Robust Market Making via Adversarial Reinforcement Learning

Extended Abstract

Thomas Spooner
Department of Computer Science
University of Liverpool
t.spooner@liverpool.ac.uk

Rahul Savani
Department of Computer Science
University of Liverpool
rahul.savani@liverpool.ac.uk

ABSTRACT

We show that adversarial reinforcement learning can be used to develop market marking strategies that are robust to adversarial and adaptively chosen market conditions.

KEYWORDS

Adversarial Reinforcement Learning; Robustness; Market Making

ACM Reference Format:
Thomas Spooner and Rahul Savani. 2020. Robust Market Making via Adversarial Reinforcement Learning. In *Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), Auckland, New Zealand, May 9–13, 2020*, IFAAMAS, 3 pages.

1 INTRODUCTION

Market makers provide liquidity by offering to buy and sell a financial instrument, and generate profits when they manage to buy and sell evenly. However, a market maker (MM) faces inventory risk, where toxic agents exploit their informational or technological advantages to preempt or cause prices to move unfavourably for the MM as it inadvertently builds a (positive or negative) inventory. This phenomenon has been the subject of much research in the optimal control [4], artificial intelligence [1], and reinforcement learning (RL) literature [10].

In this paper, we develop market making agents that are robust to *adversarial and adaptively chosen market conditions* by applying adversarial RL (ARL). We start from a well-known model of market making [2], which has been used extensively in quantitative finance [3–6]. We convert this model to a discrete-time game, with a "market player," an adversary that can be thought of as a proxy for other market participants that would like to profit at the expense of the market maker. The adversary controls the dynamics of the market environment in a zero-sum game against the MM.

We evaluate the sensitivity of RL-based strategies to three parameters of the model dynamics affecting price and execution for the MM, each of which naturally vary over time in real markets; this feature has not received much attention so far as existing works consider a single instantiation of parameters for the underlying model. We go beyond a fixed parametrisation — henceforth called the **Fixed** setting — with two extended learning settings: the **Random** setting initialises each instance of the model with (bounded) uniformly random values for the three parameters; the **Strategic** setting features the previously mentioned "market player", a learner whose objective is to *minimise* cumulative reward in a zero-sum game with the market maker. The Random and Strategic settings are more realistic than the Fixed setting and pose a significantly more difficult challenge for the market making agent. We show that MM strategies trained in each of these settings yield significantly different behaviour, and, moreover, that adversarial training has benefits beyond its immediately obvious implications.

Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), B. An, N. Yorke-Smith, A. El Fallah Seghrouchni, G. Sukthankar (eds.), May 9–13, 2020, Auckland, New Zealand.

2 TRADING MODEL

We consider a standard model of market making [2, 3] in which an MM trades a single asset of price $Z_{n+1} = Z_n + b_n \Delta t + \sigma_n W_n$, where b_n and σ_n are the *drift* and *volatility* coefficients. Randomness in this model derives from a sequence of independent Normal random variates, W_n, each with zero mean and variance Δt. The process begins with initial value $Z_0 = z$ and continues until step N is reached. At each step, the MM places limit orders about Z_n at which it is willing to buy (bid) and sell (ask), denoted by p_n^+ and p_n^-, respectively; these may be updated at each timestep at no cost.

The probability of orders being executed is dictated by market liquidity and the prices $p_n^\pm$. These interactions are modelled by independent Poisson processes, denoted by N_n^+ and N_n^- for the bid/ask sides, respectively, with intensities $\lambda_n^\pm$. The agent's inventory is then captured by the difference between these two terms, $H_n = (N_n^+ - N_n^-) \in [\underline{H}, \overline{H}]$, where H_0 is known and the values of H_n are constrained such that trading stops on the opposing side of the book when either limit is reached. The arrival intensities are defined by $\lambda_n^\pm = A_n^\pm e^{-k_n^\pm |p_n^\pm - Z_n|}$, where $A_n^\pm, k_n^\pm > 0$ describe the *rate of market order arrivals* and *distribution of volume in the book*.

In this framework, the evolution of the market maker's cash is given by the difference relation, $X_{n+1} = X_n + p_n^- \Delta N_n^- - p_n^+ \Delta N_n^+$, where $\Delta N_n^\pm \equiv N_{n+1}^\pm - N_n^\pm$. We have that the cash flow is a combination of: the profit due to executing at prices away from Z_n; and the change in value of the MM's inventory. The total value accumulated by the MM by timestep n may thus be expressed as the sum of the cash held and value invested $\Pi(X, H, Z) = X + HZ$. This is known as the *mark-to-market* (MtM) value of the MM's portfolio.

Game formulation. The proposed model can be used to define a zero-sum stochastic game between an MM and an adversary, which acts as a proxy for all other market participants.

Definition 2.1. [Market Making Game] The game has N stages. At each stage, MM chooses $p^\pm$ and the adversary b, $A^\pm$, and $k^\pm$. The stage payoff is given by expected change in MtM value of the MM's portfolio, i.e., $\mathbb{E}[\Delta\Pi]$. The total payoff paid by the adversary to MM is the sum of the stage payoffs.

In the full paper [11] we theoretically analyse Nash equilibria (NE) of the single-stage game ($N = 1$) for the two cases of variable $\{b\}$ with fixed $A^\pm$ and $k^\pm$; and variable b, $A^\pm$ and $k^\pm$, and empirically

show a correspondence between these equilibria and the solutions to the multi-stage game that we find with adversarial RL.

3 ADVERSARIAL TRAINING

ARL is used to adaptively penalise policies of the MM that are not robust, i.e., those that are susceptible to exploitation by the adversary. While there are no guarantees our ARL will reach an NE, we show that ARL: consistently converges to an approximate NE, and outperforms past approaches in terms of raw performance *and* robustness to model ambiguity. This approach is adapted from [8] to support incremental actor-critic methods and *asynchronous training*. The adversary is trained in parallel with the market maker, and uses the same state and learning algorithm:

States. The state of the environment $s = (t_n, H_n)$ contains only the current time $t_n = \frac{nT}{N} = n\Delta t$ and the MM's inventory H_n. Transitions are governed by the model dynamics in Section 2.

Rewards. The reward function is adapted from the optimisation objective pioneered by Cartea, Jaimungal, and others, to RL,

$$R_n = \Delta\Pi_n - \zeta H_n^2 - \begin{cases} 0 & \text{for } t < T, \\ \eta H_n^2 & \text{otherwise.} \end{cases} \tag{1}$$

Depending on η and ζ, this formulation can give risk-neutral or risk-adverse preferences, e.g. if $\eta > 0$ and $\zeta = 0$, then the MM is risk averse and is punished if the terminal inventory H_N is non-zero.

Algorithms. Both agents use NAC-S(λ) [13] to learn stochastic policies, using semi-gradient SARSA(λ) [9] for policy evaluation. The value functions are represented by *compatible* [7] radial basis function networks with *accumulating* eligibility traces [12].

Learning settings. We investigate three multi-stage variants of the Market Making Game, each with different restrictions on the adversary's strategy. The following three types of adversary in turn increase the freedom to control the market's dynamics:

Fixed. The adversary sets $b_n = 0$, $A_n^{\pm} = 140$ and $k_n^{\pm} = 1.5$; these are chosen to match settings in [2]. This amounts to a *single-agent learning setting with stationary transition dynamics.*

Random. Each episode has parameters chosen independently and uniformly at random from: $b_n = b \in [-5, 5]$, $A_n^{\pm} = A \in [105, 175]$ and $k_n^{\pm} = k \in [1.125, 1.875]$. These are chosen at the start of each episode and remain fixed until termination. This is analogous to *single-agent RL with varying transition dynamics.*

Strategic. The adversary chooses $b_n, A_n^{\pm}, k_n^{\pm}$ at *each intra-episode step of the game* (bounded as in Random). This is a *fully adversarial and adaptive learning environment* where, unlike e.g. [3], the source of risk is *exogenous* and *reactive* to the policy of the MM.

4 EXPERIMENTS

Training was conducted by initialising each episode with a starting time t_0 chosen uniformly at random from $[0.0, 0.95]$, starting price $Z_0 = 100$, and inventory $H_0 \in [\underline{H} = -50, \overline{H} = 50]$. Prices Z_n have fixed volatility $\sigma = 2$ between $[t_0, 1]$ with increment $\Delta t = 0.005$.

Table 1 shows the performance of MMs trained across all three learning settings and combinations of η and ζ. We find a strict ordering of strategies in terms of the Sharpe ratio ($\mathbb{E}[\Pi_N]/\mathbb{V}[\Pi_N]$), where Fixed < Random < Strategic. For example, when $\eta = \zeta = 0$, the MM in Table 1c generates approximately the same mean profit as in Table 1a, but with *half* the standard deviation. Interestingly, we note that in Table 1c the MM achieves this *with* much lower variance on terminal inventory, even in the risk-neutral case. These observations are shown to hold when the MMs are evaluated in the Random and Strategic environments, suggesting improved invariance to model specification. We also verify, using empirical best response computation, that the solutions found by ARL are (approximately) Nash equilibria of the corresponding game.

Table 1: Performance of MM policies learnt in the three different settings. Reported means and standard deviations were computed from 10^5 test episodes against a Fixed adversary.

(a) Market makers trained against the Fixed adversary.

η	ζ	Term. wealth	Sharpe	Term. inventory	Avg. spread
0.0	0.0	67.0 ± 12.0	5.57	0.56 ± 7.55	1.42 ± 0.02
1.0	0.0	61.3 ± 8.5	7.25	−0.04 ± 1.19	1.76 ± 0.02
0.5	0.0	63.0 ± 9.9	6.34	0.03 ± 1.24	1.67 ± 0.03
0.1	0.0	66.4 ± 9.4	7.06	−0.16 ± 1.86	1.42 ± 0.02
0.0	0.01	63.4 ± 6.8	9.36	0.01 ± 1.45	1.60 ± 0.02
0.0	0.001	66.1 ± 7.4	8.94	0.15 ± 2.97	1.44 ± 0.02

(b) Market makers trained against the Random adversary.

η	ζ	Term. wealth	Sharpe	Term. inventory	Avg. spread
0.0	0.0	66.7 ± 11.8	5.65	0.38 ± 7.14	1.36 ± 0.05
1.0	0.0	59.4 ± 7.6	7.79	0.02 ± 1.09	1.87 ± 0.02
0.5	0.0	62.9 ± 8.4	7.51	−0.04 ± 1.21	1.68 ± 0.02
0.1	0.0	65.8 ± 9.3	7.07	−0.18 ± 1.57	1.46 ± 0.03
0.0	0.01	64.0 ± 6.7	9.54	−0.06 ± 1.31	1.60 ± 0.02
0.0	0.001	65.9 ± 7.2	9.11	−0.07 ± 2.62	1.44 ± 0.02

(c) Market makers trained against a Strategic adversary.

η	ζ	Term. wealth	Sharpe	Term. inventory	Avg. spread
0.0	0.0	65.1 ± 6.7	9.78	−0.05 ± 1.94	1.44 ± 0.02
1.0	0.0	60.5 ± 6.8	8.88	−0.02 ± 0.97	1.75 ± 0.02
0.5	0.0	63.3 ± 6.8	9.32	−0.07 ± 1.05	1.60 ± 0.01
0.1	0.0	64.8 ± 6.7	9.72	−0.06 ± 1.37	1.49 ± 0.02
0.0	0.01	62.9 ± 6.7	9.43	−0.03 ± 1.19	1.65 ± 0.02
0.0	0.001	64.3 ± 6.5	9.85	0.0 ± 1.71	1.44 ± 0.01

5 CONCLUSIONS

We have introduced a new approach for deriving trading strategies with ARL that are robust to the discrepancies between the market model in training and testing. We show that our approach leads to strategies that outperform previous methods in terms of PnL and Sharpe ratio, and have comparable spread efficiency. This is shown to be the case for out-of-sample tests in all three of the proposed settings. In other words, our MMs are not only more robust to misspecification, but also dominate in overall performance, regardless of the reward function used. We verify empirically that ARL finds equilibria of the multi-stage stochastic game, and that in some special cases that these correspond to equilibria in the corresponding single-stage game.

REFERENCES

[1] Jacob D. Abernethy and Satyen Kale. 2013. Adaptive Market Making via Online Learning. In *Proc. of NIPS*. 2058–2066.

[2] Marco Avellaneda and Sasha Stoikov. 2008. High-frequency trading in a limit order book. *Quantitative Finance* 8, 3 (2008), 217–224.

[3] Álvaro Cartea, Ryan Donnelly, and Sebastian Jaimungal. 2017. Algorithmic Trading with Model Uncertainty. *SIAM Journal on Financial Mathematics* 8, 1 (2017), 635–671.

[4] Álvaro Cartea, Sebastian Jaimungal, and José Penalva. 2015. *Algorithmic and High-Frequency Trading*. Cambridge University Press.

[5] Olivier Guéant. 2017. Optimal market making. *Applied Mathematical Finance* 24, 2 (2017), 112–154.

[6] Olivier Guéant, Charles-Albert Lehalle, and Joaquin Fernandez-Tapia. 2011. Dealing with the Inventory Risk: A solution to the market making problem. *Mathematics and Financial Economics* 7, 4 (2011), 477–507.

[7] Jan Peters and Stefan Schaal. 2008. Natural Actor-Critic. *Neurocomputing* 71, 7-9 (2008), 1180–1190.

[8] Lerrel Pinto, James Davidson, Rahul Sukthankar, and Abhinav Gupta. 2017. Robust Adversarial Reinforcement Learning. In *Proc. of ICML*, Vol. 70. 2817–2826.

[9] Gavin A Rummery and Mahesan Niranjan. 1994. *On-line Q-learning Using Connectionist Systems*. Technical Report. Department of Engineering, University of Cambridge.

[10] Thomas Spooner, John Fearnley, Rahul Savani, and Andreas Koukorinis. 2018. Market Making via Reinforcement Learning. In *Proc. of AAMAS*. 434–442.

[11] Thomas Spooner and Rahul Savani. 2020. Robust Market Making via Adversarial Reinforcement Learning. *arXiv preprint arXiv:2003.01820* (2020).

[12] Richard S Sutton and Andrew G Barto. 2018. *Reinforcement Learning: An Introduction*. MIT Press.

[13] Philip Thomas. 2014. Bias in Natural Actor-Critic Algorithms. In *Proc. of ICML*, Vol. 32. 441–448.