A Survey of Reinforcement Learning Approaches for Market Making

Joel Rorseth JOEL.RORSETH@UWATERLOO.CA
University of Waterloo
200 University Ave W
Waterloo, ON N2L 3G1, Canada

Abstract

A key participant in financial markets around the world, the market maker holds the critical responsibility of providing liquidity through its obligation to quote buy / sell prices and execute trades. However, this task is not without risk, as the marker maker must simultaneously manage its own inventory, avoiding personal financial loss whilst ensuring liquidity for the current market demand. In this survey, we evaluate the state of the art for reinforcement learning approaches to the optimal market making problem. Over time, these approaches have strategically incorporated many landmark advancements from the reinforcement learning literature, such as adversarial reinforcement learning and deep reinforcement learning. In addition, many classic and modern market making models from the economics literature are central to their design, giving rise to diverse interpretation of a market making agent's motivations and resulting reward formulation.
Keywords: reinforcement learning, market making

1 Introduction

Following a rapid pace of computer technology innovation over the past several decades, many components and participants of global stock markets have been augmented or replaced by computers. Most notably, stock market investors have sought the aid of artificial intelligence for the purposes of improving their trading strategies, in the hopes of ultimately increasing their realized profits. Market makers, who are important participants in most stock markets, stand to benefit from these innovations in similar ways. As explained by Avellaneda and Stoikov (2008), the role of a market maker is to provide liquidity on an exchange by continuously quoting *bid* (buy) and *ask* (sell) prices for which they are prepared to buy or sell specific quantities of assets. Here, *liquidity* is defined as the availability of the assets being trading on the market. Sufficient liquidity is vital for smooth operation of any market, due to its inherent tradeoff with asset value.

In most modern financial markets, traders (regular market participants) submit bid or ask *limit orders* to a central matching engine, specifying the quantity and worst-case price at which the trader is willing to buy or sell an asset (Lim & Gorse, 2018). In this model, commonly referred to as a *(centralized) limit order book market* (or simply an *exchange*), limit orders are added to a centralized limit order book whenever the parameters of a given limit order cannot be matched with a corresponding order (a buyer matched with a seller). Unless limit orders are cancelled, they remain on the book until the matching engine is able to match it with incoming order. Most markets operate using *price / time priority* rules, meaning that outstanding orders are prioritized first by their price (matching highest bid with lowest ask) and breaking ties by time of arrival (Bessembinder, 2001). The primary alternative to a limit order book market is the *over-the-counter (OTC) market* (also known as a *dealer market*), which exhibits a decentralized design that facilitates order placement directly between market makers and investors (Ganesh et al, 2019). A market maker is often referred to as a *dealer* in the context of an OTC market, which better emphasizes the decentralized nature of their role, however the two are otherwise equivalent for the purposes of this survey.

The market making problem is concerned with the discovery of optimal actions for all possible situations a market maker may encounter. At any point in time, the market maker acts by setting *bid and ask quotes* (which together define the *bid-ask spread*), and selectively accumulating (liquidating) assets for (from) their own inventory. Logically, the market maker may characterize state using its current inventory levels or current market prices. In practice, most markets impose a minimum price change called a *tick*, which discretizes these actions. The market making problem fits naturally within a reinforcement learning framework, where these actions and states generally appear alongside a variety of others. In most works, the problem is formulated as a *partial* Markov decision process, due to (from the market maker's perspective) the hidden nature of information such as the true value of an asset or beliefs of other traders (Chan & Shelton, 2001). In general, all reinforcement learning approaches must develop a model or simulation of the dynamics of the market at hand, which will ultimately constitute its training environment. This is due to a lack of preexisting simulators in the financial domain, which would offer a standardized environment for the training and testing of such agents (Ganesh et al., 2019).

The mathematical complexity of the market making problem (as well as other financial problems) is well established. Researchers such as Lim & Gorse (2018) argue that mathematical modelling of financial markets cannot truly capture the reality of all systems and processes that comprise them. In modern markets with growing numbers of electronic or automated trades, the job of the market maker is further complicated by the increasing frequency and volume of orders. These challenges have become more severe in today's electronic high-frequency markets, which clearly motivates the use of intelligent algorithms to aid the market maker. With algorithms such as reinforcement learning, not only do market makers stand to improve their own profit, but also improve the stability of the market itself.

2 Survey

Although the publication of reinforcement learning approaches for optimal market making has been somewhat sporadic, most noteworthy approaches have evolved in lockstep with the publication of new (or renewed interest in older) market making models from the economics literature. As foretold in seminal publications such as Chan & Shelton (2001), a simple taxonomy for reinforcement learning market making models has remained, directly inspired by these early works in the economics and quantitative finance literature. Specifically, most approaches to the market making problem are classified as either *information-based* or *inventory-based*. In an inventory-based model, the market maker is primarily concerned with the management of their inventory, aiming to reduce the risk associated with holding larger amounts of inventory (Ho & Stoll, 1981). Moreover, inventory risk and order flow uncertainty are critical in the market maker's determination of the bid-ask spread. This model has been thoroughly studied in influential economics publications such as Garman (1976), Ho & Stoll (1981), and O'Hara & Oldfield (1986). On the other hand, in an information-based model, the market maker considers itself (along with other uninformed traders) as having an informational disadvantage to informed participants within the market. In contrast to the inventory-based model, this perspective suggests that the bid-ask spread could be a purely informational phenomenon, rather than a product of inventory risk (Chan & Shelton, 2001). Since most models are inventory-based at the time of this writing (Gašperov et al., 2021), we further divide all approaches by important strengths, properties, and challenges addressed using reinforcement learning techniques.

2.1 Information-Based Market Models

Widely regarded as the first application of reinforcement learning to the market making problem, Chan & Shelton (2001) proposed an adaptive information-based reinforcement learning model. In their paper, three approaches are introduced to directly leverage the well-established SARSA, actor-critic, and Monte Carlo methods respectively. Owing to the partially-observable Markov decision process (POMDP) formulation of the market making problem, convergence to a single ideal policy cannot be guaranteed for any of the three approaches. Regardless, the authors' evaluation does illustrate how after 500 training episodes, all three models will likely exhibit convergence. Although the efforts of Chan & Shelton served as an important first-step, the proposed reinforcement learning algorithms were disconnected from the underlying model, limiting it to more general applications. Further, several assumptions limit the utility of their methods in practice, such as the presumed existence of a true price process and explicit differentiation between informed and uninformed traders. Despite the relative simplicity of their model, their temporal-difference approach struggles due to partial observability and excessive noise.

Building directly upon the information-based approach of Chan & Shelton, Kim and Shelton (2002) develop a model of the order flow to directly represent the process of orders being placed over time. The resulting model eliminates the need for any assumption concerning the true price, or the existence of a true price process. This approach inherently models the reaction of market participants to the bid and ask prices set by the market maker. However, the model is rewarded only by maximizing profit, and several rigid assumptions are asserted. Orders are assumed to be stochastically generated and conditioned on market conditions. Concerning the generative aspect, arrival times and order sizes are assumed to follow a fixed gamma distribution, while the *side* (buy vs. sell) and order prices follow a normal distribution given market conditions.

While early information-based models were influential, their simplicity restricted the real-world utility, especially in modern electronic markets of recent years. Several works sought to extend these early approaches by addressing specific problems that arise from their assumptions. Mani et al. (2019) introduce a risk-sensitive reinforcement learning approach for market making, aiming to reduce risk from holding excessive inventory while increasing net profit. Their approach directly integrates the classical Glosten-Milgrom information model (Glosten & Milgrom, 1985). A market simulation is built, consisting of a single market-making agent, along with groups of informed and uninformed trading agents. The authors extend the Chan & Shelton (2001) reinforcement learning model, integrating the Double SARSA variant to learn two independent Q-value estimates (for action selection and Q-value update). The conservative nature of their risk-averse policy is offset with a Boltzmann Softmax policy, and is shown to yield high profits while maintaining low inventory. Several weaknesses do exist however, including the lack of any convergence guarantee for the Softmax operator. Further, their market formulation does not model competition from other market-makers, which hinders real-world utility. This limitation is common among information-based approaches due to their simplistic model.

2.2 High-Frequency Trading

By leveraging consistent improvements to high-speed communication technologies, modern markets now support the execution of trades in timescales as low as microseconds. Referred to as high-frequency trading, this new standard for typical trade frequencies and volume certainly increases the complexity in deriving a mathematic model of the market dynamics. While most market making research assumes a traditional low-frequency market environment, recent research has indeed considered the ramifications of the modern high-frequency environment inherent to many major stock markets. Most notably, Avellaneda & Stoikov (2008) formalize a landmark mathematical model of a modern market, which accounts for the expectation of high-frequency data (trades) that characterize modern limit order books.

Naturally, several recent reinforcement learning approaches for market making have incorporated this new standard, referred to in the literature as the Avellaneda-Stoikov model, in hopes of mitigating inventory risk more effectively. Notably, Lim & Gorse (2018) published the first practical reinforcement learning approach for high-frequency market making, formulating a discrete Q-learning algorithm. The formulation assumes Markovian state transitions, along with a fully observable environment as opposed to the typical partial variant. Citing the fact that real markets impose a minimum price change (a tick), the authors present a discrete action space to describe the market maker's chosen bid and ask quotes (relative to the *best* bid and ask), along with a state space denoting the time remaining and inventory size. Terminal rewards for the trading period T are calculated using the following form, which accounts for the constant absolute risk aversion (Babcock et al., 1993) commonly abbreviated as CARA.

$$R_T = \alpha - \exp\left(-r(C_T - i_T S_T)\right)$$

Here, α is a chosen constant, r is the risk aversion parameter, C_T is profit (loss), and S_T is the average price to immediately liquidate i_T shares. Additionally, a novel immediate reward is calculated for each timestep t, as described below

$$R_t = a(V_t - V_{t-1}) + e^{b\tau_t} sgn(|i_t| - |i_{t-1}|)$$

where a and b are constants, V_t is the agent's value, i_t is its inventory, and τ_t is the trading time remaining. Discrete Q-learning is employed directly using diminishing ϵ-greedy and learning rate.

Spooner et al. (2018) present another high-frequency market making approach, improving upon Chan & Shelton's less promising temporal-difference agent by incorporating eligibility traces. First, a simulator is built by reconstructing the market with high-frequency equities data. Learning is evaluated using several variants of the general-purpose Q-learning, SARSA, and R-learning algorithms (similar to Chan & Shelton's), ultimately finding SARSA to be optimal. While considering several reward function variants, the authors identify an asymmetrically dampened PnL reward as optimal, which is formulated as follows

$$r_i = \Psi(t_i) - \max\left[0, \eta \cdot Inv(t_i)\Delta m(t_i)\right]$$

where Ψ is the PnL function for timestep t_i, Inv is the agent's inventory, m is the mid-point of the bid-ask spread, η is a chosen scale factor, and Δ is the quoted bid-ask spread.

2.3 Risk-Sensitivity and Safety

Due to an assumed ignorance towards risks inherent to market making, early approaches such as Chan & Shelton (2001) can be characterized as *risk-neutral*. A risk-neutral agent (or model) is only concerned with maximizing the expectation of its defined return and is not concerned with the variance of the sum of discounted rewards (Mani et al., 2019). Likewise, a *risk-averse* agent is indeed concerned by this variance, and its effect on the relative importance of avoiding risky actions. These concepts are well established in the literature under the label of *risk-averse* or *risk-sensitive reinforcement learning*. Minimizing risk is critical in certain high-stakes problems such as market making, where it is crucial to minimize risk of holding too much inventory or quoting a bid-ask spread that is too wide.

To cite practical implementations, the information-based reinforcement learning approach presented by Mani et al. (2019) is specifically formulated for risk-aversion. Meanwhile, other works such as Spooner & Savani (2020) leverage the risk-sensitivity inherent to adversarial reinforcement learning, which mitigate risk by accounting for model misspecification. Mani et al. build upon the work of Mihatsch and Neuneier (2002), modifying their risk-sensitive reinforcement learning model to accommodate the proposed Double-SARSA and one-step temporal difference learning. This addition yields a higher degree of flexibility over traditional risk-neutral agents that preceded it, allowing for direct parameterization of acceptable risk via the risk sensitivity degree $k \in (-1,1)$. More specifically, their reinforcement learning formulation dynamically weighs agent transitions, using k to appropriately overweight (underweight) transitions to successor states with lower (higher) than average rewards. The resulting model is shown to reduce risk (variance) but tends to reduce profit by tolerating less risk. Ganesh et al. (2019) incorporate a similar risk sensitivity variable in their proposed agent's reward formulation, in the form of a penalty term α applied to the total *profit and loss (PnL)*. This term is effective in reducing the inventory PnL variance and excessive mid-price fluctuations.

2.4 Adversarial Robustness to Model Uncertainty

While many early works in market making literature assume complete knowledge of market conditions, true market conditions are highly unpredictable in real-world markets. To combat these inherently adversarial and uncertain market conditions, recent works have explored the application of *adversarial reinforcement learning* techniques. These agents offer the advantage of being more robust to market conditions or model misspecification, and provide inherent risk-aversion as described in the previous section.

Extending the concept of adversarial reinforcement learning, Pinto et al. (2017) formalize the notion of robustness in their definition of *robust adversarial reinforcement learning (RARL*). Spooner and Savani (2020) adapt this formalization to build an adversarial reinforcement learning approach for the market making problem. In their framework, an adversary is created (to represent all other market participants) and becomes the central opponent in a zero-sum stochastic game against the market maker. While playing this game, the market maker learns to avoid playing strategies that are not robust (which are easily exploited by the adversary), seeking convergence towards a Nash Equilibrium strategy. Unfortunately, convergence to Nash Equilibrium is not guaranteed, though convergence to approximate Nash Equilibrium is shown to be consistent. This weakness is offset by its absolute performance and robustness to model uncertainty, which improves upon that of prior approaches.

Although Pinto et al. define the framework for adversarial learning, its adaptation is subject to interpretation. While Spooner and Savani (2020) instruct the adversary to alter the simulation mode parameters, works such as Gašperov & Kostanjčar (2021) choose to perturb the market maker agent directly. In both approaches, a specific state and action space comprise the adversary, who learns to adversely affect the market maker agent on behalf of the presumed adversarial market. The Gašperov & Kostanjčar adversary seeks to displace the market maker's bid and ask quotes, Q_t^{bid} and Q_t^{ask}, though only to a limited degree. Consistent with their rivalry, the reward of the adversary is inverse to that of the market making agent

$$R'_{t+1} = -R_{t+1} = \left(M_{t+1} - Q_t^{ask}\right)\mathbf{1}\left\{Q_t^{ask} exe\right\} + \left(Q_t^{bid} - M_{t+1}\right)\mathbf{1}\left\{Q_t^{bid} exe\right\} + \lambda|I_{t+1}|$$

where for a given timestep t, R'_t and R_t are the reward for the adversary and market maker agent (respectively), M_t is the mid-price, $\mathbf{1}\{Q_t^{\cdot} exe\}$ is an indicator function for whether the bid (ask) order is executed, λ is the volatility / risk aversion parameter, and I_t is the inventory level. While this work also shows impressive results, a potential weakness is the omission of certain important information from the reward formulation, such as the price ranges defined in the state formulation.

2.5 Advantages In Deep Learning

Following recent advancements in the field of deep neural networks, classical reinforcement learning techniques have incorporated these concepts under the label of *deep reinforcement learning (DRL)*. Built upon the foundation of neural networks, deep reinforcement learning gains superior expressivity for the estimation of policies and value functions. Influential deep reinforcement learning techniques such as Deep Recurrent Q-Network (DRQN) (Hausknecht & Stone, 2015) are highly effective in integrating information over time, and have been adapted to the market making problems in works such as Kumar (2020). In this particular approach, DRQN is modified to incorporate double Q-learning and naturally employs temporal difference learning. Moreover, the evaluation presented illustrates significant improvements over the non-deep Spooner et al. (2018) approach, namely to stability and average PnL reward. On the other hand, deep neural networks have been integrated more creatively within existing reinforcement learning algorithms. Guéant & Manziuk (2019) integrate multiple deep neural networks into an actor-critic-like algorithm, which are responsible for several estimations concerning their value function and trade probabilities. The more general application of deep learning to the market making problem has procured many influential models such as Deep Hedging (Buehler et al., 2019), and has overlapped with reinforcement learning approaches on occasion.

According to Kumar (2020), reinforcement learning may be slow to learn in large state spaces or for complex control spaces, which is a motivation for similar works such as Gašperov & Kostanjčar (2021). Their framework incorporates signals from two standalone supervised learning-based signature generating units, which are fed to a DRL unit for market making. The integration of these supervised learning units brings additional advantages to deep reinforcement learning, such as the ability to leverage labelled data and to consider the sequentiality of the market making problem. Though these strengths ultimately yield improvement over several recent benchmarks, the approach is potentially weakened by its limitation to single-asset market making.

2.6 Market Making in OTC Markets

While the market dynamics for nearly all market making approaches discussed adhere to limit order book markets, many other prevalent market structures have not received much attention. Most notably, over-the-counter (OTC) markets have been overlooked within the reinforcement learning market making literature. Until the publications of certain recent works, the distinction between limit order book and OTC markets has not been directly addressed, instead modelling their reinforcement learning environments using simpler limit order book market dynamics established in early works like Chan & Shelton (2001). In contrast to the centralized matching engine in limit order book markets, market makers (dealers) in OTC markets interact directly with investors to facilitate trades (Ganesh et al., 2019). The centralized limit order book allows market makers and traders to observe all outstanding and executed transactions, however the decentralized OTC market restricts observation to trades in which the participant has been involved. Unsurprisingly, the observability of this information is a central assumption in most works, motivating the design of reinforcement learning formulations that can operate in such conditions (or perhaps more generally, partially observable market conditions).

In OTC markets, formulations generally define a relationship consisting of multiple *market makers* and multiple *investors*. Further, the interactions between a market maker and investor must be explicitly modelled in lieu of an actual limit order book process. In the recent work of Ganesh et al. (2019), a multi-agent reinforcement learning simulator is created to model the OTC market and the constituents. Within the simulator, investor agents use a probabilistic trade generation process to select a market maker agent to trade with, but greedily select the market maker with the most competitive quote. OTC market makers may reference price information from an exchange, however crucial information must be derived from previous transactions. The Ganesh et al. market maker agent observes the market share and a reference mid-price, but derives typical information such as inventory and bid-ask spread curves directly from previous trades. Otherwise, an agent's actions allow for setting the bid / ask quotes (and a unique inventory hedging fraction), and is rewarded based on a typical bid-ask spread PnL.

The exploration of market making for other types of markets is not limited to any particular asset class, although most publications in this survey have generalized the notion of asset classes altogether. While Ganesh et al. (2019) explore market making in OTC markets, Guéant & Manziuk (2019) investigate market making for the exchange of corporate bonds within OTC markets. The presented model-based actor-critic-like algorithm is specifically designed to determine bid and ask quotes over large sets of corporate bonds (perhaps a few dozen). This issue of scalability is a significant oversight in the low-dimensional settings established in most other works. The authors also argue that reinforcement learning optimization for OTC market making necessitates the use of a model, or initial estimation of model parameters. This is due to the fact that large datasets are not readily available for OTC markets (and the variety of assets traded therein), however this is not necessarily a problem for many simple limit order book markets studied in the literature.

3 Analysis

Due to the rapid pace of innovation in both artificial intelligence and electronic markets, many combinations of problems and solutions exist within the domain of optimal market making. Given the extreme variety of markets and regulations across the world, paired with the highly unpredictable nature of financial markets, no one solution may be optimal in all conditions or markets. Moreover, researchers rarely measure the performance of their models in real-world live markets, due to the obvious financial costs and related overheads. As such, authors adopt and adhere to a specific market model, and use simulations (historic data if available) to train and test their models. Moreover, early applications of reinforcement learning to market making presented simple models with simple market assumptions, leaving room for later works to mitigate these assumptions using various strategies. To establish the current state of the art in reinforcement learning approaches, we must account for the realities of the market making problem, and consider the relative capability of each model in addressing different aspects of the problem.

Reinforcement learning market making models have evolved alongside models from the economics literature, either drawing inspiration from or directly integrating a specific economic market model. While many recent publications reiterate this sentiment, Lim & Gorse (2018) state that the market model presented by Avellaneda & Stoikov (2008) has become *the* industrial standard, due to its ability to capture the high-frequency nature of modern markets. Likewise, reinforcement learning market making agents have evolved alongside advances from the reinforcement learning literature, by either employing these advances directly or by extending them. Multiple recent approaches have reduced risk and increased profit by incorporating risk-sensitive reinforcement learning or adversarial reinforcement learning, which will certainly remain at the forefront of future research. More importantly, deep reinforcement learning is proving to be essential in current state-of-the-art approaches, due to the rapid innovation occurring in the field of deep learning. Considering that the vast majority of reinforcement learning market making research has been published from 2018 onwards (Gašperov et al., 2021), it is clear that researchers are trying to keep up with recent deep reinforcement learning advances.

The most promising approaches seem to incorporate many of the models and algorithms discussed, or perhaps tackle long-standing assumptions that have restricted the utility of many models. The state of the art is clearly characterized by innovate deep reinforcement learning models such as those proposed by Ganesh et al. (2019) or Gašperov & Kostanjčar (2021). The research of Spooner & Savani has demonstrated the potential of risk-adverse and adversarial reinforcement learning, and stands to benefit from further integration with deep reinforcement learning.

Looking forward, Ganesh et al. and Guéant & Manziuk (2019) are charting new territory by generalizing the market making problem for OTC markets. While the research of market making for different market types is still relatively unexplored, many open problems persist. In general, deep reinforcement learning approaches are still few, and tend to adopt relatively primitive market models that undermine their advantages. Furthermore, many unrealistic assumptions are still common, such as the existence of a true asset price, the absence of trading costs, or ignorance of competing market makers and economic factors. Many assumptions have persisted due to insufficient training data or lack of a known solution, which is frequently noted in the literature.

4 Conclusion

While authoring this survey, I have developed considerable knowledge which spans the topics encompassed by market making and reinforcement learning. Due to the cross-disciplinary nature of this research problem, building foundational knowledge of financial markets and market making was crucial for thorough analysis, and required extensive research exceeding the scope of the selected survey publications. By reviewing these publications, I now understand the challenges and capabilities of market makers that employ reinforcement learning, and the methods by which many challenges have been overcome. Additionally, I have been exposed to several new reinforcement learning techniques, and now understand their capabilities and relevance in the market making problem. This research has educated me on the true utility of many reinforcement learning algorithms, which often prove to be useful long after their inception.

In future research, many challenges and assumptions remain to be solved. Most importantly, the continuous integration of new deep reinforcement learning approaches is sure to help address these challenges and improve efficiency, as has been demonstrated in recent research. I would recommend further investigation into the assumptions made by market models in both the economics and reinforcement learning market making literature. While several publications have focused on these assumptions, future research should continue to test potential solutions for remaining common assumptions. More specifically, emerging reinforcement learning techniques should be monitored for potential application to this challenge.

References

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

Bruce A. Babcock, E. Kwan Choi, and Eli Feinerman. Risk and probability premiums for CARA utility functions. *Journal of Agricultural and Resource Economics*, 17-24, 1993.

Hendrik Bessembinder. Price-time priority, order routing, and trade execution costs in NYSE-listed stocks. *Order Routing, and Trade Execution Costs in Nyse-Listed Stocks (June 2001)*, 2001.

Hans Buehler, Lukas Gonon, Josef Teichmann, and Ben Wood. Deep hedging. *Quantitative Finance*, 19(8):1271-1291, 2019.

Nicholas Tung Chan and Christian Shelton. An electronic market-maker. *Technical report. MIT AI Lab, AI Memo*, 2001(005), 2001.

Sumitra Ganesh, Nelson Vadori, Mengda Xu, Hua Zheng, Prashant Reddy, and Manuela Veloso. Reinforcement learning for market making in a multi-agent dealer market. *arXiv preprint arXiv:1911.05892*, 2019.

Mark B. Garman. Market microstructure. *Journal of financial Economics*, 3(3):257-275, 1976.

Olivier Guéant and Iuliia Manziuk. Deep reinforcement learning for market making in corporate bonds: beating the curse of dimensionality. *Applied Mathematical Finance*, 26(5):387-452, 2019.

Matthew Hausknecht and Peter Stone. Deep recurrent q-learning for partially observable mdps. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2015.

Thomas Ho and Hans R. Stoll. Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial economics*, 9(1):47-73, 1981.

Bruno Gašperov, Stjepan Begušić, Petra Posedel Šimović, and Zvonko Kostanjčar. Reinforcement Learning Approaches to Optimal Market Making. *Mathematics*, 9(21):2689, 2021.

Bruno Gašperov and Zvonko Kostanjčar. Market Making With Signals Through Deep Reinforcement Learning. *IEEE Access*, 9:61611-61622, 2021.

Lawrence R. Glosten and Paul R. Milgrom. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of financial economics*, 14(1):71-100, 1985.

Adlar J. Kim, Christian R. Shelton, and Tomaso Poggio. Modeling stock order flows and learning market-making from data. *Technical report. MIT AI Lab, AI Memo*, 2002(009), 2002.

Pankaj Kumar. Deep reinforcement learning for market making. *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, 1892-1894, 2020.

Ye-Sheen Lim and Denise Gorse. Reinforcement learning for high-frequency market making. *ESANN 2018-Proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 521-526, 2018.

Mohammad Mani, Steve Phelps, and Simon Parsons. Applications of Reinforcement Learning in Automated Market-Making. *GAIW: Games, Agents and Incentives Workshops, Montreal, Canada*, 2019.

Oliver Mihatsch and Ralph Neuneier. Risk-sensitive reinforcement learning. *Machine learning*, 49(2):267-290, 2002.

Maureen O'Hara and George S. Oldfield. The microeconomics of market making. *Journal of Financial and Quantitative analysis*, 21(4):361-376, 1986.

Lerrel Pinto, James Davidson, Rahul Sukthankar, and Abhinav Gupta. Robust adversarial reinforcement learning. *International Conference on Machine Learning*, 2817-2826, 2017.

Thomas Spooner, John Fearnley, Rahul Savani, and Andreas Koukorinis. Market making via reinforcement learning. *Proceedings of the 17th Inter-national Conference on Autonomous Agents and Multiagent Systems (AAMAS2018)*, 2018.

Thomas Spooner and Rahul Savani. Robust market making via adversarial reinforcement learning. *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, Special Track on AI in FinTech*, 4590-4596, 2020.