DOI:
10.1109/EAIS.2016.7502499

Document Version
Peer reviewed version

Citation for published version (APA):
Martinez Miranda, E., McBurney, P., & Howard, M. J. W. (2016). Learning Unfair Trading: A Market Manipulation Analysis from the Reinforcement Learning Perspective. In *Proceedings of the 2016 IEEE Conference on Evolving and Adaptive Intelligent Systems, EAIS 2016* (pp. 103-109). Article 7502499 Institute of Electrical and Electronics Engineers Inc.. https://doi.org/10.1109/EAIS.2016.7502499

Learning Unfair Trading: a Market Manipulation Analysis From the Reinforcement Learning Perspective

Enrique Martínez-Miranda, Peter McBurney, Matthew J. W. Howard
Department of Informatics
King's College London
Strand, London, WC2R 2LS
Email: {enrique.martinez_miranda, peter.mcburney, matthew.j.howard}@kcl.ac.uk

***Abstract*—Market manipulation is a strategy used by traders to alter the price of financial assets. One type of manipulation is based on the process of buying or selling assets by using several trading strategies, among them *spoofing* is a popular strategy and is considered illegal by market regulators. Some promising tools have been developed to detect manipulation, but cases can still be found in the markets. In this paper we model *spoofing* and *pinging* trading from a macroscopic perspective of profit maximisation, two strategies that differ in the legal background but share the same elemental concept of market manipulation. We use a reinforcement learning framework within the full and partial observability of Markov decision processes and analyse the underlying behaviour of the manipulators by finding the causes of what encourages the traders to perform fraudulent activities. Procedures can be applied to counter the problem as our model predicts the activity of the manipulators.**

***Index Terms*—Asset price manipulation, spoofing, pinging, MDP, generative model.**

I. Introduction

Market microstructure is a branch of finance concerned with the analysis of the trading process arising from the exchange of assets under a given set of rules [1]. In *double auction markets*, this exchange of assets occurs when the buy and sell sides agree on the amount to pay/receive for the trade, depending on the different trading strategies implemented by both sides; the latter being a plan of actions designed to achieve profitable returns by buying or selling financial assets [2].

While trading strategies are meant to follow the rules of regualted markets, some traders misbehave by manipulating the price of the assets being traded. For instance, some traders use strategies like the *pump and dump* to spread false information to other market participants that may affect the perceived price [3]. Other traders follow strategies like *ramping*, *wash trading*, *quote stuffing*, *layering*, and *spoofing* among others, to buy or sell the asset and artificially inflate or deflate prices and obtain profits.

Spoofing is one of the most popular strategies that uses *spoof orders* to improve the price of a given asset and is considered illegal by market regulators [4]. *Pinging* is a similar strategy used by high-frequency traders (HFTs) whereby they place orders without the intention of execution, but to find liquidity not fully displayed in the *limit order book* (LOB)[1]. It has caused controversy as it can also be viewed as a manipulative strategy [5].

In the scientific literature, studies have mainly focused to develop discriminative methods for detection. There has been little analysis of the root causes of why traders take manipulative actions, beyond the assumption that they are tempted by greater profits. However, for this to have any explanatory power, it is required to examine the interaction between market dynamics and instruments (*e.g.,* liquidity, transaction taxes, fees) come together to incentivise this behaviour. Understanding this is crucial if market regulators are to develop countermeasures to discourage or preclude fraudulent trading.

In this paper, we propose a generative model of spoofing and pinging in the context of *portfolio growth* maximisation, *i.e.,* the expected capital appreciation over time of an investment account. We use a reinforcement learning agent that simulates the behaviour of the *spoofing trader* in the context of Markov decision processes (MDP) within an environment where transitions and rewards do not change in time. We also show that a similar framework can be used to model the *pinging trader* with the introduction of partial observability (reflecting hidden states of the LOB). From this, we are able to examine the main question of this study, namely: Under what conditions are spoofing and pinging optimal strategies compared to *honest* behaviour while seeking for growth maximisation? Our results provide a granular analysis of the incentives to manipulative behaviour, identifying separate causes such as risk of a manipulative action failing, and the risk of incurring increased transaction costs. The results can be viewed as recommendations to market regulators as to how to discourage market manipulation.

II. Related Work

Research on price manipulation has been done using several approaches. Some authors have developed analytical models with the intention to investigate manipulative strategies

[1]In double auction markets, the *limit order book* is a listing of all outstanding buy and sell orders on a given asset, used by traders to make decide prices at which to place their own buy/sell orders.

under the traditional framework of *rational agents* trying to maximise their expected utility [6]–[8]. Other researchers have applied data driven approaches with the aim to present empirical evidence of stock price manipulation [9], [10].

Also, behavioural stances have been mixed with theoretical and data driven approaches to tackle with the problem of price manipulation as an intentional act [11] and investor behaviour [11]. Though this approach tries to analyse the manipulation process from the behavioural perspective, financial markets by themselves express a global sentiment from all investors that is ultimately reflected in the prices and an alternative way to analyse this problem is precisely by studying the effect of manipulation relative to the incentives present in the market, just as we propose in this paper.

Furthermore, discriminative models are intended to detect market manipulation based on empirical data. By using economic and statistical analysis it is possible to detect manipulation only after the execution, suggesting that the existence of regulatory framework may be inefficient [12]. Machine learning techniques have also been applied for detection of manipulation. Based on trading data, some authors suggest that Artificial Neural Networks and Support Vector Machines are effective techniques to detect manipulation [13]. Others suggest that a method called "hidden Markov model with abnormal states" is capable to model and detect price manipulation patterns, but further analysis is necessary [14]. Data mining methods for detecting intraday price manipulation have been used to classify and identify patterns linked to market manipulation at different time scales, but further research is needed to address the challenge on detecting the different forms of manipulation [15]. Furthermore, Naïve Bayes is a good classifier for predicting potential trades associated to market manipulation [16]. For the case of *spoofing* trading, detection can be done with the implementation of supervised learning algorithms [17], or can be identified by modelling trading decisions as MDPs and using Apprenticeship Learning to learn the reward function [18].

Though research is extensive in the area of market manipulation, few develop generative models of what encourages these economic agents to follow the disruptive strategies. Furthermore, few of them provide recommendations to regulatory entities and/or firms [19] to encourage traders to stop this harmful behaviour. Different to the discriminative models that are intended to distinguish the manipulative behaviour from other strategies, we use the (PO)MDP approach to model spoofing/pinging as it predicts the behaviour of manipulators in terms of market conditions, thus providing a powerful tool for analysis of the causes of manipulative strategies, and suggesting possible remedies for market regulators.

III. Background Trading Concepts

In this section, we provide an account background financial concepts related to manipulative trading. We start with a description of the process of trading for portfolio growth optimisation, before going on to describe market manipulation through spoofing and pinging.

A. Trading in a Bull Market

We are focused on modelling two trade-based market manipulation strategies as follows. Suppose there is a trader managing an investment portfolio on behalf of a brokerage firm with the objective of maximising trading profits through portfolio growth in the short/medium term. Suppose the agent is trading in a futures market and the portfolio consists of two different contracts, α and β, with a market full of optimism so prices are rising (a situation better known as a *bull market*).

Mathematically, the capital of the investment account at a given market *tick* $t \in [0, T]$ (where a *tick* represents the execution of a new trade in the market, either from the trader or any other participant) can be written as

$$I_t = a_t + c_t, \tag{1}$$

where $a_t = a_t^\alpha + a_t^\beta$ is the capital associated to the *market value* of the contracts α and β, and c_t is the cash to be used for future purchases of more contracts. The variable a_t changes at every *tick* since the prices of the contracts are following the *bullish* trend, while c_t changes due to cash inflows/outflows (by the sale/purchase of contracts). The net profit of the investment over a *tick window* $[0, T]$ is

$$R = G_T - \sum_{t=0}^{T} \zeta_t, \tag{2}$$

where $G_T = I_T - I_0$ is the investment growth, and ζ_t are the direct transaction costs associated to the trading of the contracts (such as exchange and government fees).

Under bull market conditions, one way in which the trader can profit from the portfolio's growth is with a simple *buy and hold* strategy, an almost risk-free strategy whereby she purchases contracts α and β and simply waits, in the long term, for the prices to rise before selling for a profit. However, the trader may, alternatively, be aiming for a higher target growth G_T^* in the short/medium term, requiring a more active strategy than the "buy and hold", *i.e.,* buying and selling contracts α and β, subject to the transaction costs ζ_t.

For this, the trader can behave in several different ways. First, the agent may trade *honestly*, *i.e.,* following all the market rules, by buying more contracts or selling them when she believes is profitable. Fig. 1 shows this behaviour with a trader starting with 1000 contracts and 10 million monetary units as the account's net capital value. While the market evolves, she takes honest actions represented by the filled triangles, reaching new levels of growth G_t and paying paying the transaction costs, ζ_t. Alternatively, the agent may act as a *manipulative* trader to control the price of the contracts in order to accelerate the growth process and quickly reach the desired level G_T^*, just as seen in the non-filed triangles of Fig. 1.[2] In either case, following the transaction, the trader ends up with a different proportion of the contracts α and

[2]In our example, the historical prices are not affected by our strategies, but we use them as a reference on how a manipulative strategy may lead to higher trading profits when compared to a honest strategy and against the buy and hold.

β, re-balancing the quantities a_t and c_t and thereby finding herself in a new level of growth G_t at a given *tick* t.

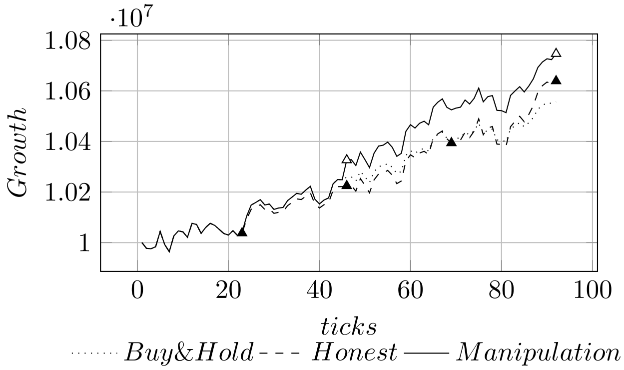

Fig. 1: A simulation of profits gained from different trading strategies during a bull market period. The data is taken from the closing prices of the market indexes *S&P 500* and *NASDAQ Composite* to simulate the contracts α and β, respectively, in the period February 27, 1995 to May 5, 1995.

Note that, the above assumes that the agent can freely transition between different levels of growth G_t by executing trades. In reality, this depends on the trades of other market participants where the buying and selling prices match those of our agent, allowing the exchange of assets to proceed. The degree to which such trades can occur is known as market *liquidity*, and must also be taken into account in models of manipulative trading. For example, if liquidity is poor in the contract to trade, actions taken by the trader can lead to no change in the level of growth G_t. A manipulator can take advantage of this situation by placing a large order that may gain the interest of other market participants and start a process of price improvement.

B. Price Manipulation by Spoofing

Spoofing is an illegal trading strategy used by traders intended to manipulate the price of a given asset by placing large orders (spoofing orders) without the intention of execution, but to give misleading information to other market participants in terms of the asset's supply and demand, thus producing a change in the price [20]. Once the price is affected, the trader cancels the spoofing order and places the real order on the opposite trading side. This real order can be either a limit (priced orders with fixed volume) or market order (non-priced orders with a fixed volume that cross the *spread*, that is, the price difference between the best buy (or *bid*) and sell (or *ask*) quotes listed in the LOB), but that depends on the spoofer's risk profile–a real limit order can be placed but may have the risk of not being executed, whereas a market order can cross the spread (thus securing execution) but the profits produced may not be as high as those gained from limit orders. For our purposes, we consider our trader only places limit orders and for the spoofer his real limit order is immediately executed before the price recovers to the previous level when being affected by the spoofing order.

C. Price Manipulation by Pinging

Pinging is a strategy implemented by HFTs by exploiting their speed advantage with the intention to *ping* the market and detect the trading pattern of *hidden liquidity* (orders that are not fully displayed in the LOB as is the case of large orders placed by institutional investors), in *dark pools* (private venues where the exchange of assets is not visible to the general public, whose prices depends on the current market prices of regulated markets), thus finding a potential interest of buying or selling [21]. In pinging, the HFT is the one that provides improved prices in her favour as she can sell/buy high/low. As mentioned in [5, see pp. 613 and references], once the HFT has detected the presence of a large order in a dark pool, she takes a small loss at first by eliminating the current liquidity in a regulated market and then she places new orders (at both, the established market and the dark pool) at better prices that, after execution, will produce profits.

IV. Trading as a Markov Decision Process

A representation of growth maximisation is provided in Fig. 2, where we changed the notation of G_t to s_t, $t \in [0,4]$, and G^*_T to s^*. There, the four growth levels correspond to holding a portfolio containing different proportions of contracts, for example, in s_1 the trader holds one contract of type α and one contract β. If the trader chooses to buy a second α contract, *i.e.,* action "Buy α" ($\uparrow$), she transitions to growth level s_2–holding two α contracts and one β contract by paying the transaction costs ζ_1. Similarly, if she then chooses to sell the second α contract, *i.e.,* action "Sell α" ($\downarrow$), she will return to growth level s_1, now paying ζ_2 costs. Additionally, while in s_2 taking actions "Buy α" ($\uparrow$) and "Sell β" ($\leftarrow$), result in no change in the level of growth. This is due to orders placed by the trader that were never filled because the price was too high/low while trying to sell/buy the β/α contract. For action "Buy β" ($\rightarrow$) the trader faces the problem of poor liquidity in the asset (the obstacle).

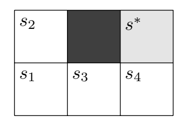

Fig. 2: Idealised representation illustrating the different levels of s_t while maximising investment growth.

A. Spoofing as a Markov Decision Process

In our representation, spoofing is illustrated as follows. Consider the case that, by taking spoofing actions, the trader can overturn the lack of liquidity in the asset β while in s_2. In Fig. 3, this corresponds to the obstacle switching from the top centre bin to the bottom centre bin, showing the effect of manipulation while purchasing more contracts. After taking action "Manipulative Buy β ($\Rightarrow$)", the trader obtains profits and finds herself in s_7, closer to s'^*. The two representations in Fig. 3 have the same levels of growth but with different conditions associated to market liquidity.

A model that fits the problem of manipulation under the representation Fig. 1 is that of a Markov Decision Process (MDP) [18], [22]. In general, an MDP is defined by the

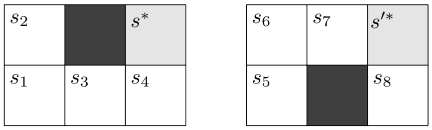

Fig. 3: Representation illustrating the effect of the spoofing in the process of investment growth maximisation.

tuple $\{S, A, T, R\}$, where S and A are sets of states and actions, respectively ($s \in S$ and $a \in A$), R is the set of rewards ($r \in R$), and T is a set of transition probabilities ($\{P(s'|s,a)\} \in T$, where $P(s'|s,a)$ represents the probability of transitioning to state s' from s after action a). Actions are taken according to the policy $\pi(s,a)$ that defines the probability of taking action a in state s.

Considering the growth s_t as the state variable, the problem for the trader is to find the best strategy for buying and selling contracts α and β, subject to the transaction costs (or rewards in the MDP model) ζ_t, in order to achieve the target short/medium term growth s^*. The complete set of states for spoofing is determined by the state representation in Fig. 3, while the actions correspond to the process of buying and selling contracts and are used by the trader to navigate within the state space. These actions are the honest ones, $A = \{\uparrow, \downarrow, \leftarrow, \rightarrow\}$ and similarly the set of manipulative actions $\mathbb{A} = \{\Uparrow, \Downarrow, \Leftarrow, \Rightarrow\}$ ("Manipulative Buy α", "Manipulative Sell α", "Manipulative Sell β", "Manipulative Buy β", respectively), and the "do nothing" action for the "buy and hold" $\mathcal{A} = \{\circ\}$, with $A = A \cup \mathbb{A} \cup \mathcal{A}$. The transition probabilities are the degree of liquidity the contracts α and β have at a given tick t; a good degree of liquidity will help the trader's orders to be filled and transition to a new level of growth, while low liquidity will restrict these transitions.

B. Pinging as a Partially Observable MDP

Similar to spoofing, pinging is illustrated in the representation of Fig. 4 by introducing the concept of observations that guide the trader on the actions to take. For example, having observation o_2 while in s_2 means there is hidden liquidity (the obstacle) in the sell side of the β contracts, so the HFT can produce profits by taking control over the prices in the regulated market while trading in the dark pool against the hidden liquidity. However, having the same observation in level s_6 means that such liquidity does not exists and taking the manipulative action may produce losses.

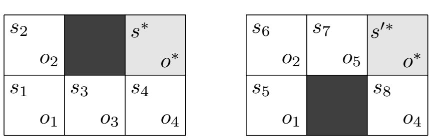

Fig. 4: Representation that illustrates pinging trading while trying to maximise the growth of the investment.

Pinging, as described in §III-C, can be modelled with a Partially Observable Markov Decision Process (POMDP) [23]. In general, the POMDP is defined by the tuple $\{S, A, T, R, \mathcal{O}, \Omega\}$, that is, the MDP tuple is extended with $\{\mathcal{O}, \Omega\}$, where $\mathcal{O}$ represents a set of observations ($o \in \mathcal{O}$) and Ω is a set of observation probabilities given states s and actions r ($\{O(o|s,r)\}$). For the POMDP, actions are taken according to the agent's belief of being on a given state and is calculated according to the observations. In our model, the observations represent the trader's detection of hidden liquidity while seeking for profits.

C. Solution

In both models, the trader has the objective to reach the goal s^* representing the maximum investment growth and, under a bull market, the highest profit comes from having the most contracts (the opposite also applies while in a *bear market* [when pessimism persist and prices tend to fall], where the trader may prefer to sell contracts). We analyse the state representations in Fig. 3 and Fig. 4 as both model a single agent's behaviour of acquiring contracts. Other grids with a more complex structure may also reproduce trading strategies, but manipulative behaviour may not emerge as an optimal control according to the simulated market conditions, thus eliminating the core of the analysis we present in this paper.

Regardless of whether manipulative trading is permitted or not, the best sequence of trading actions for the agent (optimal policy) can be determined in a straightforward manner through, for example, reinforcement learning. In this paper, for the MDP model this is achieved through simple value iteration [24] to find the optimal value function

$$V^{\pi^*}(s) = \max_{a \in A} \left[R(s,a) + \gamma \sum_{s' \in S} P(s'|s,a) V(s') \right], \quad (3)$$

where $0 < \gamma <$ is the discount factor. The POMDP formalism is intended to model states not fully observable, explaining why an observation function is needed to solve the problem. The observation function, $\Omega(a,s,o)$, is the probability of making observation o from state s after action a [25]. For POMDP's the solution is to find optimal policies with actions that maximises the value function. Based on the agent's current beliefs about the state (growth level), this value function can be represented as a system of simultaneous equations as

$$V^*(b) = \max_{a \in A} \left[\rho(b,a) + \gamma \sum_{b' \in B} \tau(b,a,b') V^*(b') \right], \quad (4)$$

where $b \in B$ is a *belief state*, $\rho(b,a) = \sum_{s \in S} b(s) R(s,a)$ are the expected rewards for the belief states; $\tau(b,a,b') = \sum_{\{o \in \mathcal{O} | b = b'\}} P(o|a,b)$, the state transition function.

The optimal value function considers the potential rewards of actions taken in the future, capturing the optimal actions that generate the most of rewards over the long-term. This argument enables us to examine the optimal actions in the (PO)MDP model ultimately determined by two factors: the reward and the transition functions.

V. Experiments

In this section we present the learnt policies from different experiments for the spoofing and pinging problems explained above. The purpose of these experiments are i) to demonstrate whether the manipulative behaviour emerges as an optimal strategy when both the manipulative and honest agents trade under the same market conditions, ii) if true, then at which extent changing the reward function can discourage this behaviour or in contrast, iii) changes in the transition dynamics affects the manipulation process. Points ii) and ii) are meant to represent market mechanisms that regulators can implement in order to disincentivise the manipulation process.

The setup of actions and costs in the trading models (ref. §IV-A and §IV-B) is as follows. All honest actions in direction of the edge of any of the states have zero costs and make the agent bounce back to the same state. The same occurs for manipulative actions, except that the obstacle switches its position (thus changing from representation). Transitioning within the different states costs the agent -1 and colliding against the obstacle has 0/-1 costs for all honest/ manipulative actions. The terminal states, s^* and s'^*, have the highest reward of $+1$, meaning the trader has reached the desired growth state. The "do nothing" action ($\circ$) has zero costs in all states, but the agent is unable to make transitions. Throughout the below, we set $\gamma = 0.95$.

A. Optimal Policy for Spoofing as an MDP

In this section, as a baseline, we first define a market where honest and manipulative traders pay the same transaction costs, and all of their actions are deterministic, *i.e.,* the traders are 100% confident about the outcome of their actions. The baseline is intended to demonstrate whether manipulation naturally emerges as an optimal strategy to produce investment growth. We then compare the baseline against the learnt policies once 'regulators' apply the proposed counter-measures.

1) Is spoofing and optimal strategy?: Here we investigate whether spoofing occurs according to the model described in §IV-A and, if so, what are the factors that encourage traders to take these actions. As mentioned above, we first model a market where all participants play under the same conditions, meaning that all trading actions pay the same transaction costs and the outcome of their actions being totally deterministic, *i.e.,* if a trader places a honest limit order then it is always executed, thus re-balancing the quantities a_t and c_t and consequently changing the value of (2).

The results of learning in the baseline scenario are shown in the first column of Table I. For each state, the optimal action is listed, and in states where multiple actions are equally optimal, the corresponding arrows are plotted.

As can be seen, spoofing actions are optimal for most of the states in our example problem. This suggests that, all things being equal, in our baseline example problem, the choice as to whether to manipulate the market or not has little impact on profit maximisation, and therefore comes down to either symmetry-breaking rules, or the preference of the trader.

TABLE I: Optimal actions for the MDP model under different conditions of the reward and transition functions.

State	Baseline	Adding fines	Adding uncertainty on liquidity	
			50% vs. 50%	10% vs. 90%
s_1	$\uparrow, \rightarrow, \Uparrow$	$\rightarrow$	$\Rightarrow$	$\Rightarrow$
s_2	$\Rightarrow$	$\downarrow$	$\Uparrow, \Leftarrow$	$\downarrow$
s_3	$\rightarrow, \Uparrow, \Rightarrow$	$\rightarrow$	$\rightarrow, \Rightarrow$	$\rightarrow, \Rightarrow$
s_4	$\uparrow, \Uparrow$	$\uparrow$	$\uparrow, \Uparrow$	$\uparrow, \Uparrow$
s_5	$\uparrow, \Uparrow, \Rightarrow$	$\uparrow$	$\uparrow$	$\uparrow$
s_6	$\rightarrow$	$\rightarrow$	$\rightarrow$	$\Rightarrow$
s_7	$\rightarrow, \Rightarrow$	$\rightarrow$	$\rightarrow, \Rightarrow$	$\rightarrow, \Rightarrow$
s_8	$\uparrow, \Uparrow$	$\uparrow$	$\uparrow, \Uparrow$	$\uparrow, \Uparrow$

2) Can financial penalties discourage spoofing?: We now focus our attention to try to discourage spoofing trading by simulating the introduction of high penalties against spoofers. These penalties can be viewed as fines or financial penalties as a warning to traders to stop the misbehaviour, otherwise they will be forced to pay them if detected. With this information in hand, we change the reward function by increasing the cost to all manipulative actions in all states up to -4.53.

In the second column of Table I we show the learnt policy for the case when financial penalties are introduced. Spoofing, in contrast to the baseline as explained in §V-A1, is no longer optimal in any of the states, suggesting that a trader with the knowledge of implementation of penalties will not take the risk of follow a disruptive trading strategy to obtain profits. It is clear that if we reduce (in absolute terms) the proposed cost for the manipulative actions, then spoofing will eventually appear once more in the learnt policy, however, this gives us an idea on the magnitude the introduction of this counter-measure has to have in order to discourage the manipulation process by spoofing.

3) Can a controlled liquidity discourage spoofing?: By "controlling the liquidity" we mean that the market mechanisms must be such that every time a high imbalance is found in the LOB (produced by the spoofer), then regulators or the market itself must find the way to quickly counter it and find a new balance. This will produce uncertainty to spoofing orders as the price impact will not be the one expected by the spoofer and consequently the price manipulation process can't proceed.

Columns two and three of Table I show the learnt policy after introducing controls over the liquidity. We used two degrees of control that produce uncertainty in the liquidity, one where the obstacle in the representation Fig. 3 has a 50%/50% chance to switch/stagnate and the second with 10%/90% chances, respectively. Both represent uncertainty in terms on how the market reacts to the spoofing order placed by the spoofer. Once more, spoofing is optimal in most of the states for the two different degrees of controls even if the price impact is not the expected one, suggesting that the spoofer can try as many times as possible until having the desired impact.

B. POMDP Model for Pinging

1) Is pinging an optimal strategy?: We now turn to analyse the pinging strategy by following the same reasoning as in spoofing. Again, we use a baseline to demonstrate whether under same conditions of transaction costs and liquidity pinging traders and honest traders emerge as optimal trading strategies from the learnt policies.

In the first column in Table 4 we show the learnt policy for the baseline in the pinging problem. In all observed states, pinging is optimal even if the hidden liquidity is not found. We can conclude that this is expected as long as the HFT can trade against other market participants in the regulated market without taking extreme risks of executing the ping orders.

TABLE II: Optimal actions for the POMDP model under different conditions of the reward and transition functions.

Observed State	Baseline	Increase transaction costs	Uncertainty on liquidity	
			50% vs. 50%	10% vs. 90%
o_1	$\Uparrow$	$\rightarrow, \uparrow$	$\uparrow$	$\uparrow, \Rightarrow$
o_2	$\Rightarrow, \rightarrow$	$\rightarrow, \downarrow$	$\rightarrow, \Rightarrow$	$\rightarrow, \Rightarrow, \downarrow$
o_3	$\Uparrow$	$\rightarrow$	$\rightarrow$	$\rightarrow$
o_4	$\Uparrow$	$\uparrow$	$\uparrow$	$\uparrow$
o_5	$\Rightarrow$	$\rightarrow$	$\rightarrow$	$\rightarrow$

2) Can we discourage pinging by increasing transaction costs?: As pinging is not illegal, instead of using the same argument of financial penalisation as it was the case for spoofing, for pinging we increase the direct transaction costs in either (or both) of the market venues where the strategy develops, as a way to discourage HFTs keep implementing the strategy. Under this scenario, we change the reward function and set a cost of -4.91 for all manipulative actions in all states.

The second column of Table II shows the results for this setup and we see that under the new conditions the learnt policies reveal pinging is not optimal in any of the observed states as it was in it's baseline. The trader can only take honest actions as pinging is paying large direct transaction costs, thus we can conjecture that pinging can still produce profits, but not large enough to cover the direct transaction costs.

3) Can a controlled liquidity discourage pinging?: Finally, a second attempt to stop pinging trading is by changing the transition function, just as in the case of spoofing as seen in §V-A3. Columns three and four in Table II shows once more the results for these changes and we see that, under mechanism that provide uncertainty to the effect of pinging trades over liquidity, pinging is still an optimal action in some of the observed states, probably because of exceptional conditions prevailing in such states. This makes us conclude one more time that a change in the reward function models the effectiveness to discourage manipulation by pinging, a similar result as in spoofing trading.

VI. Conclusions

Our results show it is promising the application of the (PO)MDP frameworks to study a real life problem as the financial markets. Our models can predict behaviours, and both the manipulative and honest trading can co-exist in regulated markets where all participants have the same direct costs. We found that both spoofing and pinging trading are optimal investment strategies while traders try to maximise the investment growth, but market regulators can discourage the use of these strategies by implementing mechanism over market liquidity, and this enforcement will be more efficient if fines are added (for spoofing) or by increasing the direct transaction costs (pinging).

However, our model works on bull market conditions and we expect to fit on bear markets if we change the side of the trading actions. Other conditions where no trends exists may produce incentives for manipulation as a way to move the market. Furthermore, in pinging HFTs have the option to avoid ping orders and analyse the predictability of the asset's order flow with the goal to infer the existence of hidden liquidity, thus saving direct transaction costs.

Further research can be focused on applying the models in real market data and more complex portfolios, and verify the effectiveness of the recommendations provided to disincentivise manipulation performed by spoofing/pinging traders.

Acknowledgment

Enrique Martínez-Miranda acknowledge support from CONACYT (384670) and *Beca Complemento SEP, 2014* and *2015* (BC-3465 and BC-4804), Mexico.

References

[1] M. O'Hara, *Market Microstructure Theory*. John Wiley & Sons, 1998.

[2] R. Pardo, *The Evaluation and Optimization of Trading Strategies*. Wiley, 2008.

[3] F. Allen and D. Gale, "Stock-price manipulation," *Review of Financial Studies*, vol. 5, no. 3, pp. 503–529, 1992.

[4] D. D. Aktas, "Spoofing," *Review of Banking and Financial Law*, vol. 33, no. 89, pp. 89–98, 2013.

[5] G. Scopino, "The (questionable) legality of high-speed "pinging" and "front running" in the futures market," *Connecticut Law Review*, vol. 47, no. 3, pp. 607–697, February 2015.

[6] R. Jarrow, "Market manipulation, bubbles, corners, and short squeezes," *Journal of Financial and Quantitative Analysis*, vol. 27, no. 03, pp. 311–336, 1992.

[7] M. K. Brunnermeier and L. H. Pederse, "Predatory trading," *The Journal of Finance*, vol. 60, no. 4, pp. 1825–1863, 2005.

[8] I. Goldstein and A. Guembel, "Manipulation and the allocational role of prices," *The Review of Economic Studies*, vol. 75, no. 1, pp. 133–164, January 2008.

[9] A. I. Khwaja and A. Mian, "Unchecked intermediaries: Price manipulation in an emerging stock market," *Journal of Financial Economics*, vol. 78, no. 1, pp. 203–241, October 2005.

[10] P. Hillion and M. Suominen, "The manipulation of closing prices," *Journal of Financial Markets*, vol. 7, no. 4, pp. 351–375, 2004.

[11] S. D. Ledgerwood and P. R. Carpenter, "A framework for the analysis of market manipulation," *Review of Law & Economics*, vol. 8, no. 1, pp. 253–295, 2012.

[12] C. Pirrong, "Detecting Manipulation in Futures Markets: The Ferruzzi Soybean Episode," *American Law and Economics Review*, vol. 6, no. 1, pp. 28–71, 2004.

[13] H. Öğüt, M. M. Doğanay, and R. Aktaş, "Detecting stock-price manipulation in an emerging market: The case of turkey," *Expert Systems with Applications*, vol. 36, no. 9, pp. 11 944 – 11 949, 2009.

[14] Y. Cao, Y. Li, S. Coleman, A. Belatreche, and T. M. McGinnity, "A Hidden Markov Model with abnormal states for detecting stock price manipulation," in *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on*, Oct 2013, pp. 3014–3019.

[15] D. Díaz-Solís, B. Theodoulidis, and P. Sampaio, "Analysis of stock market manipulations using knowledge discovery techniques applied to intraday trade prices," *Expert Systems with Applications*, vol. 38, no. 10, 2011.

[16] K. Golmohammadi, O. Zaiane, and D. R. Díaz, "Detecting stock market manipulation using supervised learning algorithms," in *Data Science and Advanced Analytics (DSAA), 2014 International Conference on*. IEEE, 2014, pp. 435–441.

[17] Y. Cao, Y. Li, S. Coleman, A. Belatreche, and T. M. McGinnity, "Detecting price manipulation in the financial market," in *Computational Intelligence for Financial Engineering Economics (CIFEr), 2104 IEEE Conference on*, 2014, pp. 77–84.

[18] S. Yang, M. Paddrik, R. Hayes, A. Todd, A. Kirilenko, P. Beling, and W. Scherer, "Behavior based learning in identifying high frequency trading strategies," in *Computational Intelligence for Financial Engineering & Economics (CIFEr), 2012 IEEE Conference on*. IEEE, 2012, pp. 1–8.

[19] M. Rossi, G. Deis, J. Roche, and K. Przywara, "Recent civil and criminal enforcement action involving high frequency trading," *Journal of Investment Compliance*, vol. 16, no. 1, pp. 5–12, 2015.

[20] E. J. Lee, K. S. Eom, and K. S. Park, "Microstructure-based manipulation: Strategic behavior and performance of spoofing traders," *Journal of Financial Markets*, vol. 16, no. 2, pp. 227–252, 2013.

[21] J. Xu, "Optimal strategies of high frequency traders," *Available at SSRN 2382378*, 2013.

[22] Y. Nevmyvaka, Y. Feng, and M. Kearns, "Reinforcement learning for optimized trade execution," in *Proceedings of the 23rd international conference on Machine learning*. ACM, 2006, pp. 673–680.

[23] A. C. E. Baffa and A. E. M. Ciarlini, "Modeling POMDPs for generating and simulating stock investment policies," in *Proceedings of the 2010 ACM Symposium on Applied Computing*. ACM, 2010, pp. 2394–2399.

[24] R. S. Sutton and A. G. Barto, *Introduction to Reinforcement Learning*, 1st ed. Cambridge, MA, USA: MIT Press, 1998.

[25] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *Artificial intelligence*, vol. 101, no. 1, pp. 99–134, 1998.