Learning-Based Trading Strategies in the Face of Market Manipulation

Xintong Wang
University of Michigan
xintongw@umich.edu

Christopher Hoang
University of Michigan
choang@umich.edu

Michael P. Wellman
University of Michigan
wellman@umich.edu

ABSTRACT

We study learning-based trading strategies in markets where prices can be manipulated through *spoofing*: the practice of submitting spurious orders to mislead traders who use market information. To reduce the vulnerability of learning traders to such manipulation, we propose two variations based on the standard *heuristic belief learning* (HBL) trading strategy, which learns transaction probabilities from market activities observed in an order book. The first variation selectively ignores orders at certain price levels, particularly where spoof orders are likely to be placed. The second considers the full order book, but adjusts its limit order price to correct for bias in decisions based on the learned heuristic beliefs. We employ agent-based simulation to evaluate these variations on two criteria: effectiveness in non-manipulated markets and robustness against manipulation. Background traders can adopt (non-learning) *zero intelligence* strategies or HBL, in its basic form or the two variations. We conduct empirical game-theoretic analysis upon simulated payoffs to derive approximate strategic equilibria, and compare equilibrium outcomes across a variety of trading environments. Results show that agents can strategically make use of the option to block orders to improve robustness against spoofing, while retaining a comparable competitiveness in non-manipulated markets. Our second HBL variation exhibits a general improvement over standard HBL, in markets with and without manipulation. Further explorations suggest that traders can enjoy both improved profitability and robustness by combining the two proposed variations.

1 INTRODUCTION

The increasing automation of trading and interconnectedness of markets have transformed the financial market from a human decision ecosystem to an algorithmic one. With trades happening on an extremely short timescale, often beyond the limit of human decision-making, algorithms, or autonomous agents, are developed to operate on behalf of human traders. They learn from new information, make decisions, and interact with each other at an unprecedented speed and complexity. Whereas automated trading and the consequent use of learning-based trading algorithms may improve efficiency in some respects, they have also made new forms of disruptive and manipulative practices possible.

In this paper, we study the strategic dynamics between traders who learn from trading actions of other market participants and a manipulator who creates artificial activities to maneuver others' pricing beliefs. We focus on a common form of order-based market manipulation, called *spoofing*. It is achieved by submitting large spurious orders that are not intended for execution, but rather to fool other traders. By feigning a strong buy or sell interest, spoof orders may persuade other traders—those who learn from market information—to believe that prices may soon rise or fall, thus altering their behavior in a way that directly moves the price. Here, both learning and manipulation rely on the *order book* disclosed by a standard financial exchange as an interface. It lists outstanding orders and reflects the aggregate supply and demand for a particular security at any given time.

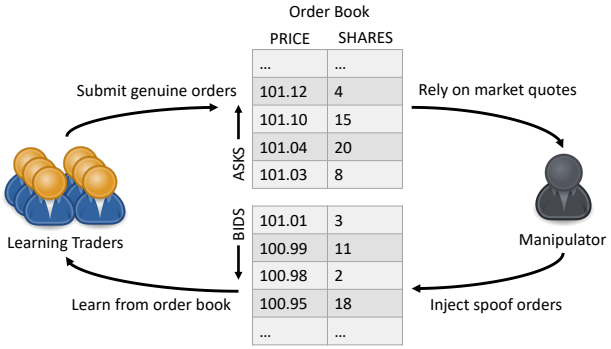


Figure 1: Overview of the game between traders who learn from order book information to trade and a manipulator who injects spoof orders to maneuver others' pricing beliefs.

In ideal markets without manipulation, there is real information to be gleaned from the order book, and thus, strategies that learn from observable market activity have an advantage over those that neglect such information. The less sophisticated non-learning strategies, however, have the advantage of being oblivious to spoofers, and thus, are not manipulable. The question we investigate is whether learning-based strategies can be designed to be similarly robust to spoofing. Specifically, we consider scenarios similar to real-world markets where traders are aware of potential manipulation, but fail to perfectly detect spoof orders in real time. We seek to identify strategies by which individual traders can learn from market information, but in less vulnerable ways.

We start with one representative learning-based trading strategy, proposed by Gjerstad [6] and referred to as *heuristic belief learning* (HBL), which learns a belief state over acceptance of hypothetical buy and sell orders from historical trading activities. HBL was previously adopted in the agent-based model of spoofing by Wang and Wellman [16], where it was shown to be susceptible to simple spoofing strategies. Here, we treat the original HBL as a baseline strategy, and propose two variations that aim to reasonably trade off learning effectiveness in non-manipulated markets for robustness against manipulation. The first variation works by selectively ignoring orders at certain price levels, particularly where spoof orders are likely to be placed to fool other traders. The second variation considers the full order book, but has the flexibility to adjust the offer price by a stochastic offset. The adjustment serves to correct

biases in learned price beliefs either caused by manipulation or the intrinsic limitation built in the belief function.

We employ agent-based simulation to evaluate the proposed variations in terms of the effectiveness in non-manipulated markets and robustness against manipulation. Our market model implements a *continuous double auction* (CDA) market where multiple background trading agents and one exploiter follow their respective strategies to trade a single security. The exploiter makes profit by first buying the underlying security at low prices and later selling at higher prices. To increase its profit, the exploiter may choose to spoof the market after its original purchase to manipulate prices up. Background traders have private values on holding long or short positions on the underlying security. They may choose to follow parameterized strategy instances from either the non-spoofable *zero intelligence* (ZI) family or the learning-based HBL family which includes its basic form and our two proposed variations.

We conduct extensive simulation over hundreds of strategy profiles across parametrically distinct market environments with and without manipulation to derive empirical equilibria. We then evaluate the strategy performances and market outcomes in equilibrium where every agent chooses its best response to both the market environment and others' behavior. Our results show that learning traders can strategically make use of the option to block orders to improve robustness against spoofing, while retaining a comparable competitiveness in non-manipulated markets. Our second HBL variation exhibits a general improvement over the baseline HBL, in markets with and without manipulation. Further explorations suggest that traders can enjoy both improved profitability and robustness by combining the two HBL variations.

Roadmap. In the next section, we present additional background on market manipulation, and discuss related work on modeling and mitigating manipulation. Section 3 describes our agent-based market model, and formally defines the two proposed strategy variations of HBL. In Section 4, we present results from extensive simulation, employing control experiments and empirical game-theoretic analysis to evaluate the two variations. Section 5 concludes.

2 RELATED WORK

2.1 Background on Market Manipulation

The US Securities and Exchange Commission formally defines market manipulation as "intentional conduct designed to deceive investors by artificially affecting the market." Spoofing, as a specific manipulation strategy, has been outlawed under the 2010 Dodd-Frank Wall Street Reform and Consumer Protection Act. Despite the regulatory enforcement, detecting manipulation in real time or even after the fact from high-volume, high-velocity order streams is challenging. Legal definitions cannot be easily translated to computer programs to direct detection, and the lack of datasets with actual order streams identified as cases of manipulation makes training a reliable detector infeasible. Moreover, though trading activities are observable, the intent and effect of misleading others behind certain activities is hard to verify purely from data. For both reasons, we follow prior literature that studies phenomena in financial markets [8, 9, 13, 16] in pursing an agent-based simulation approach to incorporate causal premises and evaluate the effects of proposed strategies.

2.2 A Computational Model of Spoofing

We build our study on an existing agent-based model of spoofing developed by Wang and Wellman [16]. The model illustrates the strategic interactions between a manipulator and two groups of background traders: those who use and do not use market information to trade. In their model, learning traders who adopt the standard HBL trading strategy can benefit social welfare, but their existence also renders a market vulnerable to manipulation. The designed spoofing strategy can effectively fool HBL traders about the market state, and make profits from their spoofed pricing beliefs. A comparison of equilibrium outcomes shows that manipulation decreases the proportion of learning traders in equilibrium and hurts market welfare.

Our work extends the prior spoofing model to study effective adjustments that can be made on learning-based trading strategies to resist manipulation. We treat the standard HBL as a baseline strategy, and explore an expanded strategy space (i.e., the two proposed variations) to find strategic adaptations that can improve its robustness while not compromising much on the learning effectiveness.

2.3 Proposals to Mitigate Market Manipulation

Due to difficulties in directly detecting manipulation, regulators and researchers seek systematic approaches to render manipulative practices uneconomical. For example, advocates propose to impose cancellation fees to increase the cost of manipulative strategies that rely on massive cancellations to avoid transaction risk [1, 10]. Opponents argue that such cancellation fees could instead make liquidity providers suffer from adverse selection and react slowly to new information [3, 5]. Wang et al. [15] propose to deter spoofing by strategically cloaking certain market information, introducing risks and difficulties for the manipulator to post misleading bids. They show that hiding certain price levels in the order book significantly diminishes the efficacy of spoofing, but can be at the cost of degrading the general usefulness of market information in non-manipulated markets. Our first strategy variation is inspired by the cloaking mechanism, but works by granting individual traders the flexibility to decide which prices to ignore.

3 MARKET MODEL AND TRADING STRATEGIES

3.1 Market Mechanism

We model the trading of a single security in a CDA market mechanism. Agents trade the security by submitting limit orders that specify the maximum (minimum) price at which they would be willing to buy (sell) some number of units. Limit order prices take on discrete integer values with a tick size of one. The market maintains a limit order book of outstanding orders, from which traders may learn at their own discretion.

Our market model is implemented in a discrete-event simulation system where time is discrete over a finite horizon T. The fundamental value r_t of the underlying security changes over time according to a mean-reverting stochastic process [2, 14]: for $t \in [0, T]$,

$$r_t = \max\{0, \kappa\bar{r} + (1-\kappa)r_{t-1} + u_t\} \text{ and } r_0 = \bar{r}, \tag{1}$$

where $\kappa \in [0, 1]$ specifies the degree to which the time series reverts back to the fundamental mean $\bar{r}$. $u_t \sim N(0, \sigma_s^2)$ represents a

systematic random shock upon the fundamental at time t, where σ_s^2 is the fundamental shock variance. This fundamental shock controls the intensity of fluctuations in the time series, and consequently influences the predictability of future price outcomes.

3.2 Agents in the Market

For the purpose of our study, we partition agents in the market into two roles: background traders and an exploiter. Multiple background traders represent investors with private preferences on holding long or short positions in the underlying security, whereas the exploiter has no private value and seeks only to profit by buying at lower prices and later selling at higher ones. We use the exploiter to control market environments with and without manipulation. In selected treatments, the exploiter can manipulate the market with spoof orders to push the price up, and thereby boost profit based on other traders' misled beliefs.

3.2.1 Private Valuation. A background trader i has a private value vector Θ_i of length $2q_{\max}$ that captures its position preference. The parameter $q_{\max}$ specifies the maximum number of units one can be long or short at any time. Element θ_i^{q+1} represents the marginal gain from buying an additional unit, given the current net position q. We generate Θ_i from a set of $2q_{\max}$ values independently drawn from $N(0, \sigma_{PV}^2)$, where σ_{PV}^2 denotes the private value variance. We then sort elements to reflect diminishing marginal utility and assign θ_i^q accordingly. The trader's overall *valuation* for a unit of the security is the sum of its private value and the fundamental value.

3.2.2 Background Agent Arrivals and Observations. Agents are allowed to enter the market multiple times throughout a trading period. Arrivals of a background trader follow a Poisson process with an arrival rate of λ_a. On each entry, the trader observes an agent- and time-specific noisy fundamental $o_t = r_t + n_t$ with the observation noise following $n_t \sim N(0, \sigma_n^2)$, where σ_n^2 represents the observation variance. The noisy observation captures investors' different perceptions of the intrinsic value of the underlying security at a given time. As this noisy observation only gives imperfect information about the fundamental, traders can benefit from considering market information, which is influenced by the aggregate observations and trading actions of all the other traders. To react to a new observation, the background trader withdraws its previous order (if not transacted) upon arrival, and submits a new single-unit order to either buy or sell as instructed with equal probability. The order price is jointly decided by the background trader's *valuation* and *trading strategy*, which we describe in detail in the next section.

3.2.3 Payoff and Surplus Calculation. We evaluate the payoffs of individual agents at the end of a trading period. A background trader's surplus is its net profits from trading plus the final valuation of holdings at T. Specifically, the market's final valuation of background trader i with final holdings H is $r_T H + \sum_{k=1}^{k=H} \theta_i^k$ for long position $H > 0$, or $r_T H - \sum_{k=H+1}^{k=0} \theta_i^k$ for short position $H < 0$. A exploiter's payoff is simply its gain or loss from trading. We calculate total surplus as the sum of all agents' payoffs, and background surplus as the sum of all background agents' payoffs.

3.3 Background Trading Strategies

3.3.1 Estimation of the Terminal Fundamental. As security holdings are evaluated at the end of a trading period, background agents maintain an estimate of the final fundamental value, updated based on information received at each market entry. Specifically, given a new noisy observation o_t, a trader estimates the current fundamental by updating its posterior mean $\tilde{r}_t$ and variance $\tilde{\sigma}_t^2$. Let t' denote the trader's preceding arrival time. We first update the previous posteriors ($\tilde{r}_{t'}$ and $\tilde{\sigma}_{t'}^2$) by taking account of mean reversion for the interval since preceding arrival ($\delta = t - t'$):

$$\begin{aligned} \tilde{r}_{t'} &\leftarrow (1-(1-\kappa)^{\delta})\bar{r} + (1-\kappa)^{\delta}\tilde{r}_{t'}\ ; \\ \tilde{\sigma}_{t'}^2 &\leftarrow (1-\kappa)^{2\delta}\tilde{\sigma}_{t'}^2 + \frac{1-(1-\kappa)^{2\delta}}{1-(1-\kappa)^2}\sigma_s^2. \end{aligned}$$

The new posterior estimates at time t are then given by:

$$\tilde{r}_t = \frac{\sigma_n^2}{\sigma_n^2+\tilde{\sigma}_{t'}^2}\tilde{r}_{t'} + \frac{\tilde{\sigma}_{t'}^2}{\sigma_n^2+\tilde{\sigma}_{t'}^2}o_t\ ;\ \tilde{\sigma}_t^2 = \frac{\sigma_n^2\tilde{\sigma}_{t'}^2}{\sigma_n^2+\tilde{\sigma}_{t'}^2}.$$

Based on the posterior estimate of $\tilde{r}_t$, the trader computes $\hat{r}_t$, its estimate at time t of the terminal fundamental r_T, by adjusting for mean reversion:

$$\hat{r}_t = \left(1-(1-\kappa)^{T-t}\right)\bar{r} + (1-\kappa)^{T-t}\tilde{r}_t. \tag{2}$$

Notice that this estimation of the terminal fundamental is a posterior update on an agent's individual observations, and thus will not be affected by other agents' trading activities.

3.3.2 Zero Intelligence. We follow prior work [13, 16] in adopting an extended and parameterized version of ZI as a representative non-learning trading strategy. Since ZI decides order prices without utilizing order book information, it is non-spoofable. The strategy has been widely adopted in agent-based finance due to its simplicity and effectiveness for market modeling [4, 7].

The ZI trader computes a limit-order price by shading its valuation with a random offset uniformly drawn from $[R_{\min}, R_{\max}]$. It further takes into account the current market quote price, controlled by a strategic surplus threshold parameter $\eta \in [0, 1]$. Before submitting a new limit order, if the ZI could achieve a fraction η of its requested surplus by accepting the most competitive order, it would take that quote by submitting an order at the same price.

3.3.3 Heuristic Belief Learning and Its Variations. The two proposed HBL variations, together with its basic form, serve as our representative learning-based trading strategies. They use transaction and order book information to decide offer prices, and thus can be affected by others' trading actions. Without directly detecting any suspicious orders, the two variations expand HBL to a larger strategy space to reduce vulnerability to manipulation. We are interested in understanding their robustness against manipulation as well as competitiveness to other trading strategies across different market environments.

We first provide a brief description on how the standard HBL works, and describe in detail its two variations. The HBL trading strategy is centered on belief functions that traders form on the basis of observed market data $\mathcal{D}$ in memory. Upon an arrival at time t, HBL estimates the probability that orders at various prices

would be accepted in the market according to the heuristic: [1]

$$f_t(\mathrm{P} \mid \mathcal{D}) = \begin{cases} \frac{\mathrm{TBL}_t(\mathrm{P}|\mathcal{D})+\mathrm{AL}_t(\mathrm{P}|\mathcal{D})}{\mathrm{TBL}_t(\mathrm{P}|\mathcal{D})+\mathrm{AL}_t(\mathrm{P}|\mathcal{D})+\mathrm{RBG}_t(\mathrm{P}|\mathcal{D})} & \text{if buying,} \\ \\ \frac{\mathrm{TAG}_t(\mathrm{P}|\mathcal{D})+\mathrm{BG}_t(\mathrm{P}|\mathcal{D})}{\mathrm{TAG}_t(\mathrm{P}|\mathcal{D})+\mathrm{BG}_t(\mathrm{P}|\mathcal{D})+\mathrm{RAL}_t(\mathrm{P}|\mathcal{D})} & \text{if selling.} \end{cases} \tag{3}$$

Based on the interpolated probabilities, HBL chooses an optimal limit price $\mathrm{P}_i^*(t)$ that maximizes its own expected surplus at the current valuation estimate. That is,

$$\mathrm{P}_i^*(t) = \begin{cases} \arg\max_p (\hat{r}_t + \theta_i^{q+1} - p) f_t(p \mid \mathcal{D}) & \text{if buying,} \\ \arg\max_p (p - \theta_i^q - \hat{r}_t) f_t(p \mid \mathcal{D}) & \text{if selling.} \end{cases} \tag{4}$$

HBL with Selective Price Level Blocking. In a normal market, traders submit orders that reflect their private observations and preferences, and learning from others' actions at no discretion helps to make informed trades and benefits price discovery. Manipulation strategies exploit such learning process, and rely on the instant order book information disclosed by standard market mechanisms. For instance, spoof orders are often placed at price levels just outside the best market quotes to mislead other investors to the largest possible extent, and are withdrawn with high probability before any market movement could trigger a trade.

To take advantage of such regular characteristics in misleading bids, our first HBL variation grants agents the flexibility to neglect limit orders at a specified price level when assembling a dataset to learn from. The strategy extends the standard HBL with a blocking parameter K, which specifies the index of a price level to ignore symmetrically from inside of the limit order book. For example, when $K = 1$, the trading agent constructs a dataset, $\mathcal{D} \setminus O_{K=1}$, by considering any order but the best bid and ask, whereas when $K = 0$, the agent learns from $\mathcal{D}$ and acts the same as the standard HBL. With this additional strategic parameter, agents may strategically exclude certain price levels where spoof orders are likely to appear. However, ignoring orders may also come at the cost of less effective learning, especially when information that conveys true insight is blocked from the belief function. In Section 4.4, we further evaluate these trade-offs.

HBL with Price Offsets. Our second HBL variation considers full order book information, and simply translates the target price $\mathrm{P}_i^*(t)$ derived by surplus maximization (4) with a random offset uniformly drawn from $[R_{\min}, R_{\max}]$. Specifically, a trader i arriving at time t with the calculated price $\mathrm{P}_i^*(t)$ submits a limit order for a single unit of the security at price

$$p_i(t) \sim \begin{cases} U[\mathrm{P}_i^*(t) - R_{\max}, \mathrm{P}_i^*(t) - R_{\min}] & \text{if buying,} \\ U[\mathrm{P}_i^*(t) + R_{\min}, \mathrm{P}_i^*(t) + R_{\max}] & \text{if selling.} \end{cases}$$

A positive offset can be viewed as a hedge against misleading information, effectively shading the bid to compensate for manipulation risk. Negative offsets increase the probability of near-term transaction, which may have benefits in reducing exposure to future spoofing. Offsets (positive or negative) may also serve a useful correction function even when manipulation is absent. In particular, negative offsets may compensate for the myopic nature of HBL optimization Eq. (4), which considers only the current bid, ignoring subsequent market arrivals and opportunities to trade additional units. Our design here is in line with prior literature [11, 12] that refines the original HBL to become more competitive.

[1] It uses the observed frequencies of *transacted* and *rejected* orders (T and R), *bids* and *asks* (B and A), and orders with prices *less* than or equal to and *greater* than or equal to P (L and G) within the HBL's memory. For example, $\mathrm{TBL}_t(\mathrm{P} \mid \mathcal{D})$ is the number of transacted bids found in memory at time t with price less than or equal to P.

3.4 Exploitation and Spoofing Strategies

We follow the CDA spoofing model [16] in using an exploitation agent (EXP) to control market environments with and without manipulation. We extend the manipulation strategy with a spoof-price-level parameter, which allows EXP to flexibly inject and maintain spoof orders at a single selected price level in the order book.

The strategy includes three stages. At the beginning of a trading period $[0, T_{\text{spoof}}]$, EXP buys as many units as possible by accepting any sell order at prices lower than the fundamental mean $\bar{r}$. During the second stage $[T_{\text{spoof}}, T_{\text{sell}}]$, if EXP does not spoof, it simply waits until the selling stage. If EXP also manipulates, it submits and maintains spoof buy orders at one tick behind a chosen price level K with some large volume $Q_{\text{sp}} \gg 1$. Spoof orders aim to artificially boost prices so that units purchased earlier in the first stage can be later sold at higher prices. The spoof-price-level parameter allows EXP to strategically respond to HBL traders who selectively block orders. In such a game, EXP may sacrifice the manipulation effect by placing spoof orders at less influential price levels that are less likely to be ignored by HBL traders. In Section 4.4, we conduct control experiments to empirically evaluate how manipulation intensity can be affected by different price levels that EXP chooses to place spoof orders. During the last stage $[T_{\text{sell}}, T]$, EXP begins to sell by accepting any buy orders at a price higher than $\bar{r}$. EXP, if it also manipulates, continues to spoof until the end of the trading period T or when all units previously purchased are sold.

4 EMPIRICAL GAME-THEORETIC ANALYSIS

We conduct agent-based simulation of the market model described in Section 3 to evaluate the proposed HBL variations with respect to both effectiveness in learning and robustness against manipulation. We explore a range of market environments varying in fundamental volatility and observation noise. A *game* is defined by a specific market environment and a strategy set from which each background trader can choose. For each game, we evaluate a wide variety of agent strategy assignments, or *profiles*. To evaluate strategies in a profile, we calculate the average payoff of agents adopting the same strategy in the profile, averaged over at least 40,000 simulations to account for stochastic effects such as the market fundamental series, agent arrival patterns, and private valuations.

As not all strategic contexts are equally relevant for evaluation, we focus on measuring trading strategies and market performance in *equilibrium*, where agents have no incentive to deviate to other strategies. Specifically, we use the simulation results to estimate a game model over the heuristic strategies, and derive approximate Nash equilibria over this restricted strategy space. Based on fixed equilibrium strategy profiles, we further perform control experiments to quantify the manipulation effect and the competitiveness of any newly introduced trading strategy. In such paired simulation instances, we control all other stochastic factors (e.g., agent arrivals, fundamental evolution, and private values), so that any change in agent behavior is caused by the experimental factor of interest.

Env	Shock σ_s^2	Observation σ_n^2
LSHN	10^5	10^9
MSMN	5×10^5	10^6
HSLN	10^6	10^3

(a) Market environments.

Strategy	ZI_1	ZI_2	ZI_3	ZI_4	ZI_5	HBL_1	HBL_2	HBL_3	HBL_4	HBL_5	HBL_6	HBL_7
K	NA	NA	NA	NA	NA	0	1	2	0	0	0	0
$R_{\min}$	0	0	0	0	0	0	0	0	-10	-20	-40	-80
$R_{\max}$	1000	1000	1000	500	250	0	0	0	0	0	0	0
η	0.4	0.8	1	0.8	0.8	1	1	1	1	1	1	1

(b) Background trading strategies.

Table 1: Market environments and background trading strategies included in empirical game-theoretic analysis.

This section is structured as follows. Section 4.1 specifies a set of parametrically defined market environments. Section 4.2 summarizes the empirical game-theoretic analysis (EGTA) methodology that we adopt to identify equilibrium solutions. In Sections 4.4, 4.5, and 4.6, we present the EGTA results on the effectiveness and robustness of the proposed HBL variations.

4.1 Market Environment

The global fundamental time series is generated according to Eq. (1) with a fundamental mean $\bar{r} = 10^5$ and a mean reversion constant $\kappa = 0.05$. Each trading period lasts $T = 10,000$ discrete time steps. We consider three environments listed in Table 1(a) that vary in fundamental shock variance σ_s^2 and observation noise variance σ_n^2. They cover representative market conditions that can affect HBL's preference on learning from market information to different extent (e.g., when the market shock is high, prices fluctuate more and become hard to predict; when observation noise is high, agents can glean only limited information from their own observations and may prefer to learn from the market's aggregated order book information.) We use *LSHN* to denote a market with low shock and high observation noise, *MSMN* a market with medium shock and medium observation noise, and *HSLN* a market with high shock and low observation noise.

For each environment, we consider two settings where background traders are provided with the two proposed HBL variations respectively, and compare equilibrium outcomes to those of a market where background traders can only choose from the standard HBL and ZI strategies. In our final set of explorations, we further offer background traders the option to combine the two variations. This altogether gives us a total of twelve market settings, hence 24 games with and without spoofing.

Our market is populated with 64 background traders and a single exploiter. Background traders arrive at the market according to a Poisson distribution with rate $\lambda_a = 0.005$ and observe a noisy fundamental o_t. Private values are drawn from a Gaussian distribution with zero mean and a variance of $\sigma_{PV}^2 = 5 \times 10^6$. The maximum number of units that they can hold at any time is $q_{\max} = 10$. Table 1(b) specifies our background trading strategy set, comprising seven parameterized instances of HBL and five of ZI.[2] Background traders can choose from this restricted strategy set to maximize payoffs.

EXP follows the strategy described in Section 3.4. If it manipulates, EXP submits spoof orders with volume $Q_{\text{sp}} = 200$ at time $T_{\text{spoof}} = 1000$. After T_{spoof}, EXP maintains its spoof orders at a tick behind a chosen price level K on the bid side of the order book to push prices up. At time $T_{\text{sell}} = 2000$, it starts to liquidate its position by selling units at prices above $\bar{r}$.

4.2 EGTA Process

We provide a brief overview of empirical game-theoretic analysis (EGTA), a methodology for performing strategy selection to find equilibria in games defined by heuristic strategy space and simulated payoff data [17]. EGTA takes an iterative process to identify candidate equilibria in subgames (games over strategy subsets), and searches for potential deviations until a candidate is confirmed. Exploration starts with subgames where all agents play a single strategy, and incrementally spread to other strategies. Equilibria from a subgame are considered as candidate solutions of the full game, and are refuted if we identify a beneficial deviation to a strategy outside the subgame set. If we examine all deviations without refuting, the candidate is confirmed. We continue to refine the empirical subgame with additional strategies and corresponding simulations until at least one equilibrium is confirmed and all non-confirmed candidates are refuted.

We model the market as a role-symmetric game, which is defined by an environment and agents representing two roles: background traders and a single exploiter. Since game size can grow exponentially in the number of players and strategies, we apply the *deviation-preserving reduction* (DPR) [18] technique to approximate large games with many agents as games with fewer players. We obtain this approximation through aggregation, which preserves payoffs from single-player deviations. To facilitate DPR, we choose values to ensure that the required aggregations come out as integers. For example, in this study, we choose 64 background traders and one exploiter, so that a market can be aggregated to a smaller one with four background traders and one exploiter; as one background trader deviates to a new strategy, the remaining 63 can be represented by three aggregate traders.

4.3 Standard HBL

We start with our baseline market environments (Fig. 4 dark grey columns) where background traders are restricted to choose from the baseline standard HBL strategy and five parametrically different ZI strategies in Table 1(b).[3] Fig. 4 (dark grey columns) shows that (1) the learning-based trading strategy is more widely preferred in environments where fundamental shock is low and observation noise is high (e.g., LSHN is the most learning-friendly environment);

[2] We explored a much wider range of background strategies in our preliminary set of experiments, and only include those that are competitive in at least one market environment in Table 1(a).

[3] Details of strategy profiles and market surpluses of all found equilibria in games with and without manipulation are available in an online appendix at https://www.dropbox.com/s/dmkaol5mypbwafy/learning-trading-strategies-icaif20-appendix.pdf.

Env	HBL	$\text{HBL}_{K=1}$	$\text{HBL}_{K=2}$	$\text{HBL}_{K=3}$	$\text{EXP}_{K=1}$	$\text{EXP}_{K=2}$	$\text{EXP}_{K=3}$	EXP
LSHN	651	643^{*}	652	651	494	$475^{*,**}$	470^{*}	468^{*}
MSMN	655	643^{*}	652^{*}	652	330	$313^{*,**}$	305^{*}	305^{*}
HSLN	645	634^{*}	646	646	221	195^{*}	202^{*}	199^{*}

Table 2: Average payoffs of learning-based trading agents and the EXP agent as they deviate from the equilibrium profiles found in Section 4.3. We deviate either HBL or EXP, one at a time, to its corresponding strategy variation. Asterisks denote statistical significance at the 1% level of the paired t-test for payoffs compared to either HBL or $\text{EXP}_{K=1}$(*) and EXP(**).

(2) the presence of spoofing generally hurts the competitiveness of the learning-based strategy and reduces background-trader surpluses. These findings confirm our hypotheses and results from prior studies [15, 16]. We next move to evaluate our main extension: the two HBL variations.

4.4 HBL with Selective Price Level Blocking

Blocking Orders in Non-manipulated Markets. Learning traders who choose to ignore certain orders face a natural trade-off between losing useful information and correctly blocking spoof orders to become immune to manipulation. We first examine, under *non-spoofing* environments, how learning effectiveness may be compromised by excluding orders at each price level. Starting with the equilibrium strategy profile of each non-spoofing market environment found in Section 4.3,[4] we perform control experiments by letting the HBL traders ignore orders from a selected price level throughout the trading period. Table 2 compares the payoffs obtained by HBL in its standard form and variations that respectively block orders at the first, second, and third price level in the order book. We find that consistently across market settings, HBL agents benefit the most by learning from market best bids and asks, and can achieve fairly similar performance even when orders at a selected level beyond the market quotes are ignored.

Placing Spoof Orders at Different Price Levels. In response to HBL traders who ignore price levels, we extend EXP to be able to strategically place spoof orders behind a chosen price level. Here, we start with the same set of equilibrium strategy profiles, and conduct control experiments to evaluate how injecting spoof orders at different levels can change the manipulation effect, even when learning agents are considering the full order book (i.e., adopting the standard HBL strategy). We measure the effectiveness of a spoofing strategy by EXP payoff as well as the market price deviation. The price deviation at a specific time is calculated as the most recent transaction price of a game with manipulation minus that of the paired game without manipulation, thus quantifying the extent to which HBL traders are spoofed. Experiments show that EXP benefits the most by spoofing behind the best bid (Table 2), and moving spoof orders to less competitive levels only reduces EXP payoff. We further confirm this weakened manipulation effect in Figure 2, which showcases market price deviations caused by different spoofing strategies in the MSMN environment. We find the market price rise diminishes as spoof orders are placed further away from the best bid.

[4]We randomly select one equilibrium if there is more than one in certain environments.

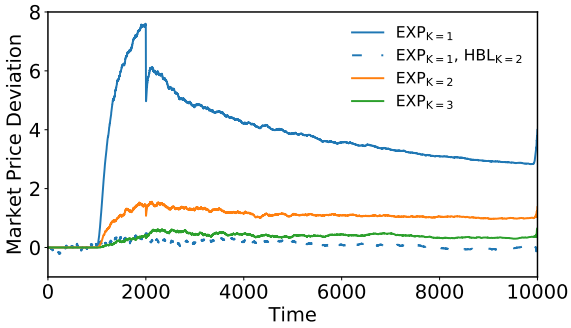


Figure 2: Market price deviations caused by spoof orders placed behind different price levels in the order book.

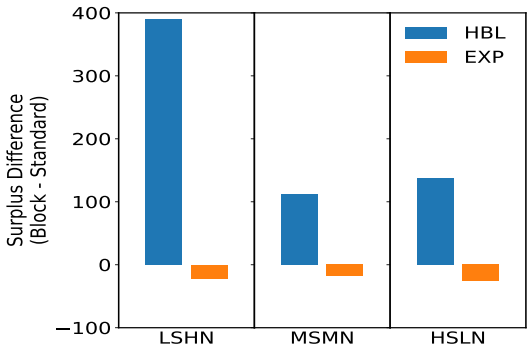
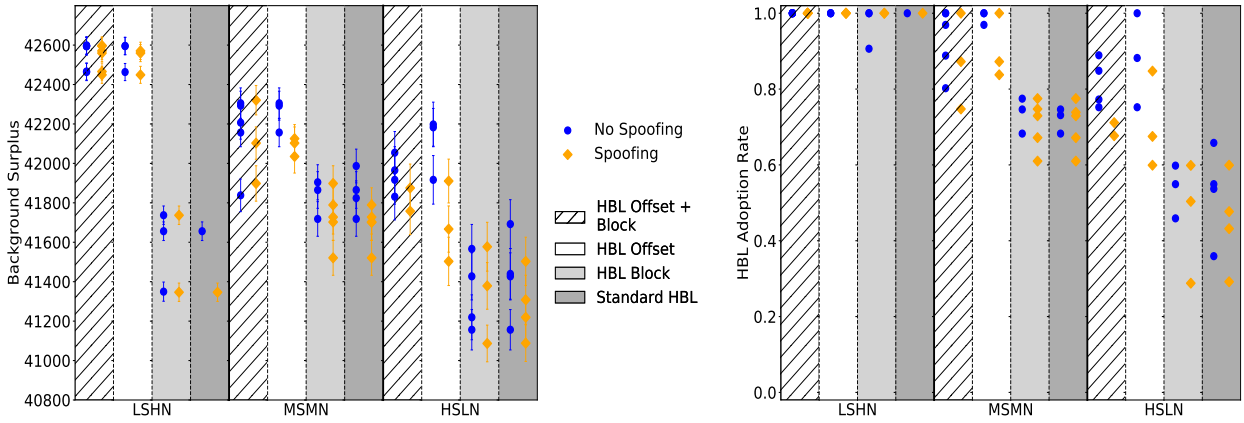


Figure 3: Correctly blocking spoof orders increases background-trader surplus and mitigates manipulation.

Re-equilibrating Games. Our preliminary explorations on a restrictive set of spoofing strategies suggest that spoof orders are more likely to appear near the market quotes to maximize manipulation effect. In response, HBL traders who adapt to the presence of spoofing may naturally block orders at such levels. Fig. 3 shows that by blocking the correct level, HBL traders can significantly increase their payoffs, and reduce the amount EXP could profit via manipulation. This mitigated manipulation effect is further verified in Fig. 2, which shows market price deviations (the dashed blue line) are close to zero.[5]

Given such beneficial payoff deviations, in the final set of experiments, we conduct EGTA to find Nash equilibria in games where background traders may choose any trading strategies from the ZI family and HBLs that block a selected price level (refer to Fig. 4 light

[5]In the dashed line, price differences are not strictly zero before spoofing (time 1000), as traders who adopt $\text{HBL}_{K=2}$ consistently block orders throughout the trading period.



(a) Background-trader surpluses achieved in equilibrium.

(b) HBL (and its variations) adoption rates in equilibrium.

Figure 4: Total background-trader surpluses and HBL strategy adoption rates achieved at equilibria across different market settings. For each market environment, we compare four settings where background traders are respectively provided with the standard HBL strategy (dark grey), HBL with selective price blocking (light grey), HBL with price offsets (white), and HBL that combines the two variations (striped). Each marker specifies one equilibrium outcome in markets with spoofing (orange) and without spoofing (blue). Error bars represent 95% confidence intervals on background-trader surpluses.

grey columns). We find that (1) adding the blocking strategic parameter does not affect the competitiveness of learning-based strategies to the ZI ones (HBL adoption rates in equilibrium remain in similar ranges as those of markets where only the standard HBL strategy is provided); and (2) the extended order blocking ability improves the learning robustness of HBL traders (compared to surplus decreases caused by manipulation in markets where background agents are restricted to the standard HBL, background-trader surpluses are no longer significantly reduced when agents can strategically block orders in the face of manipulation). In other words, background traders who learn from market information but also strategically ignore orders can have both the robustness against manipulation and a comparable effectiveness in non-manipulated markets.

4.5 HBL with Price Offsets

Exploring Different Price Offsets. Different from the first variation which strategically constructs $\mathcal{D}$, our second HBL variation instead relies on a price adjustment to adapt to different market conditions. We start with exploring a set of price offset intervals $[R_{\min}, R_{\max}]$, ranging from positive values that understate the learned offer prices (e.g., similar to price shading) to negative values that adjust prices to become more competitive. Similar to Section 4.4, we conduct control experiments basing off of equilibrium profiles found in Section 4.3, and deviate agents who adopt the standard HBL to use corresponding price offsets. Fig. 5[6] showcases in the MSMN non-spoofing environment how the HBL surpluses and the total number of transactions vary in markets where HBL traders adopt different offset intervals. We find adjusting learned prices with a range of negative offsets can be generally beneficial in our setting where agents have reentry opportunities. It increases HBL payoffs and facilitates transactions, thus improving overall price convergence in markets.

[6]HBL with positive offsets usually achieves much lower payoffs. For presentation simplicity, we cropped the surplus decrease at -200 in Fig. 5.

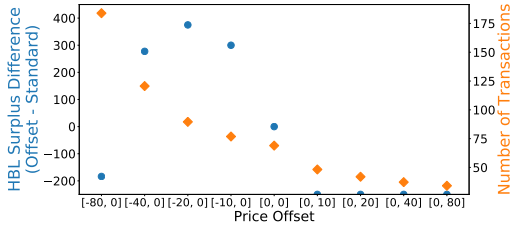


Figure 5: Average HBL surplus differences and total number of transactions in non-spoofing markets where HBL traders use different price offsets.

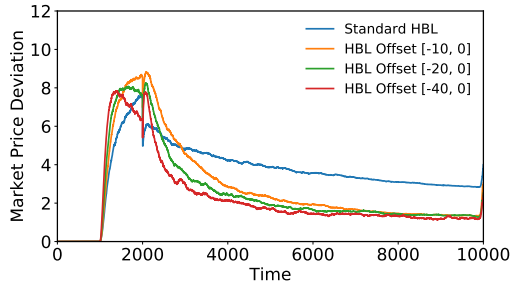


Figure 6: Market price deviations caused by spoofing in markets where HBL traders use different price offsets.

Spoofing HBL with Price Offsets. To test the effectiveness of spoofing against the new HBL variation, we further have the $\text{EXP}_{K=1}$ spoof in markets where the learning background traders respectively adopt the standard HBL, $\text{HBL}_{[-10,0]}$, $\text{HBL}_{[-20,0]}$, $\text{HBL}_{[-40,0]}$. Fig. 6 compares market price deviations caused by spoof orders in those markets. We find that though all markets experience initial price rises as a result of misled pricing beliefs, the spoofing effect tends to wear off faster in markets where HBL traders adopt negative price offsets. This may be because negative offsets promote

near-term transaction; as more transactions happen, HBL traders can glean true information from the transaction prices to construct belief functions Eq. (3), whereas the $\text{EXP}_{K=1}$ may only place spoof orders at lower prices due to the widened bid-ask spreads. Indeed, we find that markets populated with the standard HBL, $\text{HBL}_{[-10,0]}$, $\text{HBL}_{[-20,0]}$, and $\text{HBL}_{[-40,0]}$ respectively have average spoof-order prices of 99980, 99966, 99963, and 99964.

Re-equilibrating Games. Finally, we conduct EGTA in games with and without spoofing to find Nash equilibria where background traders can choose from ZI strategies and HBL variations that adjust learned prices with certain offsets (Fig. 4 white columns). Equilibrium results show that the extended price offsets tend to largely improve HBL's profitability and background-trader surpluses, in both markets with and without manipulation. Such price adjustments can especially help learning traders to better adapt to high shock environments where prices are less predictable from past observations. However, the extended offsets may not directly address manipulation and improve learning robustness against spoofing.

4.6 Combine Order Blocking and Price Offsets

Since the second HBL variation demonstrates a general improvement in both settings with and without manipulation, we augment this variation with price level blocking to reduce vulnerability to spoofing. Specifically, we extend the background trading strategy set in Table 1(b) with three strategies: $\text{HBL}^{K=2}_{[-10,0]}$, $\text{HBL}^{K=2}_{[-20,0]}$, and $\text{HBL}^{K=2}_{[-40,0]}$, which appear to be competitive in our preliminary explorations. We conduct EGTA in a similar manner across market environments with and without spoofing. Equilibrium outcomes (Fig. 4 striped columns) show that (1) compared to markets where only the standard and the price-blocking HBL are provided, HBL that combines the two variations is more widely preferred and can help to increase overall background-trader surplus in equilibrium; and (2) across all environments, background-trader surpluses in markets with and without spoofing fall roughly into the same ranges. These suggest that by combining the two proposed variations, HBL traders can enjoy both improved competitiveness and robustness against manipulation.

5 CONCLUSION

We study learning-based trading strategies by which individual traders can adopt to utilize order book information, but in less vulnerable ways when prices can be manipulated through spoofing. We explore two strategy variations based on the standard HBL strategy, which constructs a belief function from any observed trading activities. Our first HBL variation considers common characteristics of spoofing activities, and works by offering agents the flexibility to neglect limit orders at a specified price level when assembling a dataset to learn from. Our second variation learns from full order book information, and later adjusts the target price derived from surplus maximization with a random offset to correct any biases in the learning process. We employ agent-based simulation to evaluate the proposed variations in terms of their effectiveness in non-manipulated markets and the robustness against manipulation. Background traders can adopt the non-learning ZI strategies or HBL, in its basic form or the two proposed variations. We conduct extensive simulation to characterize the strategic interactions between agents in our defined empirical game model, and compare outcomes in equilibrium where agents optimally react to each other's presence.

Our analysis show that the first HBL variation offers learning traders a way to strategically block orders to improve robustness against spoofing, while achieving similar competitiveness in non-manipulated markets. Our second HBL variation exhibits a general improvement over baseline HBL, in both markets with and without manipulation. Further explorations suggest that traders can enjoy both improved profitability and robustness by combining the two HBL variations. We note that our results reflect the specific modeling and simulation choices adopted, and several factors (e.g., sampling error, restricted strategy and environment exploration) may affect our equilibrium analysis. Despite these limitations that are inherent in any complex modeling effort, we believe our proposed strategy variations and analysis can further help the design and evaluation of other strategic adjustments that individual traders could adopt to improve robustness against manipulation or other fraudulent behaviors.

REFERENCES

[1] Bruno Biais and Paul Woolley. 2012. *High Frequency Trading.* Technical Report. Toulouse University.

[2] Tanmoy Chakraborty and Michael Kearns. 2011. Market making and mean reversion. In *11th ACM Conference on Electronic Commerce*. 307–314.

[3] Thomas E. Copeland and Dan Galai. 1983. Information effects on the bid-ask spread. *Journal of Finance* 38, 5 (1983), 1457–1469.

[4] J. Doyne Farmer, Paolo Patelli, and Ilija I. Zovko. 2005. The predictive power of zero intelligence in financial markets. *Proceedings of the National Academy of Sciences* 102 (2005), 2254–2259.

[5] Thierry Foucault, Ailsa Röell, and Patrik Sandås. 2003. Market making with costly monitoring: An analysis of the SOES controversy. *Review of Financial Studies* 16, 2 (2003), 345–384.

[6] Steven Gjerstad. 2007. The competitive market paradox. *Journal of Economic Dynamics and Control* 31 (2007), 1753–1780.

[7] Dhananjay K. Gode and Shyam Sunder. 1993. Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality. *Journal of Political Economy* (1993), 119–137.

[8] Blake LeBaron. 2006. Agent-based computational finance. In *Handbook of Computational Economics*, Leigh Tesfatsion and Kenneth L. Judd (Eds.). Vol. 2. Elsevier, 1187–1233.

[9] Mark Paddrik, Roy Hayes, Andrew Todd amd Steve Yang, Peter Beling, and William Scherer. 2012. An agent based model of the E-Mini S&P 500 applied to Flash Crash analysis. In *IEEE Conference on Computational Intelligence for Financial Engineering and Economics*. 1–8.

[10] Matt Prewit. 2012. High-frequency trading: Should regulators do more. *Michigan Telecommunications and Technology Law Review* 19 (2012), 131–161.

[11] G. Tesauro and J.L. Bredin. 2002. Strategic sequential bidding in auctions using dynamic programming. *First International Joint Conference on Autonomous Agents and MultiAgent Systems* (2002), 591–598.

[12] G. Tesauro and R. Das. 2001. High-performance bidding agents for the continuous double auction. *Third ACM Conference on Electronic Commerce* (2001), 206–209.

[13] Elaine Wah, Sébastien Lahaie, and David M. Pennock. 2016. An Empirical Game-Theoretic Analysis of Price Discovery in Prediction Markets. In *26th International Joint Conference on Artificial Intelligence*. 510–516.

[14] Elaine Wah, Mason Wright, and Michael P. Wellman. 2017. Welfare effects of market making in continuous double auctions. *Journal of Artificial Intelligence Research* 59 (2017), 613–650.

[15] Xintong Wang, Yevgeniy Voroveychik, and Michael P. Wellman. 2018. A Cloaking Mechanism to Mitigate Market Manipulation. In *27th International Joint Conference on Artificial Intelligence*. 541–547.

[16] Xintong Wang and Michael P. Wellman. 2017. Spoofing the limit order book: An agent-based model. In *16th International Conference on Autonomous Agents and Multiagent Systems*. 651–659.

[17] Michael P. Wellman. 2016. Putting the agent in agent-based modeling. *Autonomous Agents and Multi-Agent Systems* 30 (2016), 1175–1189.

[18] Bryce Wiedenbeck and Michael P. Wellman. 2012. Scaling simulation-based game analysis through deviation-preserving reduction. In *11th International Conference on Autonomous Agents and Multiagent Systems*. 931–938.