweights the potential payoffs from long-lived limit orders.

5.2 Inventory and risk aversion

Inventory is also a crucial factor in managing limit orders. For example, Garriott et al. (2024) demonstrate that inventory levels and adverse selection constraints similarly affect limit order sizes. Consequently, a market maker holding a long position of 1,000 shares will place a higher value on a limit sell order compared to a market maker with a short position of 1,000 shares, assuming both market makers aim to minimize their inventory positions due to the associated risk.

We can adapt the existing RL framework to model the preferences of a risk-averse market maker with inventory considerations through two adjustments. First, we include the market maker's inventory position in the state space. Second, we modify the reward function to account for the profits related to the varying inventory position and any associated risk aversion. Specifically, equation 6 can be modified to:

$$R(\langle s,a\rangle,s') = \underbrace{(mid_{s'} - mid_s) \times Inv_s}_{\substack{\text{Profit from carrying} \\ \text{inventory}}} + \underbrace{(mid_{s'} - execPrice_s) \times Exec_{s,s'}}_{\text{Profit from execution}} - \underbrace{\lambda(|Inv_s|)}_{\text{Risk aversion}}, \tag{13}$$

where Inv_s is the market maker's inventory position in state s, $execPrice_s$ equals the price of the limit order and $Exec_{s,s'}$ is the volume of the limit order that executes during the transition from state s to s'. $\lambda(|Inv_s|)$ is a risk aversion penalty applied to the market maker's inventory position. This penalty can be defined using various risk aversion utility functions due to the flexibility of the RL framework.

5.3 Order size

In our RL framework, a risk-averse market maker (liquidity provider) trades a hypothetical order of unit size. This assumption ensures that the order is not economically meaningful and thus does not cause any permanent price impact. However, in practice, many limit orders are larger and do have an associated price impact (see Brogaard et al. (2019) and Kwan et al. (2024)). Here, we extend the RL framework by modifying the private variables in the state space to include the size of the limit order. This modification allows us to capture any permanent price impact that arises from submitting a limit order that is large enough to be economically meaningful.

While the estimation of expected profits via Q-learning remains unchanged, two modifications are required in the process. First, instead of using hypothetical limit orders of unit size, the transition matrix should incorporate appropriately sized orders that can alter the queue sizes in the state space. Second, the reward matrix must account for the volume executed, considering both partial and complete executions.

5.4 Private information

Many theoretical models assume that a trader has access to private information. Our reinforcement learning (RL) framework can be adapted to capture this feature. Specifically, we can introduce a private valuation variable to the state space, p, which represents the trader's private information. If we assume that the trader's private information is valuable, this modification will lead to limit buy orders having higher expected values when the private information indicates a potential price increase, and lower expected values when it suggests a potential price decrease.

6 Conclusion

In modern markets, where limit order submissions and cancellations constitute an overwhelming majority of trading activity, understanding the optimal management of limit orders is crucial. Despite its importance, our understanding of the dynamics of the limit order book and order management strategies remains limited due to the complexity and high dimensionality of the problem (see Parlour and Seppi (2008)).

To address this issue, we propose a recursive sequential framework for limit order management which allows us to empirically uncover the most important features contribution to the value of a limit order. In our framework, the expected value of a limit order is determined by current market conditions and future market condition expectations. The liquidity provider exercises the option to cancel a limit order if its expected value becomes negative.

Our findings reveal that the average expected value of a limit order resting at the best bid is approximately one quarter of a tick. However, this value is influenced by various market factors. Specifically, we demonstrate that queue size, order position, volatility, and order price significantly impact the expected value of a limit order. Using Mean Decreased Accuracy (MDA) to rank the importance of these variables, we find that price level is the most critical factor, followed by queue sizes, volatility, and queue position.

Finally, we show that the endogenous option to cancel is economically meaningful: On average,

this option to cancel represents 19% of a limit order's total expected value. During periods of high adverse selection risk, this option becomes even more valuable.

References

Ait-Sahalia, Y. and Saglam, M. (2023). High Frequency Market Making: The Role of Speed. *Journal of Econometrics*, Forthcoming.

Bhattacharya, A. and Saar, G. (2022). Limit Order Markets under Asymmetric Information. Working paper, Available at SSRN: https://ssrn.com/abstract=3688473.

Brogaard, J., Hendershott, T., and Riordan, R. (2014). High-frequency trading and price discovery. *Review of Financial Studies*, 27(8):2267–2306.

Brogaard, J., Hendershott, T., and Riordan, R. (2019). Price discovery without trading: Evidence from limit orders. *The Journal of Finance*, 74(4):1621–1658.

Cao, C., Hansch, O., and Wang, X. (2009). The information content of an open limit-order book. *Journal of Futures Markets*, 29(1):16–41.

Chinco, A., Clark-Joseph, A., and Ye, M. (2019). Sparse signals in the cross-section of returns. *The Journal of Finance*, 74(1):449–492.

Colliard, J.-E., Foucault, T., and Lovo, S. (2022). Algorithmic Pricing and Liquidity in Securities Markets. Working paper, Available at SSRN: https://ssrn.com/abstract=4252858.

Copeland, T. E. and Galai, D. (1983). Information effects on the bid-ask spread. *The Journal of Finance*, 38(5):1457–1469.

Dahlström, P., Hagströmer, B., and Nordén, L. L. (2023). The determinants of limit order cancellations. *Financial Review*, Forthcoming.

Dou, W. W., Goldstein, I., and Ji, Y. (2024). AI-Powered Trading, Algorithmic Collusion, and Price Efficiency. The Wharton School Research Paper.

Easley, D., López de Prado, M., O'Hara, M., and Zhang, Z. (2021). Microstructure in the Machine Age. *The Review of Financial Studies*, 34(7):3316–3363.

Engle, R. F. and Patton, A. J. (2004). Impacts of trades in an error-correction model of quote prices. *Journal of Financial Markets*, 7(1):1–25.

Fong, K. and Liu, W.-M. (2010). Limit order revisions. *Journal of Banking and Finance*, 34:1873–1885.

Foucault, T. (1999). Order flow composition and trading costs in a dynamic limit order market. *Journal of Financial Markets*, 2(2):99 – 134.

Foucault, T., Kadan, O., and Kandel, E. (2005). Limit order book as a market for liquidity. *The Review of Financial Studies*, 18(4):1171–1217.

Garriott, C., van Kervel, V., and Zoican, M. (2024). Queuing in limit-order markets.

Glosten, L. R. (1994). Is the electronic open limit order book inevitable? *The Journal of Finance*, 49(4):1127–1161.

Goettler, R. L., Parlour, C. A., and Rajan, U. (2005). Equilibrium in a dynamic limit order market. *The Journal of Finance*, 60(5):2149–2192.

Goettler, R. L., Parlour, C. A., and Rajan, U. (2009). Informed traders and limit order markets. *The Journal of Financial Economics*, 93:67–87.

Goldstein, M., Kwan, A., and Philip, R. (2023). High Frequency Trading Strategies. *Management Science*, 69(8):4413–4434.

Handa, P. and Schwartz, R. (1996). Limit order trading. *The Journal of Finance*, 51(5):1835–1861.

Hasbrouck, J. (1991). Measuring the information content of stock trades. *The Journal of Finance*, 46(1):179–207.

Kaniel, R. and Liu, H. (2006). So what orders do informed traders use? *The Journal of Business*, 79(4):1867–1913.

Kwan, A., Philip, R., and Shkilko, A. (2024). The conduits of price discovery: A machine learning approach.

Li, S., Wang, X., and Ye, M. (2020). Who provides liquidity and when? *Journal of Financial Economics, Fourthcoming*.

Lo, A. W., MacKinlay, A., and Zhang, J. (2002). Econometric models of limit-order executions. *Journal of Financial Economics*, 65(1):31 – 71.

O'Hara, M. (2015). High frequency market microstructure. *Journal of Financial Economics*, 116(2):257 – 270.

Parlour, C. and Seppi, D. (2008). Limit order markets: A survey. *Handbook of Financial Intermediation and Banking*, 5:63–95.

Parlour, C. A. (1998). Price dynamics in limit order markets. *The Review of Financial Studies*, 11(4):789–816.

Ranaldo, A. (2004). Order aggressiveness in limit order book markets. *Journal of Financial Markets*, 7(1):53–74.

Ricco, R., Rindi, B., and Seppi, D. (2020). Information, Liquidity, and Dynamic Limit Order Markets. Working paper, Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3032074.

Rosu, I. (2009). A dynamic model of the limit order book. *The Review of Financial Studies*, 22(11):4601–4641.

Rosu, I. (2020). Liquidity and information in limit order markets. *Journal of Financial and Quantitative Analysis*, page 1–48.

Sandås, P. (2015). Adverse Selection and Competitive Market Making: Empirical Evidence from a Limit Order Market. *The Review of Financial Studies*, 14(3):705–734.

Scholes, M. S. (1972). The market for securities: Substitution versus price pressure and the effects of information on share prices. *The Journal of Business*, 45(2):179–211.

Watkins, C. J. C. H. and Dayan, P. (1992). Q-learning. *Machine Learning*, 8(3):279–292.

Yao, C. and Ye, M. (2018). Why Trading Speed Matters: A Tale of Queue Rationing under Price Controls. *The Review of Financial Studies*, 31(6):2157–2183.

Yueshen, B. (2021). Queuing uncertainty of limit orders. Working paper, Available at SSRN: http://dx.doi.org/10.2139/ssrn.2336122.

Figure 1. Limit order book evolution

Figure 1 depicts the possible evolution of the limit order book from t_0 to two possible future states at t_1 (A and B). The white rectangles represent the bid volume and the grey rectangles represent the ask volume. Prices are shown on the x-axis, with the best bid at 13 and the best offer at 14 at t_0. The trader's limit order is in black and starts at the back of the queue at t_0 at price 12.

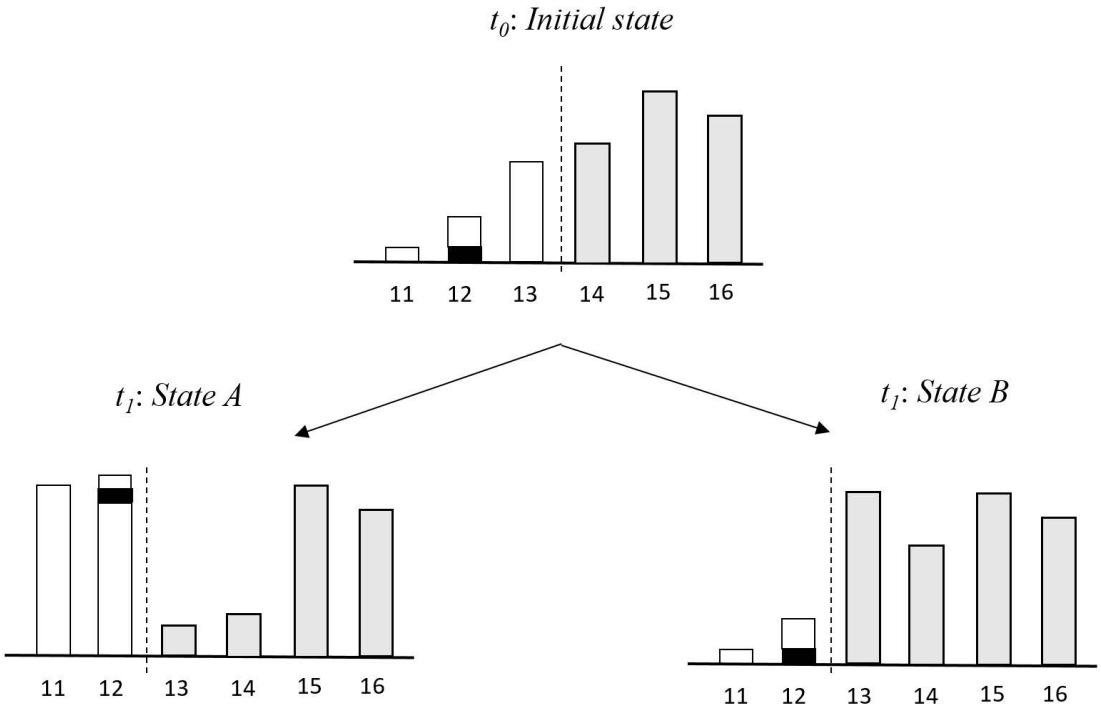


Figure 2. Traders sequential decision making process

Figure 2 depicts the time line of the liquidity provider's decision making process when monitoring their limit order. At the end of each interval, the liquidity provider observes current market conditions and decides to leave or cancel their order. This process repeats until the order is either executed or canceled.

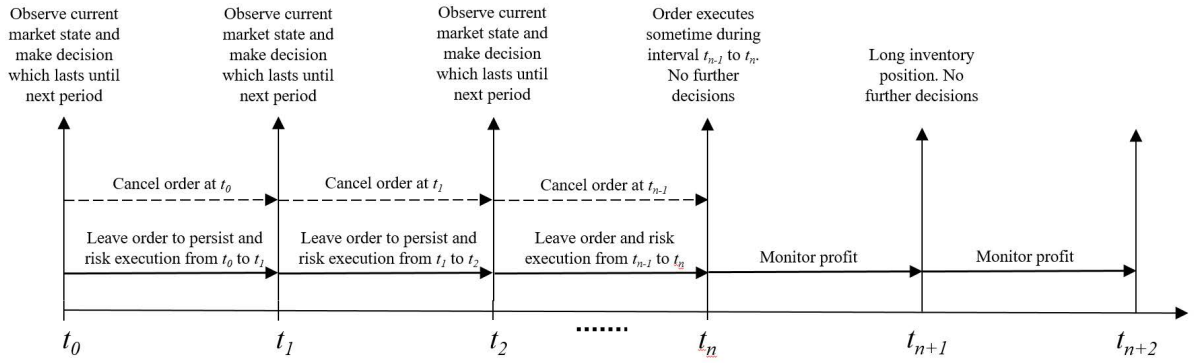


Figure 3. Boxplot of the expected value of a limit order

This figure plots a boxplot of the expected value of a limit order, estimated via our RL model. The figure contains the estimates from all 20 sample stocks. The figure depicts a boxplot for three subsamples conditional on the price level the limit order is resting at.

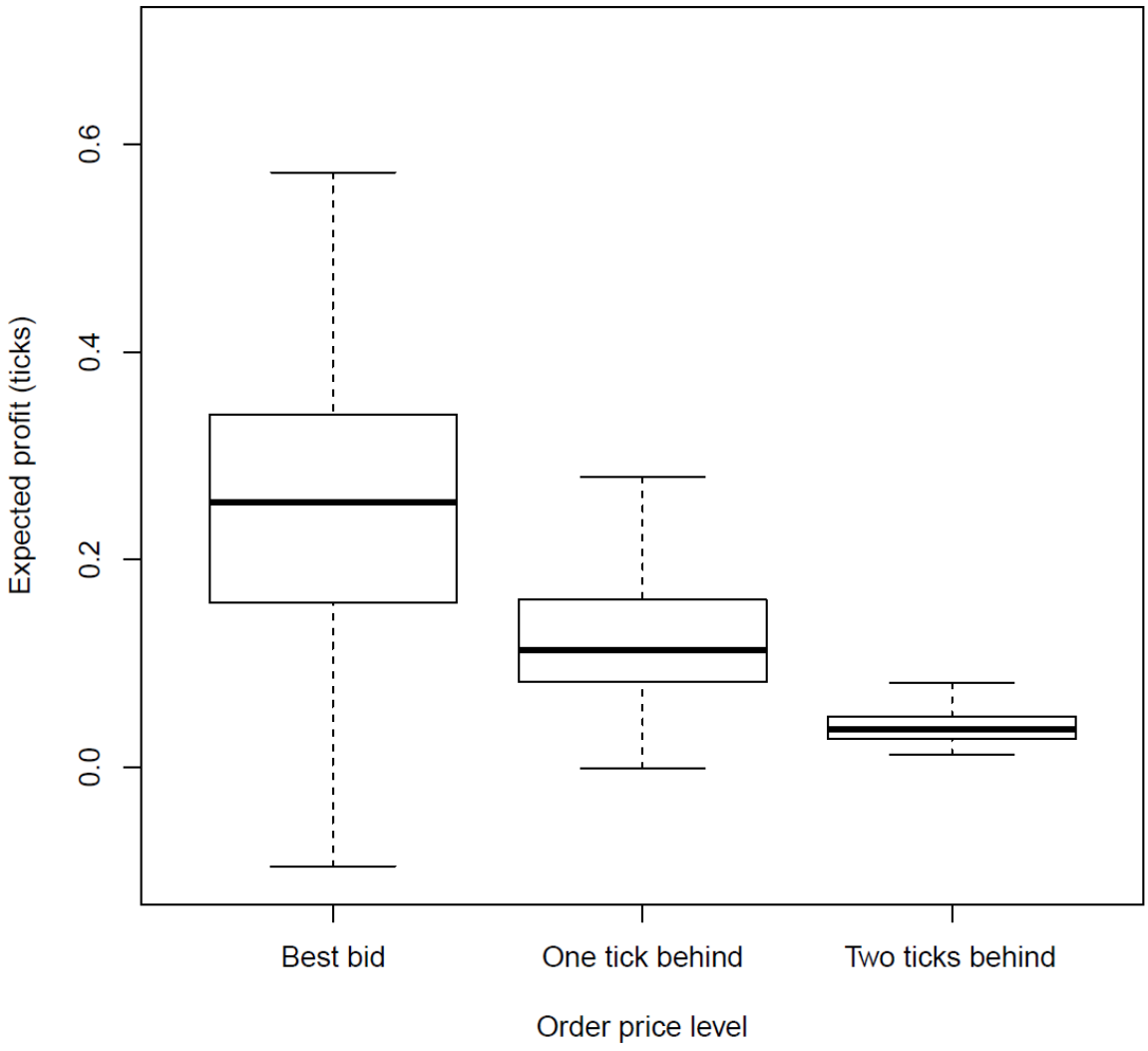


Figure 4. Volatility, bid ask spreads and the expected value of a limit order

This figure shows the predicted effects of volatility on the quoted bid ask spread, breakeven bid ask spread, and the expected value of a limit order. The top (bottom) figure in the panels depicts a low (high) volatility environment. Panel A shows a constrained stock in which the breakeven bid ask spread is always less than the one tick-mandated spread, even when volatility is high. In Panel B, when volatility is high, the breakeven bid ask spread increases beyond the quoted bid ask spread and the liquidity provider widens the quoted spread. The expected compensation to the liquidity provider (i.e., the difference between the quoted bid ask spread and the breakeven bid ask spread) is depicted in green.

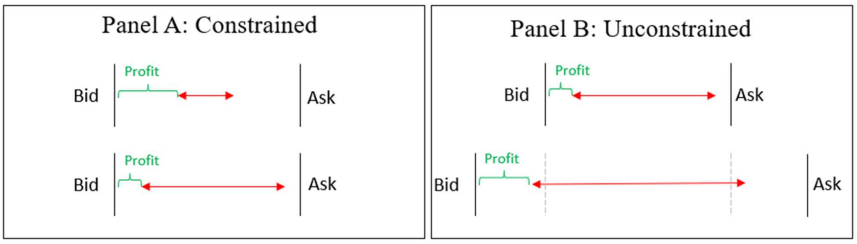


Table 1
Summary statistics

This table reports summary statistics for our sample stocks. Our sample period covers July 3, 2017 to Septemeber 29, 2017 for 20 actively traded stocks on the ASX. We report the average bid ask spread in cents (*Spread*), the average trade price in AUD (*Price*), and the average number of daily trades, order deletions and order submissions labelled *No. trades*, *No. deletions* and *No. submissions*, respectively.

	Spread	Price	No. trades	No. deletions	No. submissions
AMC	1.03	15.72	6970	9247	21142
AMP	1.01	5.11	3026	4279	9318
ANZ	1.08	29.51	11326	68791	88536
BHP	1.06	25.79	14268	20304	44994
BXB	1.04	9.40	5484	6686	16001
CBA	1.61	79.39	21498	33810	71510
CSL	2.59	129.87	16198	42372	70900
IAG	1.01	6.54	3530	5273	11142
MQG	1.97	87.20	13999	32589	57422
NAB	1.06	30.40	11714	69080	89394
NCM	1.12	21.34	10735	17888	36675
ORG	1.02	7.29	4637	6220	14191
QBE	1.08	11.14	7779	9758	22926
RIO	1.70	65.61	15955	30138	57912
STO	1.01	3.51	3255	4668	10058
SUN	1.02	13.61	7920	10994	24647
TLS	1.00	3.94	3999	4754	11186
WBC	1.08	31.62	13469	38652	62169
WOW	1.05	26.04	9208	16856	32784
WPL	1.08	29.27	11203	22482	41672

Table 2
Summary statistics

This table reports the summary statistics on the expected value of an optimally managed limit order. The first row reports summary statistics for orders placed at all price levels, whereas rows 2 to 4 report summary statistics for limit orders conditional on their price level.

Order Location	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std. dev.
All prices	-0.098	0.031	0.123	0.146	0.252	0.569	0.098
Best bid	-0.098	0.153	0.258	0.262	0.339	0.569	0.134
One tick behind best bid	0.002	0.071	0.098	0.105	0.155	0.276	0.067
Two ticks behind best bid	0.011	0.023	0.029	0.031	0.044	0.074	0.010

Table 3
Queue position and the expected value of a limit order

This table reports estimation results for the following OLS regression:

$$Q_s = \beta_1 QueuePos_s + State\ Fixed\ Effects + \epsilon_s,$$

where Q_s is the expected value of a limit order estimated via our RL model. The independent variable is $QueuePos$, with fixed effects controlling for all other variables. $QueuePos$ takes the value of 0 if the order is at the front of the queue and 1 if it is at the back of the queue. Columns 1, 2 and 3 present the regression results for subsamples in which the order rests at the best bid, one tick behind the best bid and two ticks behind the best bid, respectively. We report the mean coefficient across all sample stocks (*Mean*), along with the number of significantly positive (*No. +*) and negative (*No. -*) coefficients at the 5% significance level, out of the full sample of 20 stocks. We also report the mean R-squared value across the 20 regressions, along with the number of observations used in each regression.

	Best bid	1 behind best bid	2 behind best bid
Mean	-0.12	-0.05	-0.01
No. +	0	0	0
No. -	20	20	20
Mean R^2	0.89	0.88	0.89
No. obs. per stock	5,625	5,625	5,625

Table 4
Queue size and the expected value of a limit order

This table reports estimation results for the following OLS regression:

$$Q_s = \beta_1 q_s^{B_0} + \beta_2 q_s^{B_1} + \beta_3 q_s^{B_2} + \beta_4 q_s^{A_0} + \textit{State Fixed Effects} + \epsilon,$$

where Q_s is the expected value of a limit order estimated via our RL model, q^{B_i} is the queue size on the best bid at price level i and q^{A_0} is the queue size on the best ask. q^{B_0}, q^{B_1}, q^{A_0} take values from 0 to 4 to depict the queue size quintile, extremely short, short, normal, long, and extremely long, respectively. q^{B_2} takes values from 0 to 2 to represent the three queue size terciles (short, normal long) at the price two ticks below the best bid. Columns 1, 2 and 3 present the regression results for subsamples in which the order rests at the best bid, one level behind the best bid and two levels behind the best bid, respectively. We report the mean coefficient across all sample stocks (*Mean*), along with the number of significantly positive (*No. +*) and negative (*No. -*) coefficients at the 5% significance level, out of the full sample of 20 stocks. We also report the mean R-squared value across the 20 regressions, along with the number of observations used in each regression.

		Best bid	1 behind best bid	2 behind best bid
q^{B_0}	Mean	0.06	-0.05	-0.02
	No. +	19	0	0
	No. -	0	20	20
q^{B_1}	Mean	0.05	0.01	-0.01
	No. +	18	14	0
	No. -	2	5	20
q^{B_2}	Mean	0.04	0.03	0.00
	No. +	14	16	12
	No. -	3	4	6
q^{A_0}	Mean	-0.04	0.01	0.01
	No. +	0	19	20
	No. -	20	0	0
Mean R^2		0.88	0.89	0.83
No. obs. per stock		5,625	5,625	5,625

Table 5
Volatility and the expected value of a limit order

This table reports estimation results for the following OLS regression:

$$Q_s = \beta_1 Volatility_s + State\ Fixed\ Effects + \epsilon_s,$$

where Q_s is the expected value of a limit order estimated via our RL model for orders at the best bid. The independent variable is $Volatility$, with fixed effects controlling for all other variables. Volatility is calculated as the log of the highest traded price minus the log of the lowest traded price over the last 100 trades. $Volatility$ takes the value of 0, 1, or 2 for low, medium, and high volatility states, respectively. Column 1 presents the regression results for all stocks. Column 2 (3) presents results on the subsample of stocks that are most (least) tick constrained. We report the mean coefficient across all sample stocks (*Mean*), along with the number of significantly positive (*No. +*) and negative (*No. -*) coefficients at the 5% significance level, out of the full sample of 20 stocks. We also report the mean R-squared value across the 20 regressions, the number of stocks in each sample, along with the number of observations used in each regression.

	All stocks	Constrained	Unconstrained
Mean	0.38	-2.39	5.82
No. +	9	0	5
No. -	11	5	0
Mean R^2	0.86	0.86	0.87
No. stocks	20	5	5
No. obs. per stock	5,625	5,625	5,625

Table 6
Relative importance of variables

For each variable, k, that partially defines the market state (i.e., *Price level*, *Queue position*, queue size at best bid (*Bid size 1*), queue size one level behind best bid (*Bid size 2*), queue size two levels behind best bid (*Bid size 3*), *Ask size*, *Volatility*), this table reports the Mean Decreased Accuracy (MDA) estimated as follows:

$$MDA^k = \sum_{s=1}^{S} \left(\frac{|(Q(s, NA) - Q(s_R^k, NA))|}{Q(s, NA)} \right) / S,$$

where $Q(s, NA)$ is the expected value of a limit order while in state s and taking action NA, and $Q(s_R^k, NA)$ is the estimate associated with state s_R when variable k is randomized. For each variable k, we repeat this process 100 times and report the mean and standard deviation of the MDA.

	Price level	Queue position	Bid size 1	Bid size 2	Bid size 3	Ask size	Volatility
Mean	2.54	0.20	1.22	1.17	0.68	0.68	0.56
St. dev.	1.34	0.07	0.34	0.71	0.48	0.19	0.33

Table 7
Summary statistics for the value of the option to cancel

Table 7 reports the summary statistics on the expected value of the option to cancel a limit order. The first row reports summary statistics for orders placed at all price levels, whereas rows 2 to 4 report summary statistics for limit orders conditional on their price level.

Order Location	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	St Dev.
All prices	0.000	0.003	0.008	0.024	0.019	0.483	0.052
Best bid	0.000	0.004	0.012	0.049	0.051	0.483	0.081
One tick behind best bid	0.001	0.005	0.010	0.017	0.020	0.161	0.020
Two ticks behind best bid	0.000	0.002	0.004	0.007	0.009	0.059	0.007

Table 8
The value of the option to cancel

This table reports estimation results for the following OLS regression:

$$value\ of\ option\ to\ cancel = \beta_0 + \beta_1 q^{B_0} + \beta_2 q^{B_1} + \beta_3 q^{B_2} + \beta_4 q^{A_0} + \epsilon,$$

where the dependent variable is the option value to cancel a limit order estimated via our RL model, q^{B_i} is the queue size on the best bid at price level i and q^{A_0} is the queue size on the best ask. q^{B_0}, q^{B_1}, q^{A_0} take values from 0 to 4 to depict the queue sizes, extremely short, short, normal, long, and extremely long, respectively. q^{B_2} takes values from 0 to 2 to represent the three queue size terciles (short, normal long) at the price two ticks below the best bid. Columns 1, 2 and 3 present the regression results for subsamples in which the order rests at the best bid, one level behind the best bid and two levels behind the best bid, respectively. We report the mean coefficient across all sample stocks (*Mean*), along with the number of significantly positive (*No. +*) and negative (*No. -*) coefficients at the 5% significance level, out of the full sample of 20 stocks.

		Best bid	1 behind best bid	2 behind best bid
q^{B_0}	Mean	-0.034	-0.015	-0.019
	No. +	0	1	0
	No. -	20	19	20
q^{B_1}	Mean	-0.044	-0.032	-0.021
	No. +	1	0	2
	No. -	19	20	18
q^{B_2}	Mean	-0.034	-0.031	-0.028
	No. +	1	0	0
	No. -	19	20	20
q^{A_0}	Mean	0.014	0.025	0.035
	No. +	20	20	20
	No. -	0	0	0
	Mean R^2	0.16	0.24	0.34
	No. obs per stock	625	625	625

7 Appendix

A Transition Matrix

Transition matrix for action NA

Figure A.1 illustrates the section of the transition matrix, T, when the action is NA (i.e., leave the limit order), which is a $S \times S$ matrix that requires empirical estimation. The states $s_1(0), \ldots, s_n(0)$ reflect the n possible states when the trader has no inventory and is working an order. The states $s_1(1), \ldots, s_m(1)$ reflect the m possible states the market can exist, when the trader has a long position and is no longer working a limit order. $s^C(0)$ reflects the absorbing state once the trader cancels their order.

Figure A.1. Transtion matrix for NA

Figure A.1 depicts the $S \times S$ transition matrix for the experience tuples in which the action is to leave the resting limit order, or do nothing, NA. States $s_i(0)$ represent states when the trader is working their limit order, while states $s_j(1)$ represent states when the trader's order has been executed. The state $s^C(0)$ represents the absorbing order cancellation state.

Current state with action NA	Future State: $s_1(0)$	$s_2(0)$	...	$s_n(0)$	$s_1(1)$	...	$s_m(1)$	$s^C(0)$
$s_1(0)$	$p_{1,1}$	$\cdots$		$p_{1,n}$	$p_{1,n+1}$	$\cdots$	$p_{1,n+m}$	$p_{1,n+m+1}$
$s_2(0)$	$\vdots$	$\ddots$			$\vdots$	$\ddots$		$\vdots$
	Unexecuted				Executed			
$s_{n-1}(0)$	$p_{n-1,1}$				$p_{n-1,n+1}$			
$s_n(0)$	$p_{n,1}$	$\cdots$		$p_{n,n}$	$p_{n,n+1}$		$p_{n,n+m}$	$p_{n,n+m+1}$
$s_1(1)$	0	$\cdots$	0	0	$p_{n+1,n+1}$	$\cdots$	$p_{n+1,n+m}$	0
$\vdots$	$\vdots$	Prohibited		0	$\vdots$	Long	$\vdots$	$\vdots$
$s_m(1)$	0	0	0	0	$p_{n+m,n+1}$	$\cdots$	$p_{n+m,n+m}$	0
$s^C(0)$	0	$\cdots$	0	0	0	$\cdots$	0	1

The top left quadrant of the transition matrix, labeled "Unexecuted", contains the transition probabilities for a limit order that does not execute during the transition from one state to the next. These transition probabilities reflect the changes in market conditions and the movement of the limit order within the order book. For example, they capture the likelihood of the limit order advancing in the queue or how other market participants might respond to current market conditions. We estimate these probabilities empirically using equation (5).

The block of the transition matrix titled "Executed" contains the probability of limit order execution during the state transition. In our setup, once an order is executed, the trader has no remaining limit orders. Consequently, the trader must transition to one of m positive inventory states, $s_j(1)$, where j represents different possible states based on the public information reflected in the order book variables. Again, we estimate these probabilities empirically via (5).

After execution, the trader remains in one of the m positive inventory states and cannot submit another order. To ensure the trader does not hold another limit order while being long and remains in a positive inventory state, we define the "Prohibited" block in Figure A.1 to contain only zeros. The block labeled "Long" captures the transition probabilities for a trader who is long in one market state and transitions to another market state while remaining long; these probabilities are also estimated empirically via equation (5).

The final column of the matrix reports the probability that the trader transitions to the absorbing state by canceling their order. The absorbing nature of the state is represented by the transition probability of 1 in the bottom right of Figure A.1. If the trader is currently in the absorbing cancel state, the probability of remaining in that state in the subsequent period is 1. Given that the action for this section of the matrix is NA, we may expect the probability to enter the absorbing cancel state to be zero for all market states when the trader has a resting limit order. However, we assume that if the resting limit order transitions into an undefined state (more than three ticks from the best bid), the trader's action NA is overridden, and the order is canceled. Therefore, there can be a non-zero probability of the order being canceled, which we estimate empirically.

Transition matrix for action C

Figure A.2 illustrates the $S \times S$ section of the transition matrix, T, when the action is to cancel the resting limit order (C). Unlike the section of the transition matrix when the action is NA, this section of the transition matrix is deterministic and does not require any empirical estimation of the transition probabilities. If the trader cancels their limit order, they transition to the absorbing cancel state with certainty. Therefore, the probability of entering the absorbing cancel state, which is captured in the final column of Figure A.2, is 1 for all current states where the trader has a resting

limit order. Further, once the order is canceled, the market cannot transition to any state where the limit order still exists or executes. Thus, the "Unexecuted" and "Executed" blocks contain only zeros.

To ensure the trader only has one resting limit order at a time, we impose a restriction that any state where the trader has an inventory position or has already canceled their order cannot have another resting order. Due to this restriction, taking the action to cancel an order in a state where the trader has a long inventory position, or has canceled their order, is prohibited and has a zero probability of occurring.

Figure A.2. Transition matrix for C

Figure A.2 depicts the $S \times S$ transition matrix for the experience tuples in which the action is to cancel the resting limit order, C. States $s_i(0)$ represent states when the trader is working their limit order, whereas states $s_j(1)$ represent states when the trader's order has been executed. State $s^C(0)$ represents the absorbing order cancellation state.

Future State

Current state with action C	$s_1(0)$	$s_2(0)$	...	$s_n(0)$	$s_1(1)$	...	$s_m(1)$	$s^C(0)$
$s_1(0)$	0	...		0	0	...	0	1
$s_2(0)$	⋮	⋱			⋮	⋱		1
		Unexecuted				Executed		1
$s_{n-1}(0)$								1
$s_n(0)$	0	...		0	0		0	1
$s_1(1)$	0	...	0	0	0	...	0	0
⋮	⋮	Prohibited		0		Prohibited		0
$s_m(1)$	0	0	0	0	0	...	0	0

Full transition matrix

In Figures A.1 and A.2 we present two $S \times S$ sections of the full $2S \times S$ transition matrix, T. Specifically, Figure A.1 (A.2) is a transition matrix for all experience tuples when the action is NA (C). To generate the full transition matrix, T, we vertically stack the 2 subsections, each with dimension $S \times S$, resulting in the full transition matrix of dimension $2S \times S$. For notational convenience, we refer to $T(\langle s, a \rangle, s')$ as the probability a limit order transitions to state s' given the trader makes action a while the limit order is in state s.

B Immediate Reward

Figure B.3 shows the matrix of immediate rewards for all experience tuples that occur when the action is to do nothing, NA. If the trader's limit order remains unexecuted during the transition, the immediate reward is zero, as indicated in the upper left quadrant titled "Unexecuted". Conversely, if the trader's limit order executes, the immediate reward corresponds to the profit generated. This profit is empirically calculated using (6) and is defined as the difference between the execution price and the midpoint in the future state s'. These profits are shown in the block of Figure B.3 titled "Executed". The block of Figure B.3 titled "Long" contains the immediate profits that occur when the trader is long and the market transitions from one state to the next. We empirically estimate these immediate profits via (6), and they reflect any profit generated via a change in midpoint over a state transition.

Figure B.3. Immediate reward matrix

Figure B.3 depicts the $S \times S$ immediate reward matrix for transitioning from one state to the next. States $s_i(0)$ represent states when the trader is working their limit order, whereas states $s_j(1)$ represent states when the trader's order has been executed. State $s^C(0)$ represents the absorbing order cancellation state.

Future State

Current state with action NA	$s_1(0)$	$s_2(0)$	...	$s_n(0)$	$s_1(1)$	...	$s_m(1)$	$s^C(0)$
$s_1(0)$	0	...		0	$r_{1,n+1}$	...	$r_{1,n+m}$	0
$s_2(0)$	⋮	⋱			⋮	⋱		0
		Unexecuted				Executed		0
$s_{n-1}(0)$	0				$r_{n-1,n+1}$			0
$s_n(0)$	0	...		0	$r_{n,n+1}$		$r_{n,n+m}$	0
$s_1(1)$	0	...	0	0	$r_{n+1,n+1}$	...	$r_{n+1,n+m}$	0
⋮	⋮	Prohibited		0	⋮	Long	⋮	0
$s_m(1)$	0	0	0	0	$r_{n+m,n+1}$	...	$r_{n+m,n+m}$	0
$s^C(0)$	0	0	0	0	0	...	0	0

When the action is NA, the limit order can execute or there can be an existing long position. Either scenario can result in a non-zero immediate reward. In contrast, when the trader cancels their order (action C), the immediate reward must be zero, as no limit orders are executed and there is no inventory position. Consequently, the $S \times S$ immediate reward matrix for the action C contains only zeros.

C Illustrative example

In this appendix, we provide a simple example to illustrate the empirical estimation process of our framework via an iterative learning rule known as Q-learning defined as:

$$Q_{t+1}(s,a) = Q_t(s,a) + \alpha\Big(E[R(s,a)] + \gamma \sum_{s' \in S} T(\langle s,a \rangle, s') \max_a Q_t(s',a) - Q_t(s,a)\Big), \tag{14}$$

where α is the learning rate and t is the iteration number. The Q-learning rule is a value iteration update. Watkins and Dayan (1992) show that the Q values will converge to Q^* with probability 1 if all actions are repeatedly sampled in all states and the action-values are represented discretely.

To simplify this illustrative example, we first define a simplified state-action space. We then illustrate how to empirically estimate our simplified transition probability matrix and immediate reward matrix. We conclude with a demonstration of the Q-learning rule.

C.1 State action space

Similar to the model we formulated in Section 2, in this illustrative example, our trader has two available actions. The first action is to cancel the existing limit order (C). The second action is to do nothing (NA), allowing the existing limit order to remain in the queue. However, to simplify our example, we reduce the state space by considering only the queue size at the best bid (q^{B_0}) and best ask (q^{A_0}), ignoring the queue sizes at levels behind the best bid (q^{B_1},q^{B_2}). Moreover, we discretize queue size into only two groups, $large$ and $small$.

To further reduce dimensionality, we reduce the private state variable, queue position (Q) to two possible states: $front$ and $back$, representing whether the order is in the front half or back half of the queue, respectively. These simplifications result in a state space with 8 possible states when the trader has no inventory and is executing a limit order (m), 4 possible market states when the trader has an inventory position and is no longer executing an order (n), and 1 absorbing state

when the trader cancels their order (o). Altogether, our setup consists of a total of 13 possible unique market states (S). More formally, $m = 8$, $n = 4$ $o = 1$ and $S = 13$. We define each state as follows:

$$s_k^j(I) = [I, L, Q, q^{B_0}, q^{A_0}] = \begin{cases} s_1^f(0) = [0, 0, front, small, small] \\ s_2^f(0) = [0, 0, front, small, large] \\ s_3^f(0) = [0, 0, front, large, small] \\ s_4^f(0) = [0, 0, front, large, large] \\ s_1^b(0) = [0, 0, back, small, small] \\ s_2^b(0) = [0, 0, back, small, large] \\ s_3^b(0) = [0, 0, back, large, small] \\ s_4^b(0) = [0, 0, back, large, large] \\ s_1^X(1) = [1, X, X, small, small] \\ s_2^X(1) = [1, X, X, small, large] \\ s_3^X(1) = [1, X, X, large, small] \\ s_4^X(1) = [1, X, X, large, large] \\ s^C(0) = [0, X, X, -, -] \end{cases} \tag{15}$$

where k is an index of the public market state, which is reflected by q^{B_0} and q^{A_0}. j takes on the value of f (b) if the limit order is at the front (back) half of the queue, a value of X if the trader has an inventory position and no limit order, or a value of C if the trader has canceled their order. The X term captures our restriction that no additional limit orders can be submitted once the trader has a positive inventory position or cancels their order. I captures the trader's current inventory position and $L \in \{0, X\}$ depending on whether the trader has an order resting at the best

bid, or their order is canceled of executed.[23] For each of the 8 states, where the trader is working a limit order, the trader has the choice of making the action to do nothing, NA, or cancel their existing order, C. For the states where the trader is long or has canceled their order, they can only make action NA. With the state action space defined, the input variables for (14) are the transition matrix and immediate reward matrix, which we empirically estimate.

C.2 Transition probabilities

$T(\langle s, a\rangle, s')$ represents the probability that the limit order transitions the market to state s' under action a while in state s. For example, $T(\langle s_1^f(0), NA\rangle, s_2^f(0))$ is the probability that a limit order at the front of the queue which exists when the best bid and best ask both have short queue lengths transitions to a subsequent period where the order is still at the front half of a queue and it remains unexecuted, but market conditions have changed such that the bid volume is small and the ask volume is now large.

We compute these transition probabilities empirically using the MLE estimate defined by (5). For example, to estimate $T(\langle s_1^f(0), NA\rangle, s_2^f(0))$, we observe the subsample of observations that capture state $s_1^f(0)$ (i.e., the observations that have small queue sizes on both the bid and the ask and the limit order is at the front half of the bid). Next, we compute the proportion of observations that transition to the subsequent state $s_2^f(0)$, which is reflected by the limit order still remaining in the top half of the book, but under new market conditions (i.e., the bid queue size is small and the ask queue size is large). Table C.4, reports empirical estimates of the transition probabilities using data defined in Section 3.

Figure C.4 has a distinct structure. The upper left block of the transition matrix represents states when the trader has no inventory and completes the action of do nothing, NA. This area has a strong diagonal, which reflects that an uncanceled limit order is most likely to remain in the same state in the subsequent 100ms period. For example, observing the transition probabilities for the state $s_1^f(0)$, which reflects a resting limit order at the front half of the queue when the queue

[23] In this simplified example, L is redundant because the order can only be placed at the best bid. However, we have included L for consistency with our main analysis, in which the order can rest at multiple price levels.

Figure C.4. Transition matrix

Figure C.4 depicts the $SA \times S$ transition matrix for the experience tuple in which the action is to leave the resting limit order, NA, or cancel the order, C. States $s_i(0)$ represent states when the trader is working their limit order, whereas states $s_j(1)$ represent states when the trader's order has been executed. State $s^C(0)$ represents the absorbing order cancellation state.

		Future State												
		$s_1^f(0)$	$s_2^f(0)$	$s_3^f(0)$	$s_4^f(0)$	$s_1^b(0)$	$s_2^b(0)$	$s_3^b(0)$	$s_4^b(0)$	$s_1^X(1)$	$s_2^X(1)$	$s_3^X(1)$	$s_4^X(1)$	$s^C(0)$
Current state with action NA														
1	$s_1^f(0)$	**0.86**	0.02	0.02	0.00	0.02	0.00	0.00	0.00	**0.05**	0.00	0.01	0.00	0.01
2	$s_2^f(0)$	0.03	**0.82**	0.00	0.02	0.00	0.02	0.00	0.00	0.01	**0.07**	0.01	0.01	0.01
3	$s_3^f(0)$	0.01	0.00	**0.89**	0.03	0.01	0.01	0.01	0.00	0.01	0.00	**0.02**	0.00	0.02
4	$s_4^f(0)$	0.00	0.01	0.02	**0.90**	0.00	0.01	0.00	0.01	0.00	0.00	0.00	**0.03**	0.02
5	$s_1^b(0)$	0.06	0.00	0.01	0.00	**0.87**	0.03	0.02	0.00	**0.01**	0.00	0.01	0.00	0.01
6	$s_2^b(0)$	0.00	0.07	0.00	0.01	0.03	**0.84**	0.00	0.02	0.01	**0.01**	0.01	0.01	0.01
7	$s_3^b(0)$	0.00	0.00	0.02	0.00	0.03	0.01	**0.90**	0.03	0.00	0.00	**0.00**	0.00	0.02
8	$s_4^b(0)$	0.00	0.00	0.00	0.02	0.00	0.02	0.02	**0.92**	0.00	0.00	0.00	**0.00**	0.02
9	$s_1^X(1)$									**0.94**	0.03	0.03	0.00	0
10	$s_2^X(1)$									0.04	**0.91**	0.01	0.04	0
11	$s_3^X(1)$				0					0.03	0.01	**0.92**	0.04	0
12	$s_4^X(1)$									0.01	0.03	0.03	**0.94**	0
13	$s^C(0)$				0						0			1
Current state with action C														
14	$s_1^f(0)$													1
15	$s_2^f(0)$													1
16	$s_3^f(0)$													1
17	$s_4^f(0)$				0						0			1
18	$s_1^b(0)$													1
19	$s_2^b(0)$													1
20	$s_3^b(0)$													1
21	$s_4^b(0)$													1
22	$s_1^X(1)$													0
23	$s_2^X(1)$													0
24	$s_3^X(1)$				0						0			0
25	$s_4^X(1)$													0
26	$s^C(0)$				0						0			0

sizes on the best bid and best ask are small, there is an 86% chance the subsequent state will be the same. However, there is also a 2% chance the subsequent state is either $s_2^f(0)$ or $s_3^f(0)$, which implies either 1) the best ask has grown to become large and the market has transitioned to $s_2^f(0)$, or 2) the best bid has grown and the market has transitioned to $s_3^f(0)$.

The section of the transition matrix for transitions from state $s_i(0)$ to state $s_j(1)$ with action NA, reports the probabilities that a resting limit order executes during the transition to the subsequent state. We observe that resting limit orders at the front of the queue (rows 1-4) have a higher probability of execution than resting limit orders at the back of the queue (rows 5-6). Further, the

probability of execution for $s_2^f(0)$ is 0.1 $(0.01 + 0.07 + 0.01 + 0.01)$, which is higher than the probability of execution for any of the other states with a resting limit order. State $s_2^f(0)$ occurs when the trader has a resting limit order at the front half of the best bid and the bid queue size is small, while the ask queue size is large. Cao et al. (2009) demonstrate that when the ask volume is larger than the bid volume, aggressive sell orders are more likely to occur and prices will decrease in the near future. Therefore, it is consistent with the literature that the highest probability of execution occurs for state $s_2^f(0)$. Moreover, the strong diagonal component of this section of the transition matrix reflects that when a resting limit order executes during the transition to the subsequent period, it is most likely that the state of the order book in the subsequent period is in the same state as the current period.

Rows 9 to 12 of Table C.4 represent the transition probabilities when the trader has an inventory position. The left block of the rows take the value of zero to ensure the trader does not have additional limit orders once a long inventory position occurs. The middle block captures the probability the trader transitions to a subsequent market state with their inventory position remaining unchanged. Given the trader has no resting limit orders, we estimate these transition probabilities using only the public state variables, which in this example are the size of the best bid and ask (q^{B_0} and q^{A_0}).

As discussed in Appendix A, we do not need to estimate transition probabilities when the action is C. When the action is C, the transition probability to any state with a resting limit is 0 and the transition probability to the absorbing state is 1. Moreover, if the trader has a long position, or is already in the absorbing state, they are prohibited to make action C, as they have no order to cancel. To uphold this constraint, rows 22 to 26 all sum to zero, which ensures there is a 0 probability that action C occurs when in these states.

We note that in rows 1-8 of Figure C.5 we report non-zero values for the probability to transition to the absorbing order cancellation state, $s^C(0)$, despite the action being NA. These non-zero values maintain our assumption that if the market transitions to a state space where the resting limit is not recognized, the action NA is over ruled by action C. Specifically, in this case, the state space only contains limit orders at the best bid. Thus, if the best bid increases during the market transition,

so that the existing limit order is no longer at the best bid, the trader will be forced to cancel the order.

C.3 Immediate rewards

Next, we require the immediate reward for all possible transitions via (6). To empirically estimate the immediate reward when the trader has a long position, we take the average change in midpoint for the subset of observations that capture the correct transition from one state to the next. For example, to estimate $R(\langle s_1^X(1), NA\rangle, s_1^X(1))$, we create a subset of observations from our full sample of data by using observations when the market is in an initial state of $s_1^X(1)$ (i.e., the queue size of the best bid and ask are both small) and the subsequent market state is the same, $s_1^X(1)$. For this subset of observations, we then take the average of (6), which is the average change in midpoint price.

To estimate the immediate reward for the execution of a limit order, we use a similar approach. For example, to estimate the immediate reward for $R(\langle s_1^f(0), NA\rangle, s_1^X(1))$ we create a subset of observations that only include observations where the trader is in state $s_1^f(0)$ (i.e., the trader has a resting limit order at the front of the best bid during market conditions where the size of the best bid and ask are small) and transitions to the subsequent state $s_1^X(1)$ (i.e., the trader has a long position when the best bid and ask queue sizes are small). For this subset of observations, we use the average immediate reward, computed via (6), which is the midpoint price in the new state less the limit order's execution price.

Figure C.5 reports the empirically estimated immediate reward for all possible transitions. Figure C.5 only reports non zero values when the trader transitions to a long position. This segmentation ensures the trader only receives an immediate reward when a limit order is executed or a the trader has a long position. Otherwise, the trader receives no immediate reward.

The reported immediate rewards are the potential gains or losses that immediately occur during the transition from one market state to the next. For example, we report the immediate rewards for state $s_1^f(0)$ in row 1. When the limit order in state $s_1^f(0)$ executes and the trader transitions to

Figure C.5. Immediate reward matrix

Figure C.5 depicts the $SA \times S$ transition matrix for the experience tuple in which the action is to leave the resting limit order, NA, or cancel the order, C. States $s_i(0)$ represent states when the trader is working their limit order, whereas states $s_j(1)$ represent states when the trader's order has been executed. State $s^C(0)$ represents the absorbing order cancellation state.

		Future State												
		$s_1^f(0)$	$s_2^f(0)$	$s_3^f(0)$	$s_4^f(0)$	$s_1^b(0)$	$s_2^b(0)$	$s_3^b(0)$	$s_4^b(0)$	$s_1^X(1)$	$s_2^X(1)$	$s_3^X(1)$	$s_4^X(1)$	$s^C(0)$
Current state with action NA	1	$s_1^f(0)$								**0.32**	0.25	−0.19	−0.05	0
	2	$s_2^f(0)$								−0.30	**0.47**	−0.49	−0.00	0
	3	$s_3^f(0)$								0.20	0.16	**0.43**	0.31	0
	4	$s_4^f(0)$				0				−0.40	0.43	−0.36	**0.47**	0
	5	$s_1^b(0)$								**-0.07**	−0.02	−0.24	−0.09	0
	6	$s_2^b(0)$								−0.45	**0.33**	−0.49	−0.04	0
	7	$s_3^b(0)$								−0.18	−0.15	**-0.15**	−0.14	0
	8	$s_4^b(0)$								−0.50	0.23	−0.49	**0.16**	0
	9	$s_1^X(1)$								0	0.23	−0.20	0.08	0
	10	$s_2^X(1)$								−0.23	0	−0.84	−0.25	0
	11	$s_3^X(1)$				0				0.20	0.84	0	0.24	0
	12	$s_4^X(1)$								−0.08	0.25	−0.24	0	0
	13	$s^C(0)$				0						0		0
Current state with action C	14	$s_1^f(0)$												0
	15	$s_2^f(0)$												0
	16	$s_3^f(0)$												0
	17	$s_4^f(0)$				0						0		0
	18	$s_1^b(0)$												0
	19	$s_2^b(0)$												0
	20	$s_3^b(0)$												0
	21	$s_4^b(0)$												0
	22	$s_1^X(1)$												0
	23	$s_2^X(1)$												0
	24	$s_3^X(1)$				0						0		0
	25	$s_4^X(1)$												0
	26	$s^C(0)$				0						0		0

state $s_1^X(1)$, the immediate reward is 0.32, which implies the trader makes an immediate gain of 0.32 ticks, on average.

C.4 Estimation

We initialize our Q values, or long run expected profits forecasts, for each experience tuple to zero. Using the Q-learning rule defined by (14), we update our Q values for each experience tuple

recursively. For example, we update our estimate for $Q(s_1^f(0), NA)$ for the first iteration via:

$$Q_1(s_1^f(0), NA) = E[R(s_1^f(0), NA)] + \gamma \sum_{s' \in S} T(\langle s_1^f(0), NA \rangle, s') \max_{a_{t+1}} Q_t(s', a_{t+1}), \tag{16}$$

where the first term is the immediate profit for taking action NA which we compute via (7). The second term is the expected future profit conditional on taking action NA now. We observe the second term multiplies the probability of arriving in future state s' with the maximum Q value the trader can achieve by picking the optimal action a_{t+1} while in state s'. Because we have initialized all Q values to zero, on the first iteration, the $\max_{a_{t+1}} Q_t(s', a_{t+1})$ term in (16) will be zero for all s' and the trader will be indifferent to all choices of a_{t+1}. Thus, the second term of (16) is zero and we update our estimate for $Q(s_1^f(0), NA)$ for the first iteration as follows:

$$\begin{aligned} Q_1(s_1^f(0), NA) &= E[R(s_1^f(0), NA)] + \sum_{s' \in S} T(\langle s_1^f(0), a \rangle, s') \times R(\langle s_1^f(0), a \rangle, s') \\ &= (0.05 \times 0.32) + (0 \times 0.25) + (0.01 \times -0.19) + (0 \times -0.05) + \cdots + 0 \\ &= 0.0141 \end{aligned}$$

Applying the same process, we update the associated Q values for all experience tuples, which we report in Column 1 of Table C.1. Given the Q values were all initialized to 0, these first iteration values are the expected immediate profits.

On iteration two, the input values for our learning rule remain the same except for the Q value estimates, which are updated to the new values estimated in iteration 1. As a consequence, unlike in iteration 1, the $\max_{a_{t+1}} Q_t(s', a_{t+1})$ term in (16) will no longer be zero for all s' and the trader will have the option to pick the optimal action a_{t+1} conditional on the future state s' they transition to. For example, for the experience tuple $\langle s_1^f(0), NA \rangle$, the trader makes action NA, which can transition the trader to the future state $s_1^f(0)$ with probability 0.86. In this future state, the trader can make action NA or action C. Given the current Q value estimate for taking action NA while in state $s_1^f(0)$ is 0.0141, while the current Q value estimate for taking action C while in state $s_1^f(0)$ is 0, if the trader transitions to future state $s_1^f(0)$, it is optimal for the trader to take future action NA as this action results in a higher Q value.

An alternative scenario when it is not optimal for the trader to make future action NA occurs when the trader transitions to future state $s_1^b(0)$, which occurs with probability 0.02. In this state, the trader's future optimal action now differs, as it is optimal to take future action C and cancel. If the trader makes future action C while in future state $s_1^b(0)$, the associated current Q value, or long term profit, is zero. Whereas, if the trader makes future action NA, while in future state $s_1^b(0)$, the associated current Q value, or long term profit, is -0.0031.

This ability for the trader to select the optimal action when in a future state is the critical component of a reinforcement learning algorithm, allowing us to model a traders optimal management over the life-cycle of a limit order. Applying this logic, we update our second iteration estimate for $Q(s_1^f(0), NA)$ as follows:

$$
\begin{aligned}
Q_1(s_1^f(0), NA) &= E[R(s_1^f(0), NA)] + \gamma \sum_{s' \in S} T(\langle s_1^f(0), NA \rangle, s') \max_{a_{t+1}} Q_t(s', a_{t+1}) \\
&= E[R(s_1^f(0), NA)] \\
&\quad + \gamma T(\langle s_1^f(0), NA \rangle, s_1^f(0)) \max\{Q_t(s_1^f(0), NA), Q_t(s_1^f(0), C_0)\} \\
&\quad + \gamma T(\langle s_1^f(0), NA \rangle, s_2^f(0)) \max\{Q_t(s_2^f(0), NA), Q_t(s_2^f(0), C_0)\} \\
&\quad + \gamma T(\langle s_1^f(0), NA \rangle, s_3^f(0)) \max\{Q_t(s_3^f(0), NA), Q_t(s_3^f(0), C_0)\} \\
&\quad + \gamma T(\langle s_1^f(0), NA \rangle, s_4^f(0)) \max\{Q_t(s_4^f(0), NA), Q_t(s_4^f(0), C_0)\} \\
&\quad + \dots \\
&\quad + \gamma T(\langle s_1^f(0), NA \rangle, s_3^X) Q_t(s_3^X, NA) \\
&\quad + \gamma T(\langle s_1^f(0), NA \rangle, s_4^X) Q_t(s_4^X, NA) \\
&= 0.0141 \\
&\quad + 0.99\big(0.86 \times \max(0.0141, 0)\big) + 0.99\big(0.02 \times \max(0.0201, 0)\big) \\
&\quad + 0.99\big(0.02 \times \max(0.0106, 0)\big) + 0.99\big(0 \times \max(0.0141, 0)\big) + \dots \\
&\quad + 0.99\big(0.01 \times 0.0240\big) + 0.99\big(0.00 \times -0.005\big) \\
&= 0.0270
\end{aligned}
$$

Table C.1 reports the progression of our Q values estimates for each iteration of the learning rule. At iteration 200, the Q value estimates exhibit a minor deviation of less than 0.0001 from the value computed in the previous iteration. This stability indicates the Q-learning rule has converged and we can terminate the iterative process of the learning rule. We can observe the learning process of our estimation method via the progression of $Q(s_1^b(0), NA)$. In iteration 1, $Q(s_1^b(0), NA)$ takes on a value of -0.0031, but at termination, $Q(s_1^b(0), NA)$ is now positive at 0.0763. Recall that iteration 1 reports the expected immediate value if the order executes in the next transition, whereas our final iteration reports the expected value if the order is optimally managed up until execution or cancellation. $Q(s_1^b(0), NA)$ reflects the scenario in which the trader leaves an order at the back half

of the limit order book when both the bid and ask queue sizes are small. If this order was to execute immediately, the order likely faces adverse selection by a large incoming order, hence a negative immediate value. In contrast, if the order does not immediately execute, the trader can wait until favorable market conditions arrive, thereby giving a long term positive expected value.

Table C.1
Q-learning rule

This table shows the Q value estimates of the conditional expected value of a limit order for all experience tuples at the end of each iteration of the Q-learning rule defined by (14). The bottom row labeled *Difference*, reports the sum of the total change in estimates after each iteration.

	Iteration 1	Iteration 2	Iteration 3	...	Iteration 199	Iteration 200
$Q(s_1^f(0), NA)$	0.0141	0.0270	0.0387		0.1492	0.1492
$Q(s_2^f(0), NA)$	0.0201	0.0357	0.0477		0.0737	0.0737
$Q(s_3^f(0), NA)$	0.0106	0.0210	0.0311		0.1868	0.1868
$Q(s_4^f(0), NA)$	0.0141	0.0271	0.0389		0.1686	0.1686
$Q(s_1^b(0), NA)$	-0.0031	-0.0019	-0.0008		0.0763	0.0763
$Q(s_2^b(0), NA)$	-0.0065	-0.0050	-0.0038		0.0125	0.0125
$Q(s_3^b(0), NA)$	0.0000	0.0002	0.0006		0.0622	0.0622
$Q(s_4^b(0), NA)$	0.0000	0.0003	0.0008		0.0527	0.0527
$Q(s_1^X(1), NA)$	0.0009	0.0016	0.0022		-0.0025	-0.0026
$Q(s_2^X(1), NA)$	-0.0276	-0.0522	-0.0742		-0.2657	-0.2657
$Q(s_3^X(1), NA)$	0.0240	0.0456	0.0650		0.2287	0.2287
$Q(s_4^X(1), NA)$	-0.0005	-0.0011	-0.0017		-0.0230	-0.0230
$Q(s^C(0), NA)$	0.0000	0.0000	0.0000		0.0000	0.0000
$Q(s_1^f(0), C)$	0.0000	0.0000	0.0000		0.0000	0.0000
$Q(s_2^f(0), C)$	0.0000	0.0000	0.0000		0.0000	0.0000
$Q(s_3^f(0), C)$	0.0000	0.0000	0.0000		0.0000	0.0000
$Q(s_4^f(0), C)$	0.0000	0.0000	0.0000		0.0000	0.0000
$Q(s_1^b(0), C)$	0.0000	0.0000	0.0000		0.0000	0.0000
$Q(s_2^b(0), C)$	0.0000	0.0000	0.0000		0.0000	0.0000
$Q(s_3^b(0), C)$	0.0000	0.0000	0.0000		0.0000	0.0000
$Q(s_4^b(0), C)$	0.0000	0.0000	0.0000		0.0000	0.0000
Difference	0.1215	0.1024	0.0915		0.00017	0.00015

Table C.2 reports the converged Q value estimates for the states where the trader has a choice to either do nothing, NA, or cancel their order, C. The trader's optimal action is the action that gives the highest Q value. For example, when the market is in state s_1 and the trader has a limit order at the front half of the queue, the long run expected value is 0.1492 if the trader chooses to do nothing, and the long run expected profit is 0 if the trader chooses to cancel their order. Given these two scenarios, it is optimal for the trader to leave their order at the front half of the queue as this action provides a higher long term expected value.

Table C.2
Q-value estimates

Table C.2 reports the conditional expected value estimates for a limit order resting in four possible different market states (s_1,..,s_4) for the actions to leave the order (NA) or cancel the order (C).

	Front half		Back half	
	NA	C	NA	C
s_1 (small bid, small ask)	0.1492	0	0.0763	0
s_2 (small bid, big ask)	0.0737	0	0.0125	0
s_3 (big bid, small ask)	0.1868	0	0.0622	0
s_4 (big bid, bid ask)	0.1686	0	0.0527	0