Multi-Agent Simulation for Pricing and Hedging in a Dealer Market

Sumitra Ganesh [1] **Nelson Vadori** [*1] **Mengda Xu** [*1] **Hua Zheng** [2] **Prashant Reddy** [1] **Manuela Veloso** [1]

Abstract

Financial markets contain a rich set of multi-agent learning problems but lack simulators that could be used to research, develop and test reinforcement learning algorithms. In this paper, we demonstrate the use of agent based modeling to simulate a dealer market with two types of agents - market makers and investors. In particular, we focus on the dynamics that arise from price differentiation and risk management. We show through experiments that our simulation model is able to produce known effects in these markets such as varied price sensitivity among investors and the benefits of internalization for market makers.

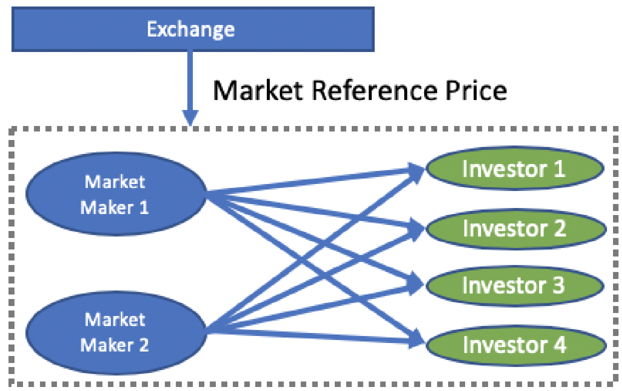

Figure 1. Dealer market and exchange

1. Introduction

Deep reinforcement learning algorithms have had considerable success in surpassing human level performance in several single and multi-agent learning problems. A key ingredient of these advances has been the availability of gaming and physics simulators which provide a ready stream of experience for developing, training and testing algorithms. The financial domain has a rich set of problems, with complex dynamics arising from interactions between heterogeneous agents, but doesn't have a corresponding suite of simulators that could be used to conduct research. Our objective is to bridge that gap and develop simulators that can be used to test and develop learning algorithms for markets. Specifically, in this paper, we focus on a particular type of market - a *dealer market* (Pagano & Röell, 1992) - and demonstrate the use of agent based simulation to model its dynamics.

A dealer market has 2 types of agents: *market-makers* and *investors*. In an electronic dealer market, market makers stream prices at which they are willing to buy and sell as a function of trade size. They act as intermediaries in the market, carrying the *inventory* they accumulate from trades and managing the resulting *risk*. Investors use this facility to execute trades that are necessary to implement their investment strategies. These markets are characterized by direct, bilateral interactions between market makers and investors that are visible only to the parties involved in the interaction. In contrast, in *auction markets* (e.g. exchanges) the participating agents submit anonymous orders to a *limit order book* which is public (Huang & Stoll, 1996), (Pagano & Röell, 1996).

Our goal is to build a multi-agent simulation of a dealer market that can be used to learn optimal policies, in particular for market maker agents, and study emergent outcomes through simulation.

An agent based simulation allows us to study scenarios with heterogeneous agents who have varied incentives and policies. This is important in a dealer market where there are a small number of agents (∼1000s), and the direct interaction makes it valuable for an agent to learn about the other agents' beliefs and preferences. For example, market makers can strategically customize pricing for different investors. A market maker agent could also seek to capture specific portions of the investor trade flow by being marginally more competitive than the next best market maker in the segment.

Previous works on building agent based simulations of fi-

[*]Equal contribution [1]JPMorgan AI Research, New York, USA [2]JPMorgan Quantitative Research, New York, USA. Correspondence to: Sumitra Ganesh <sumitra.ganesh@jpmorgan.com>, Prashant Reddy <prashant.reddy@jpmorgan.com>.

36[th] International Conference on Machine Learning, Workshop on AI in Finance: Applications and Infrastructure for Multi-Agent Learning, Long Beach, CA, 2019.

nancial markets have largely focused on continuous double auction (i.e. limit order) markets (Wah et al., 2017), (Das, 2008), (Palmer et al., 1994). Among the works that consider dealer markets, the focus has been on studying emergent price formation (Das*, 2005) and factors that impact it, e.g.news (Izumi & Ueda, 1999). In contrast, our focus is on studying price differentiation and risk management (hedging), and the interplay between these two important aspects of a market maker's overall strategy in a dealer market.

Our key contribution in this paper is the formulation of an electronic dealer market as a multi-agent simulation, with a comprehensive structure of rewards for market makers. Our experiments show that the agent based simulation is able to produce effects that are commonly observed in these markets - variations in price sensitivity of investors and benefits of a diverse trade flow for market makers. We also implement a realistic hedging policy based on (Almgren & Chriss, 2001) and study its effect on market maker rewards; this study further validates our reward formulation.

2. Electronic Dealer Markets

In this section, we formulate the problem of modeling electronic dealer markets in a multi-agent setting. We define (i) the agents in an electronic dealer market, and (ii) their incentive structures and interactions.

2.1. Market Makers and Investors

In an electronic dealer market, a market maker agent continuously streams prices to all market participants, including other market makers. However, it can provide differentiated pricing depending on the participant. We consider a particular mechanism of price differentiation called *tiering*. In tiering, the market maker assigns each market participant to one of a fixed number of tiers - each tier then receives a different pricing from the market maker. The price (or *spread*) for each tier is also a function of the trade size, with a higher price for larger trades, which carry a greater risk.

Unlike exchanges or auction markets, in a dealer market the price at which a trade is executed is only known to the parties involved in the trade. We consider a dealer market where no information about executed trades, anonymized or delayed, is published and made available to all market participants.

How then, do the market makers decide on how to set their pricing curves? The market makers source reference pricing information from an exchange limit order book (See Appendix A.1 for details). We assume that the trading activity in the dealer market does not impact exchange prices. This is a highly simplified view of the interaction and dependence between the dealer and exchange markets, but it lets us focus on our primary objective to study price differentiation and risk management in a dealer market.

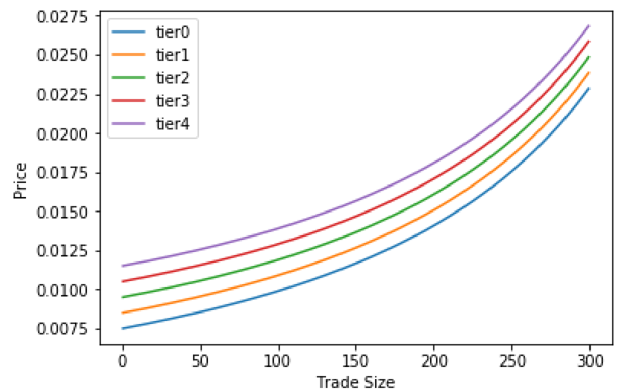

Figure 2. Sample price curves for different tiers

In transacting with other participants, the market maker accumulates a *net position* e.g. if the market maker buys $\$100\mathrm{mm}$ and sells $\$50\mathrm{mm}$, it results in a net position of $\$50\mathrm{mm}$ for the market maker. This net position exposes the market maker to the risk of fluctuaons in the market price of the underlying asset. The goal of the market maker is to optimally manage the risk exposure from its net position while earning revenue from trades with investors.

Some of the risk exposure is organically mitigated by investor trades in the opposite direction - this effect is commonly known as *internalization*. However, the market maker could also choose to actively reduce risk by *hedging* - trading with another market maker and incurring a cost. In general, the market maker has to trade off the risk associated with carrying a net position over a period of time against the cost of reducing the risk by actively hedging.

Investor agents receive prices from all market makers and can choose the market maker to execute with; they are primarily trying to minimize their trade execution costs and typically select the market maker with lowest price.

We assume that each investor has an internal investment process that generates trades. Investors vary in the size, frequency, time horizon and sophistication of their trades. The heterogeneity of investors, coupled with price differentiation by market makers, implies that not all investors have access to the same level of liquidity in this market and that we are likely to see a range of price sensitivities.

2.2. Definition of Agents: Observations, Actions, Rewards

We consider a market with M market maker agents, N investors and K tiers (we assume for simplicity of notation that all market makers use the same number of tiers to differentiate price). We assume there is a single underlying asset being traded in the market. We use the following definitions:

Notation

- Let $\tilde{v}_{i,j,t}$ be the trade executed by agent i with agent j at time t, from agent i's perspective (i.e. $\tilde{v}_{i,j,t} = -\tilde{v}_{j,i,t}$). Buys are positive; sells are negative.
- $\tilde{v}_{i,t}$ denotes the (sum of) trades executed by agent i at time t; i.e. $\tilde{v}_{i,t} := \sum_j \tilde{v}_{i,j,t}$.
- For simplicity, the trade size is represented by v, where $v = |\tilde{v}|$.
- $u_{i,j,t}$ denotes the tier assigned by market maker i to participant j at time t; $u_{i,j,t} \in \{0, .., K-1\}$ where K is the number of tiers and $u = 0$ denotes the best tier.
- P_t denotes the reference mid-price from the exchange (See Appendix A.1 for details)
- $S_{ref,t}(v)$ denotes the market (exchange) reference price curve at time t, where v is the trade size
- $s_{i,t}(v, u)$ denotes the price curve(s) streamed by market maker i at time t; it is a function of trade size v and (discrete) tier u

By convention, both $S_{ref,t}$ and $s_{i,t}$ are expressed relative to the mid price P_t, and are also commonly referred to as the *spread*. For example, the absolute price for executing a trade of size v at price (or spread) $S_{ref,t}(v)$ would be $S_{ref,t}(v) + P_t$.

2.2.1. Market Maker Agent

We summarize below the state, action and rewards structure for market maker agent i at time t.

State / Observations

- History of trades executed $\{\tilde{v}_{i,j,t'}\}_{t'<t, j\neq i}$.
- Net position $z_{i,t} = \sum_{t'<t} \tilde{v}_{i,t'}$
- Pricing from other market makers $s_{j,t}(v, u_{j,i,t})$ where $j \in \{1, .., M\}$ and $j \neq i$
- Market reference price curve $S_{ref,t}(v)$
- Market reference mid price P_t

Actions

- Tiering: Assign each market participant to a tier i.e. determine $u_{i,j,t}$ for all $j \in \{1, ..., M + N\}$ where $j \neq i$
- Pricing: Determine price curve for each tier i.e. $s_{i,t}(v, u = k)$ for $k \in \{0, .., K-1\}$
- Hedging: Determine hedge trade v_{hedge} to execute and the market maker to execute with

Rewards

- Spread revenue: Positive rewards the market maker directly earns from a trade with another participant. For a trade of size v with participant j, the reward is $s_{i,t}(v, u_{i,j,t}) * v$ since the market maker charges a price of $s_{i,t}(v, u_{i,j,t})$
- Position revenue: Additional rewards associated with a trade which are driven by the market mid price fluctuation between the time of trade t and the time the market maker unwinds it (assumed to be a fixed time $t + t_m$). The position revenue for a trade $\tilde{v}$ at time t is $(P_{t+t_m} - P_t) * \tilde{v}$ and added to the rewards at time $t + t_m$. It could be positive or negative.
- Hedging cost: If the the market maker chooses to execute a hedging trade of size v_{hedge} with market maker j, it incurs a cost of $s_{j,t}(v_{hedge}, u_{j,i,t}) * v_{hedge}$
- Risk cost: We also assume there is a risk cost which penalizes the market maker for any adverse impacts of market price movements on its net position. This is computed as $\min(z_{i,t} * (P_t - P_{t-1}), 0)$.

2.2.2. Investor Agent

We summarize below the state, action and rewards structure for investor agent j at time t. In this formulation, we assume an investor has only one trade at each time step and executes it immediately by picking a market maker to trade with.

State / Observations

- Trade to execute: trade size and direction
- Streamed prices from all market makers $s_{i,t}(v, u_{i,j,t})$, $i \in \{1, ...M\}$
- Market reference price curve $S_{ref,t}$ and mid price P_t

Actions

- Select a market maker to execute trade with

Rewards

- Execution cost: The investor incurs a cost in executing trade v with selected market maker i, which is equal to the spread revenue earned by the market maker $s_{i,t}(v, u_{i,j,t}) * v$

3. Simulation Design

The multi-agent formulation detailed above provides the general specifications for how agents interact in a dealer market, the information they observe and the rewards that drive behavior. To simulate a specific instantiation of this multi-agent system we also need to address the following questions:

- What *agent attributes* distinguish individual agents? How do we define a heterogeneous *population of agents* using these attributes?
- What *policies* do agents follow (heuristic/learned) to decide which action to take at each time step?
- How are *exogenous inputs* generated or sourced?

For the simulations in this paper, we focus on (i) generating a heterogeneous population of investors differentiated by the trades they generate, and (ii) developing realistic baseline (heuristic) policies for market maker agents. All investor agents follow the same policy which is to select the market maker with the lowest price for the trade they wish to execute. We describe, in the sections below, the methodology used to generate investor trades and the heuristic policies implemented for the market maker agents.

There are 2 exogenous inputs to our simulation - market reference price curve $S_{ref,t}(v)$ and market mid price P_t. For the simulations presented in this paper, we simulate the exogenous inputs using the methodology described in the Appendix A.1.

3.1. Trade generation process for investor agent

The trade generation process for an investor is specified by probability distributions for trade size, direction and arrival. At each time step, investor j has a trade arrive with probability p_j^{trade}. We assume the trade size is log-normally distributed with parameters $(\mu_j^{trade}, \sigma_j^{trade})$ and that the investor buys with probability p_j^{buy}. Thus the attributes (p_j^{trade}, μ_j^{trade}, σ_j^{trade}, p_j^{buy}) together specify the trade generation process for investor agent j.

We define *sophisticated investors* as those who are more likely than chance to correctly predict direction of market mid price P_t over next t_m time steps (defined by an attribute $q_j \in [0.5, 1)$), resulting in negative Position revenue for the market maker. In practice, we simulate these by using an oracle that reveals the direction of price movement over the next t_m time steps to investor agent j with a probability determined by q_j. If the investor agent gets a look ahead from the oracle, it adjusts its p_j^{buy} for that time step accordingly; otherwise the agent is equally likely to buy/sell.

3.2. Heuristic policies for market maker agent

In our simulations, we define agent behaviors through heuristic policies which use the agent's observations to determine the action to take at each time step. These heuristic policies allow us to study the emergent dynamics of the system and validate that we are able to observe known cause-effect cycles under reasonable assumptions of agent behavior. Our goal is to eventually use the simulator as a multi-agent environment to learn policies for one or more agents; the heuristic policies provide both baselines for comparison and a way to specify validated behaviors for non-learning agents.

The market maker has 3 policy components which determine how it sets the pricing curve for each tier (pricing policy), how it assigns each market participant to a tier (tiering policy) and how it hedges the risk exposure resulting from trades (hedging policy).

3.2.1. Tiering Policy

The tiering policy specifies how the market maker assigns each market participant to a tier. This is typically based on a tiering metric ψ which captures how lucrative it is for the market maker to trade with a specific participant. Looking at the reward structure for the market maker agent, Spread and Position revenue are directly attributable to trading activity with a specific participant and are a good heuristic measure of the participant's profitability.

We define the *revenue* from a trade to be the sum of its spread and position revenue and consider two related tiering metrics for a participant: (i) *Average Yield* which is the average revenue per unit trade volume, and (ii) *Revenue rate* which is the average revenue per time step.

The yield for a trade is computed by normalizing its revenue by the size of the trade. The average yield metric for a participant is an exponentially weighted average of the yields for its past trades. The revenue rate for the participant is then calculated as the product of the average yield and the average volume traded by the participant per time step.

At each time step t, the tiering policy for market maker i is computed as follows:

- Update the tiering metric $\psi_{j,t}$ for each participant j based on trades between the participant and market maker i
- Sort $\{\psi_{j,t}\}_{j=1:M+N, j\neq i}$ into K quantiles or tiers; the highest quantile corresponds to the best tier.

3.2.2. Pricing Policy

For each market maker, the pricing curve $s(v, u)$ represents the price streamed to investors of tier u to enter a trade of size v (we remind the reader that this price is expressed relatively to the mid price P_t). The pricing policy takes as a single input the reference price curve of the underlying asset $S_{ref,t}(v)$, which is known by all participants and specifies the cost of trading a size v on the reference exchange, and generates a pricing curve $s(v, u)$ for each tier u by applying a deformation α and a tiering penalty $s_{tier}(u)$ as follows:

$$s(v, u) := S_{ref}(0) \left(\frac{S_{ref}(v)}{S_{ref}(0)} \right)^{\alpha} + s_{tier}(u).$$

The reference market spread s_0 is the difference between the best ask and bid prices on the reference exchange (cf. Appendix A.1). The half reference market spread $S_{ref}(0) = \frac{1}{2}s_0$ represents the additional cost of trading at the best ask (resp. bid) vs. trading at the mid price on the exchange, and can be seen as a base value for $s(v, u)$. This base value is then scaled by the quantity $\frac{S_{ref}(v)}{S_{ref}(0)}$ that represents the

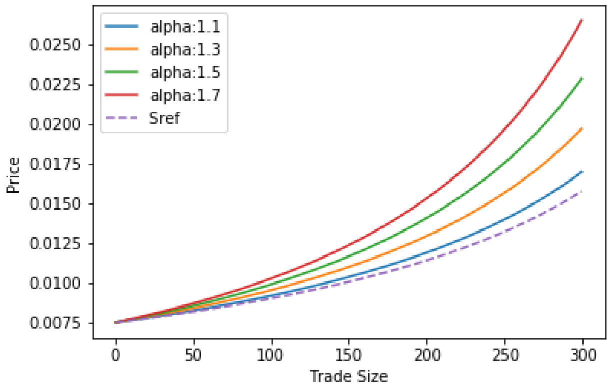

Figure 3. Sample price curves for different values of α

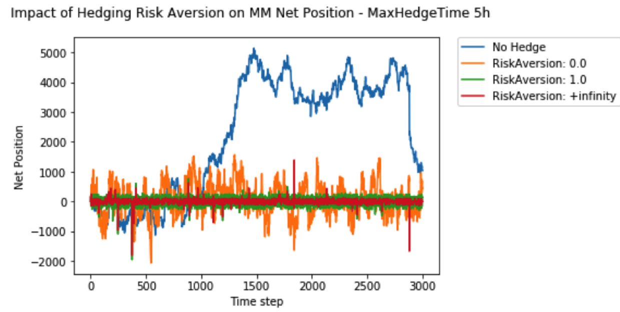

Figure 4. Impact of hedging on market maker's net position, $N_{max} = 5hrs$ - 1 Timestep value=15mins

exchange's volume penalty factor capturing the fact that trading large sizes is done at a higher spread. The market maker further modulates this volume penalty factor via the coefficient α. Finally, it is assumed that the tier penalty applied by the market maker is of the form $s_{tier}(u) = \delta_{tier} \cdot u$, where δ_{tier} is a constant that doesn't depend on the tier u. α and δ_{tier} are the 2 parameters driving the pricing policy and may vary among market makers.

We build on the analysis in (Bouchaud et al., 2002) to derive $S_{ref}(v)$ as a function of the reference spread s_0 and the total liquidity available in the exchange (See Appendix A.1 for details). The total liquidity available in the exchange is chosen empirically so that it is always greater than the maximum trade size generated by the investors and maximum absolute net positions of the market makers. It can alternatively be seen as the probability that the market maker is able to trade at the best ask (resp. bid). The reference market spread s_0 is simulated, at each time step, as a normal random variable of mean 0.015% and stdev 0.005% (we further enforce s_0 to be no less than 0.002% and no more than 0.05%).

3.2.3. Hedging Policy

When the market maker agent has a non zero net position at time t, he needs to hedge so as not to be impacted by the future underlying market price move $(P_u)_{u \geq t}$. The latter is uncertain, which could generate losses for the market maker if it takes no action. If the market maker is long (i.e. its net position is positive), it will sell to reduce its net position, otherwise it will buy. For the market maker agent i, the price at which he can hedge a size v_{hedge} is given by $c(v_{hedge}) := \min_{j \neq i} s_{j,t}(v_{hedge}, u_{j,i,t})$. The heuristic approach we choose for hedging is inspired by the paper (Almgren & Chriss, 2001): since $c(v_{hedge})$ is an increasing function of trade size v_{hedge}, the market maker needs to make a trade-off between:

- hedging immediately a large quantity, in which case he will pay a high price for trading a large volume (cf. hedging cost term of the market maker's reward), but on the other hand he will be less subject to the mid price move (cf. risk cost term of the market maker's reward).
- hedging immediately a small quantity, in which case he will pay a low price for trading a low volume, but on the other hand he will be more subject to the mid price move.

At each time t, the goal for the market maker agent i is to model the above trade-off and derive an optimal size v_{hedge} to hedge. We consider the following inputs of the strategy:

- $z_{i,t}$ is the net position of market maker i at time t.
- $c(v) := \min_{j \neq i} s_{j,t}(v, u_{j,i,t})$ is the price at which market maker i can hedge a size v.
- we assume that the mid price increment over the time period $[t_k, t_{k+1})$ is equal to σZ_k, where $Z_k \sim N(0,1)$ are i.i.d., and σ is market maker i's estimation of the mid price volatility over 1 time period.

Let N_{max} be the maximum number of time periods that the market maker agent gives himself to completely hedge its position. He can decide to hedge everything sooner than N_{max}, but not later. N_{max} is typically of the order of a few hours. Let x_k be the fraction of $z_{i,t}$ that is being hedged at time t_k, and y_k be the remaining fraction of $z_{i,t}$ on $[t_k, t_{k+1})$ $(k = 0..N_{max} - 1)$. x_k and y_k are linked by $x_k = y_{k-1} - y_k$, where $y_{-1} := 1$ for notational convenience and $y_{N_{max}-1} = 0$ by definition of N_{max}. The total cost C (renormalized by $z_{i,t}$) of such a strategy is given by:

$$C = \sum_{k=0}^{N_{max}-1} x_k c(z_{i,t} x_k) - \sigma \sum_{k=0}^{N_{max}-1} y_k Z_k$$

C is a normal random variable with mean $E(C) = \sum_{k=0}^{N_{max}-1} x_k c(z_{i,t} x_k)$ and variance $var(C) = \sigma^2 \sum_{k=0}^{N_{max}-1} y_k^2$, and the corresponding Value-at-Risk at confidence level $p \in [0,1]$, $VaR_p := F_C^{-1}(1-p)$

is given by:

$$VaR_p = E(C) + \gamma\sqrt{var(C)}$$

where $\gamma := \mathcal{N}^{-1}(1-p)$, F_C^{-1} (resp. $\mathcal{N}^{-1}$) denotes the inverse cumulative distribution function of C (resp. inverse normal c.d.f.). VaR_p has the practical meaning that the cost C will always be lower than VaR_p with a probability $1-p$. According to (Almgren & Chriss, 2001), a strategy (x_k) is efficient if - for fixed p - it minimizes VaR_p. In our context, γ is a parameter that represents the risk-aversion of the market maker:

- γ small ($\sim$ 0) means the market maker does not attribute importance to the mid price move and as a consequence is not risk averse.
- γ large ($\sim>$ 2) means the market maker is worried about the mid price move and as a consequence is very risk averse.

Strategy: once γ and N_{max} have been chosen by the market maker, he first solves for the schedule (x_k) that minimizes VaR_p, and then take the amount to hedge for the current time period to be $v_{hedge} = z_{i,t}x_0$.

In figure 4 we show, for one simulation, the impact of γ on a market maker's net position. As expected, the higher γ, the closer the net position is to 0.

4. Experiments

We conducted experiments with 10 investors and 2 market makers. Every time step in the simulation corresponds to 15 min. We simulate a 24-hour, highly liquid market. The 10 investors had μ^{trade} ranging from 0.5 to 7 and trade frequencies ranging from every 2 hours to 2 days. All investors are equally likely to buy/sell at each time step.

We conducted two validation experiments which test whether we observe a variation in price sensitivity among investors and the internalization effect described in Section 2.1. We also varied the parameters of the heuristic hedging policy to study the effect on market maker rewards.

4.1. Investor price sensitivity

In this experiment, we tested the following hypothesis:

Hypothesis: The heterogeneity of investors, combined with price differentiation by market makers in a non-anonymous market, implies that investors will exhibit a range of sensitivities to change in pricing.

Metric: We define price sensitivity in the following manner:

Market share: Market share for market maker i with investor j is defined as the percentage of investor j's trades that are executed with market maker i.

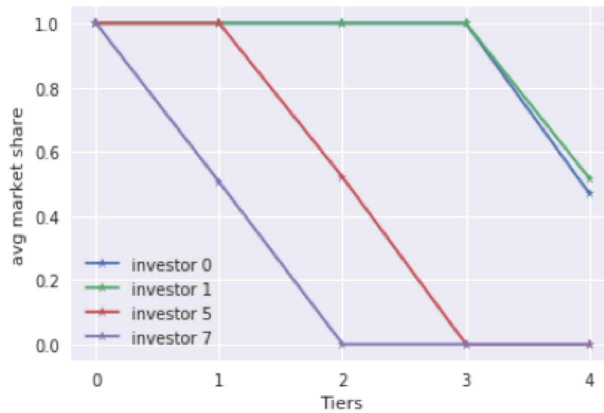

Figure 5. Investor Price Sensitivity - Fixed Tiers. Larger tiers have higher pricing. Comparing with Table 2, we see that the market share is 100 % when market maker 2 tiers the investor better than the competitor, falls to 50% when the tiering is equal, and goes to 0 when it is worse.

Price Sensitivity: Price sensitivity of investor j to market maker i's pricing is defined as the change in market share when market maker i changes the tier for investor j from k to $k+1$; a larger value of k indicating a higher price.

Experiment Design: In each simulation run, market maker 2 picks a particular investor j and places it in tier k for the length of the simulation. We ran 300 simulations for each combination of investor and tier and measured the market share of the investor with market maker 2. To isolate the effect of tier changes, both market makers have the same pricing curve for a given tier. We also set all investors to have the same trade frequency so that the only investor attribute that impacts its tiering is trade size.

Results: In our first set of experiments, both market makers assign a fixed tiering to investors that doesn't change based on investor trades. As seen Table 2, the investors with larger trade sizes, e.g. investor 5 and 7, are tiered well by the competing market maker. Hence, as market maker 2 starts increasing the tier for these investors, it loses market share rapidly (Figure 5). In contrast, for the smaller investors (investor 0 and 1), the market maker can increase the tier several steps before seeing a drop off in market share.

In our second set of experiments, the market makers used the tiering policy based on average revenue per unit time described in Section 3.2.1. The tiering from the competing market maker is now time-varying and depends on investors trades. However, we observe a similar effect (Figure 6), where the larger investors are more sensitive to changes in tiering than smaller investors.

Discussion: The above experiments validate our hypothesis of varied price sensitivity among investors.

Table 1. Summary of investors in Figure 5 - average trade size of investors and tier with the competing market maker

INVESTOR ID	AVG. TRADE SIZE	TIER
INVESTOR 0	0.5	4
INVESTOR 1	1.0	4
INVESTOR 5	3.5	2
INVESTOR 7	4.5	1

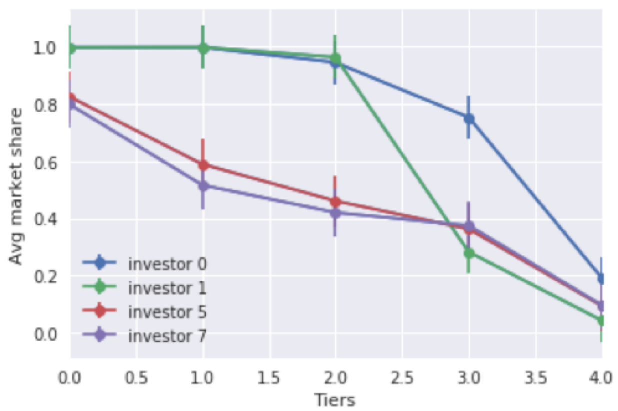

Figure 6. Investor Price Sensitivity - Tiering policy based on average revenue per unit time. Bars indicate standard deviation over 300 simulations.

- Price sensitivity of an investor depends on two factors (i) attributes of the investor (ii) tiering policies of other market makers
- The experiment above could be considered as a brute force exploration that reveals information about competitor tiering policies by measuring the price sensitivity of investors. A market maker agent could employ smarter policies that learn to infer competitor policies and do more efficient exploration using the observations available to it (the market maker doesn't actually observe market share with a specific investor).

4.2. Internalization Effect

Hypothesis: As a market maker captures a larger proportion of trades across all investors, it is able to reduce its net position organically because it sees a greater diversity of investor trades. This is called the *internalization effect*. It is beneficial for the market maker because it lowers risk and hedging costs.

Metric: We measure the internalization effect for a market maker by looking at the net position at time t normalized by the total volume of trades up to that point in time.

$$\frac{|z_t|}{\sum_{t'=0}^{t} v_{t'}} \tag{1}$$

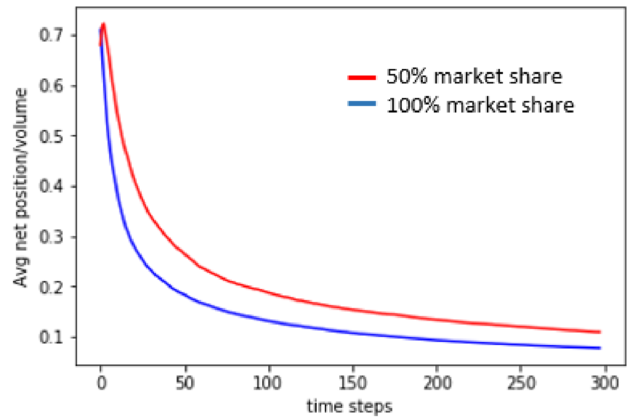

Figure 7. Internalization Effect

where $v_{t'}$ is total volume of trades executed by the market maker at time t'. This internalization metric lies between 0 and 1 - where 0 denotes perfect internalization with stream of trades that perfectly cancel each other out, and 1 denotes no internalization i.e. all trades are in the same direction (buy/sell). In our experiments, we compute the average of this metric over multiple simulations.

Experiment Design: To test the above hypothesis, we compared 2 scenarios: (i) both market makers have the same pricing curve and hence 50% market share (ii) market maker 2 has a strictly lower pricing curve and captures 100% market share. For both scenarios, we ran 1000 simulations. To isolate the effect of the price curve shift, and the resulting increase in trade flow, we considered the simplified case where both market makers do not tier. Also both market makers do not hedge - so any reduction in net position is purely due to the internalization effect.

Results: Figure 7 shows the internalization metric at each time step, averaged over 1000 simulations. As expected, the internalization metric is lower with a 100% market share vs 50%. However, in both scenarios there is internalization - it is just slower with lower market share.

Since all investor agents in our population are equally likely to buy/sell, any portion of the trade flow will show an internalization effect. However, if the internalization is too slow it will not be useful because the market maker is exposed to a large risk for a longer period of time, and will accordingly incur a large cost of risk.

Discussion: While the experiment above confirms that we do observe the internalization effect in our agent based simulation, we need further experiments to test with a larger population of investor agents. We would also like to study the effect of different investor populations and pricing/tiering policy. A smart market maker could use pricing and tiering as a tool to shape its investor flow in a way that increases

Table 2. Internalization metric at 96 time steps (1 day)

INTERNALIZATION METRIC	50% SHARE	100% SHARE
75TH PERCENTILE	0.28	0.20
50TH PERCENTILE	0.16	0.11
25TH PERCENTILE	0.08	0.06

diversity of trade flow and the internalization benefit, while maintaining market share.

4.3. Impact of risk aversion on market maker costs

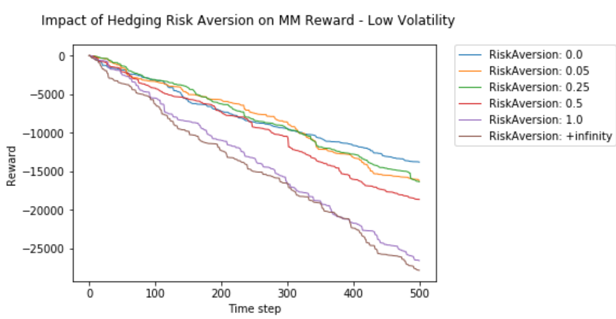

Figure 8. Effect of risk aversion γ on costs: Low volatility scenario. Rewards shown are cumulative and only include hedging and risk costs. Mid-price P_t is a geometric brownian motion with volatility 10%.

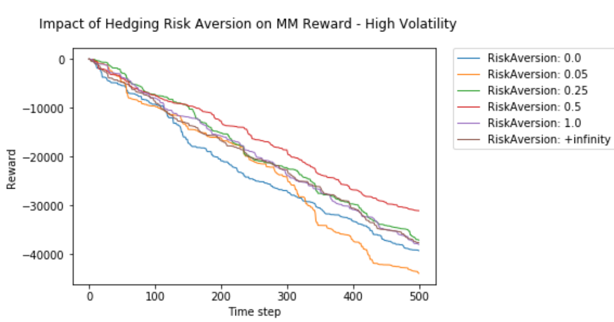

Figure 9. Effect of risk aversion γ on costs: High volatility scenario. Rewards shown are cumulative and only include hedging and risk costs. Mid-price P_t is a geometric brownian motion with volatility 30%.

As described in Section 3.2.3, the hedging policy for a market maker depends on the risk aversion parameter γ and the maximum hedge time N_{max}. The market maker hedges more or less aggressively depending on the level of risk aversion and the maximum hedge horizon he gives himself to entirely hedge his position. Note that in the absence of any cost or penalties for carrying risk, there is actually no incentive for the market maker to hedge (N_{max} very large, $\gamma = 0$), since any hedging would only serve to incur hedging costs without impacting the risk cost (zero).

In our reward design, the risk cost penalizes the market maker for any adverse impacts of the mid-price (P_t) movements on its net position. Ideally, a smart market maker would learn the optimal risk aversion γ and N_{max} from the rewards; however, these are fixed parameters in our heuristic hedging policy.

In this experiment, we fix $N_{max} = 5\text{hrs}$. We empirically study the effect of varying γ on the market maker's total costs - hedging and risk costs - under high and low mid-price volatility.

In the limiting case where the mid-price has a volatility of 0, there should be no incentive for the market maker to hedge anything since there would be no risk in holding the underlying asset. On the other hand, as the mid-price volatility grows, the optimal behavior should tend to $\gamma = +\infty$ as it would become too risky to hold a non zero net position. In the low volatility scenario (Figure 8), a low risk aversion of 0-0.25 indeed produces the lowest cumulative costs, and risk aversions greater than 1 produce the highest costs as expected. In contrast, in the presence of higher mid-price volatility (Figure 9), a risk aversion of 0.5 produces the lowest cumulative costs, and the highest costs are produced by the low risk aversions of 0-0.05. This is because the increased cost of hedging (with higher γ) is outweighed by the reduction in the high risk costs generated by the large mid-price movements.

5. Conclusions and Future Work

In this paper, we formulate a multi-agent simulation of a dealer market, with a detailed reward structure for the market maker agent. We show, through our experiments, that our simulation model is able to produce (i) a variation in price sensitivity among investors, and (ii) an internalization effect for market makers through the netting of trade flows, when its market share increases. We also vary the risk aversion parameter for our heuristic hedging policy and show that our reward structure drives expected results in different market scenarios.

Our experiments in this paper demonstrate the nuanced interplay between price differentiation and risk management policies for market maker agents, which make the joint learning of optimal policies an interesting learning problem for future research. To make our simulation model a more robust platform for learning, we plan to build on the work presented in this paper to (i) scale up the simulation model to thousands of agents, and (ii) use real data to calibrate populations of agents.

References

Almgren, R. and Chriss, N. Optimal execution of portfolio transactions. *Journal of Risk*, 3:5–39, 2001.

Bouchaud, J.-P., Mézard, M., and Potters, M. Statistical properties of stock order books: empirical results and models. *Quantitative Finance*, 2(4):251–256, 2002.

Cont, R. and Mueller, M. S. A stochastic partial differential equation model for limit order book dynamics. *Available at SSRN: https://ssrn.com/abstract=3366536*, 2019.

Das*, S. A learning market-maker in the glosten–milgrom model. *Quantitative Finance*, 5(2):169–180, 2005.

Das, S. The effects of market-making on price dynamics. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 2*, pp. 887–894. International Foundation for Autonomous Agents and Multiagent Systems, 2008.

Huang, R. D. and Stoll, H. R. Dealer versus auction markets: A paired comparison of execution costs on nasdaq and the nyse. *Journal of Financial Economics*, 41(3):313357, 1996.

Izumi, K. and Ueda, K. Analysis of dealers processing financial news based on an artificial market approach. *Journal of Computational Intelligence in Finance*, 7:23–33, 1999.

Pagano, M. and Röell, A. Auction and dealership markets: what is the difference? *European Economic Review*, 36 (2-3):613–623, 1992.

Pagano, M. and Röell, A. Transparency and liquidity: a comparison of auction and dealer markets with informed trading. *The Journal of Finance*, 51(2):579–611, 1996.

Palmer, R. G., Arthur, W. B., Holland, J. H., LeBaron, B., and Tayler, P. Artificial economic life: a simple model of a stockmarket. *Physica D: Nonlinear Phenomena*, 75 (1-3):264–274, 1994.

Wah, E., Wright, M., and Wellman, M. P. Welfare effects of market making in continuous double auctions. *Journal of Artificial Intelligence Research*, 59:613–650, 2017.

A. Appendix

A.1. Simulation of market reference price curve

The market reference price curve $S_{ref}(v)$ is assumed to be known by all participants and specifies the cost of trading a size v on the reference exchange (relatively to the mid price), which operates like a limit order book (cf. figure 10). Such a structure is characterized by 2 lists of 2-tuples (n^a_k, a_k), (n^b_k, b_k), $k \geq 0$, where:

- (orders available on the ask side) one can buy at most a size of n^a_k at price level a_k, where (a_k) is an increasing sequence and a_0 is called the ask price, i.e. the lowest price at which you can buy (but only at most a size of n^a_0).
- (orders available on the bid side) one can sell at most a size of n^b_k at price level b_k, where (b_k) is a decreasing sequence and b_0 is called the bid price, i.e. the highest price at which you can sell (but only at most a size of n^b_0).

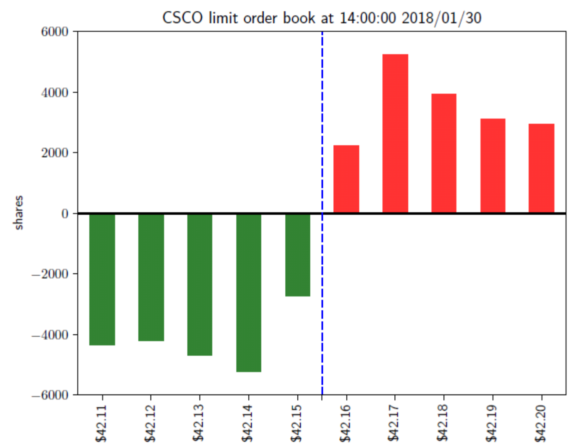

FIGURE 1. Snapshot of the NASDAQ limit order book for CISCO shares (Jan 30, 2018), displaying outstanding buy orders (green) and sell orders (red) awaiting execution at different prices. The highest buying price ($ 42.15 in this example) is the *bid* and the lowest selling price ($ 42.16) is the *ask*.

Figure 10. Illustration of a LOB as displayed in (Cont & Mueller, 2019)

For example, if you want to buy a total size of v, the related cost will be given by:

$$(v - \sum_{k=0}^{n^*-1} n^a_k) a_{n^*} + \sum_{k=0}^{n^*-1} a_k n^a_k$$

$$n^* := inf\{n : v < \sum_{k=0}^{n} n^a_k\}$$

The reference market spread is defined as $s_0 := a_0 - b_0$, and the reference market mid price is given by $P_t = \frac{a_0+b_0}{2}$. In the following we will assume that $S_{ref}(v)$ is the same on the bid and ask side, so we continue the reasoning with the ask side.

In order to derive a realistic behavior for $S_{ref}(v)$, we assume that in the limit order book, all trades have unit size, are independent, and arrive at price level $\frac{s_0}{2} + x$ with a probability density function $f(x) \propto (\frac{s_0}{2} + x)^{-\lambda}$, $\lambda = 1.6$, similarly to the analysis done in (Bouchaud et al., 2002).

Note that we work with a continuous price-scale $\frac{s_0}{2} + x, x \geq 0$ for analytical tractability. After a total of v_{max} trades have arrived (v_{max} then represents the total market liquidity on the ask side), we denote $V(x)$ the number of trades at price level $< x$. We have immediately that $V(x)$ is distributed according to a binomial distribution with parameters v_{max} and $p(x) := \int_0^x f(y)dy$, from which we derive that $E(V(x)) = v_{max}p(x)$.

In order to buy a trade size v, the related additional cost $c(v)$ with respect to the reference market mid price is given by:

$$c(v) = \int_0^{x^*} (\frac{s_0}{2} + x)dV(x),$$

where x^* is implicitly given by:

$$V(x^*) = v.$$

We approximate x^* by the quantity $\widetilde{x^*}$ that solves $E(V(\widetilde{x^*})) = v$, i.e. $\widetilde{x^*}$ represents the average price level at which we have to go up to when buying a trade size v. From there we derive the approximated expected cost:

$$S_{ref}(v) = E(c(v)) \approx E\left[\int_0^{\widetilde{x^*}} (\frac{s_0}{2} + x)dV(x)\right]$$

$$= v_{max} \int_0^{\widetilde{x^*}} (\frac{s_0}{2} + x)f(x)dx,$$

where we have used that $dE(V(x)) = v_{max}f(x)dx$. Using the specific form of f, we get immediately that:

$$S_{ref}(v) = -\frac{s_0}{2}\frac{1}{\widetilde{v}}\ln(1 - \widetilde{v}) \text{ if } \lambda = 2$$

$$S_{ref}(v) = \frac{s_0}{2}\frac{\omega}{\widetilde{v}}(1 - (1 - \widetilde{v})^{\frac{1}{\omega}}) \text{ if } \lambda \neq 2, \lambda > 1$$

$$\widetilde{v} = \frac{v}{v_{max}} \in [0, 1), \qquad \omega = \frac{\lambda - 1}{\lambda - 2}$$