Behavior Based Learning in Identifying High Frequency Trading Strategies

Steve Yang, Mark Paddrik, Roy Hayes, Andrew Todd, Andrei Kirilenko, Peter Beling, and William Scherer

***Abstract*— Electronic markets have emerged as popular venues for the trading of a wide variety of financial assets, and computer based algorithmic trading has also asserted itself as a dominant force in financial markets across the world. Identifying and understanding the impact of algorithmic trading on financial markets has become a critical issue for market operators and regulators. We propose to characterize traders' behavior in terms of the reward functions most likely to have given rise to the observed trading actions. Our approach is to model trading decisions as a Markov Decision Process (MDP), and use observations of an optimal decision policy to find the reward function. This is known as Inverse Reinforcement Learning (IRL). Our IRL-based approach to characterizing trader behavior strikes a balance between two desirable features in that it captures key empirical properties of order book dynamics and yet remains computationally tractable. Using an IRL algorithm based on linear programming, we are able to achieve more than 90% classification accuracy in distinguishing high frequency trading from other trading strategies in experiments on a simulated E-Mini S&P 500 futures market. The results of these empirical tests suggest that high frequency trading strategies can be accurately identified and profiled based on observations of individual trading actions.**

Keywords:Limit order book, Inverse Reinforcement Learning, Markov Decision Process, Maximum likelihood, Price impact, High Frequency Trading.

I. Introduction

MANY FINANCIAL MARKET PARTICIPANTS now employ algorithmic trading, commonly defined as the use of computer algorithms to automatically make certain trading decisions, submit orders, and manage those orders after submission. By the time of the "Flash Crash" (On May 6, 2010 during 25 minutes, stock index futures, options, and exchange-traded funds experienced a sudden price drop of more than 5 percent, followed by a rapid and near complete rebound), algorithmic trading was thought to be responsible for more than 70% of trading volume in the U.S. ([3], [9], [10], and [15]). Moreover, Kirilenko et al. [15] have shown that the key events in the Flash Crash have a clear interpretation in terms of algorithmic trading.

A variety of machine learning techniques have been applied in financial market analysis and modeling to assist market operators, regulators, and policy makers to understand the behaviors of the market participants, market dynamics, and the price discovery process of the new electronic market phenomena of algorithmic trading([1], [2], [3], [4], [5], [6], and [8]). We propose modeling traders' behavior as a Markov Decision Process (MDP), using observations of individual trading actions to characterize or infer trading strategies. More specifically, we aim to learn traders' reward functions in the context of multi-agent environments where traders compete using fast algorithmic trading strategies to explore and exploit market microstructure.

Our proposed approach is based on a machine learning technique ([20], [21], and [23]) known as Inverse Reinforcement Learning (IRL) ([11], [12], [13], [18], and [22]). In IRL, one aims to infer the model that underlies solutions that have been chosen by decision makers. In this case the reward function is of interest by itself in characterizing agent's behavior irregardless of its circumstances. For example, Pokerbots can improve performance against suboptimal human opponents by learning reward functions that account for the utility of money, preferences for certain hands or situations, and other idiosyncrasies [17]. Another objective in IRL is to use observations of the traders' actions to decide ones' own behaviors. It is possible in this case to directly learn the reward functions from the past observations and be able derive new policies based on the reward functions learned in a new environment to govern a new autonomous process (apprenticeship learning). In this paper, we focus our attention on the former problem to identify trader's behavior using reward functions.

The rest of the paper is structured as follows: In Section 2, we define notation and formulate the IRL model. In Section 3, we first propose a concise MDP model of the limit order book to obtain reward functions of different trading strategies, and then solve the IRL problem using a linear programming approach based on an assumption of rational decision making. In Section 4, we present our agent-based simulation model for E-Mini S&P 500 futures market and provide validation results that suggest this model replicates with high fidelity the real E-Mini S&P 500 futures market. Using this simulation model we generate simulated market data and perform two experiments. In the first experiment, we show that we can reliably identify High Frequency Trading (HFT) strategies from other algorithmic trading strategies using IRL. In the second experiment, we apply IRL on HFTs and show that we can accurately identify a manipulative HFT strategy (Spoofing) from the other HFT strategies. Section 5 discusses the conclusion of this study and the future work.

Mr. Yang, Mr. Paddrik, Mr. Hayes, Mr. Todd, Dr. Beling, and Dr. Scherer are with the Department of Systems and Information Engineering at University of Virginia (email: {yy6a, mp3ua, rlh8t, aet6cd, pb3a, wts}@virginia.edu). Dr. Kirilenko is with the Commodity Futures Trading Commission (email: akirilenko@cftc.gov).

II. Problem Formulation - Inverse Reinforcement Learning Model

The primary objective of our study is to find the reward function that, in some sense, best explains the observed behavior of a decision agent. In the field of reinforcement learning, it is a principle that the reward function is the most succinct, robust and transferable representation of a decision task, and completely determines the optimal policy (or set of policies) [22]. In addition, knowledge of the reward function allows a learning agent to generalize better, since such knowledge is necessary to compute new policies in response to changes in environment. These points motive our hypothesis that IRL is a suitable method for characterizing trading strategies.

A. General Problem Definition

Lets define a (infinite horizon, discounted) MDP model first. Let $M = \{\mathcal{S}, \mathcal{A}, \mathcal{P}, \gamma, \mathcal{R}\}$, where:

$s \in \mathcal{S}$ where $\mathcal{S} = \{s_1, s_2, ..., s_N\}$ is a set of N states;
$\mathcal{A} = \{a_1, a_2, ..., a_k\}$ is a set of k possible actions;
$\mathcal{P} = \{P_{a_j}\}_{j=1}^{k}$, where P_{a_j} is a transition matrix such that $P_{sa_j}(s')$ is the probability of transitioning to state s' given action a_j taken in state s;
$\gamma \in (0, 1)$ is a discount factor;
$\mathcal{R}$ is a reward function such that R or $R(s, a)$ is the reward received given action a is taken when in state s.

Within the MDP construct, a trader or an algorithmic trading strategy can be represented by a set of primitives $(\mathcal{P}, \gamma, \mathcal{R})$ where $\mathcal{R}$ is a reward function representing the trader's preferences, $\mathcal{P}$ is a Markov transition probability representing the trader's subjective beliefs about uncertain future states, and γ is the rate at which the agent discounts reward in future periods. In using IRL to identify trading strategies, the first question that needs to be answered is whether $(\mathcal{P}, \gamma, \mathcal{R})$ is identified. Rust [7] discussed this identification problem in his earlier work in economic decision modeling. He concluded that if we are willing to impose an even stronger prior restriction, stationarity and rational expectations, then we can use non-parametric methods to consistently estimate decision makers' subjective beliefs from observations of their past states and decisions. Hence in formulating the IRL problem in identifying trading strategies, we will have to make two basic assumptions: first, we assume the policies we model are stationary; second, the trading strategies are rational expected-reward maximizers.

Here we define the *value function* at state s with respect to policy π and discount γ to be $V_\gamma^\pi(s) = E[\sum_{t=0}^{\infty} \gamma R(s^t, \pi(s^t))|\pi]$, where the expectation is over the distribution of the state sequence $\{s^0, s^1, ..., s^t\}$ given policy π (superscripts index time). We also define the $Q_\gamma^\pi(s, a)$ for state s and action a under policy π and discount γ to be the expected return from state s, taking action a and thereafter following policy π. And then we have the following two classical results for MDPs (see, e.g., [20], [19]):

Theorem 1: (Bellman Equations) Let an MDP $M = \{\mathcal{S}, \mathcal{A}, \mathcal{P}, \gamma, \mathcal{R}\}$, and a policy $\pi : \mathcal{S} \to \mathcal{A}$ be given. Then, for all $s \in \mathcal{S}, a \in \mathcal{A}, V_\gamma^\pi$ and Q_γ^π satisfy:

$$V_\gamma^\pi(s) = R_\pi(s, \pi(s)) + \gamma \sum_{j \in S} P_{s\pi(s)}(j) V_\gamma^\pi(j), \forall s \in \mathcal{S} \quad (1)$$

$$Q_\gamma^\pi(s, a) = R_\pi(s, \pi(s)) + \gamma \sum_{j \in S} P_{sa}(j) V_\gamma^\pi(j), \forall s \in \mathcal{S} \quad (2)$$

Theorem 2: (Bellman Optimality) Let an MDP $M = \{\mathcal{S}, \mathcal{A}, \mathcal{P}, \gamma, \mathcal{R}\}$, and a policy $\pi : \mathcal{S} \to \mathcal{A}$ be given. Then, π is an optimal policy for M if and only if, for all $s \in \mathcal{S}$:

$$V_\gamma^{\pi^*}(s) = \max_{a \in \mathcal{A}} [R_\pi(s, \pi(s)) + \gamma \sum_{j \in \mathcal{S}} P_{s\pi(s)}(j) V_\gamma^\pi(j)], \forall s \in \mathcal{S} \quad (3)$$

The Bellman Optimality condition can be written in matrix format as follows:

Theorem 3: Let a finite state space $\mathcal{S}$, a set of $a \in \mathcal{A}$, transition probability matrix P_a and a discount factor $\gamma \in (0, 1)$ be given. The a policy given by π is an optimal policy for M if and only if, for all $a \in \mathcal{A} \backslash \pi$, the reward R satisfies:

$$(P_\pi - P_a)(I - \gamma P_\pi) R \succeq 0 \quad (4)$$

B. Linear Programming Approach to IRL

The IRL problem is, in general, highly underspecified, which has led researchers to consider various models for restricting the set of reward vectors under consideration. The only reward vectors consistent with an optimal policy π are those that satisfy the set of inequalities in Theorem 3. Note that the degenerate solution $R = 0$ satisfies these constraints, which highlights the underspecified nature of the problem and the need for reward selection mechanisms. Ng and Russel [11] advance the idea choosing the reward function to maximize the difference between the optimal and suboptimal policies, which can be done using a linear programming formulation. We adopt this approach, maximizing:

$$\sum_{s \in \mathcal{S}} [Q_\gamma^\pi(s, a') - \gamma \max_{a \in \mathcal{A} \backslash a'} Q_\gamma^\pi], \forall a \in \mathcal{A} \quad (5)$$

Putting theorem 4 and 5 together, we have an optimization problem to solve to obtain a reward function under an optimal policy:

$$\begin{aligned} \max_R [\sum_{s \in \mathcal{S}} \beta(s) - \lambda \sum_{s \in \mathcal{S}} \alpha(s)] \\ s.t. \\ \alpha(s) \succeq \beta(s), \forall s \in \mathcal{S} \\ (P_\pi - P_a)(I - \gamma P_\pi) R \succeq \beta(s), \forall a \in \mathcal{A}, \forall s \in \mathcal{S} \\ (P_\pi - P_a)(I - \gamma P_\pi) R \succeq 0 \end{aligned} \quad (6)$$

In summary, we assume an ergodic MDP process. In particular, we assume the policy defined in the system has a proper stationary distribution. And we further assume that trader's trading strategies are rational expected reward maximizers. There are specific issues regarding the non-deterministic nature of trader's trading strategies when dealing with empirical observations, and we will address them later in the next section.

C. Key Modeling Issues

One of the key issues that arise in applications of IRL or apprenticeship learning to algorithmic trading is that the trader under observation may not appear to follow a deterministic policy. In particular, a trader observed in the same state on two different occasions may take two different actions, either because the trader is following a randomized policy or because the state space used in the model lacks the fidelity to capture all the factors that influence the trader's decision. To address the issue of non-deterministic policies, we need to first understand the relationship between a deterministic policy versus non-deterministic policy under the assumption we made earlier. We use notation MD for Markov deterministic policy, and MR for Markov non-deterministic policy. We can establish the relationship between the optimality of a deterministic policy versus a non-deterministic policy through the following proposition ([17]):

Proposition: For all $v \in V$ and $0 \leq \gamma \leq 1$:

$$\sup_{d \in D^{MD}} \{R_d + \gamma P_d v\} = \sup_{d \in D^{MR}} \{R_d + \gamma P_d v\}, \forall d \in \mathcal{A} \quad (7)$$

Policies range in general from deterministic Markovian to randomized history dependent, depending on how they incorporate past information and how they select actions. In the financial trading world, traders deploy different trading strategies where each strategy has a unique value proposition. We can theoretically use cumulative reward to represent the value system encapsulated in the various different trading strategies. For example in a simple keep-or-cancel strategy for buying one unit, the trader has to decide when to place an order and when to cancel the order based on the market environment (can be characterized stochastic processes) to maximize its cumulative reward under the constraint of the traders' risk utility and capital limit. This can be realized in a number of ways. It can be described as a function R(s) meaning when the system is in state s the trader is always looking for a fixed reward. This notion of value proposition drives the traders to take corresponding optimal actions according to the market conditions. However due to the uncertainty of the environment and the random error of the measurement in the observations, a deterministic policy could very likely be perceived to have a non-deterministic nature.

Based on the proposition or equation 7, the optimal value attained by a randomized policy is the same as the one attained by a deterministic policy, and there exists an optimal value and it is unique in V. Therefore, we know that the supremum value obtained from all policies can be used to recover an equivalent optimal stationary deterministic policy. Essentially we are looking for an optimal deterministic stationary policy which achieves the same optimal value as the non-deterministic policy. This guarantees the learning agent to obtain a unique reward function that achieves the optimal value. The merit of this approach is that the reward function will be unique for a specific set of observations. We will not be concerned about whether the trader's real policy is deterministic or not. This is especially useful in the problem where we attempt to identify traders' trading strategies based on a series of observations.

III. A MDP Model for Limit Order Book

Cont et al. ([24], and [25]) make the claim that order flow imbalance and order volume imbalance have the strongest link with the price changes. It seems that these two variables can best capture the limit order book dynamics. It has been proven effective in modeling buy-one-unit and make-the-spread strategies by Hunt, et al. [27] where three price levels have shown significantly good resemblance to the real market characteristics. Other financial market microstructure studies also provide strong evidence of using order book imbalance to represent the market supply and demand dynamics or information asymmetry ([5], [8], [6], [14] and [26]). Based on this evidence, we choose two bid/ask volume imbalance variables to capture the market environment, and we choose position/inventory level as a private variable of the trader. In summary, we use three sensory variables to characterize the environment in which the traders operate. Now we can define state $s = [TIM, NIM, POS]^T$, and each variable takes following discrete values:

TIM - volume imbalance at the best bid/ask: {-1, 0, 1};
NIM - volume imbalance at the 3rd best bid/ask: {-1, 0, 1};
POS - position status: {-1, 0, 1}.

When the variable takes value 0 (in neutral state), it means that the variable takes mean (μ) value within $\mu \pm 1.96\sigma$; when the value is above $\mu + 1.96\sigma$, we define it as high; and when the value is below $\mu - 1.96\sigma$, we define it as low. Essentially we have two external variables: TIM and NIM. Variables TIM and NIM inform the traders whether volume imbalance is moving toward sell side (low), neutral, or toward buy side (high), as well as the momentum of the market price movement. The private variable POS informs traders whether his or her inventory is low, neutral or high. All three variables are very essential for algorithmic traders to make their trade decisions. We also define a set of actions that correspond to traders' trading choices at each state a = {PBL, PBH, PSL, PSH, CBL, CBH, CSL, CSH, TBL, TBH, TSL, TSH}, and each value is defined in TABLE I.

We assume a highly liquid market where market orders will always be filled, and we apply the model to a simulated order book where both limit orders and market orders are equally present.

IV. Experiments

In this section, we conduct two experiments using the MDP model defined earlier to identify algorithmic trading strategies. We use the six trader classes defined by Kirilenko et. al. [15], namely High Frequency Traders, Market Makers, Opportunistic Traders, Fundamental Buyers, Fundamental Sellers and Small Traders. In general, HFTs have a set of distinctive characteristics, such as, very high activity volume throughout a trading day, frequent modification of orders,

TABLE I

ACTION DEFINITION.

Action Code	Action Description
1	PBH - place buy order higher than the 3rd best bid price
2	PBL - place buy order lower than the 3rd best bid price
3	PSH - place sell order higher than the 3rd best ask price
4	PSL - place sell order lower than the 3rd best ask price
5	CBH - cancel buy order higher than the 3rd best bid price
6	CBL - cancel buy order lower than the 3rd best bid price
7	CSH - cancel sell order higher than the 3rd best ask price
8	CSL - cancel sell order lower than the 3rd best ask price
9	TBH - Trade buy order higher than the 3rd best bid price
10	TBL - Trade buy order lower than the 3rd best bid price
11	TSH - Trade sell order higher than the 3rd best ask price
12	TSL - Trade sell order lower than the 3rd best ask price

maintenance of very low inventory levels, and an agnostic orientation toward long or short positions. Market Makers are short horizon investors who follow a strategy of buying and selling a large number of contracts to stay around a relatively low target level of inventory. Opportunistic Traders sometimes behave as Market Makers buying and selling around a target position, and sometimes they act as Fundamental Traders accumulating long or short positions. Fundamental Buyers and Sellers are net buyers and sellers who accumulate positions in one single direction in general. Small Traders are the ones who have significant less activities during a typical trading day.

In the first experiment, we are interested in separating High Frequency Trading strategies from Market Making and Opportunistic Trading strategies in the simulated E-Mini S&P 500 futures market. From Figure (b) in Fig. 1, we see that the behaviors of the Fundamental Buyers/Sellers are distinctively different from the other algorithmic traders. It is clear that classification between the HFTs and these classes of trading strategies are relatively trivial. We therefore devote our attention to separate High Frequency Trading strategies from the Market Marking and the Opportunistic Trading strategies. We will start this section with a description of the design of our agent-based simulation for E-Mini S&P 500 futures market [33]. We will then use the data generated from this simulation as observations to recover the reward functions of different kinds of trading strategies, and we apply various classification methods on these trading strategies in the reward space to see whether we can accurately identify the different trading strategy classes. In the second experiment, we will focus on a specific High Frequency Trading strategy called Spoofing, and try to separate this trading strategy from the other High Frequency Trading strategies. In general, we test the hypothesis that reward functions can be used to effectively identify High Frequency Trading strategies in both within-group and across-group situations.

A. Simulated E-Mini S&P 500 Futures Market

When simulating a system it is convenient to decompose the system into its basic parts. A financial market can be understood as a set of market participants, a trading mechanism, and a security. Agent-based models have a similar structure and include a set of agents, a topology and an environment. Through this framework it is possible to describe market participants as a set of agents with a set of actions and constraints, the market mechanism as the topology, and the exogenous flow of information relevant to market as the environment [32].

Using this framework, the simulation is tuned to replicate the same market conditions and variables as that of the nearest month E-Mini S&P 500 futures contract market. The agents in the model reflect closely the classes of participants observed in the actual S&P 500 E-mini futures market and the market mechanism is implemented as an electronic limit order book (see Fig. 1). Each class of participants is then characterized by their trade speed, position limit, order size distribution, and order price distribution. All these characterizations are based on the order book data from the E-Mini S&P 500 futures contracts provided by the Commodity Futures Trading Commission (CFTC). (see TABLE II).

TABLE II

TRADER GROUP CHARACTERIZATION

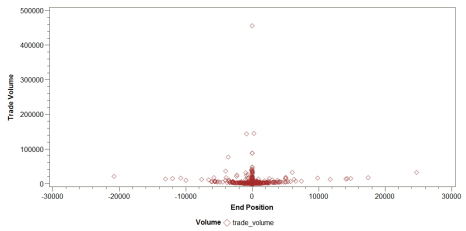
Trader Type	Number of Traders	Speed of Order	Position Limits	Market Volume
Small	6880	2 hours	$-30 \sim 30$	1%
Fundamental Buyers	1268	1 minute	$-\propto\sim\propto$	9%
Fundamental Sellers	1276	1 minute	$-\propto\sim\propto$	9%
Market Makers	176	20 seconds	$-120 \sim 120$	10%
Opportunistic	5808	2 minutes	$-120 \sim 120$	33%
HFTs	16	0.35 seconds	$-3000 \sim 3000$	38%

TABLE III

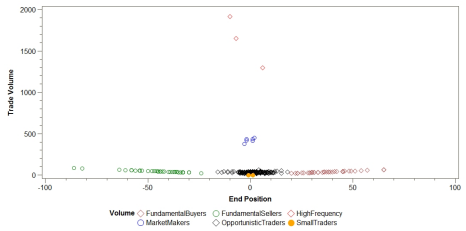
TRADER GROUP VALIDATION

Trader Type	Simulated Volume	Actual Volume	Rate-Simulated Cancellation	Rate-Actual Cancellation
Small	1%	1%	40%	$20-40\%$
Fundamental Buyers	10%	9%	44%	$20-40\%$
Fundamental Sellers	10%	9%	44%	$20-40\%$
Market Makers	10%	10%	35%	$20-40\%$
Opportunistic	31%	33%	50%	$40-60\%$
HFTs	38%	38%	77%	$70-80\%$

After the model is simulated there are two stages of validation. The first stage consists of a validation of the basic statistics for each set of agents, such as arrival rates, cancellations rates, and trade volume (see TABLE III). The values observed in the simulation are compared to data of



(a) Actual E-Mini S&P 500 traders



(b) Simulated E-Mini S&P 500 traders

Fig. 1. E-Mini S&P 500 traders' end-of-day position vs. trading volume.

participants in the actual market. The second stage of validation consists of verifying that the price time-series produced by the simulation exhibits "stylized facts" (Kullmann,1999 [28]) that characterize financial data. These include heavy tailed distribution of returns[1] (Appendix A Fig. 7), absence of autocorrelation of returns[2] (Appendix A Fig. 8), volatility clustering[3] (Appendix A Fig. 9), and aggregational normality[4] (Appendix A Fig. 10). The detailed simulation validation results can be found in the work done by M. E. Paddrik, et al. [33].

B. Identify HFTs from Market Making and Opportunistic Trading Strategies

Using the IRL model that we formulated above, we learn the corresponding reward functions from 18 simulation runs where each run consists of approximately 300,000 activities including orders, cancellations, and trades. We then use the different classification methods on the rewards to see how well we can separate the HFTs from the other two different trading strategies.

From Fig. 2, we see that reward space has a very succinct structure, which tends to confirm the observations made in ([22], and [34]) that policies are generally noisier than reward functions. We also observe that the reward function converges faster than the policy as observation time increases. In addition to the lack of robustness in policy space, the lack of portability of learned policies is another important drawback in the use of policies to characterize trading strategies. Furthermore, the fact that actions are notional makes it unclear how one could use policies to measure differences

[1]The empirical distributions of financial returns and log-returns are fat-tailed.

[2]There is no evidence of correlation between successive returns.

[3]Absolute returns or squared returns exhibit a long-range slowly decaying autocorrelation function.

[4]As the time scale increases, the fat-tail property diminishes and the return distribution approaches Gaussian distribution.

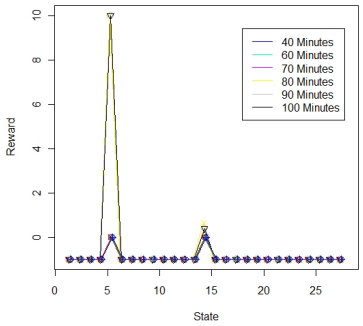


Fig. 2. **Reward Space Convergence** For a series of observations of a particular trader, as time interval increases, the reward at state 5 converges from 10 to 0, and the reward at state 14 converges from 0.66 to 0. At all the other states, the reward remains at -1.

among trading strategies. Hence, our study focuses attention on reward space. Using Principal Component dimension shrinkage method, we are able to compare the two trading strategies in a three dimensional space visually. Fig. 3 and Fig. 4 show a clear separation of the High Frequency Trading strategies from the other two classes of trading strategies.

Three different classification methods are then applied on the learned reward functions. From the comparison (Table IV) of the results of the three different classification methods, i.e. Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), and Multi-Gaussian Discriminant Analysis (MDA). The two non-linear methods perform better than the linear one. It can be seen from the visualization reward distributions. The highest classification accuracy achieved by all three methods is 100%. In general, all of them achieved relatively high accuracy in the range between 95% and 100%. The sensitivity (i.e. true positive) is in the range between 89% and 94%. The specificity (i.e. true negative) is in general better, and it is 100% across all three classification methods.

TABLE IV
TRADING STRATEGY CLASSIFICATION RESULTS

High Frequency Traders vs. Opportunistic Traders			
	LDA	QDA	MDA
Accuracy	97%	100%	97%
Sensitivity	94%	100%	94%
Specificity	100%	100%	100%
High Frequency Traders vs. Market Makers			
	LDA	QDA	MDA
Accuracy	95%	97%	95%
Sensitivity	88%	94%	88%
Specificity	100%	100%	100%
Opportunistic Traders vs. Market Makers			
	LDA	QDA	MDA
Accuracy	70%	75%	83%
Sensitivity	39%	100%	72%
Specificity	100%	100%	94%

The results using this model for separating Opportunistic Traders vs. Market Makers are not as good compared with those between HFT vs. Market Making and HFT vs. Op-

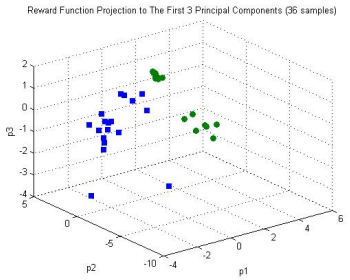


Fig. 3. Reward space clustering between High Frequency Trading strategies vs. Opportunistic Trading strategies

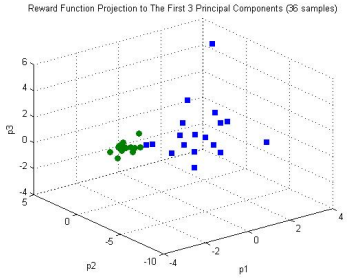


Fig. 4. Reward space clustering between High Frequency Trading strategies vs. Market Marking Trading strategies

portunistic strategies (TABLE IV). From the classification results, we can see that MDA classification performed the best and achieved 83% accuracy, 72% sensitivity, and 94% specificity. However, this result is expected in that the current order book model is specifically targeted at characterizing High Frequency Trading strategies. In order to achieve better results between Opportunistic and Market Making strategies, we will have to consider other factors that can best characterize the Opportunistic Trader's behaviors. Further study of the these two classes' behaviors will be critical in improving the classification performance between these two classes' of trading strategies.

C. Identify A Spoofing Strategy from Other HFTs

In this section, we are interested in one particular manipulative strategy in the High Frequency Trading paradigm: Spoofing, which sometimes is referred to as "Hype and Dump" manipulation ([29], [30], and [31]). Both empirical and theoretical evidence show that the manipulators can profit from this manipulative trading practice. In this scheme, the manipulator artificially inflates the asset price through promotion in order to sell at the inflated price, or deflates the asset price through false hype in order to buy at the deflated price. One concrete example of this trading strategy is illustrated in Fig. 5 A. Suppose a trader intends to sell 5 shares of an asset, he first submits a large limit-buy order with a bid at or below the current market price making the buy side of the order book seem large. Based on the market information infusion process or supply-demand theory, the market price will tend to move higher. And the spoofing trader will then submit a market-sell order and consequently cancels the original buy order as it is illustrated in Fig. 5 B.

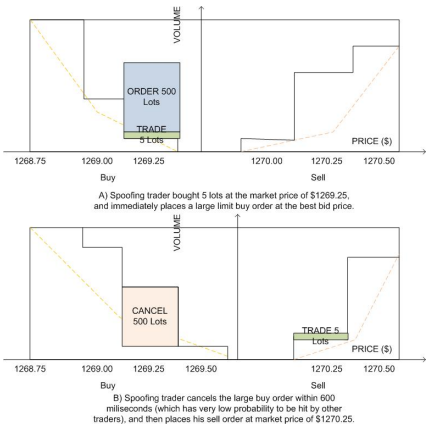


Fig. 5. Market microstructure-based manipulation example: buy spoofing

This manipulative practice is illegal under the U.S. securities law, yet it has been frequently discovered in both equity and futures markets. Our simulated spoofing trading strategy is based on our observations on a futures market where a trader repeatedly exercised the spoofing pattern over a month period. Due to the nature of the CFTC investigation, we will not be able to disclose the specifics for publications, but we are able to capture the deterministic nature of their strategy in the simulation. Specifically, in our discrete time agent-based simulation model, we design a spoofing agent as one of the HFTs except that it deploys additional trading plots: first they engage in a signaling game and then a trading game. In the signaling stage, the spoofing agent places a large buy order at the best bid price. After 600 milliseconds (it is designed with relative to the speed of HFT's cancellation rate), it transitions into the trading stage where they cancel the original limit order and places a market order. Since the trader is a HFT, in order to maintain the constraint of his inventory, the trader will have to spoof and trade on the other side of the book at certain point.

As we have done for the general simulation, we run 18 times of the simulation to generate 18 market instances. And then we randomly select 18 samples for all the general HFT trading strategies, and select 18 samples for the Spoofing trading strategy for IRL. We then obtain 36 reward functions with labels and apply three classification methods on these samples, and obtain results in TABLE V. From these results, we see that we can identify the Spoofing strategy from the other High Frequency Trading strategies with at least 92% accuracy. We also observe again that the non-linear classification rule works better in general.

V. Conclusions

The primary focus of this paper is to use Inverse Reinforcement Learning method to capture the key characteristics of the High Frequency Trading strategies. From the results using a linear programming method for solving IRL with simulated E-Mini S&P 500 futures market data, we attain

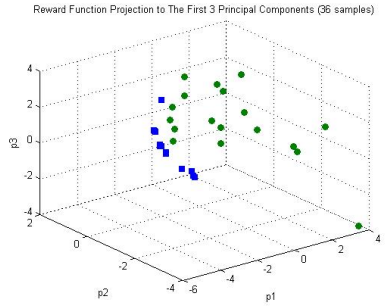


Fig. 6. Reward space clustering between High Frequency Trading strategies vs. the Spoofing strategy

TABLE V

SPOOFING TRADING STRATEGY VS. OTHER HFT CLASSIFICATION RESULTS

Market Makers vs. Opportunistic Traders

	LDA	QDA	MDA
Accuracy	92%	97%	97%
Sensitivity	100%	100%	100%
Specificity	83%	94%	94%

a high identification accuracy ranging between 95% and 100% for the targeted trading strategy class, namely High Frequency Trading from Market Making and Opportunistic strategies. We also show that the algorithm can accurately (between 92% and 95%) identify a particular type of HFT spoofing strategy from other HFT strategies. And we also argue that the reward space is better suited for identification of trading strategies than the policy space.

We investigate and address the issues of modeling algorithmic trading strategies using IRL models such as, addressing non-deterministic nature of the observed policies in learning, constructing efficient MDP models to capture order book dynamics, achieving better identification accuracy in reward space, etc. With a reliably validated agent based market simulation, we capture the essential characteristics of the algorithmic trading strategies. The practical implication of this research is that we demonstrate that the market operators and regulators can use this behavior based learning approach to perform trader behavior based profiling, and consequently monitor the emergence of new HFTs and study their impact to the market.

Here is a list of future research to be done:

- Apply both the linear programming approach and maximum likelihood approaches to the simulated trading strategies and the Spoofing data collected from the actual market, and compare the results of these two approaches in terms of identification accuracy.
- Create simulation agent based on reward functions learned from the actual market observations, and study the new trading strategy's impact to the market quality.

REFERENCES

[1] J. HASBROUCHK, D. J. SEPPI, *Common factors in prices, order flows and liquidity*, Journal of Financial Economics, 59 (2001), 383-411.

[2] V. PLEROU, P. GOPIKRISHNAN, X. GABAIX, AND H. E. STANLEY, *On the Origin of Power Law Fluctuations in Stock Prices*, Quantitative Finance 4, (2004).

[3] T. HENDERSHOTT ET AL., *Algorithmic Trading and Information*, NET Institute Working Paper, No. 09-08 (2008).

[4] J. GATHERAL, *No-dynamic-arbitrage and market impact*, Quantitative Finance,10 (2010) p. 749.

[5] J. HASBROUCHK, *Measuring the information content of stock trades*,Journal of Finance, 46 (1991) pp. 179-207.

[6] C. JONES, G. KAUL, AND M. LIPSON, *Transactions, volume, and volatility*, Review of Financial Studies, 7 (1994) pp. 631-651.

[7] J. RUST, *Structural Estimation of Markov Decision Processes*, Handbook of Econometrics, V IV, p 3082-3139 (1994).

[8] J. KARPOFF, *The relation between price changes and trading volume: A survey*, Journal of Financial and Quantitative Analysis, 22 (1987), p. 109.

[9] J. BROGAARD, *High Frequency Trading and its Impact on Market Quality*, Ph.D. thesis, Northwestern University, (2010).

[10] T. HENDERSHOTT, C. M. JONES, AND A. J. MENKVELD, *Does Algorithmic Trading Improve Liquidity?*, Journal of Finance, V. 66 p.1-33, (2011).

[11] A. Y. NG AND S. RUSSEL, *Algorithms for inverse reinforcement learning*, In Proc. ICML, 663-670 (2000).

[12] P. ABBEEL, AND A. Y. NG, *Apprenticeship learning via inverse reinforcement learning*, in ICML '04 Proceedings of the twenty-first international conference on Machine learning (2004).

[13] U. SYED, AND R. E. SCHAPIRE, *A Game-Theoretic Approach to Apprenticeship Learning*, NIP, 2007.

[14] Z. EISLER, J. P. BOUCHAUD, AND J. KOCKELKOREN, *The price impact of order book events: market orders, limit orders and cancellations*, Quantitative Finance Papers 0904.0900, arXiv.org, Apr. 2009.

[15] A A. KIRILENKO, A. S. KYLE, M. SAMADI, AND T. TUZUN, *The Flash Crash: The Impact of High Frequency Trading on an Electronic Market*, SSRN Working Paper, 2010.

[16] D. BILLINGS, D. PAPP, J. SCHAEFFER, AND D. SZAFRON, *Opponent modeling in Poker*, In AAAI, pages 493498, Madison, WI, 1998. AAAI Press.

[17] M. L. PUTERMAN, *Markov Decision Process: Discrete Stochastic Programming*, John Wiley and Sons, Inc. New York. (1994).

[18] S. RUSSELL, *Learning Agents for Uncertain Environments (Extended Abstract)*, Proceedings of the Eleventh Annual Conference on Computational Learning Theory. ACM Press (1998).

[19] DIMITRI P. BERTSEKAS, *Dynamic Programming and Optimal Control*, Athena Scientific (1995).

[20] R. S. SUTTON, AND A. G. BARTO, *Reinforcement Learning: An Introduction*, The MIT Press Cambridge, Massachusetts (1998).

[21] R. S. SUTTON, A. G. BARTO, AND R. J. WILLIAMS, *Reinforcement Learning is Direct Adaptive Optimal Control*, Presented at 1991 American Control Conference, Boston, MA June 26-28, 1991.

[22] D. RAMACHANDRAN, AND E. AMIR, *Bayesian Inverse Reinforcement Learning*, In Proc. IJCAI, 25862591 (2007).

[23] B. D. ZIEBART, A. MASS, J. A. BAGNELL, AND A. K. DEY, *Maximum Entropy Inverse Reinforcement Learning*, In Proceedings of the Twenty-Third AAAI on Artifical Intelligence (2008).

[24] R. CONT, A. KUKANOV, AND S. STOIKOV, *Order book dynamics and price impact*, IEOR Department, Columbia University, New York. Working paper, 2010.

[25] R. CONT, S. STOIKOV, AND R. TALREJA, *A stochastic model for order book dynamics*, SSRN Working Paper, 2011.

[26] A. A. OBIZHAEVA, AND J. WANG, *Optimal Trading Strategy and Supply/Demand Dynamics*, SSRN eLibrary, 2005.

[27] H. HULT, AND J. KIESSLING, *Algorithmic trading with Markov chains*, Doctoral thesis, Stockholm University, Sweden 2010.

[28] L. KULLMANN, J. TOYLI, A. KANTO, AND K. KASKI, *Characteristic times in stock market indices*, Physica A: Statistical Mechanics and its Applications, 1999, 269, 98-110.

[29] R. K. AGGARWAL, AND G. WU, *Stock Market Manipulation-Theory and Evidence*, AFA 2004 San Diego Meetings, 2003.

[30] J. MEI, G. WU, AND C. ZHOU, *Behavior Based Manipulation: Theory and Prosecution Evidence*, SSRN Working Paper 2004.

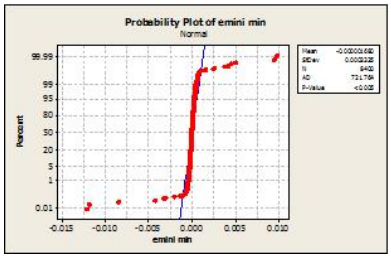
[31] N. EREN, AND H. N. OZSOYLEV, *Hype and Dump Manipulation*, AFA 2008 New Orleans Meetings Paper (2008).

[32] C. M. MACAL, AND M. J. NORTH, *Tutorial on agent-based modelling and simulation*, Journal of Simulation, 4(3), 151162. Nature Publishing Group.

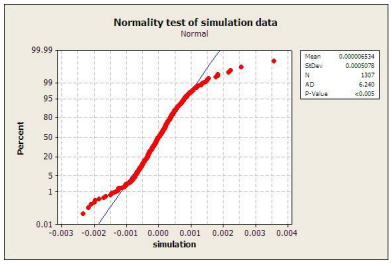
[33] M. E. PADDRIK, R. HAYES JR., A. TODD, S. YANG, W. SCHERER, AND P. BELING, *An Agent Based Model of the E-Mini S&P 500 and the Flash Crash*, SSRN Working Paper 2011.

[34] Q. QIAO, AND P. BELING, *Inverse Reinforcement Learning with Gaussian Process*, Proceedings of 2011 American Control Conference, San Francisco, CA, 2011.

APPENDIX A

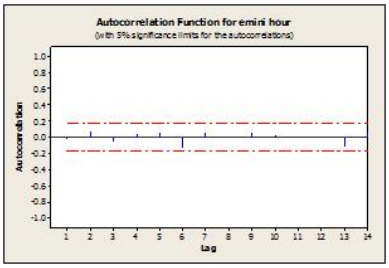


(a) Actual E-Mini S&P 500

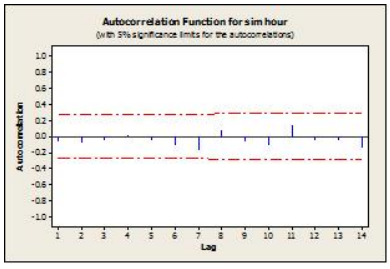


(b) Simulated E-Mini S&P 500

Fig. 7. **E-Mini S&P 500 Heavy Tailed Distribution of Returns** From panel (a) and (b), we see normality tests of returns for both actual and simulated E-Mini S&P 500 show deviation from Gaussian distribution toward both tails.

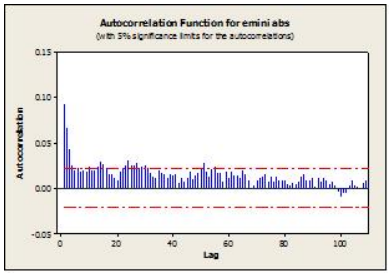


(a) Actual E-Mini S&P 500

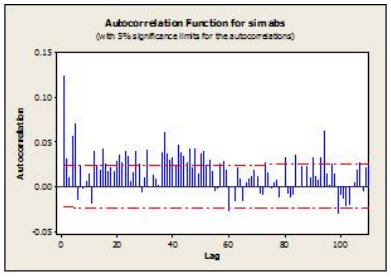


(b) Simulated E-Mini S&P 500

Fig. 8. **E-Mini S&P 500 Absence of Autocorrelation of Returns** From panel (a) and (b), we see autocorrelation of returns for both actual and simulated E-Mini S&P 500 are all close to zero within 95% confidence level.

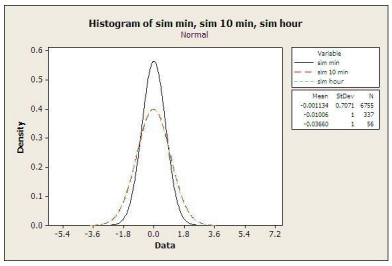


(a) Actual E-Mini S&P 500

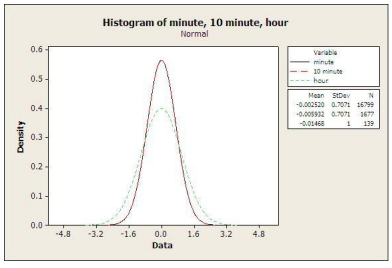


(b) Simulated E-Mini S&P 500

Fig. 9. **E-Mini S&P 500 Autocorrelation Clustering** From panel (a) and (b), we see returns decay slowly for both actual and simulated market. Even though there are few lags outside the 95% confidence lines, the simulation decaying pattern closely resembles that of the actual market as lag increases.



(a) Actual E-Mini S&P 500



(b) Simulated E-Mini S&P 500

Fig. 10. **E-Mini S&P 500 Aggregational Normality** As shown in panel (a) and (b), returns approaches to Gaussian distribution as the time scale increase for both actual and the simulated market.