Machine learning in a dynamic limit order market.

R. Philip*[1]

[1]Discipline of Finance, The University of Sydney

September 22, 2021

Abstract

We use a novel machine learning approach to tackle the problem of limit order management. Applying our framework to data, we show that the most important variable for a trader to consider is the price level of their order, followed by the queue sizes of the order book, volatility and finally queue position. Further, we show the option to cancel a limit order is valuable and contributes approximately 15% of a limit order's total expected value. This paper takes an important step towards describing pervasive features and dynamics that exist in financial markets.

Keywords: Limit order markets, machine learning, queue size, optimal limit order
JEL Classification: G10, G20, D40

*Corresponding author. *E-mail:* richard.philip@sydney.edu.au

This paper has benefited from the comments of Michael Brolley, James Brugler, David Cimon, Bjorn Hagstromer, Amy Kwan, Tom McInish, Ryan Riordan, Gideon Saar, Andriy Shkilko, Vincent van Kervel, Ester Felez Vinas, Yajun Wang, Ying Wu, Chen Yao, Marius Zoican, and the audiences at the SFS Cavalcade Asia Pacific and the FIRN annual meeting. Thank you to the participants at The University of Sydney Business Analytics seminar series and Wilfrid Laurier University seminar series and discussions with the team at Vivienne Court Trading for their insightful comments.

1 Introduction

A limit order trader faces several difficult decisions. First, the trader must decide the price level to submit their order. After order submission, as market conditions change, the trader must decide if they should cancel or update their resting limit order, to manage adverse selection risk and execution uncertainty. These decisions are non-trivial; the dimensionality of the problem is extremely large and decisions are path dependent. With limit orders becoming ever more prevalent in the modern trading era, a thorough understanding of this decision making process is important for academics, regulators and practitioners alike.

Despite the complexity of the problem, theory has made significant strides in modeling 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) and Rosu (2020) 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. The advantage of these theoretical models is that they provide an understanding of how agents optimally trade and the factors a trader must consider when making decisions under a set of strict assumptions. Further, by relaxing certain assumptions, we gain a deeper understanding into how different channels affect trading behavior. However, due to the complexity of the problem, to gain analytical tractability, these assumptions are often oversimplifying, which can come at the cost of realism (see Parlour and Seppi (2008)). Thus, it is important to empirically verify the predictions made by these theoretical models. Unfortunately, due to a lack of technology, many of the predictions made by theory have not been empirically tested.

In this paper, we present a novel machine learning approach that enables us to empirically test which factors are important in a trader's limit order management process, as predicted by theory. Specifically, for over 18,000 unique market states, we empirically estimate the expected profit of a resting limit order conditional on optimal management over its life cycle. As a consequence, our framework produces one of the richest sets of expected profit estimates, conditional on a broad range of potential market states. The variables that define our market states include the price level the limit order rests at, the shape of the order book, the queue position of the limit order and volatility. Our technique allows us to uncover several features. First, we determine when it is optimal to leave or cancel a resting limit order for different market conditions. Second, we rank the relative importance of different variables that define our market states. Finally, we identify the market conditions when the endogenous option to cancel is most valuable and uncover pervasive features and dynamics that exist in financial markets.

Recently, Moritz and Zimmermann (2019) demonstrate how ML can help a portfolio manager optimally combine multiple factors to estimate expected returns. However, limit order management has additional complexity. Not only must the trader consider how to optimally combine multiple factors, her trades or behavior may have market impact that can change these factors. Moreover, as market conditions or factors vary, she can revise her expectations and cancel the order if required. Thus, the trader has a sequential decision making process that is path dependent, akin to when one should exercise an American option.

To address this additional complexity, we cast limit order management as a sequential Markovian decision process within a reinforcement learning (RL) framework. RL is a type of machine learning that enables an agent to learn the optimal action, given the current environment, via trial and error using feedback from the agent's own actions and experiences. In our setup, at short periodic time intervals, our trader faces the same decision: to leave or cancel their resting limit order. This decision making process repeats until the trader's limit order executes or is canceled. For each periodic decision, our trader maximizes expected profit and leaves (cancels) their limit order if the order has a positive (negative) expected profit conditional on current market conditions and conditional on the future optimal management of the limit order. Thus, our framework captures the endogenous option to cancel based on the trader's future expectations. As a result, the conditional expected profit at time t is a recursive estimate based on all future conditional expected profits and their corresponding likelihoods. To overcome the recursive nature of the problem, we empirically estimate the conditional expected profit via a value iterative update function, known as Q-learning.

The key estimate in our trader's decision making process is the limit order's conditional expected profit. We draw on the existing theoretical literature to determine the variables or market conditions that could affect the conditional expected profit of a limit order. Parlour (1998) provides theoretical arguments that strategic traders should consider queue lengths on both sides of the limit order book. Further, Yueshen (2014), Li et al. (2020) and Yao and Ye (2018) argue that there is an advantage to being at the top of the queue, due to the time priority rule. Last, Foucault (1999) finds that volatility is a main determinant for limit order management. Using these concepts, we define a state space for a bid order, which considers the lengths of the queues on the first three levels of the bid side of the order book and the length of the queue on the best ask price. The bid limit order can sit at the best bid, one tick behind the bid, or two ticks behind the bid. We also consider the limit order's position within the queue and volatility. For tractability, we estimate a model in which we discretize our states, resulting in a state space of 18,001 unique market states. At any point in time, the limit order exists in one of the market states, which then transitions to a different market state in the future. Because our model is completely data driven, our framework provides the flexibility to use alternate variables to define the state space.

Application of our technique to empirical data confirms several theoretical predictions. First, consistent with Yueshen (2014), Li et al. (2020) and Yao and Ye (2018), we find that queue priority is advantageous. Further, we find the benefits of favorable queue priority is more pronounced as the order moves closer to the best price, at which execution is most likely.

Next, consistent with Parlour (1998), we show that the larger the queue size behind a resting limit order, the higher the expected profitability of the order, and the larger the queue size in front of a resting limit order, the lower the expected profitability of the order. We also find that an increase in queue size on the opposite side of the book decreases (increases) the expected profit 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. 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, adverse selection outweighs execution probability. In contrast, for orders resting behind the best price, execution probability outweighs adverse selection.

Last, Foucault (1999) predicts volatility has two opposing forces on a limit order's profitability.

The first force suggests an increase in volatility *decreases* the expected profit of a limit order, via an increase in the risk 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. Due to discrete price ticks, we demonstrate 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 profit. In this scenario, liquidity providers do not widen spreads to compensate for the increased picking off risk. In contrast, for less tick constrained stocks, we find that an increase in volatility increases the expected profit of a limit order as liquidity providers are willing to widen their spreads as compensation for the increase in picking off risk. Because price levels are discrete, liquidity providers widen spreads to price levels that over compensate, rather than under compensate, for the increase in losses due to picking off risk.

Using a ML technique known as Mean Decrease in Accuracy (MDA), we rank the relative importance of the different variables that define our market states. We find that the price level at which the limit order rests is the most important variable for a trader to consider. Following price level, the sizes of the queues at different price levels are next most important, then volatility and last the queue position of the order.

Our technique also allows us to determine how valuable is the option to cancel a limit order and under what market conditions is the option most valuable. We find the option to cancel represents 15% of a limit order's total expected value, on average. However, during periods of high *ex-ante* adverse selection risk, which we proxy by order book pressure, this option becomes even more valuable.

The advantage of our approach is four-fold. First, rather than directly applying ML to a problem, which can obscure economic intuition (see Chinco et al. (2019)), we cast optimal limit order management within a theoretical framework. By doing so, we are able to identify economically important variables for limit order management and rank their relative importance. Second, while our technique does model limit order management, we differ from traditional theory models in that our approach is completely driven by empirical data and does not require an equilibrium for all market particpants. Thus, our empirical approach is similar to Hollifield et al. (2004) as we remove the need for assumptions about trader behavior or market dynamics. By removing assumptions about trader behaviour, we determine optimal order management under real market conditions, where traders may not necessarily behave rationally, or follow a stylized set of assumptions. This feature allows us to determine if traders behaviors are consistent with existing theories. Third, similar to Goettler et al. (2005) and Goettler et al. (2009), our approach can handle a state-action space with large dimensionality, which enables us to capture a wide range of market states and price levels in the limit order book. Last, as in Goettler et al. (2009), our limit order's expected profit estimates are conditional on the future endogenous option to cancel. Thus, we can estimate the option value of cancelling a limit order.

We provide three main contributions to the literature. First, O'Hara (2015) highlights that in the modern era, markets and trading have changed, with limit orders now playing a more crucial role. As a consequence of this change, O'Hara (2015) issues 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. Second, our technique allows us to empirically test several theoretical predictions. Moreover, using a ML technique know as Mean Decreased Accuracy (MDA), we are able to rank the relative importance of variables that theory has identified as important to a limit order trader. Third, we provide new insights on the value of the option to cancel a limit order. While order cancellations have grown significantly in recent years, the literature has not fully considered the value a trader should place on the option to cancel.

We are not the first to use RL for solving a trader's objective function. Nevmyvaka et al. (2006) and Bertsimas and Lo (1998) demonstrate the efficacy of RL in solving the problem a trader faces when required to execute a large block over a pre-defined time period. However, our trader is not forced to execute a large order and therefore has a significantly different objective function. Our trader's goal is to optimally manage their limit orders such that they make the most profitable opportunistic trades based on current market conditions. Thus, our objective function allows us to estimate the expected profit of a limit order and identify key variables in a traders decision making process.

2 Method

2.1 Intuition

Consider a trader who wants to optimally manage their limit orders to ensure that only limit orders with a positive expected value execute. The dynamics of the limit order book make this task non-trivial, as the trader must constantly monitor her resting limit orders and cancel them if they are expected to lose money. To achieve this task, the trader must estimate the expected profit of a limit order conditional on the current state of the market *and* future optimal order management over the order's life cycle.

Estimating the expected profit of a limit order conditional on future optimal management requires the trader to consider the evolution of market conditions and their likelihoods of occurring. The trader must consider the evolution of market conditions until two points in time, 1) when the order executes, or 2) when the order is canceled. However, the decision to cancel an order is endogenous and should occur when the limit order has a negative expected profit.

Figure 1 provides an illustration of 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 price and ask price is 13 and 14, respectively, resulting in a bid ask spread of 1. Assume a trader submits a limit buy order at t_0 at a price of 12 (one tick behind the best available bid), which we depict as a black rectangle in Figure 1.

[Insert Figure 1]

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 has evolved to one of only two possible market 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 and thus, our trader's order has moved 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. Rather, a large sell limit order at 13 has been submitted, removing the bid at 13 which existed at t_0.

If the volume available on the bid side of the order book is significantly larger (smaller) than the volume available on the ask side of the order book, the midprice is more likely to increase (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 is much larger than the volume available on the bid, indicating the price is likely to decline in the future and the order would be adversely selected.

The expected profit of the limit order submitted at t_0, if left unmonitored, is the expected value in State A and B multiplied by their respective probabilities of occurring. Therefore, if the probability of transitioning to State B is much higher than the probability of transitioning to State A, then it is possible that the expected profit of the limit order submitted at t_0, when left unmonitored, is negative. However, if we allow monitoring and cancellation of the limit order, then the expected profit of the order becomes positive as the trader would cancel the order if the market transitions to State B, which results in a profit of 0, whereas the trader would leave the limit order if the market transitions to State A, where the order has a positive expected value. In this oversimplified example, it is evident that the option to cancel can change an order from having a negative expected profit, to a positive expected profit.

In this illustrative example, we make two tenuous assumptions. First, we assume the market can only transition to two possible states after the trader submits their order. In reality, the market can transition to an almost infinite number of states. Second, we arbitrarily assert the limit order has a positive expected value when in State A, whereas the limit order has a negative expected value when in State B. To acquire accurate estimates of the expected value, we must estimate the expected value of the limit order while in State A and B (time t_1) if optimally managed over its life cycle, which is the exact same problem we are trying to solve at t_0. To overcome these two limitations, we use a recursive state space technique known as reinforcement learning (RL) which can handle many states and capture the recursive nature of the problem.

2.2 Reinforcement Learning

Typically in a 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. The agent maximizes this 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

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

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 has executed.

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:[1]

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

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 profit the limit order will make 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 profit equals any immediate profit for taking action a plus the expected long run profit 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.[2]

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

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

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

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 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 trader cancels their order. If the trader's limit order executes, the trader continues to monitor market conditions to observe the long term profit of the executed order.

[Insert Figure 2]

This recursive decision making system allows the trader to expose the same limit order for multiple consecutive intervals, during which time she can monitor the order's queue position and market conditions. If at any point in time the order appears to have a high chance of adverse selection (measured by a negative expected profit), 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, we are able to make better estimates on the likelihood of transitioning to future market conditions as dramatic changes in market conditions 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 action is made at the beginning of the interval. To ensure that the trader's limit order is only at price levels in our state space, if the market transitions to a state where the trader's resting limit order is no longer in the state space, then the action NA is overruled by action C, which means the resting limit order is canceled. This overruling forces the trader to cancel resting limit orders if the best bid and offer has moved far 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 is a consideration for trader's strategic behavior. 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).[3] 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).

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

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

is either 0 (no position) or 1 (long). We also include a variable, L, that captures the price level the traders limit order is resting at. We let L take on the value of $i \in 1, 2, 3$ if the trader has a resting limit order submitted at level i of the order book. Finally, some argue there is an advantage to being at the top of queue, as the order has time priority (see Yueshen (2014), Li et al. (2020) and Yao and Ye (2018)). For this reason, 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 quantiles, which we define as *top*, *top-middle*, *middle*, *middle-back* and *back*.

Last, to ensure we estimate the expected profit of a single limit order in isolation, we include a state which 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

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

where

$$\begin{aligned}
&I \in \{0, 1\}\\
&L \in \{0, 1, 2, X\}\\
&Q \in \{top, top\text{-}middle, middle, middle\text{-}back, back, X\}\\
&q^j \in \{ELo, Lo, No, Sh, ESh\}, \forall j \in \{B_0, B_1, B_2, A_0\}\\
&V \in \{Low, Med, Hi\}
\end{aligned}$$

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.[4] In contrast, when the trader has no inventory and is working their limit order, the state is defined by the public variables plus the private variables (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.[5] 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 transition probability estimates. 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 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 $T(\langle s_i, NA \rangle, s_j)$

[4] 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.

[5] 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.

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.[6]

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 profit generated during the transition from the current state to the next. Therefore, if the trader has a positive inventory when in state s, the immediate reward for the transition from state s to s' is the observed change in midpoint during the transition. 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. If a trader has no inventory in state s but their order executes during the transition from s to s', then the immediate reward is the midpoint price observed in state s' less the limit order's execution price. Formally, if we define the midpoint price in state i as mid_i, then the immediate reward from making action a while in state s that results in a transition to state s' is

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

$$R(\langle s,a\rangle,s') = (mid_{s'} - mid_s) \times I_s + (mid_{s'} - execPrice_s) \times Exec_{s,s'}, \tag{6}$$

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.[7] 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 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 statis-

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

tics for the 20 sample stocks we use, ranging from the lowest price stock of Santos (STO), with a price of approximately at 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 ticks to 2.59 ticks.

[Insert Table 1]

The ITCH data contains full order book data with nanosecond time stamping and allows us to fully reconstruct the complete order book at all prices levels, including all individual resting limit orders and their queue position. We prepare the data for our analysis as follows. First, we reconstruct the limit order book so that we can replay the market over the course of a trading day. Second, for each trading day, we create 210,000 consecutive intervals of length 100ms, with the first interval starting at the beginning of continuous trading at 10:10 and the last interval ending at 16:00, when continuous trading ceases.

At the beginning of each interval, we assume there are a series of hypothetical limit orders located at various price levels and positions in the queue. Specifically, for consistency with our RL model, we assume hypothetical bids, for one share, are located at the prevailing best bid, one tick behind the best bid and two ticks behind the best bid. Moreover, at each of these price levels we assume there is a hypothetical order at the top of the queue, three quarters, half way, and one quarter towards the top of the queue and one at the very back of the queue.[8]

Next, using the granularity of the data, we track these hypothetical limit orders over the next 100ms and determine if any of them would have executed.[9] If the hypothetical limit order did not execute during the 100ms interval we track the order's progression by knowing which real orders executed ahead of the hypothetical order and which real orders were canceled and submitted over the interval, thereby allowing us to identify the location of the hypothetical order in the order book.

[8] We assume each order is only one share to ensure the order does not have an economically meaningful impact. Moreover, the price and queue location for the hypothetical orders is chosen to ensure our observations span the state space defined by our RL framework.

[9] We assume a hypothetical order would have executed if a real order located in the queue behind the hypothetical limit order executes during the 100ms interval, or, if a trade occurs during the interval, at a price worse than the hypothetical limit orders price.

At the end of each interval, for each hypothetical order, we record information on the state space of the order at the beginning and at the end of the interval. Specifically, at the start of the interval, we record the volatility, initial queue position and the total volume available at the first three best bid prices and the best ask price at the beginning of the interval.[10] At the end of the interval, we record whether the order executes or not. If the order does not execute, we report the order's new queue position. Further, regardless of whether the order executes, we record the volatility, total volume available at the first three best bid prices and the best ask price measured at the end of the interval.

With the extracted information, we know each order's initial starting state and its state at the end of the interval, thereby enabling us to estimate the required 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 ($q^{B_0}, q^{B_1}, q^{B_2}$ and q^{A_0}), each with five possible values (except q^{B_2}, which only has three states for tractability reasons). We also include the public volatility state which can take on three possible values. We also have two private state variables based on 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 have an inventory position, and 1 absorbing cancel state. Collectively, this means we have $m = 16,875$, $n = 1,125$ and $o = 1$, resulting in 18,001 unique states.[11] In the following subsections, we investigate the effect of each variable on the expected profit of a limit order.

[10] We measure volatility as the highest traded price minus lowest traded price over the last 100 trades.

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

4.1 Price levels

The relation between a resting limit order's price and expected profit is not immediately clear due to two opposing forces; 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 probability (see Handa and Schwartz (1996)).

Figure 3 presents a boxplot of expected profit 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 profit 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 who accommodate purchases (sales) should be compensated with a higher (lower) price than the fundamental value (see Scholes (1972)).

[Insert Figure 3]

Table 2 reports the summary statistics for our expected profit 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 profit is highest at 0.319 ticks. When an optimally managed limit order moves away from the best bid, its expected profit drops to 0.202 ticks when it is one tick behind the best bid, and drops further to 0.071 ticks, when it is two ticks behind the best best bid. Similarly, the variance in a limit order's conditional expected profit decreases as the order moves away from the best price. The expected profit of an order located at the best bid has a standard deviation of 0.213 ticks, but this value drops to only 0.066 ticks when the order is two behind the best bid.

[Insert Table 2]

The observation that the mean expected profit and variance of expected profit decreases as the order moves further away from the best bid or offer may provide an explanation for why the majority of order cancellations occur at the best bid or ask (see Fong and Liu (2010)). Intuitively, a trader has no incentive to cancel a limit order resting far from the best price. This order should have a small positive expected profit as it has little execution risk and may gain favorable queue priority in the future. If the market moves towards the resting limit order, the probability of execution increases and the expected profit of the order could turn negative, and at this point, the trader should evaluate the option to cancel their order.

4.2 Queue Position

In this section, we investigate the effect of a limit order's queue position on the order's expected profit. Some argue that there is an advantage to being at the top of the queue, due to the time priority rule (see Yueshen (2014), 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 will 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 profit 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 profit 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. To isolate just the effect of queue position, we use fixed effects for all other variables that define our state space

Table 3 presents the mean coefficient across all 20 sample stocks for orders resting at the best bid, one tick behind the best bid and two ticks behind the best bid in columns 1,2 and 3, respectively. Table 3 also reports the number of stocks with a statistically positive or negative coefficient. Our results provide strong evidence that queue priority is advantageous for a limit order trader. The coefficient for queue position is negative and significant across all 20 sample stocks and all price levels, suggesting that the further back a limit order's position in the queue, the lower the limit order's expected profit. The magnitude of the coefficients suggest that queue position has an economically meaningful effect on the expected profit of a limit order. For example, an orders expected profit on the best bid would drop by 0.12 ticks if it went from the top of the queue ($QueuePos_s = 0$) to the back of the queue ($QueuePos_s = 1$), which consists of almost half the average value of a limit order resting on the best bid (0.319 ticks).

We also observe that the magnitude of the mean coefficient decreases as the order moves to price levels further from the best bid. Specifically, for orders at the best bid, the mean coefficient is -0.12, whereas for orders 1 level behind the best bid, the mean coefficient is -0.05, with the mean coefficient even more attenuated at -0.01 for orders two levels behind the best bid. This result suggests that queue priority becomes more important 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 (2014), Li et al. (2020) and Yao and Ye (2018). Similarly, our results provide support for Lo et al. (2002), who postulate that the simulated profits generated from placing a network of buys and sell limit orders reported in Handa and Schwartz (1996) may be over-inflated due to an assumption that the network of orders are placed at the top of the queue, thereby not fully considering the importance of queue priority.

4.3 Queue sizes

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 profit of a limit order. To investigate the relation between queue sizes and expected profit, we estimate the following regression for orders at different price levels:[12]

[12] 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.

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

where Q_s is the expected profit 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 just 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 best bid, Two ticks behind 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 profitability of the order. Conversely, the larger the queue size *in front* of a resting limit order, the lower the expected profitability 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 price level the order is resting at increases the limit order's expected profit. 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, an increase in queue sizes at price levels ahead of the resting limit order decreases the order's expected profit. 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. Similarly, 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 for orders resting two levels behind the best bid.

[Insert Table 4]

Our findings manifest in two ways. First, 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.

Second, a limit order with more volume in front of the order has a lower execution probability. This is because the volume in front of the order creates 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, and consequently the price is likely to move away from the order, resulting in no execution.

The relation between a limit order's expected profit and the amount of 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 profit 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 loses from adverse selection more than the expected profits from increased execution probability. In contrast, when the order is behind the best bid, expected profits from increased execution probability outweighs the expected loses from adverse selection. Specifically, the magnitude of the midpoint movements due to order book pressure is typically not more than two ticks. Therefore, when a limit order is behind the best bid and there is downward book pressure, the order's execution probability is high but adverse selection risk is low, as the midpoint is unlikely to move lower than the order. Moreover, if the downward order book pressure persists after the first downtick, the order is canceled, before it is adversely selected.

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 profitability of the order, and the larger the queue size *in front* of a resting limit order, the *lower* the expected profitability 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 loses 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.

4.4 Volatility

In this section, we investigate the effect of volatility on the expected profit of a limit order. While existing theoretical models shed some light on the relation between volatility and the expected profit 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 would decrease the expected profit of a limit order. Specifically, Foucault (1999) predicts that when volatility increases, the probability of a limit order being picked off and the losses that ensue are larger, which would decrease the expected profit of a limit order. The second force is a reaction to the first force; to compensate for the higher likelihood of adverse selection and corresponding reduction in expected profit of a limit order, liquidity providers widen the bid ask spread when volatility increases (see Copeland and Galai (1983) and Foucault (1999)). If the prices that limit orders could be submitted to was continuous, then liquidity providers would widen the spread to exactly offset the expected losses due to the increase in picking off risk. However, in reality price levels are discrete and liquidity providers may not always be able to set the spread to perfectly offset the effect of increased volatility. Because of this price discretization, volatility could have a positive or negative effect on the expected profit of a limit order. To determine the effects of volatility on the expected profit of a limit order, we perform the following regression:

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

where Q_s is the expected profit of a limit order in state s and $Volatility_s$ is the volatility in state s. To isolate just the effect of volatility, we use fixed effects for all other variables that define our state space. Given that we are primarily interested 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 across all sample stocks of 0.38. While this average coefficient suggests that volatility increases the expected profit of a limit order, the results are less clear upon closer inspection of the individual coefficients for each stock. Specifically, 9 out of our 20 sample stocks have a positive coefficient, while 11 out of 20 have a negative coefficient. Thus, the effect of volatility on the expected profit of a limit order at the best bid is not consistent across all stocks.

[Insert Table 5]

Given the opposing forces volatility may have on the expected profit of a limit order identified in Foucault (1999), we hypothesize that the effect differs depending on whether the stock is tick constrained. For stocks that are typically tick constrained, we propose that an increase in volatility has a negative effect on the expected profit of a limit order, as liquidity providers need not widen their spreads to compensate for the increase in picking off risk. In contrast, for stocks that are less tick constrained, we predict that an increase in volatility increases the expected profit of a limit order. For these stocks, liquidity providers are willing to widen their spreads as compensation for the increase in picking off risk during high volatility periods. Because of the discrete price levels, liquidity providers post orders at prices that over compensate rather the under compensate for the increased losses due to picking off risk. Thus, an increase in volatility increases the expected profit of a limit order.

To test if volatility has different effects on the expected profit of a limit order for tick constrained and unconstrained stocks, we analyze two stock sub samples. Table 5, column 2 reports results for the quartile of most tick constrained stocks, while column 3 reports results for the quartile of stocks that are least tick constrained. Consistent with our hypothesis, we find that volatility decreases the expected profit of a limit order for stocks that are tick constrained. In contrast, for stocks that are less tick constrained we find that volatility increases the expected profit of a limit order.

Taken together, our results support Foucault (1999), who predicts that volatility can affect the expected profit of a limit order via two channels. First, Foucault (1999) suggests that an increase in volatility increases adverse selection (i.e., decreases the expected profit of a limit order). Due to price discretization, this channel is most prevalent in tick constrained stocks; in this subsample, we find an increase in volatility decreases the expected profit of a limit order. The second channel identified in Foucault (1999) is in response to the first channel: Liquidity providers demand compensation for the increased risk of adverse selection by widening bid ask spreads (i.e., an increase in volatility increases the expected profit of a limit order). Again, due to price discretization, this channel is most prevalent in less tick constrained stocks. For these less constrained stocks, an increase in volatility increases the expected profit of a limit order.

4.5 Variable importance

The results so far show that price levels, queue sizes, queue position and volatility all affect the expected profit of a limit order. In this section, we determine the importance of these variables using a technique found in the machine learning literature know as Mean Decreased Accuracy (MDA), which has more recently been used in finance by Easley et al. (2019). In our setting, MDA measures the decrease in accuracy of the forecast expected profit of a limit order if one of the variables defining our states is measured with error.

Estimating the MDA requires two parameters. The first parameter is the true expected profit of a resting limit order, $Q(s, NA)$, which we estimate via the RL model. The second parameter is the randomized expected profit of a resting limit order, $Q(s_R^k, NA)$ which we estimate by randomizing one of the 7 variables that define the state space while holding all other variables constant. The randomized expected profit, $Q(s_R^k, NA)$, is the expected profit associated with the randomly chosen state, s_R^k, that is created by randomizing variable k. Using these two parameters, we can 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 profit estimates that is caused when one variable is measured with error. Thus, the larger a variable's MDA, the more important that variable is for determining the expected profit of a limit order. For each variable, k, we estimate the MDA and repeat this process 100 times.

Table 6 reports the mean and standard deviation of the MDA for each variable. We find that the most important variable that drives the expected profitability of a limit order is the price level at which the limit order rests. Our results suggest that the next most important variables are the queue sizes on the same side of the order book that the limit order rests. We find that the importance for queue size decreases as the queues move further away from the best bid. More specifically, we find that the size of the queue 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 the queue size 2 ticks behind the best bid (MDA = 0.68). After queue sizes on the same side of the book as the order, we find the next most important variable is queue size on the opposite side of the order book (MDA = 0.68), then volatility (MDA = 0.56) and last queue position (MDA = 0.2).

[Insert Table 6]

4.6 The option to cancel

Despite the prevalence of order cancellations, the option to cancel has received little attention in the literature. Accordingly, in this section, we contribute to the literature by investigating the value of the option to cancel and the market conditions when this option is most valuable. We re-estimate a constrained version of our RL model, which restricts the trader to only one action, NA, which implies the trader is unable to cancel their order. Thus, the estimated Q values from this restricted model is the expected profit of a limit order that is not optimally managed. To determine the value of the option to cancel, we compute the difference between the Q value of the unrestricted model, which contains the option to cancel, and the Q value of the restricted model, which has no option to cancel.

Table 7 reports the summary statistics on the value of the option to cancel. The first row reports summary statistics for limit orders in any market state, while rows 2 to 4 report summary statistics for limit orders conditional on their price level. The value of the option to cancel a limit order on the best bid is 0.049 ticks, on average. Thus, on average, the option to cancel a limit order from the best level is worth approximately 15% of the total value of an optimally managed limit order.[13] This finding suggests the endogenous option to cancel a limit order contributes an economically meaningful amount towards the total expected profit of an optimally managed limit order.

[Insert Table 7]

[13] Table 2 reports a mean value of an optimally managed limit order at the best bid of 0.31 ticks.

Table 7 also suggests the value of the option to cancel is heavily skewed towards certain market conditions. Specifically, for orders resting at any price level, we observe the means are substantially higher than the corresponding medians, indicative of a large right skew in the data.

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*. Thus, 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. Thus, 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:

$$value\,of\;option\,to\,cancel = \beta_0 + \beta_1 q^{B_0} + \beta_2 q^{B_1} + \beta_3 q^{B_2} + \beta_4 q^{B_3} + \epsilon. \tag{12}$$

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 to decrease the value of the option to cancel. Similarly, we expect an increase in volume on the *opposite* side of the order 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. In other words, the more volume on the same side as the limit order, the lower the value of the option to cancel. Whereas, the more volume on the opposite side of the limit order, the higher the value of the option to cancel.

[Insert Table 8]

5 Conclusion

While limit order markets have become the dominant trading mechanism, we know little about the dynamics of the limit order book and order management strategies due to the complexity of the problem (see Parlour and Seppi (2008)). Understanding limit order management is relevant for academics, practitioners and those who regulate our financial markets. For example, to set a maximum order-to-trade ratio or a minimum resting time, one must first understand how to optimally manage a limit order to determine what are reasonable values.

We propose a recursive sequential framework for limit order management within a machine learning model. One innovation is modeling the endogenous choice to cancel a limit order. In our framework, the option to cancel a limit order is exercised if the expected profit of the limit order becomes negative. The expected profit of a limit order is a function of the current market conditions *and* expectations about future market conditions and their likelihoods of occurring.

Our ML approach empirically confirms the theoretical predictions on the variables, or market conditions, that are important for a limit order trader. Specifically, we show that queue size, order priority, volatility and order price have an economically meaningful effect on the expected profit of a limit order. Moreover, using Mean Decreased Accuracy (MDA) to rank the relative importance of these variables, we find that price level is the most importance variable, followed by queue sizes, volatility and queue priority. Further, we show that the endogenous option to cancel is important: On average, this option to cancel represents 15% of a limit order's total expected profit. During periods of high adverse selection risk, this option becomes even more valuable.

6 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, 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.

$$
\begin{array}{cc|cccc|ccc|c|}
 & & \multicolumn{8}{c}{\text{Future State}} \\
 & & s_1(0) & s_2(0) & \cdots & s_n(0) & s_1(1) & \cdots & s_m(1) & s^C(0) \\
\hline
 & s_1(0) & p_{1,1} & \cdots & & p_{1,n} & p_{1,n+1} & \cdots & p_{1,n+m} & p_{1,n+m+1} \\
 & s_2(0) & \vdots & \ddots & & & \vdots & \ddots & & \vdots \\
\text{Current state with} & & & \text{Unexecuted} & & & & \text{Executed} & & \\
\text{action } NA & s_{n-1}(0) & p_{n-1,1} & & & & p_{n-1,n+1} & & & \\
 & s_n(0) & p_{n,1} & \cdots & & p_{n,n} & p_{n,n+1} & & p_{n,n+m} & p_{n,n+m+1} \\
\hline
 & s_1(1) & 0 & \cdots & 0 & 0 & p_{n+1,n+1} & \cdots & p_{n+1,n+m} & 0 \\
 & \vdots & \vdots & \text{Prohibited} & & 0 & \vdots & \text{Long} & \vdots & \vdots \\
 & s_m(1) & 0 & 0 & 0 & 0 & p_{n+m,n+1} & \cdots & p_{n+m,n+m} & 0 \\
\hline
 & s^C(0) & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 \\
\hline
\end{array}
$$

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 capture the evolution of market conditions and the limit order's movement. For example, the transition probabilities capture the likelihood of the limit order's progression up the queue of the limit order book, or how other market participants are likely to react to current market conditions. We estimate these values empirically via (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 further limit orders. As a result, the trader must transition to one of m positive inventory states, $s_j(1)$, where j captures different possible states based on the public information reflected in the order book variables. Again, we estimate these values empirically via (5).

Following execution, the trader must remain in one of the m positive inventory states and is unable to submit another order. To ensure the trader does not have another limit order once they are long, and remains in a positive inventory state, we specify the block titled "Prohibited" in Figure A.1 to contain only zeros. The block titled "Long" captures the transition probabilities for a trader who is long in one market state and transitions to another market state where they continue to be long, which requires empirical estimation via (5).

The final column reports the probability 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, then the probability they are in the absorbing cancel state in the subsequent period is 1. Given 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 should the resting limit order transition into an undefined state (more than three ticks from the best bid), the trader's action NA is overruled and the order is canceled. Thus, there can be a non zero probability the order is canceled, which we empirically estimate.

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, if 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, we restrict any state where the trader has an inventory position, or has already canceled their order, to not having another resting order. Because of this restriction, taking the action to cancel an order when 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

$$\begin{array}{l|cccc|ccc|c|}
 & s_1(0) & s_2(0) & \cdots & s_n(0) & s_1(1) & \cdots & s_m(1) & s^C(0) \\
s_1(0) & 0 & \cdots & & 0 & 0 & \cdots & 0 & 1 \\
s_2(0) & \vdots & \ddots & & & \vdots & \ddots & & 1 \\
 & & \text{Unexecuted} & & & & \text{Executed} & & 1 \\
s_{n-1}(0) & & & & & & & & 1 \\
s_n(0) & 0 & \cdots & & 0 & 0 & & 0 & 1 \\
\hline
s_1(1) & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \text{Prohibited} & & 0 & & \text{Prohibited} & & 0 \\
s_m(1) & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0
\end{array}$$

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 contains the matrix of immediate rewards for all experience tuples that occur when the action is NA. If the trader's limit order is unexecuted during the transition, then the immediate reward is zero, which is shown in the upper left quadrant. In contrast, if the trader's limit order executes, then the immediate reward is the profit generated. We empirically compute the immediate profit via (6), which is the difference between the execution price and the midpoint in the future state s' and is 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

$$
\begin{array}{c|cccc|ccc|c}
 & s_1(0) & s_2(0) & \cdots & s_n(0) & s_1(1) & \cdots & s_m(1) & s^C(0) \\
s_1(0) & 0 & \cdots & & 0 & r_{1,n+1} & \cdots & r_{1,n+m} & 0 \\
s_2(0) & \vdots & \ddots & & & \vdots & \ddots & & 0 \\
 & & \text{Unexecuted} & & & & \text{Executed} & & 0 \\
s_{n-1}(0) & 0 & & & & r_{n-1,n+1} & & & 0 \\
s_n(0) & 0 & \cdots & & 0 & r_{n,n+1} & & r_{n,n+m} & 0 \\
\hline
s_1(1) & 0 & \cdots & 0 & 0 & r_{n+1,n+1} & \cdots & r_{n+1,n+m} & 0 \\
\vdots & \vdots & \text{Prohibited} & & 0 & \vdots & \text{Long} & \vdots & 0 \\
s_m(1) & 0 & 0 & 0 & 0 & r_{n+m,n+1} & \cdots & r_{n+m,n+m} & 0 \\
\hline
s^C(0) & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0
\end{array}
$$

When the action is NA, the limit order can either execute or there can be an existing long position. Either of these scenarios can result in a non-zero immediate reward. 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 the $S \times S$ immediate reward matrix, when the action is 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:

$$Q_{t+1}(s,a) = Q_t(s,a) + \alpha\Big(E[R(s,a)] + \gamma \sum_{s' \in S} T(\langle s,a \rangle, s') \max_{a} Q_t(s',a) - Q_t(s,a)\Big), \tag{13}$$

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 available action is to cancel the existing limit order (C). The second available action is to do nothing (NA), which leaves the existing limit order in the queue. However, to simplify our illustrative example, we reduce the state space and only consider the queue size at the best bid (q^{B_0}) and best ask (q^{A_0}) and ignore the queue sizes at levels behind the best bid (q^{B_1},q^{B_2}). Moreover, we discretize queue size into only two categories, which we define as *large* and *small*. To further reduce dimensionality, we reduce the private state variable, queue position (Q) to only two possible states, $front$ and *back*, representing whether the order is in the front half or back half of the queue, respectively. These reductions result in a state space of 8 possible states when the trader has no inventory and is executing a limit order, 4 possible market states when the trader has an inventory position and is no longer executing an order, and 1 absorbing state which occurs when the trader cancels their order. Collectively, our setup has a total of 13 possible unique market states. More formally, $m = 8$, $n = 4$ $o = 1$ and $S = 13$ and we define each state as

$$s_k^j(I) = [I, L, Q, q^{B_0}, q^{A_0}] = \begin{cases} s_1^f(0) = [0, 0, front, small, small] \\ s_2^f(0) = [0, 0, front, small, large] \\ s_3^f(0) = [0, 0, front, large, small] \\ s_4^f(0) = [0, 0, front, large, large] \\ s_1^b(0) = [0, 0, back, small, small] \\ s_2^b(0) = [0, 0, back, small, large] \\ s_3^b(0) = [0, 0, back, large, small] \\ s_4^b(0) = [0, 0, back, large, large] \\ s_1^X(1) = [1, X, X, small, small] \\ s_2^X(1) = [1, X, X, small, large] \\ s_3^X(1) = [1, X, X, large, small] \\ s_4^X(1) = [1, X, X, large, large] \\ s^C(0) = [0, X, X, -, -] \end{cases} \tag{14}$$

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 cancelled 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. 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 cancelled their order, they can only make action NA. With the state action space defined, the input variables for (13) 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. 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												
			$s_1^f(0)$	$s_2^f(0)$	$s_3^f(0)$	$s_4^f(0)$	$s_1^b(0)$	$s_2^b(0)$	$s_3^b(0)$	$s_4^b(0)$	$s_1^X(1)$	$s_2^X(1)$	$s_3^X(1)$	$s_4^X(1)$	$s^C(0)$
1	Current state with action NA	$s_1^f(0)$	**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								**0.94**	0.03	0.03	0.00	0
10		$s_2^X(1)$									0.04	**0.91**	0.01	0.04	0
11		$s_3^X(1)$									0.03	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
14	Current state with action C	$s_1^f(0)$	0								0				1
15		$s_2^f(0)$													1
16		$s_3^f(0)$													1
17		$s_4^f(0)$													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								0				0
23		$s_2^X(1)$													0
24		$s_3^X(1)$													0
25		$s_4^X(1)$													0
26		$s^C(0)$	0								0				0

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 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. 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.

$$
\begin{array}{lll|cccccccc|cccc|c}
 & & & \multicolumn{13}{c}{\text{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) \\
 & 1 & s_1^f(0) & & & & & & & & & \mathbf{0.32} & 0.25 & -0.19 & -0.05 & 0 \\
 & 2 & s_2^f(0) & & & & & & & & & -0.30 & \mathbf{0.47} & -0.49 & -0.00 & 0 \\
 & 3 & s_3^f(0) & & & & & & & & & 0.20 & 0.16 & \mathbf{0.43} & 0.31 & 0 \\
 & 4 & s_4^f(0) & \multicolumn{8}{c|}{0} & -0.40 & 0.43 & -0.36 & \mathbf{0.47} & 0 \\
 & 5 & s_1^b(0) & & & & & & & & & \mathbf{-0.07} & -0.02 & -0.24 & -0.09 & 0 \\
\text{Current state with} & 6 & s_2^b(0) & & & & & & & & & -0.45 & \mathbf{0.33} & -0.49 & -0.04 & 0 \\
\text{action } NA & 7 & s_3^b(0) & & & & & & & & & -0.18 & -0.15 & \mathbf{-0.15} & -0.14 & 0 \\
 & 8 & s_4^b(0) & & & & & & & & & -0.50 & 0.23 & -0.49 & \mathbf{0.16} & 0 \\
\hline
 & 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) & \multicolumn{8}{c|}{0} & 0.20 & 0.84 & 0 & 0.24 & 0 \\
 & 12 & s_4^X(1) & & & & & & & & & -0.08 & 0.25 & -0.24 & 0 & 0 \\
\hline
 & 13 & s^C(0) & \multicolumn{8}{c|}{0} & \multicolumn{4}{c|}{0} & 0 \\
\hline
\hline
 & 14 & s_1^f(0) & & & & & & & & & & & & & 0 \\
 & 15 & s_2^f(0) & & & & & & & & & & & & & 0 \\
 & 16 & s_3^f(0) & & & & & & & & & & & & & 0 \\
\text{Current state with} & 17 & s_4^f(0) & \multicolumn{8}{c|}{0} & \multicolumn{4}{c|}{0} & 0 \\
\text{action } C & 18 & s_1^b(0) & & & & & & & & & & & & & 0 \\
 & 19 & s_2^b(0) & & & & & & & & & & & & & 0 \\
 & 20 & s_3^b(0) & & & & & & & & & & & & & 0 \\
 & 21 & s_4^b(0) & & & & & & & & & & & & & 0 \\
\hline
 & 22 & s_1^X(1) & & & & & & & & & & & & & 0 \\
 & 23 & s_2^X(1) & & & & & & & & & & & & & 0 \\
 & 24 & s_3^X(1) & \multicolumn{8}{c|}{0} & \multicolumn{4}{c|}{0} & 0 \\
 & 25 & s_4^X(1) & & & & & & & & & & & & & 0 \\
\hline
 & 26 & s^C(0) & \multicolumn{8}{c|}{0} & \multicolumn{4}{c|}{0} & 0
\end{array}
$$

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 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 (13), 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:

$$Q_1(s_1^f(0), NA) = E[R(s_1^f(0), NA)] + \gamma \sum_{s' \in S} T(\langle s_1^f(0), NA \rangle, s') \max_{a_{t+1}} Q_t(s', a_{t+1}), \tag{15}$$

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 (15) will be zero for all s' and the trader will be indifferent to all choices of a_{t+1}. Thus, the second term of (15) 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 (15) 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:

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

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 minor deviations of less than 0.0001 from the values 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 profit if the order executes in the next transition, whereas our final iteration reports the expected profit 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 were to execute immediately, the order likely faces adverse selection by a large incoming order, hence a negative immediate profit. In contrast, if the order does not immediately execute, the trader can manage the order until favorable market conditions, thereby giving a long term positive expected profit.

Table C.1
Q-learning rule

This table shows the Q value estimates of the conditional expected profit of a limit order for all experience tuples at the end of each iteration of the Q-learning rule defined by (13). 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 profit 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 profit.

Table C.2
Q-value estimates

Table C.2 reports the conditional expected profit 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

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
Expected profit summary statistics

This table reports the summary statistics on the expected profit 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	-0.695	0.053	0.155	0.197	0.302	1.566	0.179
Best bid	-0.695	0.164	0.325	0.319	0.465	1.566	0.213
One tick behind	-0.615	0.122	0.191	0.202	0.277	1.089	0.124
Two ticks behind	-0.662	0.023	0.056	0.071	0.103	0.939	0.066

Table 3
Expected profit and queue position

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 profitability of a limit order estimated via our RL model. The independent variable is $QueuePos$, with fixed effects controlling for all other variables. 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 out of the full sample of 20 stocks.

	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
R-Square	0.89	0.88	0.89
No. obs	5625	5625	5625

Table 4
Expected profit and queue size

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 profitability 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. 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 out of the full sample of 20 stocks.

		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
R-Square		0.88	0.89	0.83
No. obs		5625	5625	5625

Table 5
Expected profit and volatility

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 profitability of a limit order estimated via our RL model. The independent variable is $Volatility$, with fixed effects controlling for all other variables. 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 out of the sample stocks. The table reports the number of sample stocks for each column (*No. stocks*)

	All stocks	Constrained	Unconstrained
Mean	0.38	-2.39	5.82
No. +	9	0	5
No. -	11	5	0
R-Square	0.86	0.86	0.87
No. stocks	20	5	5

Table 6
Variable relative importance

For each variable ,k, that partially defines the market state (i.e., *Price level, Queue position, Bid size 1, Bid size 2, 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 profit 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 profit 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. 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 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-squared	0.16	0.24	0.34
	Mean No. obs	625	625	625

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 and offer at t_0 being 13 and 14, respectively. The trader's limit order is in black and starts at the back of the queue at t_0 at price 12.

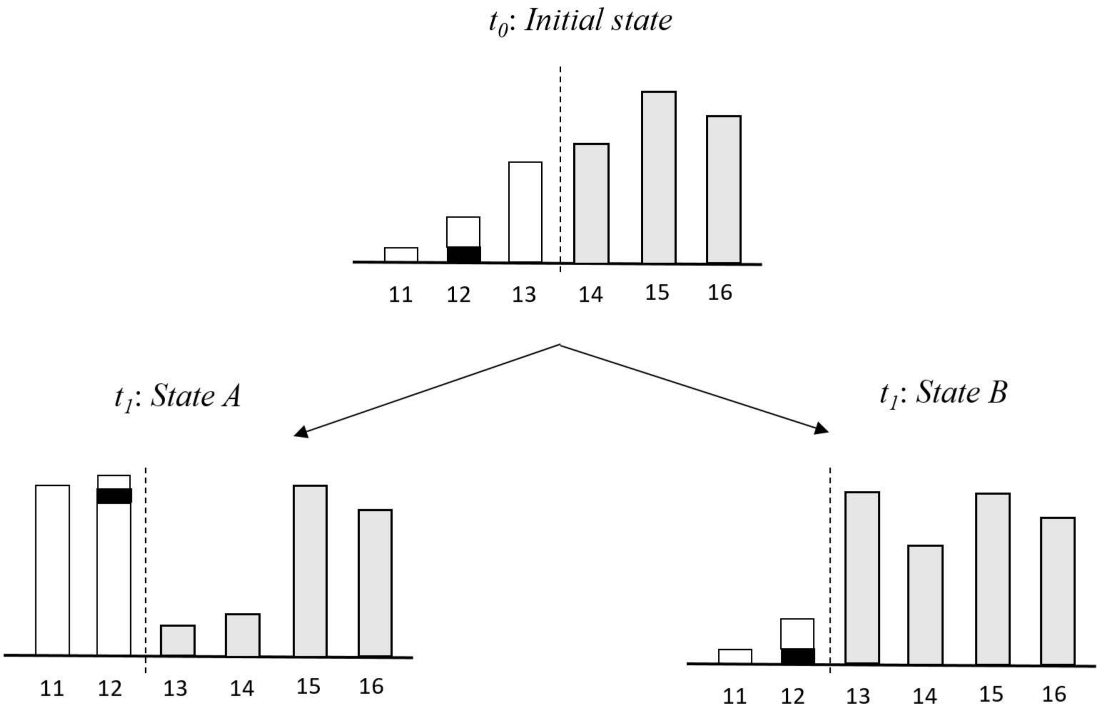
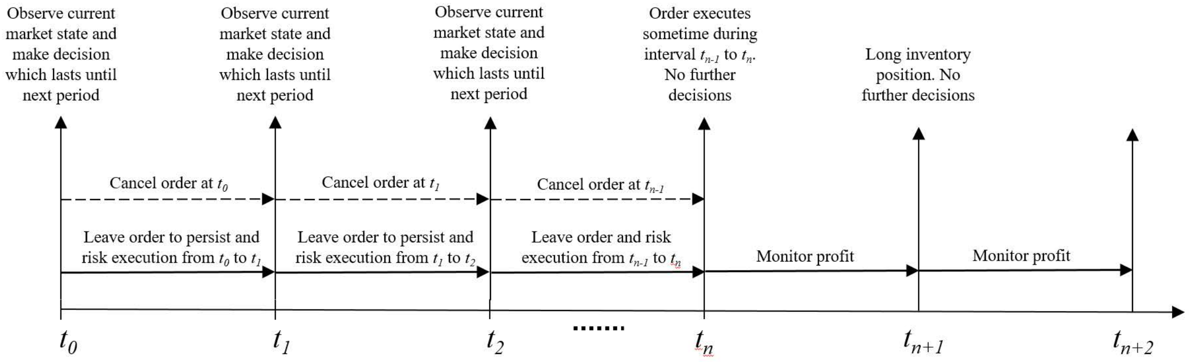


Figure 2. Traders sequential decision making process

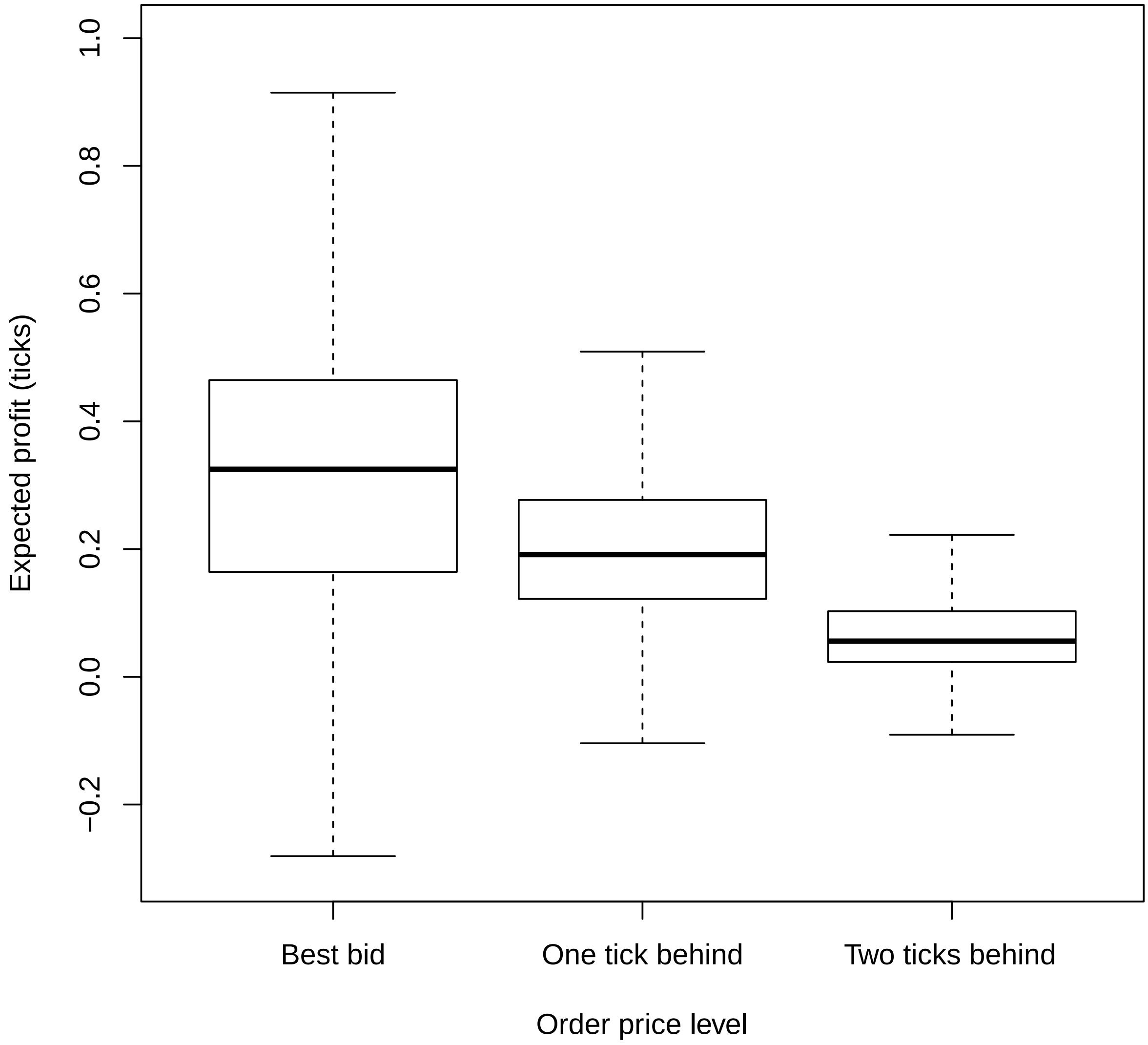
Figure 2 depicts the time line of the trader's decision making process when monitoring their limit order. At the end of each interval, the trader observes current market conditions and decides to leave or cancel their order. This process repeats until the order is either executed or cancelled.



Observe current market state and make decision which lasts until next period
Observe current market state and make decision which lasts until next period
Observe current market state and make decision which lasts until next period
Order executes sometime during interval t_{n-1} to t_n. No further decisions
Long inventory position. No further decisions
Cancel order at t_0
Cancel order at t_1
Cancel order at t_{n-1}
Leave order to persist and risk execution from t_0 to t_1
Leave order to persist and risk execution from t_1 to t_2
Leave order and risk execution from t_{n-1} to t_n
Monitor profit
Monitor profit
t_0
t_1
t_2
....
t_n
t_{n+1}
t_{n+2}

Figure 3. Boxplot of expected profit

This figure plots a boxplot of the expected profit 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.



References

Bertsimas, D. and Lo, A. (1998). Optimal control of execution costs. *Journal of Financial Markets*, 1(1):1–50.

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

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.

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

Easley, D., de Prado, M. L., O'Hara, M., and Zhang, Z. (2019). Microstructure in the machine age. Working paper, Available at SSRN: `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3345183`.

Fong, K. and Liu, W.-M. (2010). Limit order revisions. *Jounal 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.

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.

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.

Hollifield, B., Miller, R. A., and Sandås, P. (2004). Empirical analysis of limit order markets. *The Review of Economic Studies*, 71(4):1027–1063.

Li, S., Wang, X., and Ye, M. (2020). Who provides liqudity 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.

Moritz, B. and Zimmermann, T. (2019). Tree-Based conditional portfolio sorts: The relation between past and future stock returns. Working paper, Available at SSRN: `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2740751`.

Nevmyvaka, Y., Feng, Y., and Kearns, M. (2006). Reinforcement learning for optimized trade execution. *ICML 2006 - Proceedings of the 23rd International Conference on Machine Learning*, pages 673–680.

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.

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.

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. (2014). Queuing uncertainty in limit order market. Working paper, Available at SSRN: `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2336122`.