Strategic trading and manipulation: machine learning in limit order markets*

Xinyi Deng and Xue-Zhong He

University of Technology Sydney, Business School, Australia

Abstract

This paper introduces Q-learning, a novel machine learning technique, as a learning tool to a dynamic limit order market to examine how order book information and learning affect strategic trading behaviour of bounded rational traders. In equilibrium, informed traders unambiguously favour limit (market) orders when the magnitude of mispricing is small (large), while uninformed traders tend to chase market orders from the informed. Interestingly, by anticipating a mispricing reversal when a small-in-size positive (negative) mispricing is accompanied by high depth imbalance at the best bid (ask), informed traders manipulate the market by "deviating" from their predictable trading behaviours. Instead of preferring limit buys (sells) with balanced orders at the best quote, the informed use market buys (sells) to trigger market buys (sells) from the uninformed to enhance their subsequent execution probability and profitability of limit sells (buys). Consequently, the uninformed experience a profit reduction in trend-chasing market orders. The findings provide insight on order book information channels for strategic trading and highlight the impact of order choices, particular for uninformed traders, on market quality.

JEL classification: C63, G14, D44, D82, D83

Keywords: Machine learning, limit order market, asymmetric information, strategic trading, market manipulation, liquidity, price discovery

---

* Address: University of Technology Sydney, PO Box 123 Broadway, NSW 2007, Australia. Fax: +61 295147726
Email addresses: xinyi.deng@student.uts.edu.au (Deng), tony.he1@uts.edu.au (He).

1. Introduction

Limit order markets (LOM) have become the dominant form of financial market organization. In contrast to a quote-driven market, where designated market makers (DMMs) take the other side of the informed and noise trader's market orders, an order-driven market is characterized by "democratized" liquidity provision. Investors, informed or not, can provide immediacy by submitting limit orders, or consume it by using market orders. The central question in LOM is how investors make their order choice decision. For model tractability, existing rational expectations equilibrium (REE) models of LOM have oversimplified uninformed order choices by assuming them as stochastically drawn from distributions (Chakravarty and Holden, 1995) or dependent on exogenous parameters, such as private values (Goettler et al., 2009) and time preference (Rosu, 2020). Hence, these models are stylized enough such that they are limited in generating explanations and testable predictions of differences in informed and uninformed trading behaviours as documented in empirical order aggressiveness literature (see Doung et al., 2009; Chiu et al., 2016). The limitation is nontrivial given that both informed and uninformed order choices are essential to understand aggregate patterns in order flow dynamics (e.g., Ellul et al., 2007) and how new information is impounded into prices (e.g., Anand et al., 2005; Bloomfield et al. 2005).

In this paper, we address this limitation by implementing *Q-learning*, a novel machine learning technique, to fully endogenize informed and uninformed order choices and numerically solve for equilibrium in a dynamic limit order book that is a direct methodological descendant of Chiarella et al. (2015) and He and Lin (2020). With Q-learning, traders form estimates about the expected cumulative payoffs of combinations of the order book state and the order choice, via trial and error using feedback from repeated experiences, and the generated estimates are Q-values. Easing oversimplification of uninformed order choices with Q-learning, we unveil information channels underly strategic trading, which is characterized by predictable trading behaviors of informed and uninformed traders, and even deliberate and profitable "violation" of some predictable patterns by the informed, i.e., informed manipulation.

We also shed some light on market quality consequences of the strategic interaction between informed and uninformed traders.

There are two additional advantages of applying Q-learning in our model. First, it demonstrates the possibility and potential of using reinforcement learning techniques as alternative ways of modelling agents' belief updating in market microstructure studies rather than the traditional Bayesian updating rule. Bayesian updating assumes agents have adequate knowledge of model priors (e.g., the joint distribution for fundamental value and order flow). Reinforcement learning techniques do not make such an assumption and represent one step toward realism. Second, with Q-learning traders who maximize expected cumulative payoffs of current and future periods (i.e., Q-values), our model captures strategic trading in a dynamic order choice problem – a trader not only factors in the impact of future trader on his current order's payoff as in existing dynamic LOB models (e.g., Goettler et al., 2009; Rosu, 2020; Ricco et al., 2020; Chiarella et al., 2015; He and Lin, 2020), but also considers the impact of his current order on future market conditions and further on all his future orders' payoffs. By considering the two impacts simultaneously, Q-learning traders trade strategically to maximize their expected lifetime utility. Our setting is also different from He and Lin (2020) by considering the cost of delayed execution.

The model delivers three main results. First, equilibrium is learnable in a dynamic LOM with information asymmetry. Measured by Q-value criteria and average reward criteria, belief convergence of traders is achieved. The converged model is able to replicate a selection of statistical regularities documented in empirical LOM literature, including hump-shaped mean depth profiles (Bouchaud et al., 2002), absence of autocorrelations of returns (Cont, 2001), and slow decaying autocorrelations of absolute returns (Cont, 2001; Schnaubelt et al, 2018). In equilibrium, informed (uninformed) traders tend to consume (provide) liquidity in aggregate and on average.

Second, in equilibrium, the informed and the uninformed, who share the same learning mechanism, demonstrate systematic differences in trading behaviours conditional on order book information, fundamental volatility, and informed trading level. Compatible with equilibrium strategies in existing REE models of dynamic LOM (see Foucault,1999; Kaniel and Liu, 2006; Rosu, 2020), informed traders are most reliant on the information relevant to fundamental value to determine their trade/no trade, buy/sell and limit/market decisions, have higher trading interests when the mid-price is not equal to the fundamental value, tend to buy (sell) when the mid-price is lower (higher) than the fundamental value, and submit more market (limit) orders when the deviation of mid-price from the fundamental value is large (small), reflecting increased sensitivity to execution risk. Uninformed traders, due to the lack of information advantage, learn to "chase the trend" and are more prone to place buy (especially market buy) orders following a market buy. Since informed and uninformed traders have different order aggressiveness strategies conditional on spread and depth, informed trading is resiliency improving: informed traders submit more aggressive limit (market) orders when spread widens (narrows) and prefer limit (market) orders when same-side cumulative depth is large (small), the uninformed also do, but less significantly.

Increased fundamental volatility increases information advantage for the informed and adverse selection for the uninformed, increases (reduces) market orders and reduces (increases) aggressive limit orders for the informed (uninformed), reducing market liquidity. Increased informed trading, on the one hand, leads informed traders to undercut each other using more aggressive limit but not more market orders, a *competition effect*. On the other hand, it causes more efficient prices, reduces market order profits, a *weakening information effec*t. Since the two effects reinforce each other, increased informed trading reduces market orders and increases aggressive limit orders for the informed, harming informed welfare. It also reduces market orders and aggressive limit orders for the uninformed, improving uninformed welfare. Consequently, increased informed trading improves price efficiency and market liquidity. Our work is complementary to Rosu (2020) by showing that *increased intertemporal competition* between the informed traders can have a liquidity improving effect when they compete in liquidity provision, rather than the liquidity deteriorating effect discussed in Rosu (2020).

Lastly, and perhaps most interestingly, informed traders learn to trade like a deviant, strategically “violate” their predictable trading patterns, and also exploit uninformed traders’ predictable trading patterns. Informed manipulation naturally emerges as a result of the strategic interaction between informed and uninformed traders. Informed traders anticipate a later mispricing reversal when a small-in-size positive (negative) mispricing is accompanied by high depth imbalance at the best bid (ask): given uninformed traders’ preference for market buy (sell) following a market buy (sell), informed traders go against their preference for limit buys (sells), and react strategically by using market buys (sells) to trigger uninformed market buys (sells) and enhance the execution probability and profitability of later informed traders’ limit sells (buys). This strategy of the informed is both manipulative and collusive, because informed traders confuse uninformed traders by taking “wrong” action when facing make-take decisions, and sacrifice their own profit difference between limit and market buys (sells) in exchange for profit increases of later arriving informed traders.

1. 1 Related literature

This paper is most relevant to the literature of order choice problems in static and dynamic limit order markets with information asymmetry (see Parlour and Seppi (2008) and Rosu (2012) for excellent surveys). We simulate a dynamic limit market model, in which informed and uninformed traders optimally choose between trading or not trading, buying or selling, and market or limit orders.

Goettler et al. (2009) state that it is a challenging issue to find an analytic solution for LOB models that incorporate relevant frictions like asymmetric information, discrete prices, and asynchronous arrival of investors. Consequently, existing static and dynamic LOB models with information asymmetry are prone to oversimplify the order choice problems for uninformed traders.

For static models, Chakravarty and Holden (1995) posit that the profit-maximizing informed trader can choose the order size, order type (market/limit and buy/sell) and the limit order price, while draw the uninformed’s order from random distributions. In Kaniel and Liu (2006), the uninformed have a positive probability of becoming impatient and submit only market orders, and the informed and the patient uninformed determine the order aggressiveness optimally. For dynamic models, Goettler et al. (2009) numerically solve a continuous-time game featured by the endogenous cancellation and endogenous information acquisition, in which agents are endowed with positive, negative or zero private values. The uninformed's buy/sell and limit/market decision rely on the private value: uninformed trader with a large positive (large negative) private value is more inclined to submit a market buy (market sell). For Rosu (2020), equilibrium limit order submission behaviours of the uninformed are deterministically determined by exogenously given waiting costs. In these papers, the order choice problem of uninformed traders is either not explicitly modelled or highly dependent on exogenous parameters (private value, patience). Relaxing these possibly unrealistic assumptions, we contribute to this line of research by establishing causality from information asymmetry to differences between the informed trader's and the uninformed trader's limit order submission strategies. He and Lin (2020) also have this merit, but this paper brings depth imbalance at and beyond the best quote into the order choices, generating richer results.

More importantly, relaxing possibly unrealistic assumptions in previous theoretical studies allows us to generate testable predictions and provide explanations for the empirical literature on order aggressiveness by different trader types (see Duong et al. (2009) and Chiu et al. (2016) for institutional and individual traders; see Beber and Caligo (2005) for high PIN and low PIN trading periods). Doung et al. (2009) find that, for large-cap stocks, institutional traders place more aggressive orders under high volatility periods and profit from picking off stale quotes while individual traders trade less aggressively under high volatility periods. Chiu et al. (2016) find that institutional traders demonstrate crowding-out effects and increase limit order submissions when own side liquidity is relatively sparse than the other side, while individual traders increase market order submissions when own side liquidity is relatively sparse than the other side. The two findings cannot be explained by existing theoretical studies and are consistent with our simulation evidence.

This paper is also related to the theoretical and empirical literature on informed trading's impact on liquidity in limit order market - both static dimension of liquidity, which are depth and spread, and the dynamic dimension of liquidity, which is resiliency. For empirical studies, Kempf et al. (2009) examine the electronic limit order market XETRA using market order imbalance to represent the unobservable information variable, and conclude that the impacts of informed trading on both spread resiliency and depth resiliency are strongly negative. Menkoff et al. (2010) study the interdealer forex market for Russian rubles using the trading activity and the trading location of the dealer as two proxies for information, and show that limit order submission rate of the informed is much more positively responsive to a drop in spread, an increase in depth or an increase in cumulative depth than the that of the uninformed. They argue that informed trading is thus resiliency improving. The two empirical studies are constrained by the extent to which their proxies represent the information. Our work is different from them, because our regression analysis on the simulated data has exact identification of the informed traders and the uninformed traders.

For theoretical studies, Rosu (2020) shows that an increase in the fraction of informed traders always improves spread, has no effect on price impact, and improves resiliency. Nevertheless, he provides no further explanations on possible underlying channels that drive the improvement in resiliency. Our work is thus complementary to Rosu (2020) in two ways – First, by showing that the resiliency improving effect is caused by the differences between equilibrium liquidity provision strategies of the informed trader and the uninformed trader, and more fundamentally, driven by information asymmetry; Second, by showing that more intensive intertemporal competition between informed traders can be liquidity improving (in terms of spread and depth), rather than liquidity deteriorating. Ricco et al. (2020) also depict cases when increases in adverse selection lead to narrowing spreads and increased inside depths, but they focus on the impact of increases in asset volatility – an increase in volatility increase adverse selection, render the informed trader's information more valuable, and increase the informed trader's trading interest, which, however, migrates to the aggressive limit orders rather than the market orders.

This paper further contributes to the literature on market manipulation (surveyed in Vives, 2010; Putninš, 2012; 2020). The growing literature categorizes manipulation techniques into three forms: (i) action-based manipulation, which involves taking actions that change the value (or the perceived value) of the asset (Vila, 1989), e.g., a company manager can divest a factory to depress the stock value; (ii) information-based manipulation, which involves spreading misleading information or rumours (Benabou and Laroque, 1992; Van Bommel, 2003; Eren and OZsoylev, 2006); (iii) trade-based manipulation, which involves influencing stock price purely through trading (John and Narayanan, 1997; Brunnermeier, 2000; Huddart et al., 2001; Shino, 2021). Our work ananlyzes the third category of manipulation conducted by the informed. In existing REE literature on informed trade-based manipulation, a trader with long-lived information are incentivized to perform manipulation to sabotage other participants' technical analysis if faced with ex-post trade disclosure requirements (John and Narayanan, 1997; Huddart et al., 2001), or future public announcements about the fundamental (Brunnermeier, 2000). In these paper, an informed trader manipulates to hide or enhance his own information advantage, either by occasional trading in the opposite direction of his information (John and Narayanan,1997; Shino, 2021), or by adding noise components to trades (Huddart et al., 2001), i.e., choosing "wrong" action when facing buy/sell or amount decision. In contrast, in our setting, an informed trader acts collusively and manipulates to enhance the profit of other later arriving informed traders, by choosing "wrong" action when facing limit/market order decision.

Lastly, our work also contributes to the emerging literature of applying machine learning in economics and finance (see Varian (2014), Athey and Imbens (2019) and Athey (2018) for excellent surveys). Recent studies are mainly empirical, and use machine learning to extract unstructured information and construct novel variables like investor sentiment (Renault, 2017) or focus on prediction and apply machine learning to test theories that imply stock return predictability (Easley et al. 2019; Bryzgalova et al., 2019; Gu et al., 2020). Though a simulation study, our work unveils the promising future of applying machine learning to theory, especially the possibility of integrating reinforcement learning with market microstructure theory framework. In particular, using reinforcement learning to update the agent's belief in our dynamic LOB model rather than the traditional Bayesian updating rule enables us to relax a set of strict assumptions like the agent's perfect knowledge of the model's probability structure and avoid using time preference and private value parameters. Fewer parametrizations can help reveal the importance and consequence of information itself, which matters particularly in the context of market microstructure studies.

The rest of the paper proceeds as follows. Section 2 introduces Q-learning and the order book information classifier system into a dynamic limit order market with informed and uninformed traders. Section 3 defines the concepts of the numerical equilibrium and evaluates the model against a selection of stylized facts. Section 4 demonstrates endogenous liquidity provision and investigates the role of order book information in order choices and strategic trading. Section 5 analyzes the informed traders' and uninformed traders' trading behaviours under different volatility regimes and different informed trading levels. Section 6 illustrates informed traders' manipulative behaviours, i.e., strategic and deliberate "violation" of their predicatable trading patterns, and uninformed traders' reaction. Section 7 concludes.

2. The model

We consider a dynamic limit order market model of trading a single risky asset, which is motivated by Goettler et al. (2009), Chiarella et al. (2015), and He and Lin (2020). The innovations to the fundamental value of the asset v_t follow a Poisson process with parameter θ. If an innovation takes place, the fundamental value increases or decreases by Δ ticks with equal probability.

There are N risk-neutral traders who enter the market randomly following a Poisson process at rate λ. Among them, N_I are informed, N_U are uninformed. When arriving the market, informed trader knows the current fundamental value v_t, while uninformed trader only knows the lagged fundamental value $v_{t-\tau}$, where τ is a positive integer. The only difference between informed and uninformed traders is their knowledge of the fundamental value. Different from Goettler et al. (2009) that trader can only trade one share in his lifetime, repeatedly visits the market if no execution, and leaves the market forever after an execution, our Q-learning traders can repeatedly visit the market despite all his prior executions and no executions.

2. 1. Traders' order choices

The set of available actions for any trader is related to his choices about trading or no trading, market or limit order, order direction (buy or sell), and limit order aggressiveness. When re-entering the market, trader cancels his last limit order if unexecuted, and optimally chooses an action that maximizes his expected cumulative payoffs given the trading history H_t, the current state of the limit order book, and his type.

Formally, the types of buy orders submitted can be defined as follows. A market buy order (*mb*) is a request to complete the transaction immediately at the best ask. An extremely aggressive limit buy order (*ealb*) lies within the spread and is posted at $a_t - 1$, a price that is one tick below the best ask a_t. A moderately aggressive limit buy order (*alb*) lies within the spread and is posted at $b_t + 1$, a price that is one tick above the best bid b_t. An ordinary limit buy order (*lb*) is at the best bid b_t. An unaggressive limit buy order (*ulb*) is at $b_t - 1$, one tick below the best bid. The market sell order (*ms*), extremely aggressive limit sell order (*eals*), moderately aggressive limit sell order (*als*), and ordinary limit sell order (*ls*) can be defined analogously. A trader can also choose not to trade (*nt*). In summary, trader's action space contains 11 actions $\mathcal{A} = \{mb, ealb, alb, lb, ulb;\ ms, eals, als, ls, uls;\ nt\}$.

2. 2. Order book information state space

A limit order book $L_t = \{l_t^i\}_{i=1}^{\infty}$ contains the history of order book information at time t, l_t^i, consisting of a backlog of unexecuted limit orders at each discrete price level, p^i, with the standard price-time priorities for limit order execution. To characterize order book information, the continuous state space of the limit order book is discretized as a finite set of states denoted as $\mathcal{C} = \{(s, E(v_t), R, b, a, d^a - d^b, D^a - D^b, LT)^j\}_{j=1}^{J}$, the spread condition s, the expected fundamental value $E(v_t)$, Rosu's signal $R = \frac{|\bar{v}_t - p^m|}{s}$, the current best bid b, best ask a, the depth imbalance measured at the best ask and bid $d^a - d^b$, the cumulative depth imbalance measured at the whole sell-side and buy-side $D^a - D^b$, the last trade direction LT (buyer-initiated or seller-initiated). Note that the expected fundamental value equals to v_t for the informed and equals to $v_{t-\tau}$ for the uninformed. Rosu's signal measures the level of mispricing observed, in which $\bar{v}_t = v_t$ for the informed and $\bar{v}_t = \frac{1}{N}\sum_{t'=t-(L-1)}^{t} p_{t'}$, a moving average price of the past L periods, for the uninformed.

For computational tractability, state variables are further discretized using the classifier system developed in Chiarella et al. (2015) and He and Lin (2020). Appendix A1 reports the classified rules (CRs) of the classifier system. A feasible state a trader may encounter is a vector of the values of state variables. For instance, a possible value of s_t is "30", where "3" denotes that the current spread is larger than 2, and "0" means the order book is not empty; a possible value of $\bar{v}_t$ is "1" denotes the expected fundamental value is higher than the mid-price; a possible value of Rosu's signal $R_t = \frac{|\bar{v}_t - p^m|}{s}$ is "1" denotes that the mispricing measure is between 0 and 0.5; a possible value of b_t (or a_t) is "0" denotes that the current bid (or ask) equals to the last bid (or ask); a possible value of $d_t^a - d_t^b$ (or $D_t^a - D_t^b$) is "1", where 1 denotes that the depth at the best ask (or at the whole sell-side) is larger than the best bid (or at the whole buy-side); a possible value of LT is "1", where 1 denotes that the last trade is buyer-initiated; the resulting state vector is thus "301100111". After applying the classifier rules, the final discretized state space is composed of 14580 (6×3×5×3×3×3×3×2) feasible states.

For dynamical strategic trading in LOM with such large number of state variables, it becomes extremely challenging to have an analytic solution. In this paper, we solve the trading game numerically by formalizing the limit order book's evolution through time as a Markov Decision Process (MDP) and applying Q-learning algorithm (Watkins, 1989), a reinforcement learning technique, to obtain the traders' equilibrium beliefs about the expected cumulative payoffs of possible state-action combinations.

2.3. Trader objective and Q-learning

The Markov Decision Process of the limit order book is characterized by a 4-tuple $(\mathcal{C}, \mathcal{A}, \mathbb{P}(c'|c,a), \rho(c,a))$. As defined before, $\mathcal{C}$ is the set of feasible states of the order book, and $\mathcal{A}$ is a trader's action space. $\mathbb{P}(c'|c,a)$ is an unobserved Markov probability transition matrix, which determines new order book state $c' \in \mathcal{C}$ at the trader's next entry given his current action $a \in \mathcal{A}$ and the current order book state $c \in \mathcal{C}$. $\rho(c,a)$ is an underlying reward rate that is constant for each state-action combination. $\rho(c,a)$ and the time it takes to transfer from c to c' (i.e., the time elapsed since the trader's current entry and until his next entry) jointly determine the reward r observed by the trader when he re-enters the market.

By viewing the order book's evolution through time as a MDP, we can formulate trader's dynamic optimization problem. Each trader has a type (I, β), where I represents his information type (informed or not). As in Goettler (2009), β is a discount rate that reflects the cost of market monitoring until a limit order's execution and the opportunity costs like delaying trades in other assets if the trader is implementing a portfolio strategy. The reward r for a trader whose buy order executed at time t before his re-entry is the difference between the fundamental value v_t and the price he paid, or the order profit. The reward r for an executed sell order can be defined analogously. The reward associated with no execution or no order placement is 0. Formally, when re-entering the market, conditional on his last order placement decision, an trader receives:

$$r = \begin{cases} v_t - p_t & \textit{if he executes a buy order at } t \\ p_t - v_t & \textit{if he executes a sell order at } t \\ 0 & \textit{if no execution or no submission} \end{cases} \quad (1)$$

Let $Q_\pi(c,a)$ represent the discounted cumulative rewards that can be expected (i.e., expected cumulative payoffs) if action a is taken in state c, and a given strategy π is followed thereafter:

$$Q_\pi(c,a) = \sum_{c' \in \mathcal{C}} \Pr(c'|c,a) \int_0^\infty \int_0^{\Delta t} e^{-\beta s} \rho(c,a)\, ds dF(\Delta t) + \sum_{c' \in \mathcal{C}} \Pr(c'|c,a) \int_0^\infty e^{-\beta \Delta t} Q_\pi(c', a') dF(\Delta t) \quad (2)$$

where $Pr(c'|c,a)$ is an element of the transition matrix $\mathbb{P}$, and Δt is the random time between the trader's two consecutive entries, of which the CDF is denoted as $F(\Delta t)$. Since both the informed and uninformed enter the market randomly following a Poisson process at a rate λ, Δt follows an exponential distribution with mean $1/\lambda$. For the optimal strategy π^*, we have:

$$Q^*(c,a) = \max_{\pi}\left(\sum_{c'\in\mathcal{C}} \Pr(c'|c,a) \int_0^{\infty}\int_0^{\Delta t} e^{-\beta s}\rho(c,a)\, ds dF(\Delta t) + \sum_{c'\in\mathcal{C}} \Pr(c'|c,a) \int_0^{\infty} e^{-\beta\Delta t} Q_{\pi}(c',a') dF(\Delta t)\right)$$

$$= \sum_{c'\in\mathcal{C}} \Pr(c'|c,a) \int_0^{\infty}\int_0^{\Delta t} e^{-\beta s}\rho(c,a)\, ds dF(\Delta t) + \sum_{c'\in\mathcal{C}} \Pr(c'|c,a) \int_0^{\infty} e^{-\beta\Delta t} \max_{a'\in\mathcal{A}} Q^*(c',a') dF(\Delta t) \qquad (3)$$

$$= \mathbb{E}\left(\int_0^{\infty}\int_0^{\Delta t} e^{-\beta s}\rho(c,a)\, ds dF(\Delta t) + \int_0^{\infty} e^{-\beta\Delta t} \max_{a'\in\mathcal{A}} Q^*(c',a') dF(\Delta t)\middle| C_t = c, A_t = a\right) \qquad (4)$$

Eq.(3) and Eq.(4) are two forms of Bellman optimality equations for the trader's continuous time dynamic optimization problem. If Markov probability transition matrix $\mathbb{P}$ and the underlying reward rate $\rho(c,a)$ were known, the value function $Q^*(c,a)$ can be explicitly solved from Eq.(3) or Eq.(4) for any state-action combination (c,a), as there is the same number of equations and unknowns. Without such perfect knowledge, both informed traders and uninformed trader are thus assumed to learn their equilibrium beliefs about expected cumulative payoffs and optimal strategies conditional on their information sets using Q-learning, a model-free reinforcement learning algorithm.

Based on Eq.(4), the Q-learning iteration procedure includes the following four steps. (i) Arbitrarily initialize the value function Q^* for every state, every action. (ii) After observing the current state c, a trader chooses the best action a given his current belief with probability $1-\varepsilon$, and chooses each inferior action with probability $\varepsilon/10$ (there are 11 feasible actions). The trader trembles to avoid local optima.(iii) Upon next entry, the trader observes reward r and the new state $c' \in \mathcal{C}$, and updates his belief for the value function following the rule depicted by Eq.(5), in which Δt is the realized value of time between the two entries. $\frac{1-e^{-\beta\Delta t}}{\beta}\rho(c,a)$ equals to the observed reward r. k is the number of times action a has been chosen in state c by traders who have the same type as him, and α is the learning rate. (iv) Repeat the previous two steps until traders' beliefs converge.

$$Q^{(k+1)}(c,a) = Q^{(k)}(c,a) + \alpha\left(\frac{1 - e^{-\beta\Delta t}}{\beta}\rho(c,a) + e^{-\beta\Delta t}\max_{a'\in\mathcal{A}} Q^{(k)}(c',a') - Q^{(k)}(c,a)\right) \tag{5}$$

3. Equilibrium

This section illustrates the equilibrium concept in our trading game, characterized by Q-value convergence and average reward convergence criteria, using a benchmark parametrization. The benchmark equilibrium results allow us to investigate informed and uninformed traders' liquidity provision strategies conditional on various order book states in later sections. In equilibrium, informed traders have motivations to trade on mispricing signals. The converged model demonstrates statistical properties consistent with empirical data including hump shaped mean depth profiles, absence of autocorrelations of returns, and slow decaying autocorrelations of absolute returns.

3. 1. A benchmark

For illustrative purpose, we regard each trading period as 1 minute and 360 trading periods as a 6-hour trading day, and set benchmark parameter values as follows. We choose the total number of traders populated in the market to be N = 1000, with N_I = 150 are informed and N_U = 850 are uninformed. Informed and uninformed traders do not differ in trading speed, with the same returning rate of 1/60 and make an order choice 6 times every trading day. The uninformed's information lag τ = 180, which corresponds to half a trading day. We set the tick size to 1, and the initial fundamental value to v_0 = 5000 ticks (say, i.e., \$50). Expected time between innovations about fundamental value is set to 1 minute, i.e., θ = 1. After an innovation occurs, the fundamental value will either go up or down by Δ = 4 ticks with equal probability.

As for Q-learning belief updating, the learning rate is set to $max(0.0003, \frac{1}{n+1})$ for all traders, where n represents the number of trading rounds (the "trading round" concept here is equivalent to the "training episode" concept in machine learning literature, each round consists of 360,000 trading periods). All traders, despite their information types, have the same continuous discount rate of β = 0.05 and tremble rate of ε = 0.01.

3. 2. Convergence criteria

We consider that a numerical equilibrium is reached if both the convergence of Q-values and the convergence of traders' average rewards (order profits) are satisfied. Intuitively, the convergence of traders' estimates of expected cumulative payoffs (Q-values) mirrors the fixed point problem in REE models. Moreover, the convergence of traders' average rewards is a more stringent criterion than the convergence of Q-values given the dynamic feedback mechanism between trading behaviours and limit order book: It requires that both the informed's and uninformed's strategies stabilize, and it also requires that the order book settles into equilibrium states, such that the traders can obtain equilibrium rewards from each interaction with the order book. Formally, the equilibrium concept and the two corresponding in-sample convergence criteria are defined as follows.

Definition 1 *An numerical equilibrium of the limit order market under Q-learning, characterized by informed and uinformed traders' beliefs about their own cumulative payoffs and corresponding strategies, is considered to be reached after the n-th trading round for sufficiently large n when the following two criteria are satisfied:*

- *Q-value criteria: The correlation of Q-values of active trading strategies between the n-th and the (n+1)-th rounds reaches 0.999;*
- *Average reward criteria: The correlation of traders' average rewards between the n-th and the (n+1)-th rounds reaches 0.999.*

We simulate the model under the benchmark parametrization for 100 rounds and check for the two convergence criteria along with the training. The convergence results are reported in Figure 1.

[Figure 1]

In Figure 1, the horizontal axis represents the number of trading round n, the vertical axis represents the correlation coefficient of Q-values (blue solid line) and the correlation coefficient of average rewards (green dashed line) over two adjacent trading rounds. Panel A shows the convergence of the informed's learning. Panel B shows the convergence of the uninformed's learning. On Q-value criteria, the uninformed converges to their equilibrium beliefs at round 25 with a correlation coefficient of 99.92%, while the informed converges to their equilibrium beliefs at round 41 with a correlation coefficient of 99.91%. Uninformed traders learn faster because there are more of them in the trading crowd (850 out of 1000). As expected, the average reward criteria is satisfied at round 60, which are 19 rounds later than when all traders have formed their optimal strategies.

3. 3. Order book statics and stylized facts

After the benchmark model reaches the equilibrium, we fix traders' Q values, disallow the tremble and simulate for another 1000 trading days (3,600,000 trading periods). We report overall frequencies of all order book state variables in Table 1, except for $E(v_t) - p_t^m$ and Rosu's signal, which suggests the order books on the buy side and the sell side are symmetric and on average quite balanced over the whole 1000 trading-day simulation.

[Table 1]

Since informed and uninformed traders mainly differ in observations of $E(v_t) - p_t^m$ and Rosu's signal due to information asymmetry in terms of fundamental value, we report frequencies of the two state variables by trader types separately in Table 2.

[Table 2]

The informed 's observations on $E(v_t) - p_t^m$ justifies the usage of mid-price as a proxy for contemporal fundamental value in the existing empirical literature. For him, $E(v_t)$ equals to v_t, and he is observing $v_t - p_t^m > 0$ and $v_t - p_t^m < 0$ of roughly the same probability, i.e., 49.57% and 49.66%, meaning that the current mid-price is very close to the current fundamental value. On the other hand, for the uninformed, $E(v_t)$ equals to v_{t-180}, and he is observing $v_{t-180} - p_t^m > 0$ and $v_{t-180} - p_t^m < 0$ of relatively different probability, i.e., 48.96 and 49.79, meaning that the current mid-price is relatively far from the lagged fundamental value 180 periods ago.

The relatively large differences between the uninformed trader's and the informed trader's observation frequencies of Rosu's signal indicate that Rosu's signal might be an inaccurate mispricing signal for the uninformed trader (since he is estimating $\bar{v}_t$ using moving average past prices, while for the informed $\bar{v}_t = v_t$).

We now evaluate our model based on a selection of stylized facts documented in empirical LOM literature, including (i) hump-shaped order book (ii) absence of autocorrelations of returns (iii) slow decaying autocorrelations of absolute returns.

During 360,000 trading periods simulated using the converged benchmark model, we record the order book every 100 trading periods. The average order book shape is then calculated using the mean depths of the 3600 snapshots. We record the price series period by period and calculate the log returns accordingly. The resulting average order book shape, the autocorrelations of returns, and the autocorrelations of absolute returns are depicted in Figures 2-4.

[Figure 2]

[Figure 3]

[Figure 4]

As shown in Figure 2, our order book has a "hump" located at one tick away from the best bid (ask), with average depth increases from the best bid (ask) to the second-best bid (ask), then subsequently decreases. In particular, the average depth at the best bid (ask) is 4.38 (4.44), at second best bid (ask) is 5.09 (4.93). Figure 3 suggests that autocorrelations (ACs) of returns are only negative for the initial 3 lags, which reflects the bid-ask bounce, and then quickly approach 0. Figure 4 indicates that the ACs of absolute returns is slowly decaying from 0.26 to around 0.05 even with 30 lags, which reflect long memory. All three stylized effects are reproduced in the paper.

4 Order choices and order book states

In this section, we first characterize the equilibrium liquidity provision/consumption roles of traders using unconditional probabilities for each order type, and then show how order book information impact uninformed and informed traders' trading behaviours (especially limit order submissions) differently, and hence the consequent implication on market quality. In our model, informed (uninformed) traders place more (less) aggressive quotes out of their liquidity supply when same side best-quote depth is large, consistent with empirical findings of Aitken et al.(2007) and Chiu et al. (2016) about institutional and individual traders order aggressiveness strategies, and haven't been depicted by theoretical studies yet.

Panels A and B in Table 3 report the unconditional and conditional probabilities of trader's order choices based on order book states. Since the buy side and the sell side are quite symmetric, we only report on the buy-side.

[Table 3]

4. 1. Liquidity consumption and provision

On liquidity provision and consumption, based on unconditional probabilities reported in the last rows of the two panels in Table 3, we can see that informed traders are submitting more market orders (24.41%) than limit orders (9.71%), and uninformed traders less market orders (7.84%) than limit orders (14.40%). Also, the fraction of market (limit) orders informed traders contributing to all orders submitted by all trader types is 15.24% (6.06%), and the fraction of market (limit) orders uninformed traders contributing to all orders submitted by all trader types is 27.74% (50.95%). On average, informed traders mainly consume liquidity, while uninformed traders mainly provide liquidity. Nevertheless, informed traders can switch to endogenous liquidity provision when the spread is large or the mispricing signal is large.

4. 2. Trade/no-trade decision

The unconditional probability of *nt* for informed traders is 14.71%, and that of uninformed traders is 28.14%, meaning that uninformed traders choose to not trade for more than half of the time due to their information disadvantage. Our results indicate that $E(v_t) - p_t^m$, $\frac{|\bar{v}_t - p^m|}{s}$ are more important for informed traders to decide between trading and not trading than for uninformed traders, while uninformed traders are more reliant on bid trend and ask trend than informed traders do. Additionally, the impacts of $d_t^a - d_t^b$ on informed traders' and uninformed traders' trading interests are of opposite direction (it motivates informed traders but discourages uninformed traders to trade).

We now elaborate on informed and uninformed traders' differences in trade/no-trade decisions conditional on order book information using Table 3. When $E(v_t) - p_t^m$ changes from 0 to greater than (smaller than) 0, the probability of *nt* for informed traders drops from 46.60% to 15.65% (13.27%), while the probability of *nt* for uninformed traders does not decrease as much (from 47.92% to 28.19% or 27.60%).

When $\frac{|\bar{v}_t - p^m|}{s}$ increases from (0, 0.5] to (3.5, + ∞), the probability of *nt* for the informed monotonically decreases from 47.21% to 6.24%, while the probability of *nt* for uninformed traders first decreases from 28.81% to 25.84% then increases back to 29.73%. This is very intuitive because uninformed traders do not profit as informed traders do from observed fundamental directional movements or mispricing enlarging due to inaccurate observations.

When $d_t^a - d_t^b$ changes from 0 to greater than (smaller than) 0, the probability of *nt* for informed traders drops from 15.95% to 14.52% (14.14%), while the the probability of *nt* for uninformed traders increases from 27.84% to 28.01% (28.47%). A possible explanation is that they have a better ability than the uninformed to profit from picking off stale quotes during sudden directional changes of fundamental value. More specifically, $d_t^a - d_t^b > 0$ is very likely to be observed when the fundamental value's long-run decreasing trend ceases and suddenly starts to rise, during which times, an informed traders' most possible action type is a market buy (Prob($mb|d_t^a - d_t^b > 0$ *all trading oppurtunies of informed*) = 31.17%), while the uninformed traders are most likely to sell. Additionally, later sections suggest that Prob(*mb*|*informed buyer*) is smaller during $d_t^a - d_t^b > 0$ times than $d_t^a - d_t^b \leq 0$ times, which should be another force at work, i.e., the reduced execution risk, and should be compatible with findings here.

As for bid trend and ask trend, uninformed is more reliant on the two to profit from trade since the lack of correct directional information about the fundamental value, their trading interests unambiguously increase with any directional movements in bid trend or ask trend, while the same is not true for the informed traders.

4. 3. Buy/sell decision

Our results indicate that $E(v_t) - p_t^m$ is most important for informed traders' buy/sell decisions, while uninformed traders' buy/sell decisions are more reliant on bid trend and ask trend than that of informed traders. Additionally, the impacts of $d_t^a - d_t^b$ on informed traders' and uninformed traders' buy/sell decisions are of opposite directions. To elaborate on informed and uninformed traders' differences in buy/sell decisions conditional on order book information, we calculate conditional probabilities of traders' buy orders out of all orders under feasible states, and results are reported in table 4.

[Table 4]

Table 4 indicates that informed traders learn almost perfectly to exploit their private information advantage, and buy with a probability of 99.28% when the asset is undervalued ($E(v_t) - p_t^m > 0$).

The uninformed traders learn to chase the trend due to their lack of private information. They are more reliant on the directional movements in Bid trend and Ask trend than the informed traders do: the informed trader has almost equal probabilities of buying and selling when the order book is moving up (Bid trend > 0: Prob(buy|informed and trade) = 50.05%; Ask trend > 0: Prob(buy|informed and trade) = 50.89%), while the uninformed traders are inclined to buy when the order book is moving up, sell when the order book is moving down.

Interestingly, informed traders' probability of buying equals to 64.84% when the depth at the best ask is large, combined with the fact that they are buying using market orders as suggested by Table 3, as discussed before, we argue that they are able to infer from the imbalance at the inside bid/ask about when there are stale quotes standing at the best bid/ask. This, again, is because they have the correct directional information about the fundamental value. Uninformed traders are not able to infer stale quotes from the imbalance at the inside bid/ask, in fact, they have to infer directional information about the fundamental value from $d_t^a - d_t^b$ due to the lack of private information advantage.

4. 4. Market/limit order decision

Our results indicate that market/limit order decisions of the informed and the uninformed have different responses to spread, Rosu's signal and depth at inside quote level and the cumulative level. A possible underlying mechanism should be information asymmetry. To elaborate on this, we calculate conditional probabilities of limit orders out of all orders under the four discussed states, and the results are reported in table 5.

[Table 5]

In equilibrium, both informed and uninformed traders consume liquidity when it's ample and supply liquidity when it's scarce. When spread increases from = 1 to > 2, the informed trader's limit order submission rate monotonically increases from 16.04% to 52.50%. In other words, endogenous informed liquidity provision emerges when spread surpasses 2. As spread increases from = 1 to > 2, the uninformed trader's limit order submission rate monotonically increases from 57.55% to 65.49%. Both of them are more willing to submit limit orders because the price of immediacy is high, but the response of uninformed traders is not as strong, since for them there is another force of opposite direction at work – the spread widening could reflecting the increase in adverse selection risk, or the intensification of information disadvantage.

In terms of Rosu's signal, the Q-learning informed trader learns to play a threshold strategy as in the REE model of Rosu (2020). His limit (market) order submission rate monotonically decreases (increases) from 94.85% (5.15%) to 24.96% (75.01%) when observed mispricing increases, reflecting increased sensitivity to execution risk. In particular, in Rosu (2020) the informed trader would deterministically and optimally submit a buy market order if he observes mispricing above a threshold, and deterministically and optimally submit a buy limit order when he observes mispricing below the threshold (but positive). In our model, we also detect a "threshold", when $\frac{|\bar{v}_t - p^m|}{s}$ moves below the (1.5, 2.5] region, the informed trader's limit order submission rate surpasses 50% and reaches 62.98%, and endogenous informed liquidity provision emerges. On the other hand, when $\frac{|\bar{v}_t - p^m|}{s}$ moves beyond the (1.5, 2.5] region, the informed trader's market order submission rate surpasses 50% and reaches 58.71%, and he favors market order. The uninformed trader fails to learn such a threshold strategy. When the mispricing signal increases from (0, 0.5] to (3.5, +∞), limit order submission rate decreases at first from 69.18% to 56.62% then increases back to 64.59%. Due to his information disadvantage, his observed mispricing is inaccurate, and he won't treat it as an equivalent to increases in the implicit cost of non-execution.

Though informed traders submit more (less) market orders when same side depths at inside quote level and the cumulative level are large (small), uninformed traders only submit more (less) market orders only when same side depth at the cumulative level is large (small). When $d_t^a - d_t^b < 0$, the informed buyer's limit order submission rate is 22.50%, lower than the $d_t^a - d_t^b = 0$ (32.26%) and the $d_t^a - d_t^b > 0$ (26.84%) cases; When $D_t^a - D_t^b < 0$, the informed buyer's limit order submission rate is 27.60%, lower than the $D_t^a - D_t^b = 0$ (29.23%) and the $D_t^a - D_t^b > 0$ (44.79%) cases; As for the uninformed trader, when the depth at whole sell side is smaller than the depth at whole buy side ($d_t^a - d_t^b < 0$), his limit order submission rate is 64.66%, lower than the $d_t^a - d_t^b = 0$ case (65.51%), but higher than the $d_t^a - d_t^b > 0$ case (43.59%. Nevertheless, negligible because the $D_t^a - D_t^b$ has a very low occurrence frequency of 1.13%). We argue that the two reasons why informed and uninformed traders' market/limit order decisions and limit order aggressiveness have different responses to depth at the inside level are: (i) private information moves the informed trader's trade-off between price risk, execution risk, and adverse selection risk more toward the execution risk side than the uninformed trader since he has a higher implicit cost of non-execution; (ii) the informed trader have correct directional information about the fundamental value, so he can learn better about execution probability than the uninformed trader do.

Further, as shown in Table 6, when same side depth at inside quote level is large, informed traders are increasing their ALO usage (both *ealb* and *alb*) out of all limit orders, while uninformed traders are decreasing their ALO usage out of all limit orders. This reflects that uninformed traders can interpret large same side inside quote depth as signals for future adverse movements in fundamental values, and is consistent with

Aitken et al.(2007) and Chiu et al. (2016)'s empirical findings of institutional (possibly informed) and individual traders' (possibly uninformed) order aggressiveness strategies.

[Table 6]

4. 5. Liquidity provision and information asymmetry

We now further validate the different responses of informed traders' and uninformed traders' market/limit order decisions react to order book information differently using Logistic regression, demeaned Logistic regression with intercept, and demeaned OLS regression with intercept. The results are consistent with the last subsection and suggest that informed traders are resiliency improving. The regressions are inspired by Menkoff et al.(2010), and are mainly different from them in the sense that our simulated data has exact identification of informed traders and uninformed traders while they use dealer trading activity and dealer trading location as proxies for information.

The regressions are conducted for the informed traders and uninformed traders, respectively, using the 360,000 trading period simulated data generated by the converged benchmark model. The dependent variable is a dummy variable that takes 1 when the order is a limit order and takes 0 when the order is a market order. The independent variables are the 8 discretized order book state variables. The demeaning is carried out on the 8 state variables by subtracting their means. Similar to Menkoff et al. (2010), we make directional adjustments on the Depth imbalance variable such that it takes 1 when $d_t^b - d_t^a > 0$ (< 0) is faced by buyer (seller), it takes 0 when $d_t^b - d_t^a = 0$ is faced by buyer(seller), and it takes -1 when $d_t^b - d_t^a < 0$ (> 0) is faced by buyer (seller). In other words, its increases reflect the increases in the relative magnitudes of the depth of a trader's own side compared to the other side. The Cumulative depth imbalance variable is adjusted similarly.

We also adjust the Bid trend variable such that it takes 1 when $b_t - b_{t-1} > 0$ (< 0) is faced by a buyer (seller), takes 0 when $b_t - b_{t-1} = 0$ is faced by a buyer(seller), and takes -1 when $b_t - b_{t-1} < 0$ (> 0) is faced by a buyer (seller). In other words, its increases represent that the order book is moving up for the buyer and moving down for the seller. The Ask trend variable is adjusted similarly.

The $E(v_t) - p_t^m$ variable (Expected fundamental) is adjusted such that the adjusted variable takes 1 when $E(v_t) - p_t^m = \pm 1$, and takes 0 when $E(v_t) - p_t^m = 0$. In other words, it is adjusted such that it reflects whether there is any directional movements in fundamental value (0 = no movement; 1 = with movement). The direction of the Last trade direction variable is adjusted such that it takes 1 when a buyer (seller) observes a market buy(sell), and it takes -1 when a buyer (seller) observes a market sell (buy). The regression results are reported in Table 7.

[Table 7]

The first two columns are results from the Logistic regression. The third and the fourth columns are results from demeaned Logistic regression with intercept, and the last two columns are results from the demeaned OLS regression with intercept. We focus on the Logistic regressions without intercept, since the demeaned regressions deliver quite consistent results.

The Spread variable's coefficient for the informed (uninformed) is 0.5509 (0.4087), reflecting that both the informed traders and uninformed trader will shift to limit orders in response to higher costs of immediacy. The smaller magnitude of the Spread variable's coefficient for the informed is driven by the fact that the lack of private information moves the uninformed trader's trade-off between price risk, execution risk, and adverse selection risk more toward the adverse selection risk side: They would have a tendency to interpret rises in spreads as intensifications of the adverse selection risk.

The Expected fundamental variable's coefficient for the informed is -1.0587, and the Expected fundamental variable's coefficient for the uninformed is 0.5092, reflecting that, when the informed (uninformed) observe any directional movements in the fundamental value (lagged fundamental value), they are inclined to profit by using market orders (unaggressive limit orders).

The Bid trend variable's coefficient for the informed (uninformed) is -0.1329 (-0.0242), and the Ask trend variable's coefficient for the informed (uninformed) is -0.1311(-0.0277). When informed and uninformed buyers observe that an order book is moving up, they tend to use market orders to improve their execution probability. The

Depth variable's coefficient for the informed (uninformed) is -0.2861 (0.1221), and the Cumulative depth variable's coefficient for the informed (uninformed) is -0.0456 (-0.0315). For all these four variables, the informed trader's responses are more negative than the uninformed because the information advantage renders the informed traders more sensitive to execution risk than the uninformed traders.

The Last trade direction variable's coefficient for the informed (uninformed) is -0.1994 (-0.1311), reflecting "diagonal effect", a well-documented stylized fact, whereby a market buy (sell) is more likely to be followed by a market buy (sell). For the Last trade direction variable, the uninformed trader's response is less negative because when an uninformed seller observes a market buy, he has a tendency to interpret it as a rise in the fundamental only known to the informed, i.e., an intensification of adverse selection risk, which discourages him from using a limit sell.

Most importantly, the informed traders are resiliency improving according to the coefficients of Spread, Depth imbalance, and Cumulative depth imbalance: when spread enlarges or same side cumulative depth decreases, both informed and uninformed increase their limit order usage, but informed traders demonstrate much stronger responses. Additionally, when the same side depth at best quote decreases, the informed increases limit order usage, but the uninformed increases market order usage.

5. Fundamental volatility and informed trading

We now show how the informed traders and the uninformed trader differ in their limit order submission strategies when responding to the changes in fundamental volatility and the proportion of informed traders, and the ensuing impact on market quality.

5. 1. Limit order submission and volatility

We leave all the parameters in the benchmark parametrization unchanged and only alter the volatility level. The volatility regimes are $\delta = \{2, 4, 6, 8\}$. We apply the same convergence criteria as defined in Section 3.2, and then fix Q-values, disallow the tremble, and simulate for another 3600000 trading periods.

[Table 8]

Table 8 reports the limit order submission strategies conditional on volatility changes. As discussed by Copeland and Galai (1983), limit order traders give market order traders a timing option. When the volatility increases, the value of the option increases, and spread, in general, should enlarge, and the cost of immediacy increases. Consequently, the uninformed trader submits less MO (because it's more expensive) and submits more ALO: When the volatility of fundamental value increases from 2 to 8, the uninformed trader's usage of MO monotonically decreases from 34.41% to 32.65%, so he is increasing his limit order submission rate in general. In particular, his usage of ALO monotonically increases from 18.64% to 23.33%. Though the increased adverse selection imposed by informed traders should induce him to reduce the usage of ALO, the adverse selection effect is outweighed by the cost of immediacy effect.

Now, as volatility increases, the private information the informed trader posses again moves his trade-off between price risk, execution risk, and adverse selection more toward the execution risk side than he used to be when the volatility is at its lowest level. His information is more valuable, and his cost of non-execution is higher. Consequently, the informed trader increases his usage of MO due to increased sensitivity to execution risk, reduces his usage of ALO (both due to increases in MO and to avoid future adverse selection), and does not vary his usage of NLO much.

As the volatility monotonically increases from 2 to 8, Prob (spread > 2) increases from 28.42% to 47.39%. The informed should be more responsible for the spread widening since they increase MO submission and reduce ALO submission. The uninformed, compared to the informed, is liquidity improving because of their increased ALO submission. The finding of volatility increases leading to informed's (uninformed's) increased (reduced) usage of MO and reduced (increased) usage of ALO, is different from He and Lin (2020), in which both informed and uninformed use more ALO, and is also different from Goettler et al. (2009), in which both informed and uninformed traders submit more conservative limit orders.

5. 2. Limit order submission and informed trading

We leave all the parameters in the benchmark parametrization unchanged and only alter the fractions of informed and uninformed traders. $N_I : N_U =${0.100:0.900, 0.125:0.875, 0.150:0.850, 0.200:800}. We apply the same convergence criteria as defined in Section 3.2, fix Q-values, disallow the tremble, and simulate for another 360,000 trading periods. We focus on the case when mispricing is low, i.e., $\frac{|\bar{v}_t - p^m|}{s} \in (0,0.5]$.

[Table 9]

Table 9 reports the limit order submission strategies conditional on informed trading level changes. In the $\frac{|\bar{v}_t - p^m|}{s} \in (0,0.5]$ region, since the informed trader observes the accurate and rather small mispricing information, his sensitivity to execution risk becomes rather low. Therefore he drastically decreases his market order usage compared to other mispricing scenarios and becomes "de facto" market makers. Take the benchmark case as an example, the informed trader's unconditional limit order submission rate (out of all his orders) is 28.74%, while at the low mispricing level scenario his limit order submission rate becomes 94.84%. When informed trading levels increase from 10% to 20% at the low mispricing time, the price discovery improves and decreases from 0.58 to 0.29, the informed undercut each other and compete by using more ALO (increase its usage from 38.00% to 59.97%) but not more MO, i.e., they compete in liquidity provision. For informed traders, there is also a weakening information effect, caused by more informed trading and more efficient prices, which reduces market order profit and augment the ALO usage increase brought by the competition effect. The uninformed traders respond to the informed share increases by submitting less ALO (decrease its usage from 31.14% to 16.08%) to avoid being picked off due to intensified information disadvantage. The spread decreases from 1.47 to 1.36, and depth at the best quote increases from 3.16 to 5.21 because the effect of increased intertemporal competition between informed liquidity providers outweighs the increased adverse selection risk imposed on uninformed traders. The standard intuition that liquidity deteriorates given more adverse selection is violated. In this case, the informed trader is liquidity improving due to their increased usage of ALO.

As for welfare consequences, using the mean Q-value as the welfare measure, increases in informed trader proportion decrease their welfare from 2.10 to 1.78, because they can extract less rent from their information advantage. And the uninformed trader's welfare becomes less negative and increases from -0.21 to -0.06 , because he free rides the price efficiency improvement brought by the intensified informed trader competition.

In summary, in this subsection, we show that increases in informed trading intensify competition among the informed and adverse selection for the uninformed, reducing market orders and increasing (reducing) aggressive limit orders for the informed (uninformed), reducing (improving) social welfare for the informed (uninformed), improving price efficiency and market liquidity. Different from Rosu (2020), in which competition effect leads to larger information decay and larger slippage component of the spread, and hence deteriorating liquidity, we show that intertemporal competition between the informed traders can have a liquidity improving role if informed traders compete by submitting more aggressive limit orders.

6. Manipulative behaviours of informed traders

As previously shown in Table 5, informed traders unambiguously favour limit orders (market orders) when mispricing is low (high). It is thus intriguing to investigate when and why would informed traders, faced with low mispricing, go against their natural tendency of using limit orders and employ market orders instead.

Our analysis suggests that the seemingly irrational market order usage turns out to be partially attributable to informed traders' collusive manipulation. The intuition can be expressed as follows. Taking buy-side as an example, when small positive mispricing is accompanied by high buy-side depth imbalance ($d^a - d^b < 0$) rather than zero depth imbalance($d^a - d^b = 0$), informed traders could anticipate a mispricing reversal in the near future (mispricing direction changes from $v_t > p_t^m$ to $v_t < p_t^m$), which can possibly be caused by upward price changes following buying pressure. Given that the probability of uninformed traders placing market buy following a market buy is 10.15% (shown in Panel B of Table 3), the highest among all types of uninformed orders, informed traders might react by strategically using market buys to trigger uninformed market buys, sacrifice current profit difference between limit and market buys in exchange for enhanced execution probability and profitability of later informed traders' limit sells, and collusively manipulate for the good of informed traders as a group.

To justify this intuition, when the actual mispricing varies from small positive ($v_t > p_t^m$ and $\frac{|v_t - p^m|}{s} \leq 1.5$) to large positive ($v_t > p_t^m$ and $\frac{|v_t - p^m|}{s} > 3.5$), we compare and contrast the following statistics of high buy-side depth imbalance condition with those of zero depth condition, for the informed and uninformed group respectively: (i) the probability of market buy placement at current depth, i.e., $P^{\text{placement}}$(MB|at current depth); (ii) the probability of limit sell placement between same-group market buy at current depth and a subsequent same-group market buy, i.e., $P^{\text{placement}}$(LS|after MB at current depth); (iii) the probability of limit sell execution in the 20 orders interval after observing current depth, i.e., $P^{\text{execution}}$(LS|after current depth); (iv) the profit per trade (PPT) of LS in the 20 orders interval after observing current depth; (v) the profit per order (PPO) of LS between same-group MB at current depth and a subsequent same-group MB; (vi) PPO of uninformed MB between informed MB and a subsequent informed MB; (vii) the probability of uninformed MB placement in the 20 orders interval after observing current depth, i.e., $P^{\text{placement}}$(MB| after current depth). For the uninformed, statistics (vi) and (vii) are unique to them, and their statistics (i)-(v) are presented and discussed in Appendix A1.

Blue (yellow) solid lines on the left-hand side of Figure 5 Panel A report statistics (i)-(v) of informed traders under high buy-side depth imbalance condition (zero depth imbalance condition). Blue (yellow) solid lines on the left-hand side of Figure 5 Panel B report statistics (vi)-(vii) of uninformed traders under high buy-side depth imbalance condition (zero depth imbalance condition). Blue (yellow) dotted lines on the right-hand side of Figure 5 report the difference between these statistics of high buy-side depth imbalance condition and zero depth imbalance condition.

[Figure 5]

We first examine informed trader behaviours shown in Panel A. For $P^{placement}$(MB|at current depth), statistic (i), of informed traders, both blue and yellow solid lines monotonically rise as mispricing enlarges, reflecting the ceteris paribus effect in subsection 4.4 that greater mispricing increases informed traders' implicit cost of non-execution; the blue solid line lies on top of the yellow solid line, somehow reflecting the ceteris paribus effect in subsection 4.4 that deeper depth imbalance at the best bid/ask encourages same-side informed traders' to jump the queue. The difference between statistic (i) under high buy-side depth imbalance and zero depth imbalance conditions monotonically decreases with mispricing from 0.86% to 0.31%. Current informed traders increase MB usage if small positive mispricing is accompanied by high buy-side depth imbalance rather than zero depth imbalance.

For $P^{placement}$(LS| after MB at current depth) and the corresponding PPO, statistics (ii) and (v), of informed traders, both blue and yellow solid lines monotonically decrease as mispricing enlarges. Intuitively, the larger the current positive mispricing is, the more likely it is that limit sells placed after the current period are going to lose. Yellow solid lines are significantly higher than (rather close to) corresponding blue solid lines when mispricing is no larger than 3.5 (greater than 3.5), somehow reflecting the ceteris paribus effect in subsection 4.4 that deeper depth imbalance at the best bid/ask generally tilts informed traders more toward limit rather than market orders. The difference between statistic (ii) under high buy-side depth imbalance and zero depth imbalance monotonically decreases with mispricing from 1.66% to 0.03%. Later informed traders increase LS usage following the current informed MB if small positive mispricing is accompanied by high buy-side depth imbalance rather than zero depth imbalance.

Further, when Rosu's mispricing signal is no larger than 1.5, the difference between statistic(v) (statistic(iv)) under high buy-side depth imbalance and zero depth imbalance reaches the highest of 1.80 ticks (1.20 ticks); and the difference between $P^{execution}$(LS|after current depth), statistic (iii), under high buy-side depth imbalance and zero depth imbalance conditions reaches the highest of 10.80%. According to Panel A, informed traders, faced with low mispricing, despite their natural tendency to use limit orders under low mispricing, indeed appear to use market buys to try to trigger uninformed market buys after observing high buy-side depth imbalance because they anticipate a later mispricing reversal, and thereby increase the execution probability and profitability of later informed traders' limit sells. Though low mispricing level combined with high buy-side depth imbalance are informative about future mispricing reversal for the informed, according to statistics (i)-(v) of the uninformed discussed in Appendix A1, the uninformed do not observe the actual mispricing level and are not able to extract such mispricing reversal related information.

We now examine how uninformed traders respond to informed manipulation by investigating Panel B. In terms of PPO of uninformed MB after an informed MB-statistic (vi), the yellow solid line is above (below) the blue solid line for relatively low (high) actual mispricing, and both of two lines monotonically increase. The high buy-side depth imbalance condition, compared to the zero depth imbalance condition, reinforces uninformed traders' tendency to interpret an informed MB as positive mispricing that will proceed into the future: strong buying pressure means current price is relatively low. An (informed) market buy observing uninformed trader would thus use more market buy order at high buy-side depth imbalance than zero depth imbalance, a *chasing* effect. When mispricing is low and mispricing reversal is more likely, informed traders have a stronger motive to use market buys to mislead uninformed traders, the *chasing* effect thereby results in a higher chance to get fooled and a lower statistic (vi) for high buy-side depth imbalance observing uninformed traders than zero depth imbalance observing peers. When mispricing is high, the misleading motive weakens and informed market buys are more "genuine", the *chasing* effect thereby results in a higher chance to trade in the right direction and a higher statistic (vi) for high buy-side depth imbalance observing uninformed traders than zero depth imbalance observing peers. As for the two blue dotted lines, when mispricing increases, the difference between uninformed MB placement probability after high buy-side depth imbalance and zero depth imbalance monotonically decreases from 0.67% to -0.21%, and the difference between PPO of uninformed MB after an informed MB at high buy-side depth and zero depth imbalance monotonically increases from – 1.50 ticks to 2.80 ticks. This indicates that high depth-imbalance and low mispricing observing informed traders successfully trick uninformed traders into using more market buys than they should have. Further, informed traders' limit sell profit rise is at least partially achieved via reducing trend-chasing uninformed traders' market buy profit.

So far, what we have discussed is the scenario where mispricing direction does change from $v_t > p_t^m$ to $v_t < p_t^m$. But it is worth pointing out that even if small positive mispricing does not result in mispricing reversal, i.e., $v_t > p_t^m$ persists, informed traders, when faced with *high buy-side depth imbalance condition* rather than the *zero-depth imbalance condition*, might still have a stronger incentive to deviate from LB and use MB at the current period, since the priorly discussed *chasing* effect causes an increase in trend-chasing MB usage of the uninformed and pushes up the price, leading to a reduction in uninformed trend-chasing MB profits.

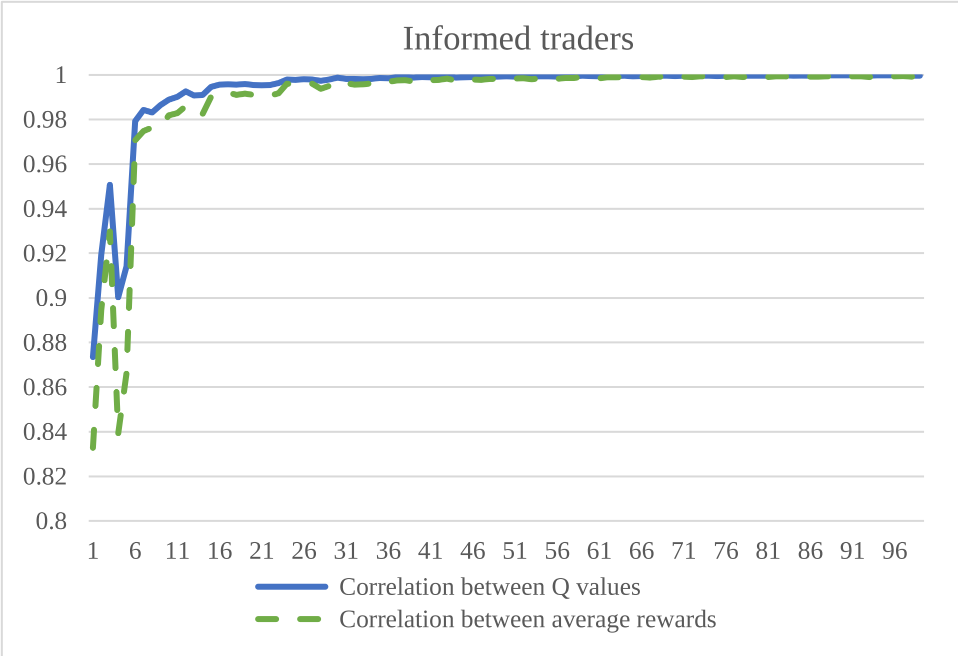
Trade-based manipulation has been widely studied by the rational expectations literature (e.g., Huddart et al., 2001; John and Narayanan,1997; Shino, 2021). In existing REE models, an informed trader endowed with long-lived information and faced with ex-post disclosure requirements seek to garble the information conveyed by his trade by either adding noise components to trades at each period but the last (Huddart et al., 2001) or by occasional trading in the opposite direction of their information (i.e., contrarian trading. John and Narayanan,1997; Shino, 2021). In these models, an informed trader, who wants to preserve its own information advantage longer, trick uninformed traders by choosing "wrong" action when facing buy/sell decision or amount decision. Different from them, we present a novel form of informed manipulation originated from the inherent differences in informed and uninformed make-take strategies, where an informed trader, who acts collusively and is willing to sacrifice his current profit in exchange for profit increases of later arriving informed traders, trick uninformed traders by choosing "wrong" action when facing limit/market decisions.

7. Conclusion

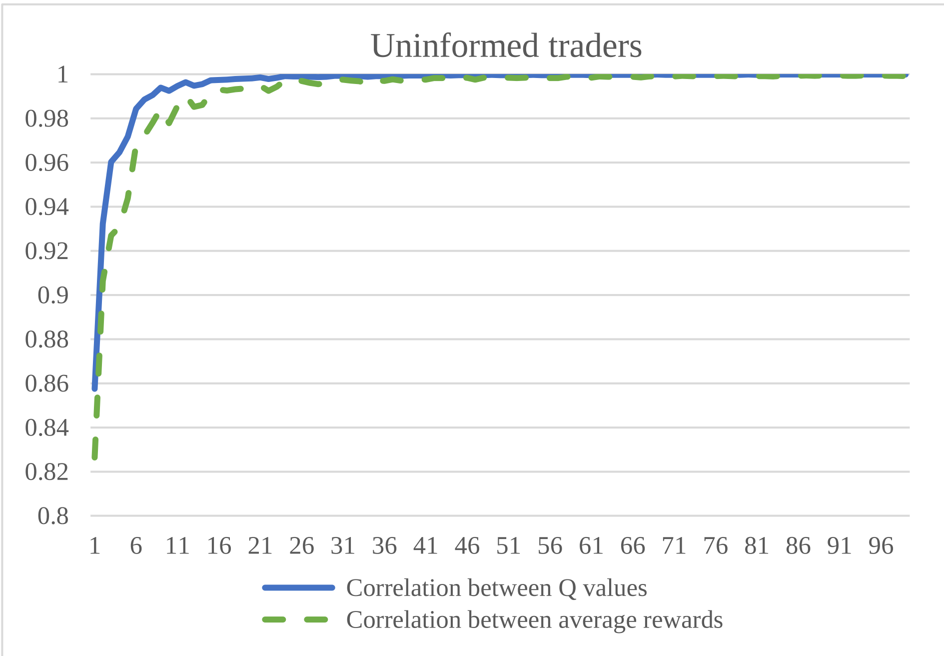
We develop a dynamic limit order book informed and uninformed traders who learn to trade via Q-learning. Q-learning enables us to fully endogenize order choices. In general, it is promising to integrate reinforcement learning with market microstructure theory framework and use it as an alternative belief updating rule because it enables us to relax a set of strict assumptions like the agent's perfect knowledge of model priors.

Trial-and-error learning of bounded rational agents from order book information gives rise to strategic trading, of which a key component is predictable trading behaviours. With information advantage, the informed are most reliant on the sign of mid-price's deviation from fundamental value to determine their trade/no trade, buy/sell decisions. Due to information disadvantage and learning, the uninformed "chase the trend" and are more prone to place buy (especially market buy) orders following a market buy. The informed submit more aggressive limit (market) orders when the spread widens (narrows); the uninformed also do, but less significantly. The informed are thereby resiliency improving. The informed unambiguously favour limit (market) orders when the magnitude of mispricing is small (large), though the uninformed do not. In addition, uring the strategic interaction, both informed and uninformed order choices are indispensable to understand market quality consequences: Increasing informed trading reduces market orders and increases (reduces) aggressive limit orders for the informed (uninformed), reducing (improving) social welfare for the informed (uninformed), improving price efficiency and market liquidity.

Strategic trading in equilibrium is further characterized by informed manipulation, where informed deviants deliberately "deviate" from their own and exploit uninformed traders' predictable trading behaviours. Given uninformed traders' tendency to "chase the trend", informed traders, who anticipate a mispricing reversal when observing both small-in-size positive (negative) mispricing and high depth imbalance at the best bid (ask), react by strategically going against their own preference for limit buys and using market buys to trigger uninformed market buys, sacrifice current profit difference between limit and market buys in exchange for enhanced execution probability and profitability of later informed traders' limit sells, and collusively manipulate for the good of informed traders as a group. The novelty of manipulative trading in our setting is that informed traders take the "wrong" action not when faced with buy/sell or amount decision but when faced with make-take decision.



(A) Informed traders



(B) Uninformed traders

Figure 1. Convergence of traders' learning

The figure shows the convergence results of the benchmark model. Panel A and B show the convergence of the informed's and the uninformed's learning, respectively. The horizontal axis corresponds to the number of trading rounds, and each round consists of 360,000 trading periods. For Q-value convergence criteria, uninformed traders' learning converges faster than informed traders' learning (uninformed convergence: round 24; informed convergence: round 40); For average reward convergence criteria, the two types of traders' learning simultaneously at round 59.

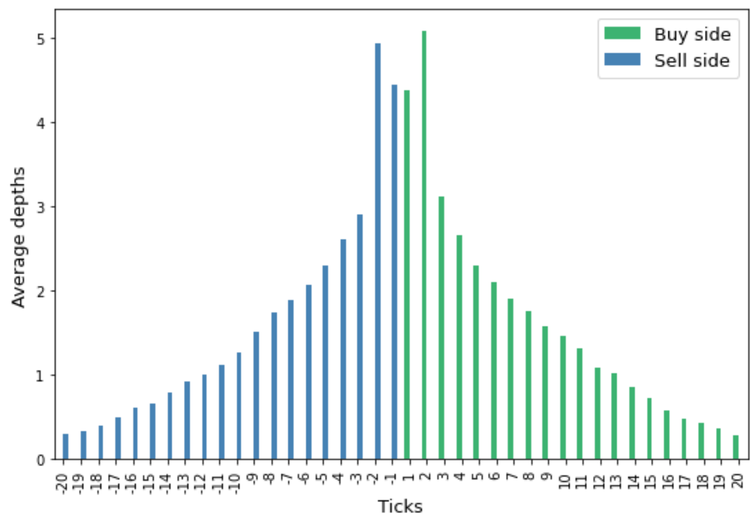


Figure 2. Hump-shaped order book

The figure shows the order book's mean depth profile based on 360,000 trading period simulated data generated by the converged benchmark model. Green bars represent the buy side, and blue bars represent the sell side. The depth of each price level is recorded every 100 trading periods for the 20 best quotes on both sides.

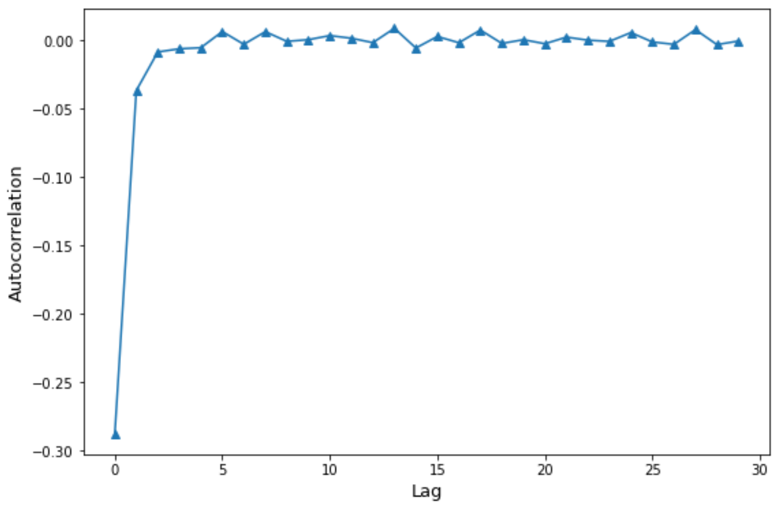


Figure 3. Absence of autocorrelations of returns

The figure shows the absence of autocorrelations (ACs) of price returns in the converged model. The ACs of the log returns are only negative for initial 3 lags, which reflects the bid-ask bounce, and then quickly approach 0.

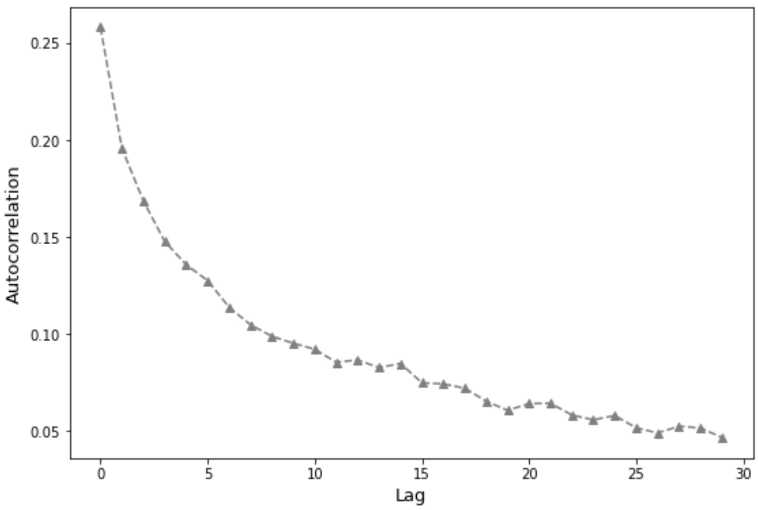
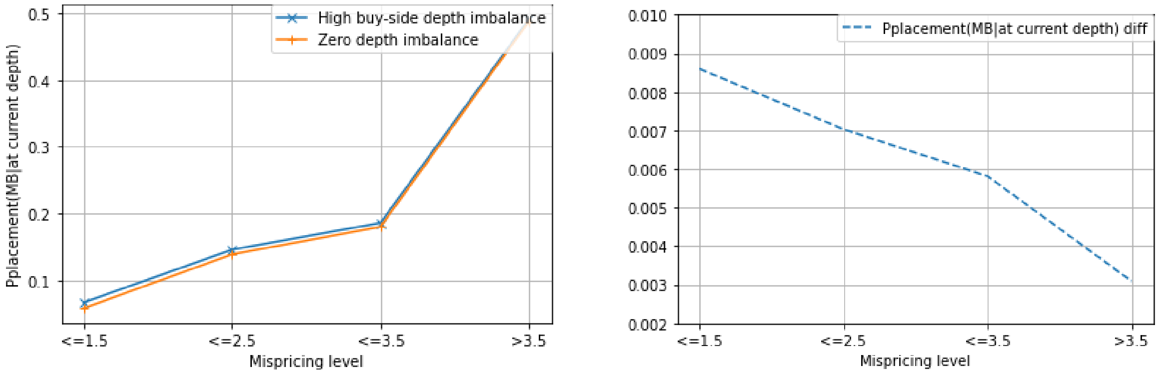


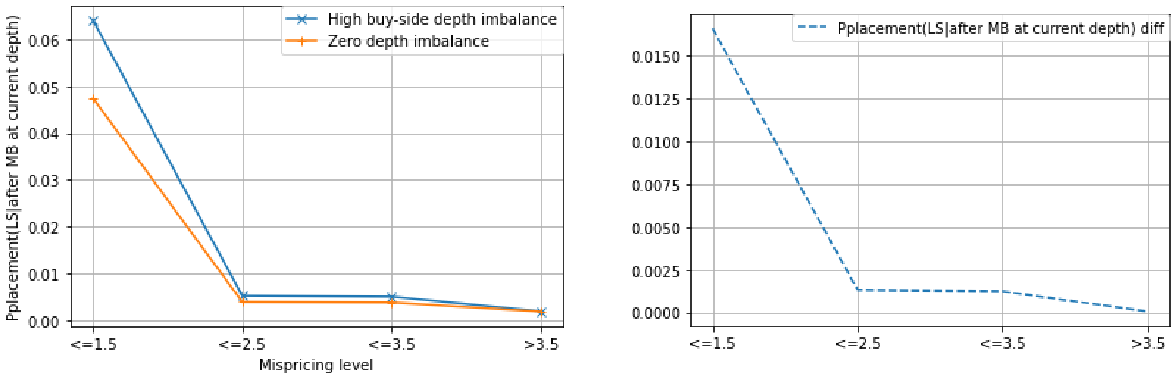
Figure 4. Slow decaying autocorrelations of absolute returns

The figure shows that the slow decaying ACs of absolute returns in the converged model. The ACs of the absolute returns are decreasing from 0.26 to around 0.05 even with 30 lags, which reflects long memory.

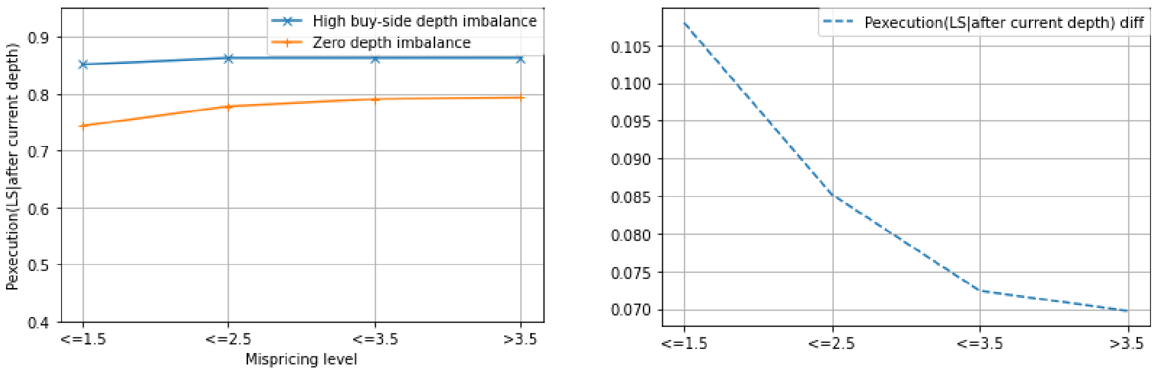
(i) $P^{placement}(MB|\text{at current depth})$



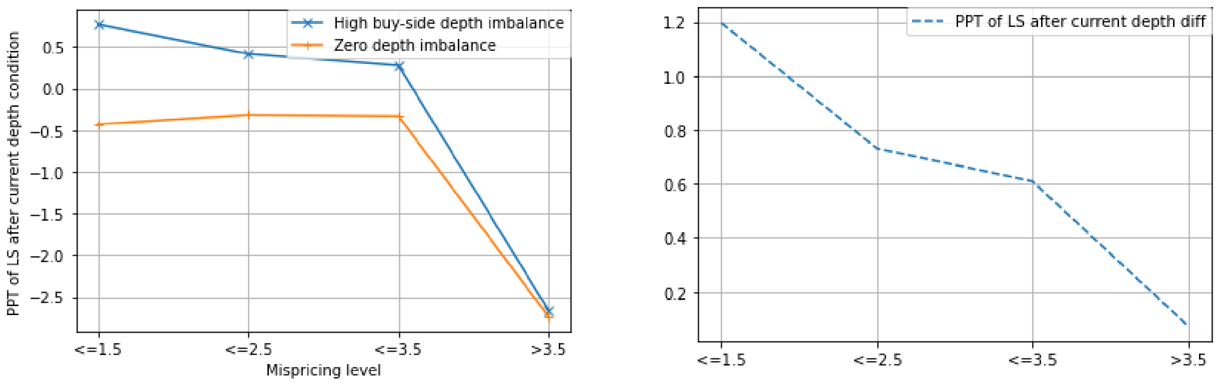
(ii) $P^{placement}(LS|\text{after MB at current depth})$



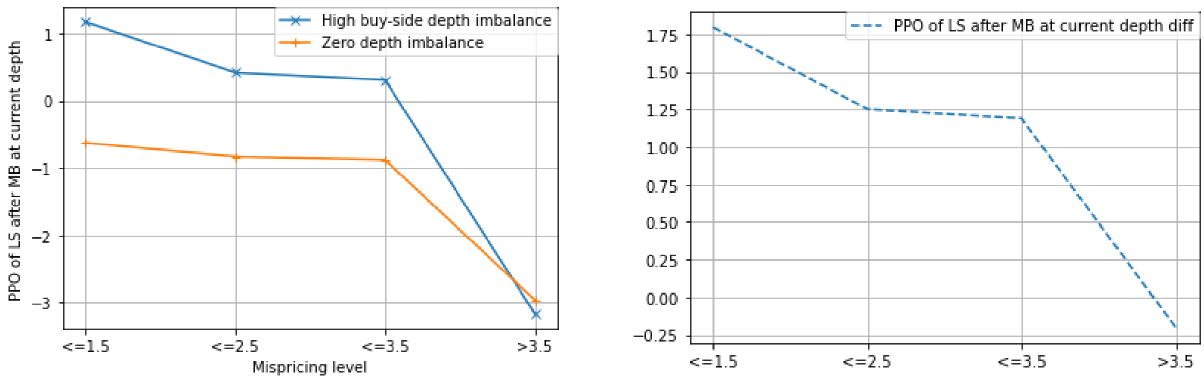
(iii) $P^{execution}(LS|\text{after current depth})$



(iv) PPT of LS after current depth



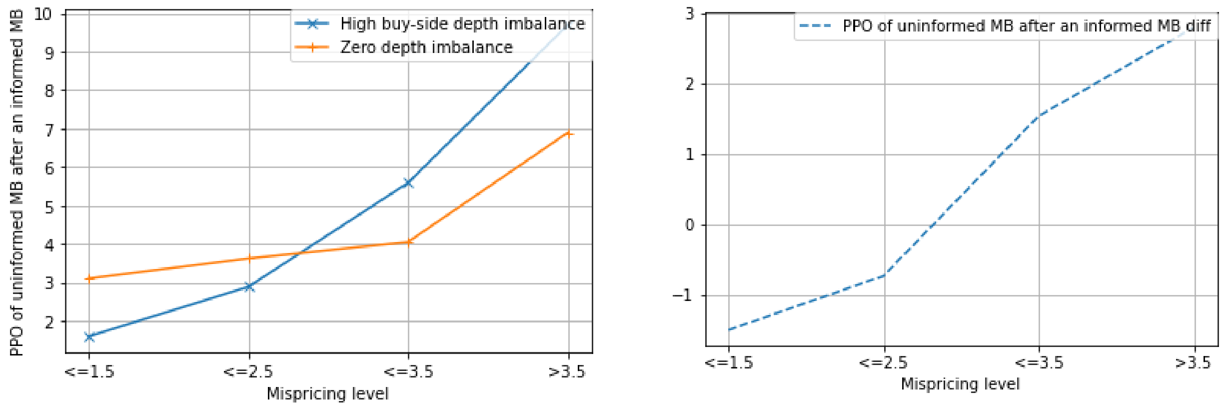
(v) PPO of LS after MB at current depth



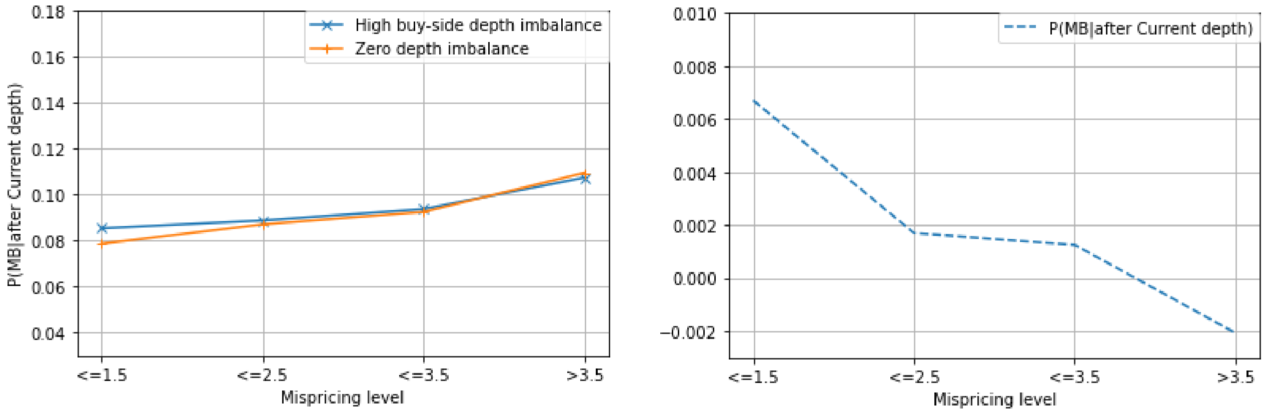
(A) Informed traders

Figure 5. Collusive behaviours of informed trader (continued on next page)

(vi) PPO of uninformed MB after an informed MB



(vii) P(MB|after current depth)



(B) Uninformed traders

Figure 5. Collusive behaviours of informed traders

Table 1. Order book state frequencies for all traders in the benchmark model

The table shows the frequencies (in percentage) of feasible values of state variables Spread s_t, Bid trend $b_t - b_{t-1}$, Ask trend $a_t - a_{t-1}$, Depth imbalance $d_t^a - d_t^b$, and Cumulative depth imbalance $D_t^a - D_t^b$ for all traders based on 360,000 trading period simulated data generated by the converged benchmark model. The frequencies of Bid trend, Ask trend, Depth imbalance, Cumulative depth imbalance, and Last trade direction reflects the symmetry of buy sides and sell sides.

Spread: s_t				Bid trend: $b_t - b_{t-1}$			Ask trend: $a_t - a_{t-1}$		
= 1	= 2	> 2	emp.	> 0	< 0	= 0	> 0	< 0	= 0
53.01	17.17	27.94	1.87	4.05	4.09	91.86	4.39	4.39	91.24

Depth imbalance: $d_t^a - d_t^b$			Cumulative depth imbalance: $D_t^a - D_t^b$			Last trade direction: LT_t	
> 0	< 0	= 0	> 0	< 0	= 0	> 0	< 0
37.71	38.81	23.48	49.05	49.82	1.13	49.44	50.56

Table 2. Frequencies of $E(v_t)$ and $|\bar{v}_t - p_t^m|/s_t$ for informed and uninformed traders in the benchmark model

The table shows frequencies (in percentage) of feasible values of state variables Expected fundamental value $E(v_t)$, Rosu' signal $|\bar{v}_t - p_t^m|/s_t$ for informed and uninformed traders, respectively, based on 360,000 trading period simulated data generated by the converged benchmark model. The relatively large differences between uninformed and informed traders' observation frequencies of Rosu's signal indicate that it might be an inaccurate mispricing signal for uninformed traders.

Expected fundamental value: $E(v_t)$			Rosu's signal: $|\bar{v}_t - p_t^m|/s_t$				
$> p_t^m$	$< p_t^m$	$= p_t^m$	⩽0.5	⩽1.5	⩽2.5	⩽3.5	> 3.5
49.57	49.66	0.77	5.59	12.40	8.17	6.26	67.58
48.96	49.79	1.25	33.02	29.47	11.90	6.41	19.19

Table 3. The relationship between order choices and order book states in the benchmark model

The table shows the unconditional and conditional probabilities of order choices based on feasible values of all state variables. Panel A (B) shows order choices of informed (uninformed) traders. Given the symmetry, only buy side probabilities are reported. No trading probabilities are scaled by two for ease of comparison.

Panel A: Informed traders

Conditional probability (%)

	Spread: s_t				Expected fundamental value: $E(v_t)$			Rosu's signal: $\lvert\bar{v}_t - p_t^m\rvert/s_t$					Bid trend: $b_t - b_{t-1}$		
Order type	= 1	= 2	> 2	emp.	$> p_t^m$	$< p_t^m$	$= p_t^m$	⩽0.5	⩽1.5	⩽2.5	⩽3.5	> 3.5	> 0	< 0	= 0
mb	32.32	21.91	12.23	4.84	49.04	0.11	0.28	0.13	4.27	14.90	16.35	32.01	27.21	6.57	25.08
ealb	-	9.89	6.38	11.50	7.40	0.05	0.17	0.12	1.22	5.67	5.51	4.04	1.75	11.86	3.42
alb	-	-	4.27	8.39	2.56	0.03	0.57	1.38	4.07	2.52	1.87	0.57	0.53	3.18	1.24
lb	4.49	1.71	2.01	5.07	6.56	0.13	0.74	0.62	1.89	1.26	2.30	4.16	5.81	3.45	3.21
ulb	1.67	1.17	0.85	2.79	2.65	0.13	0.83	0.25	0.55	1.04	0.98	1.71	2.12	1.21	1.36
nt	10.61	14.85	22.4	14.89	15.65	13.27	46.60	47.21	35.99	24.58	22.07	6.24	12.62	15.85	14.75

	Ask trend: $a_t - a_{t-1}$			Depth imbalance: $d_t^a - d_t^b$			Cumulative depth imbalance: $D_t^a - D_t^b$			Last trade direction: LT_t	
Order type	> 0	< 0	= 0	> 0	< 0	= 0	> 0	< 0	= 0	Buy	Sell
mb	14.87	23.10	24.88	31.17	18.11	23.96	22.94	26.21	9.01	33.50	15.53
ealb	10.28	1.47	3.51	6.10	2.18	2.34	3.83	3.65	0.50	3.59	3.80
alb	2.88	1.57	1.21	1.20	1.13	1.70	1.16	1.43	1.22	0.91	1.66
lb	3.44	3.55	3.32	5.40	1.30	3.35	3.16	3.46	4.61	3.79	2.88
ulb	1.38	2.50	1.34	2.14	0.65	1.40	1.33	1.45	0.98	1.68	1.10
nt	17.72	12.20	14.68	14.52	14.14	15.95	14.69	14.40	28.89	15.46	13.97

Unconditional probability (%)

mb: 24.41	*ealb*: 3.70	*alb*: 1.29	*lb*: 3.33	*ulb*: 1.39	*nt*: 14.71

Panel B: Uninformed traders

Conditional probability (%)

Order type	Spread: s_t				Expected fundamental value: $E(v_t)$			Rosu's signal: $\lvert \bar{v}_t - p_t^m \rvert / s_t$					Bid trend: $b_t - b_{t-1}$		
	= 1	= 2	> 2	emp.	$> p_t^m$	$< p_t^m$	$= p_t^m$	≤0.5	≤1.5	≤2.5	≤3.5	> 3.5	> 0	< 0	= 0
mb	9.61	8.59	4.29	3.85	9.21	6.60	1.47	6.69	8.27	9.45	10.96	7.11	11.49	6.31	7.74
ealb	-	7.68	4.18	3.80	3.37	1.78	0.45	3.43	2.79	2.35	1.37	1.23	1.84	5.03	2.48
alb	-	-	4.46	3.80	1.99	0.66	0.06	2.25	1.29	0.65	0.42	0.47	1.50	2.29	1.27
lb	7.07	4.39	3.69	3.35	9.08	2.18	0.09	4.77	5.84	6.02	6.28	6.14	8.13	4.13	5.55
ulb	5.96	4.23	3.52	3.45	9.15	0.76	0.03	4.38	4.92	5.21	5.36	5.57	5.16	3.88	4.97
nt	27.65	25.22	30.55	33.05	28.19	27.60	47.92	28.81	27.59	26.38	25.84	29.73	23.47	23.87	28.54

Order type	Ask trend: $a_t - a_{t-1}$			Depth imbalance: $d_t^a - d_t^b$			Cumulative depth imbalance: $D_t^a - D_t^b$			Last trade direction: LT_t	
	> 0	< 0	= 0	> 0	< 0	= 0	> 0	< 0	= 0	Buy	Sell
mb	14.18	10.21	7.46	8.30	7.04	8.40	7.54	7.92	16.84	10.15	5.58
ealb	8.77	1.40	2.34	2.69	2.43	2.57	2.54	2.61	1.37	3.30	1.84
alb	3.83	1.26	1.21	1.15	1.30	1.60	1.39	1.25	1.03	1.50	1.14
lb	4.28	5.90	5.64	5.14	6.25	5.24	5.51	5.66	5.81	6.16	5.04
ulb	3.94	5.32	4.96	3.79	6.16	4.73	4.89	4.98	4.79	4.95	4.91
nt	21.89	21.51	28.70	28.01	28.47	27.84	27.98	28.50	19.88	27.81	28.48

Unconditional probability (%)

mb: 7.84 *ealb*: 2.56 *alb*: 1.32 *lb*: 5.59 *ulb*: 4.93 *nt*: 28.14

Table 4. Percentage of buy orders under different states

The table shows conditional probabilities of traders' buy orders out of all orders based on feasible values of state variables Expected fundamental value $E(v_t)$, Bid trend $b_t - b_{t-1}$, Ask trend $a_t - a_{t-1}$, Depth imbalance $d_t^a - d_t^b$, Cumulative depth imbalance $D_t^a - D_t^b$, and Last trade direction LT_t. Informed traders learn almost perfectly to exploit their information advantage, and uninformed traders chase the trend.

	Expected fundamental value: $E(v_t)$			Bid trend: $b_t - b_{t-1}$			Ask trend: $a_t - a_{t-1}$			Depth imbalance: $d_t^a - d_t^b$			Cumulative depth imbalance: $D_t^a - D_t^b$			Last trade direction: LT_t	
	$> p_t^m$	$< p_t^m$	$= p_t^m$	> 0	< 0	$= 0$	> 0	< 0	$= 0$	> 0	< 0	$= 0$	> 0	< 0	$= 0$	Buy	Sell
Informed	99.28	0.61	38.22	50.05	38.46	48.68	50.89	42.59	48.49	64.84	32.59	48.09	45.90	50.84	38.67	62.93	34.65
Uninformed	75.21	26.77	50.26	53.01	41.38	51.27	62.23	42.25	50.72	47.92	53.82	50.83	49.67	52.12	49.54	58.68	42.99

Table 5. Percentage of limit orders under different states

The table shows conditional probabilities of traders' limit orders out of all orders based on feasible values of state variables Spread s_t, Rosu's signal $|\bar{v}_t - p_t^m|/s_t$, Depth imbalance $d_t^a - d_t^b$, Cumulative depth imbalance $D_t^a - D_t^b$. Informed traders increase their market order usage when mispricing is large, same side depth at the best quote level, and the cumulative level is large. Uninformed traders increase their market order usage when same side depth at the cumulative level is large. Both informed and uninformed traders respond to spread widening by submitting more limit orders, but informed traders are more responsive than uninformed traders.

Panel A: Informed traders

	Spread: s_t				Rosu's signal: $\lvert\bar{v}_t - p_t^m\rvert/s_t$					Depth imbalance: $d_t^a - d_t^b$			Cumulative depth imbalance: $D_t^a - D_t^b$		
Limit/All orders	= 1	= 2	> 2	emp.	⩽0.5	⩽1.5	⩽2.5	⩽3.5	> 3.5	> 0	< 0	= 0	> 0	< 0	= 0
Buyers	16.04	34.92	52.50	83.66	94.78	64.38	41.29	39.48	24.69	32.26	22.50	26.84	29.23	27.60	44.79
Sellers	15.95	36.82	53.16	86.27	94.90	61.93	41.28	35.88	25.22	22.09	33.04	27.66	27.57	30.17	49.67
Both sides	15.99	35.86	52.86	85.16	94.85	62.98	41.29	37.62	24.96	28.69	29.61	27.27	28.33	28.86	47.78

Panel B: Uninformed traders

	Spread: s_t				Rosu's signal: $\lvert\bar{v}_t - p_t^m\rvert/s_t$					Depth imbalance: $d_t^a - d_t^b$			Cumulative depth imbalance: $D_t^a - D_t^b$		
Limit/All orders	= 1	= 2	> 2	emp.	⩽0.5	⩽1.5	⩽2.5	⩽3.5	> 3.5	> 0	< 0	= 0	> 0	< 0	= 0
Buyers	57.55	65.49	78.70	78.93	68.93	64.21	60.10	55.09	65.34	60.61	69.63	62.70	65.51	64.66	43.59
Sellers	57.93	64.97	79.34	80.31	69.43	64.70	59.06	58.17	63.82	68.99	61.22	63.37	63.89	66.29	52.36
Both sides	57.74	65.23	79.01	79.57	69.18	64.45	59.58	56.62	64.59	64.97	65.75	63.03	64.70	65.44	48.02

Table 6. Percentage of ALO under different depth imbalance levels

The table shows conditional probabilities of traders' aggressive limit orders out of all limit orders based on feasible values of the state variable Depth imbalance $d_t^a - d_t^b$. Informed (uninformed) traders increase (reduce) their ALO/LO ratio when same side depth at the best quote level is large.

	Depth imbalance: $d_t^a - d_t^b$		
ALO/Limit orders	> 0	< 0	$= 0$
Informed buyers	49.21	63.01	45.94
Uninformed buyers	30.08	23.11	29.49
Informed sellers	58.91	49.25	55.67
Uninformed sellers	22.41	29.08	28.21

Table 7. Impact of private information on liquidity provision

The table shows the regression results of traders' market/limit order decisions on order book information. Directional adjustments are made to variables Expected fundamental, Bid trend, Ask trend, Depth imbalance, Cumulative depth imbalance, and Last trade direction. Informed traders are resiliency improving according to the coefficients of Spread, Depth imbalance, and Cumulative depth imbalance: when spread enlarges or same side cumulative depth decreases, both informed and uninformed increase their limit order usage, but informed traders demonstrate much stronger responses. Additionally, when the same side depth at best quote decreases, the informed increases limit order usage, but the uninformed increases market order usage.

	Logistic regression		Demeaned Logistic regression		Demeaned OLS regression	
	Informed	Uninformed	Informed	Uninformed	Informed	Uninformed
Spread	0.5509	0.4087	0.6823	0.3834	0.1836	0.1056
	(220.170)	(341.984)	(222.370)	(205.649)	(290.210)	(262.746)
Expected fundamental	-1.0587	0.5092	-1.8776	0.5241	-0.3710	0.1305
	(-6.257)	(273.793)	(-11.206)	(275.618)	(-22.245)	(321.787)
Rosu's signal	-0.3886	0.0035	-0.1106	-0.0085	-0.0176	0.0015
	(-358.913)	(5.051)	(-28.671)	(-8.801)	(-27.541)	(6.413)
Bid trend	-0.1329	-0.0242	-0.1336	-0.0228	-0.0258	-0.0231
	(-15.242)	(-4.466)	(-15.396)	(-4.212)	(-10.733)	(-21.965)
Ask trend	-0.1311	-0.0277	-0.1301	-0.0295	-0.0296	-0.0198
	(-15.372)	(-5.994)	(-15.330)	(-6.387)	(-18.080)	(-19.210)
Depth imbalance	-0.2861	0.1221	-0.2931	0.1218	-0.0578	0.0338
	(-95.196)	(71.465)	(-97.162)	(71.255)	(-106.464)	(88.980)
Cumulative imbalance	-0.0456	-0.0315	-0.0498	-0.0330	-0.0085	-0.0031
	(-18.096)	(-21.037)	(-19.698)	(-22.000)	(-18.567)	(-9.613)
Last trade direction	-0.1994	-0.1311	-0.1996	-0.1337	-0.0460	-0.0355
	(-75.075)	(-83.864)	(-74.839)	(-85.120)	(-94.281)	(-104.195)
Intercept	None	None	-0.8344	0.6678	0.3172	0.6439
			(-128.393)	(414.384)	(293.569)	(811.870)
No. of observations	847326	2098132	847326	2098132	847326	2098132
Pseudo/ Adjusted R^2	0.1165	0.03694	0.1305	0.03769	0.348	0.140

Table 8. Order aggressiveness conditional on volatility levels

The table shows the order aggressiveness strategies of informed and uninformed traders conditional on volatility changes. When fundamental volatility increases, uninformed traders decrease MO and increase ALO, informed traders increase MO and decrease ALO.

Volatility of fundamental value	2	4	6	8
Prob (Spread $\leqslant$ 2)	71.58	57.42	52.93	52.61
Informed order choices				
Market order %	73.81	74.79	76.00	76.11
Aggressive limit order %	15.68	15.39	14.13	12.99
Nonaggressive limit order %	10.51	9.82	9.86	10.90
Uninformed order choices				
Market order %	34.41	33.91	33.15	32.65
Aggressive limit order %	18.64	22.71	23.27	23.33
Nonaggressive limit order %	46.95	43.39	43.58	44.02

Table 9. Order aggressiveness conditional on informed trading levels

The table shows order aggressiveness strategies of informed and uninformed trader conditional on informed trading levels. When informed trading level increases, uninformed traders decrease MO and decrease ALO, and informed traders decrease MO and increase ALO due to intensified intertemporal competition.

Proportion of informed traders	10%	12.5%	15%	17.5%	20%
Market quality					
Quoted spread	1.47	1.44	1.41	1.40	1.36
Depth at best bid	3.16	3.51	3.65	4.83	5.21
Price discovery $\lvert p_t - v_t \rvert / v_t$ %	0.58	0.48	0.37	0.33	0.29
Welfare					
Informed trader	2.10	2.03	1.94	1.93	1.78
Uninformed trader	-0.21	-0.15	-0.09	-0.07	-0.06
Informed order choices					
Market order %	5.89	5.19	5.15	4.66	4.20
Aggressive limit order %	38.00	49.37	52.32	53.78	59.97
Nonaggressive limit order %	56.11	45.44	42.52	41.56	35.84
Uninformed order choices					
Market order %	31.22	31.56	30.82	28.34	27.40
Aggressive limit order %	31.14	28.11	26.01	19.47	16.08
Nonaggressive limit order %	37.64	40.33	43.17	52.19	56.52

Appendix

A1. Uninformed traders' inability to infer mispricing reversal

The uninformed cannot extract mispricing reversal related information and do not demonstrate manipulative behaviours resemble informed traders'. To illustrate this, we look at Figure A1. For $P^{placement}$(MB|at current depth), statistic (i), of uninformed traders, blue and yellow solid lines do not display monotonicity, an artefact driven by the assumption that uninformed traders have no access to actual mispricing. The blue solid line is below the yellow solid line, consistent with the ceteris paribus effect in subsection 4.4 that deeper depth imbalance at the best bid/ask decreases same-side uninformed traders' tendency to submit market orders.

More importantly, the difference between $P^{placement}$(MB|at current depth) of high buy-side depth imbalance and zero depth imbalance fluctuates around -0.70%, reflecting that uninformed traders do not discriminate between various depth imbalance and mispricing combinations when it comes to market buy placement probability at current period.

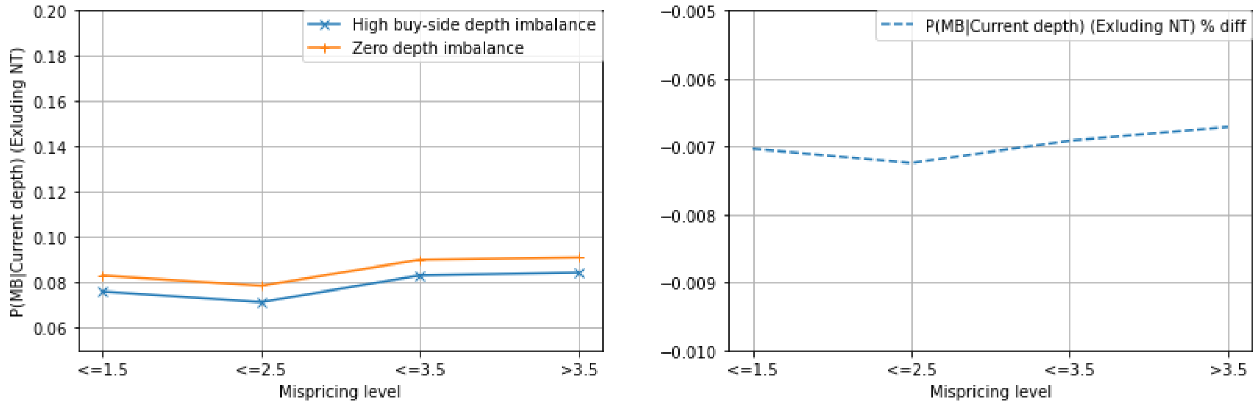
For $P^{placement}$(LS| after MB at current depth), statistic (ii), of uninformed traders, blue and yellow solid lines again have no monotonical patterns because uninformed traders do not observe actual mispricing. The blue solid line is below the yellow solid line, consistent with the ceteris paribus effect in subsection 4.4 that deeper depth imbalance at the best bid/ask decreases other-side uninformed traders' tendency to submit limit orders. This ceteris paribus effect can be justified by the corresponding PPO of $P^{placement}$(LS| after MB at current depth). As seen from statistic (v) in panel B, when the actual mispricing is greater than 3.5, accounting for more than 60% of the simulated sample, the PPO of high buy-side imbalance is lower than that of zero depth imbalance, and the difference has a nonnegligible magnitude of -1.15 ticks.

More importantly, the difference between $P^{placement}$(LS| after MB at current depth) of high buy-side depth imbalance and zero depth imbalance fluctuates around -0.60%, reflecting that uninformed traders do not discriminate between various depth imbalance and mispricing combinations when it comes to limit sell placement probability after the current uninformed MB.

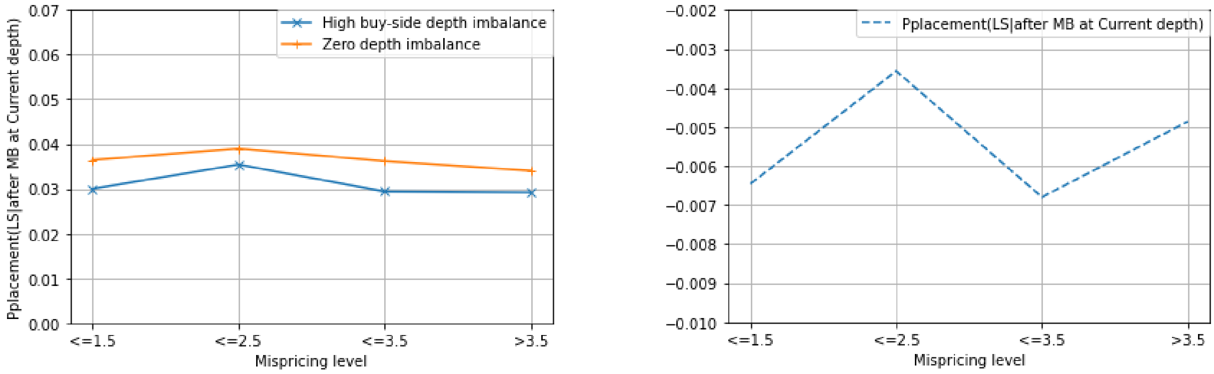
Furthermore, when Rosu's mispricing signal is no larger than 1.5, the difference between $P^{execution}$(LS|after current depth), statistic (iii), of uninformed traders under high buy-side depth imbalance and zero depth imbalance is not at its highest value. The same applies for statistic (iv)-PPT of uninformed LS after current depth, and statistic (v)-PPO of uninformed LS after uninformed MB at current depth.

To sum up, uninformed traders cannot extract mispricing reversal related information, and do not vary current MB usage and future LS usage across different depth imbalance and mispricing combinations

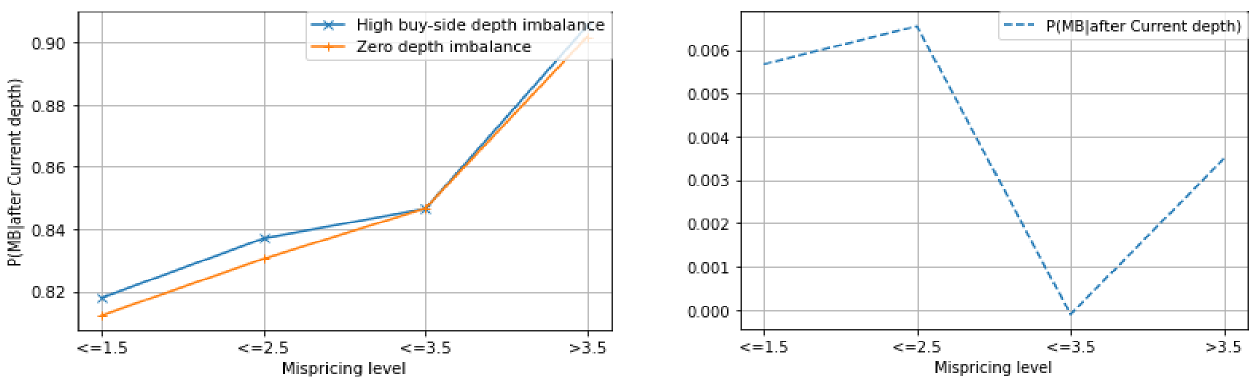
(i) P(MB|Curerent depth)



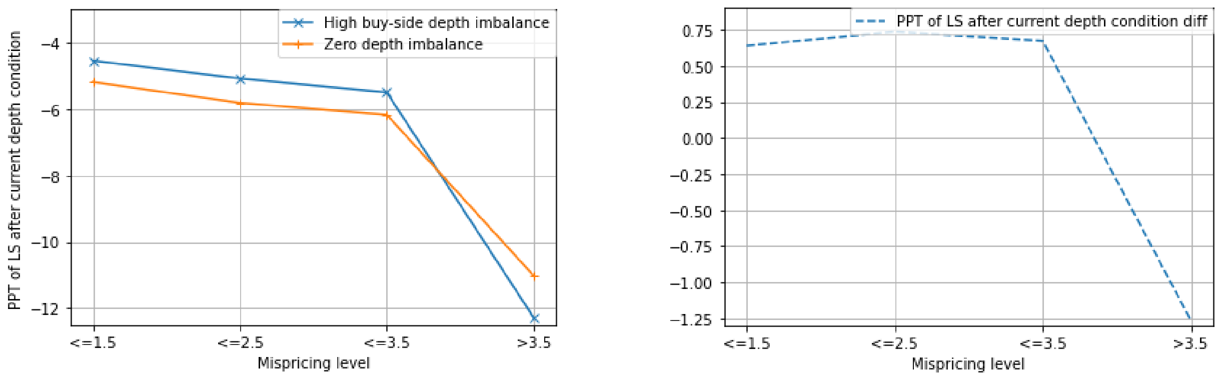
(ii) $P^{placement}$(LS|after MB at current depth)



(iii) $P^{execution}$(LS|after current depth)



(iv) PPT of LS after current depth condition



(v) PPO of LS after MB at current depth

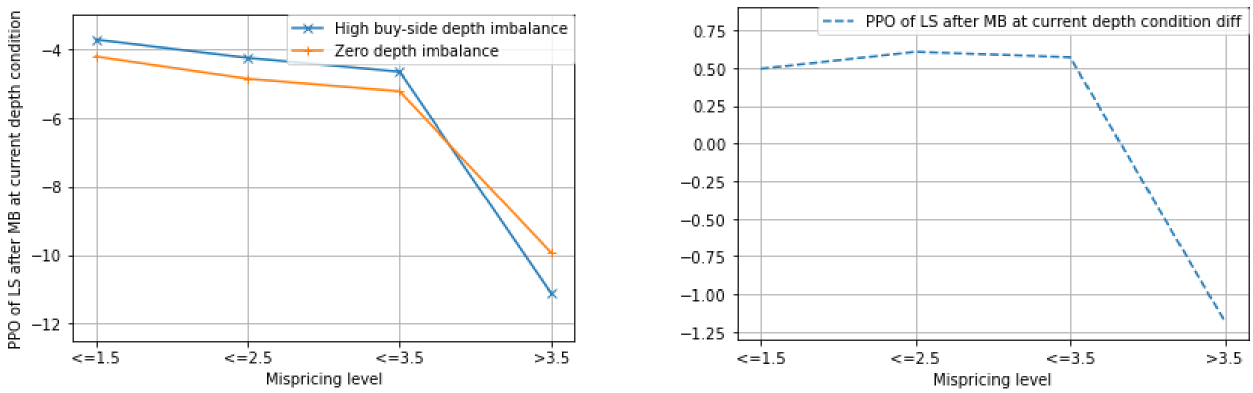


Figure A1. Uninformed traders' inability to infer mispricing reversal

Table A1. classified rules (CRs) for state variable discretization

This table presents eight classification rules (CRs) in the classifier system based on the spread, the expected fundamental value, the mispricing signal, order book movements, depth imbalances, and the last trade direction. Using the classifier system, the continuous state space is transformed into a discrete one that contains 14,580 possible states.

Classified rules	**Possible values**
$Spread\ s_t$	Current spread is equal to 1 Current spread is equal to 2 Current spread is higher than 2 Empty on buy side (emp^{+}) Empty on sell side (emp_{-}) Empty on both sides (emp_{-}^{+})
$Expected\ fundamental$ $E(v_t) - p_t^m$	Expected fundamental is higher than mid-price Expected fundamental is lower than mid-price Expected fundamental equals to mid-price
$Rosu's\ signal$ $\lvert\bar{v}_t - p_t^m\rvert / s_t$	Mispricing signal is in range [0, 0.5] Mispricing signal is in range (0.5, 1.5] Mispricing signal is in range (1.5, 2.5] Mispricing signal is in range (2.5, 3.5] Mispricing signal is in range (3.5, $+\infty$)
$Bid\ trend$ $b_t - b_{t-1}$	Current bid is higher than last bid Current bid is lower than last bid Current bid equals to last bid
Ask trend $a_t - a_{t-1}$	Current ask is higher than last ask Current ask is lower than last ask Current ask equals to last ask
Depth imbalance $d_t^a - d_t^b$	Depth at best ask is higher than the opposite side Depth at the best ask is higher than the opposite side Depth at the best ask is equal to the opposite side
Cumulative depth imbalance $d_t^a - d_t^b$	Depth at the sell side is higher than the opposite side Depth at the sell side is higher than the opposite side Depth at the best ask is equal to the opposite side
Last trade direction LT_t	Last market order is a buy order Last market order is a sell order

References

Aitken, M., Almeida, N., Harris, F. H. D., & McInish, T. H. (2007). Liquidity supply in electronic markets. *Journal of Financial Markets*, *10*(2), 144-168.

Anand, A., Chakravarty, S., & Martell, T. (2005). Empirical evidence on the evolution of liquidity: Choice of market versus limit orders by informed and uninformed traders. *Journal of Financial Markets*, *8*(3), 288-308.

Athey, S. (2018). The impact of machine learning on economics. In *The economics of artificial intelligence: An agenda*. University of Chicago Press.

Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, *11*, 685-725.

Beber, A., & Caglio, C. Order submission strategies and information: Empirical evidence from the NYSE. Working paper.

Benabou, R., & Laroque, G. (1992). Using privileged information to manipulate markets: Insiders, gurus, and credibility. *The Quarterly Journal of Economics*, *107*(3), 921-958.

Bloomfield, R., O'hara, M., & Saar, G. (2005). The "make or take" decision in an electronic market: Evidence on the evolution of liquidity. *Journal of Financial Economics*, *75*(1).

Bouchaud, J. P., Mézard, M., & Potters, M. (2002). Statistical properties of stock order books: empirical results and models. *Quantitative finance*, *2*, 251-256.

Brunnermeier, M. K. (2000). Buy on rumours-sell on News: a manipulative trading strategy. Working paper.

Bryzgalova, S., Pelger, M., & Zhu, J. (2019). Forest through the trees: Building cross-sections of stock retufrns. Working paper.

Chakravarty, S., & Holden, C. W. (1995). An integrated model of market and limit orders. *Journal of Financial Intermediation*, *4*(3), 213-241.

Chiarella, C., He, X., & Wei, L. (2013). Learning and evolution of trading strategies in limit order markets. *Journal of Economic Dynamics and Control*, *61*, 245-268.

Chiu, J., Chung, H., & Wang, G. H. (2014). Intraday liquidity provision by trader types in a limit order market: Evidence from Taiwan index futures. *Journal of Futures Markets*, *34*(2), 145-172.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues.

Duong, H. N., Kalev, P. S., & Krishnamurti, C. (2009). Order aggressiveness of institutional and individual investors. *Pacific-Basin Finance Journal*, *17*(5), 533-546.

Easley, D., de Prado, M. L., O'Hara, M., & Zhang, Z. (2019). Microstructure in the machine age. Working paper.

Ellul, A., Holden, C. W., Jain, P., & Jennings, R. (2007). Order dynamics: Recent evidence from the NYSE. *Journal of Empirical Finance*, *14*(5), 636-661.

Eren, N., & Ozsoylev, H. N. (2006, November). Hype and dump manipulation. Working paper.

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

Goettler, R. L., Parlour, C. A., & Rajan, U. (2009). Informed traders and limit order markets. *Journal of Financial Economics*, *93*(1), 67-87.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical Asset Pricing via Machine Learning, *The Review of Financial Studies*, *33*(5), 2223–2273.

Harris, L. (2003). *Trading and exchanges: Market microstructure for practitioners*. OUP USA.

He, X., & Lin, S. (2019). Reinforcement Learning in Limit Order Markets. Working paper.

Huddart, S., Hughes, J. S., & Levine, C. B. (2001). Public disclosure and dissimulation of insider trades. *Econometrica*, *69*(3), 665-681.

John, K., & Narayanan, R. (1997). Market manipulation and the role of insider trading regulations. *The Journal of Business*, *70*(2), 217-247.

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

Kempf, A., Mayston, D., & Yadav, P. K. (2009). Resiliency in limit order book markets: A dynamic view of liquidity. Working paper.

Menkhoff, L., Osler, C. L., & Schmeling, M. (2010). Limit-order submission strategies under asymmetric information. *Journal of Banking & Finance*, *34*(11), 2665-2677.

Parlour, C. A., & Seppi, D. J. (2008). Limit order markets: A survey. In *Handbook of financial intermediation and banking*.

Putniņš, T. J. (2012). Market manipulation: A survey. *Journal of Economic Surveys*, *26*(5), 952-967.

Putniņš, T. J. (2020). An overview of market manipulation. *Corruption and Fraud in Financial Markets: Malpractice, Misconduct and Manipulation*, *13*.

Renault, T. (2017). Intraday online investor sentiment and return patterns in the US stock market. *Journal of Banking & Finance*, *84*, 25-40.

Ricco, R., Rindi, B., & Seppi, D. J. (2020). Information, Liquidity, and Dynamic Limit Order Markets. Working paper.

Rosu, I. (2012). Order choice and information in limit order markets. In *Market Microstructure: Confronting Many Viewpoints*.

Rosu, I. (2020). Liquidity and information in limit order markets. *Journal of Financial and Quantitative Analysis*, *55*(6), 1792-1839.

Schnaubelt, M., Rende, J., & Krauss, C. (2019). Testing stylized facts of bitcoin limit order books. *Journal of Risk and Financial Management*, *12*(1), 25.

Shino, T. (2021). Price manipulation, dynamic informed trading, and the uniqueness of equilibrium in sequential trading. *Journal of Economic Dynamics and Control*, *125*.

Van Bommel, J. (2003). Rumors. *The journal of Finance*, *58*(4), 1499-1520.

Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, *28*(2), 3-28.

Vila, J. L. (1989). Simple games of market manipulation. *Economics Letters*, *29*(1), 21-26.

Vives, X. (2010). *Information and learning in markets: the impact of market microstructure*. Princeton University Press.

Watkins, C. J. C. H. (1989). *Learning from delayed rewards* (Doctoral dissertation, King's College, Cambridge).