Market Based Mechanisms for Incentivising Exchange Liquidity Provision

W. Gawlikowicz[1], B. Mannerings[1], T. Rudolph[1], D. Šiška[1,2]

[1] Vega Protocol,

[2] School of Mathematics, University of Edinburgh.

30th April 2020

Draft for Vega and selected reviewers

Abstract

Exchanges and other trading venues must be able to reliably offer sufficient liquidity to meet traders' demand, and, in turn, benefit from higher revenue and faster growth when markets are liquid. Traders enjoy lower volatility, tighter spreads, and more efficient pricing in liquid markets. A key problem for exchanges is how to attract liquidity providers and retain their support in all market conditions. This is commonly approached through individual business agreements with market makers whereby a bespoke contract is negotiated for specific obligations and rewards. Such approaches require a central intermediary that profits from liquidity provision to administer, and typically fail to align the incentives of exchanges and liquidity providers as markets grow. This is costly, slow, and scalability is limited by the exchange's resources, contacts, and expertise.

This paper develops mechanisms for creating open, automated and scalable *liquidity markets*. We describe more formal methods to quantify liquidity and discuss various approaches to determine its price. In so doing, we introduce a novel way to structure liquidity commitments, along with a mechanism based on a financial bond with penalties for under-provision to maximise market makers' adherence to their obligations. We also investigate mechanisms to allocate rewards derived from trading fees between market makers, so as to incentivise desirable-but-risky behaviours such as market creation and early commitment of liquidity. We complement this work with several agent based simulations exploring the proposed mechanisms.

1 Introduction

Since the advent of electronic computers in the last century, more and more aspects of running a financial exchange have been automated. Indeed, for most of us the idea of keeping a limit order book by hand seems absurd. Additionally, more and more of the actual trading, including market making, is carried out algorithmically, confirming what Black wrote some fifty years ago [1].

However, there are aspects of financial exchanges that have so far eluded full automation. In particular the contractual relationship between exchanges and market makers still relies on bespoke legal agreements. This reduces the pool of liquidity available to compete for the provision of market making as a service and increases the cost of running the exchange. Both of these in turn decrease the efficiency of the market (the exchange needs to collect higher fees to cover these costs).

If the whole mechanism of the exchange is automated (it is all software) then a significant portion of its revenue (generated by charging fees for trading) should be split between two types of participants: those running the software (operators) and those providing liquidity (market makers). Determining the relative apportioning of incentives between market makers should take into account the value that their specific provision of liquidity has provided, both in terms of its timing, competitiveness, and longevity.

1.1 Motivation

A key problem for exchanges is attracting and rewarding liquidity providers. The entrenched model is to establish business partnerships with market makers, and negotiate obligations (provision of order book volume at some bid/ask) and rewards; typically in the form of fee rebates. These relationships are governed by business contracts making them a non-scalable and expensive solution limited by the business development capacity of the exchange. Furthermore, these agreements are typically non-transferable, have limited responsiveness to market conditions and do not align the incentives of owner/operators of a market and liquidity providers. This is particularly noticeable in the case where market makers (akin to early venture in a startup), invest resources into bootstrapping a new market's liquidity, yet typically gain no explicit benefit from the market's commercial success. This paper develops mechanisms for creating automated and scalable *liquidity markets*. We describe various price determination methods for liquidity provision and how to divide rewards between market makers so as to incentivise desirable behaviours such as early commitment of liquidity to a market. We further outline various ways to measure liquidity provision on an order book, and in so doing introduce a novel way to structure liquidity obligations, along with an automated financial method for penalising market makers who fail to meet the obligations.

We have built our framework assuming that attracting liquidity is occurring in a competitive environment where each market (for a single financial product) is competing with other markets to attract liquidity supply. To put it in a different way, each liquidity provider has the choice of markets to which they supply liquidity and are therefore going to rationally select markets that reward them more highly. We also assume that each market may have multiple market makers and that the rewards are derived from the *all fee income* paid by traders of that market.

To optimally incentivise liquidity providers, it is essential to understand the value, to a given market, of liquidity provision on an order book. The authors know of no established way to do this and this paper explores multiple approaches. Black [1] defines a liquid market as one that is both continuous and efficient. Intuitively, these characteristics increase when order book volume is supplied closer to the "true market price". This paper develops various approaches to evaluating relative liquidity profiles on an order book. We extend this to consider various levers that may be tailored according to a market's liquidity requirements, e.g. firmness, valuing long term provision, etc.

It is crucial to note that there are at least two classes of markets that one should consider: spot markets and derivatives markets (by the latter we mean any that allow trading on margin, e.g.

short positions). For a spot market, liquidity is important in attracting business, efficient price discovery, aiding information flows, etc. However a market that allows margin trading needs to protect itself against potentially insolvent participants. See e.g. [7, Section 6] One of the key parts of the protection is the ability to close out an insolvent participant while protecting the position of the counterpart. In order to be able to do this there must be sufficient volume on the order book to execute such a trade (albeit possibly at a loss). Thus for a derivatives market, sufficient liquidity provision is key for safe operation.

1.2 Literature review

Existing academic research on liquidity focuses on modelling the impact of information flow and interactions between market participants on order book composition.

Grossman and Stiglitz [3] model a market consisting of a risk-free asset with a known, constant return and a risky asset with a return modelled as a sum of two random variables, one of which can be observed at a fixed cost. There are two types of agents in the market: informed traders who choose to incur the cost to reduce variance of the risky asset as observed by them, and uninformed traders who make decisions solely based on observed prices. Furthermore, the agents observe returns realised by the other market players and decide if they want to switch from being informed to uninformed or vice versa. In [3], the authors have shown that, for a special case of agent's with constant absolute risk-aversion utility functions and normally distributed random variables, the equilibrium price distribution exists and can be calculated, and that a number of conjectures that they have formulated can be shown formally. These relate to the impact of the information content of the system on the market price and traded volumes.

Kyle [4] prescribes a similar type of market with one notable addition - market maker is modelled as a separate type of an agent. The market is modelled as a sequence of two step auctions. In the first step, informed trader submits the market order with the quantity based on private observation of the liquidation value of the asset and own trading history. Uninformed traders submit market orders that are uncorrelated with that of the informed agent or own trading history. In the second step, market maker sets prices conditional on quantities traded by other market participants such that the market clears. Market maker is unable to distinguish between the two other types of traders. Market making is assumed to be perfectly competitive - market maker chooses a pricing rule such that the expected profit is zero. The authors in [4] show that a pricing rule linear in the observed traded quantities is optimal and leads to the existence of the equilibrium price. One of the measures of liquidity that the authors consider is the market depth - the order flow required to move a price by one unit of measure. They show that it's proportional to the ratio of quantity traded by uninformed traders to the value of the information held by the informed traders.

Glosten and Milgrom [5] model interaction between price asymmetry and the size of the bid-ask spread, and the volumes traded. The model assumes that market maker sets a single bid and a single ask price per unit of stock. Once the prices are set one of the traders arrives at the market at a random time and decides whether to buy or sell one unit, or do nothing. The market maker is then free to revise the prices and the process continues. Like in the models described above, there are both informed and uninformed traders in the market. All agents are assumed to be risk-neutral. Additionally, market maker is perfectly competitive and incurs no transaction costs - expected profit from each trade is zero. The authors in [5] show that even with the above assumptions the bid-ask spread still arises under their model as a purely informational phenomenon. The bid prices will decrease and ask prices will increase as the proportion of informed traders increases. Moreover, the authors were able to derive a bound on the size of the spread and show that there can be occasions when all trading ceases as no uniformed traders are willing to trade in the presence of too much insider information. Lastly, it's been shown that on average the spread in the model decreases as the traded volume increases.

While the above academic publications were important stepping stones in the analysis of market liquidity, they do not offer any explicit solution to the problem that we're trying to solve.
An interesting approach towards rationalising and automating liquidity provision has been put

forward by Hummingbot [8].
The authors propose a liquidity marketplace built around the *Spread Density Function*. The liquidity buyer specifies a monotonically decreasing function $\rho(s)$ of spread supported on $[0, s_{\max}]$, where $s_{\max}$ is the maximum spread at which rewards for market makers will still be provided. Additionally, the total monthly budget B and number of seconds T defining the frequency with which order book snapshots will be taken get specified.
The total payout available per snapshot is then:

$$b := \frac{B \cdot T \cdot 12}{365.25 \cdot 24 \cdot 3600} .$$

The sum of weighted orders per snapshot is given by $W := \sum_{|s|<s_{\max}} \nu_s \rho(s)$, where ν_s is the aggregate volume of all orders at the spread level s. The payout for market maker m at that spread level is then:

$$R_{s,m} := b \frac{\nu_{s,m}\rho(s)}{W} .$$

The total compensation for market maker m per snapshot is thus given by:

$$b_m := \sum_{|s|<s_{\max}} R_{s,m} .$$

The approach outlined above provides clear rules for interaction between market makers and exchanges and as such is a noteworthy innovation. While the frequent order book sampling and market making reward attribution addresses the problem of market makers withdrawing liquidity at times of high volatility to some extent, it does not fully preclude it. As already mentioned, while it may be an acceptable, albeit undesirable state of affairs for spot exchanges, it is potentially fatal for derivatives exchanges which rely on liquidity for their risk management measures. Thus, we proceed with our analysis of the liquidity provision problem and ways addressing it.

2 Dynamic liquidity rewards

The goal is to set up a market mechanism that optimises the amount of liquidity provision such that liquidity incentives increase when liquidity is under-supplied, and decrease when there is sufficient liquidity in the market. Markets are assumed to potentially have multiple market makers, each of whom can decide which market to supply liquidity to. The mechanism is based on rewards and penalties outlined below.

Market makers are rewarded from the revenue derived by an exchange through the fees charged on trades. Typically both sides of a trade are charged a fee and then rebates are given to market makers if they are involved in the trade. This fee amount is usually expressed in either basis points (bps) or as a percentage of the trade's notional value at the point of trade. One way or another, the fee has cash value and the amount can be split between various participants to motivate desired behaviour.

A spot exchange will highly value market makers who are involved in trades. Hence the mechanism should reward limit orders that are hit resulting in a trade. This rewards market makers for the competitiveness of their pricing.

An exchange allowing margin trading (derivatives exchange) will rely on liquidity depth for closing out delinquent traders. Thus it will choose to reward providing guaranteed liquidity at all times, based on an appropriate measure, see Section 3.

While legal contracts can be used to enforce the obligations, we propose an economic approach where market makers commit a financial bond (or stake) for providing liquidity, which is slashed if they fail to meet their liquidity obligations. The size of the stake will imply a level of liquidity provision commitment.

These commitments may be to provide prices and or respond to prices on the order book. For example, a market maker may be required to maintain an amount of volume, proportional to their stake bid and offered within 15% of the best bid / offer mid price for 85% of the time. In some illiquid markets, the market makers may be required to simply respond to a price placed on the order book with an appropriately competitive counter price.

Here we propose to first fix a measure of liquidity, see Section 3, e.g. by taking λ given by (1) or (6). A market maker committing to provide the liquidity level $\lambda_{\text{committed}}$ will then have to deposit a bond (stake) with the exchange typically calculated as

$$S = \lambda_{\text{commtited}} \cdot \text{scaling constant}\,,$$

where the scaling constant will depend on the liquidity measure and the market in question.

The exchange will then fix a time period τ which could be anything from several seconds to hours or days. If any time during that period volume of limit orders provided by the market maker results in liquidity $\lambda_{\text{provided}}$ which is lower than $\lambda_{\text{committed}}$ a penalty will be applied to market makers' stake. A number of reasonable penalisation strategies can be devised based on the specific market. Typically the penalty should be a fraction of the stake

$$\text{Penalty} = \text{Penalty fraction} \cdot S$$

with $0 < \text{Penalty fraction} < 1$.

The overall income from trading on a given market at an instant of time is the volume at the time multiplied by the trading fee. Market makers will choose to participate (commit a bond and provide liquidity) if the share of the return they are getting is sufficient reward for the capital they are contributing and risk they are taking. In what we are proposing, the market makers are rewarded by obtaining a fraction of the entire fee income. Thus to increase their income they would like to increase the fee or increase the traded volume. However an increase in fee is likely to lead to a decrease in volume and vice versa. We see that the key is then to allow the market makers to jointly set the fee at an appropriate level. It is also clear that different market makers are likely going to have different opinions on what the appropriate level is.

2.1 Voting based mechanism

Each market maker can submit their desired fee: f_i for $i = 1, \dots, n$ with n the number of market makers. Each market maker also has a stake committed S_i (with the resulting liquidity commitment $\lambda_i^{\text{committed}} = S_i/\text{scaling constant}$).

The trading fee is then a simple weighted average

$$f = \frac{1}{S}\sum_{i=1}^{N} f_i \cdot S_i\,,$$

where S is the total committed stake i.e. $S = \sum_{i=1}^{n} S_i$.

2.2 Radical market method

Upon inception of a market there is only one market maker, providing stake S_{old} and setting the market fee f_{old}. During each time period τ the fee either stays as before, or if another market maker chooses to enter the market by providing an additional stake ΔS, then we have $S_{\text{new}} = S_{\text{old}} + \Delta S$ and the fee is adjusted as

$$f_{new} = \frac{S_{\text{new}} - \Delta S}{S_{\text{new}}} \cdot f_{\text{old}}\,.$$

A possible stake and fee evolution is given in Table 1. A more complete agent based simulation is provided in Section 4.1.

Period index	Added Stake	Total Stake	Fee
1	100	100	1%
2	0	100	1%
3	100	200	0.5%
4	300	500	0.2%
5	-200	300	0.33%
6	-200	100	1.0%
7	-50	50	2%

Table 1: Possible fee evolution responding to stake in the radical market method.

2.3 Offer stack meeting liquidity demand

Let us start by trying to estimate liquidity demand in a given market. The simplest way to do this is to consider recent trading activity. One could, for instance, use a moving weighted average of volume of recent trades.

However, a lack of trading should not be interpreted as necessarily meaning there is low liquidity demand. Markets that have very wide pricing (and no trading) may be demonstrating a need for more competitive liquidity provision, since that which is provided is not priced where the demand is.

In the case of derivatives markets, the open interest captures the potential size of defaulting positions that the exchange is bearing at a point in time. The exchange may require immediate access to liquidity in order to close out traders when they approach a risk of bankruptcy. Hence, in this situation, open interest can be taken as an estimate of liquidity demand. Again there is a problem: a derivative market with no open interest doesn't necessarily imply that there is no demand for liquidity. However, if the aim of the exchange in attracting liquidity is primarily to mitigate risk then this may be a very reasonable measure.

Once the liquidity demand has been established, we then have the following relationship

$$\text{Liquidity demand} \longrightarrow \text{Required committed liquidity} \longrightarrow \text{Required market making stake.}$$

Translating required committed liquidity into required stake is just a multiplication using scaling factor. We propose that translating "liquidity demand" (in whichever measure) into "required committed liquidity" best achieved via an affine transformation:

$$\lambda^{\text{required}} := \text{Required committed liquidity} = \text{Scaling factor} \cdot \text{Liquidity demand} + \text{Additive factor.}$$

The n different market makers now submit bids with stake and proposed fee: (S_i, f_i). Assume we have sorted so that they are increasing in f_i (so f_1 is the lowest offered trading fee, f_n the highest). Since stake can be directly translated into committed liquidity we may view this also as $(\lambda_i^{\text{committed}}, f_i)$. Let us define

$$\lambda_k^{\text{cumulative}} := \sum_{i=1}^{k} \lambda_i^{\text{committed}}, \;\; k = 1, \ldots, n\,.$$

The market trading fee is then set by first calculating $k^* := \min\{k = 1, \ldots, n : \lambda^{\text{required}} \leq \lambda_k^{\text{cumulative}}\}$ and then taking the fee to be $f = f_{k^*}$. In other words: we take the liquidity offers of the market makers willing to provide the most competitive trading fees, then we start adding up their committed liquidity and the trading fee is that proposed by the market maker who's committed liquidity meets or just exceeds the required liquidity.

2.4 Distributing Fees

We have mentioned earlier that the trading fees should be distributed between those providing the market infrastructure (operators) and those who provide liquidity (market makers). Since creating a new liquid market where there was none before is typically expensive, the market makers who provide liquidity since inception should receive higher rewards than those who join a liquid, successful venture. On the other hand, having more market making capital committed is generally a good thing (it may drive down fees, and it will increase market resilience). So late entrants need to be incentivised to join.

Hence we propose that the share of fees going to various market makers so that each market maker $i = 1, \ldots, n$ gets proportion p_i given by

$$p_i := \frac{\phi(t - T_i) \cdot S_i}{\sum_{j=1}^{n} \phi(t - T_j) \cdot S_j},$$

where t is the current time,[1] S_i is the stake committed by a market maker i at time T_i (in the past) and where $s \mapsto \phi(s)$ is a bounded, increasing function of time e.g. logistic

$$\phi(s) = \frac{1}{1 + e^{-k \cdot (s - s_0)}}.$$

Here k fixes the steepness of the curve and s_0 the time it takes to go from 0 the midpoint value of $\frac{1}{2}$. The reason to take a bounded function is to make sure that late entrants will, eventually, be assigned enough weight. If we allowed an unbounded function then the late entrants will never catch-up with the early ones - and so they will not have any incentive to join.

Figure 1 shows the resulting fee split between four market makers staking the same overall amount in annual, semi-annual, quarterly and monthly arrears over the course of one year. The example uses a logistic function with parameters $k = 8, s_0 = 0.5$.

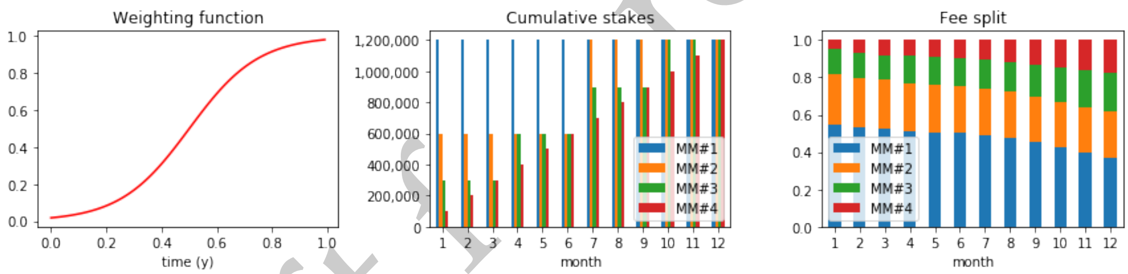

Figure 1: Fee split for a 1y market with $k = 8, s_0 = 0.5$ and four market makers following different staking schedules.

3 Measuring liquidity provision

In this section we consider several possible ways of measuring liquidity. In each case, it must be possible to calculate the liquidity from the current state of the order book.

3.1 Limit order book description

At any time, the state of orders on the order book can be described in terms of the volume $V(t, p)$ of orders waiting at price level p on a grid with mesh size given by the "tick" size θ. When time is fixed or plays no role then we will write simply $V(p)$. Following the usual convention we will use negative volumes (i.e. $V < 0$) for *buy* orders and positive volumes for *sell* orders. The best bid price (best buy offer) is $s^b(V) := \max\{p > 0, V(p) < 0\}$. The best ask price (best sell offer) is

[1] We write a difference in our notation but could in principle count time using various conventions e.g. ignoring periods when a market is shut.

$s^a(V) := \min\{p > 0, V(p) > 0\}$. We will assume that $0 < s^b(V) < s^a(V) < \infty$. This give the *mid price* $S = \frac{1}{2}\left(s^a(V) + s^b(V)\right)$ and the *bid-ask spread* $s^a(V) - s^b(V)$.

We can also take an alternative but related view which describes the order book as volume U at a distance x from the mid-price so that

$$U(x) = V(S + x).$$

If the tick-size θ is small then we can also adopt a continuous approximation to the order book which is now described in terms of *density* $v = v(p)$ given by

$$v(p) \approx \frac{V(p)}{\theta} \quad \text{and} \quad u(x) = v(S + x).$$

Note that we assume that $p \in (-\infty, \infty)$ as on some order books p would not be a price but instead for example a bond yield which can be negative.

3.2 Probabilistically weighted volume

A simplest possible measure of liquidity would be to calculate $\int_{-\infty}^{\infty} |u(x)|\, dx$ i.e. to sum up the entire volume on the book. The problem with this naive approach is that it counts equally all the volume regardless of how far away from the mid price it is.

Let us assume that we have a stochastic model for the mid price at a future time $\tau > 0$, denoted S_τ, which provides us with the probability density for S_τ. Let us denote the density by $f_S = f_S(x)$. Then we could measure the order book liquidity (or equivalently the liquidity amount provided by a single participant) as

$$\lambda(u) := \int_{-\infty}^{\infty} |u(x)| f_S(x)\, dx.$$

This could however report high liquidity even if there were e.g. only sell orders on the book. A more sensible measure of order book liquidity is thus

$$\lambda(u) := \min\left(\int_{-\infty}^{0} |u(x)| f_S(x)\, dx, \int_{0}^{\infty} |u(x)| f_S(x)\, dx\right). \tag{1}$$

On a market where at a certain (large and unlikely) price movement triggers an auction one may only wish to account for volume that is not as far away as to trigger an auction. Specifically if $x_{\text{min}} < 0$ and $x_{\text{max}} > 0$ are the levels that trigger the auction then

$$\lambda(u) := \min\left(\int_{x_{\text{min}}}^{0} |u(x)| f_S(x)\, dx, \int_{0}^{x_{\text{max}}} |u(x)| f_S(x)\, dx\right). \tag{2}$$

Notice that as long as at two points x and x' we have $f_S(x) = f_S(x')$ then equal volume at x and x' provides equal amount of liquidity according to this measure, regardless of the distance of x and x' from the mid. This is counterintuitive to most peoples' understanding of liquidity and hence most likely undesirable. To rectify this let us write F_S for the cumulative density function[2] and let

$$p_S(x) := \begin{cases} \displaystyle\int_{x}^{\infty} f_S(y)\, dy = 1 - F_S(x), & \text{if } x > 0, \\ \displaystyle\int_{-\infty}^{x} f_S(y)\, dy = F_S(x), & \text{if } x \le 0. \end{cases}$$

A moment reflection reveals that $p_S(x)$ is the probability that volume at point x traded at the next time step τ (according to the model given by f_S). Thus another reasonable way to measure liquidity is

$$\lambda(u) := \min\left(\int_{x_{\text{min}}}^{0} |u(x)| p_S(x)\, dx, \int_{0}^{x_{\text{max}}} |u(x)| p_S(x)\, dx\right). \tag{3}$$

[2] i.e. $F_S(x) = \int_{-\infty}^{x} f_S(y)\, dy$.

3.3 Liquidity in terms of slippage

We will measure liquidity in connection with "slippage" i.e. the difference between the theoretical mid market price and the actual volume-weighted price of a hypothetical trade. In particular we will say that we have **volume Λ at percentage** p if one can trade volume Λ while suffering percentage slippage less than p. This method has the following advantages:

i) We can communicate this to participants in an easy fashion: you can trade volume, say 1000 whilst suffering slippage of no more than 10%.

ii) It is easy to attribute across different market makers - we can see exactly who provides how much of the liquidity.

The main disadvantage is that we do not obtain a single number by default: if there are several participants with limit orders on the book who provides more liquidity? The one with big volume far from the mid or the one with less volume but close to the mid? Of course once you fix the percentage then you have a well defined single number (take the lower of buy volume and sell volume achievable at that percentage).

Let $u = u(x)$ denote volume at distance x from mid-price S. Clearly $u(S) = 0$ (if there was volume at mid then a trade would have occurred removing said volume). We will have $u < 0$ denoting selling interest (offers) and $u > 0$ buying interest (bids).

Consider a buy market order for volume M_{buy}. The worst price level it will reach is $x_{\max} = x_{\max}(M_{\text{buy}})$ given by

$$\int_m^{x_{\max}} u(y)\,dy = M_{\text{buy}}\,.$$

The "achieved price" or volume-weighted price is

$$P_{\text{achieved}}(M_{\text{buy}}) = \frac{1}{x_{\max}(M_{\text{buy}}) - S}\int_S^{x_{\max}(M_{\text{buy}})} y\,u(y)\,dy\,.$$

We can then see that the "percentage slippage" from mid that this order suffered is

$$S_{\text{buy}}(M_{\text{buy}}) = \frac{P_{\text{achieved}}(M_{\text{buy}})}{S} - 1\,.$$

Definition. Buy volume at percentage p is defined as the biggest buy market order that can be made suffering slippage at most p% from the mid price S. This is

$$\Lambda^p_{\text{buy}} := \max\Big(V : S_{\text{buy}}(M) \le p\Big) = \frac{1}{S}\left(\frac{\frac{1}{x_{\max}(M_{\text{buy}})-S}\int_S^{x_{\max}(M_{\text{buy}})} y\,u(y)\,dy}{S} - 1\right)\,. \tag{4}$$

We can now consider sell market orders analogously. Given a sell order with volume M_{sell} we know that the worst price level it will reach is $x_{\min} = x_{\min}(M_{\text{sell}})$ given by

$$\int_{x_{\min}}^S u(y)\,dy = M_{\text{sell}}\,.$$

The achieved price or volume-weighted price is

$$P_{\text{achieved}}(M_{\text{sell}}) = \frac{1}{x_{\min}(M_{\text{sell}}) - S}\int_{x_{\min}(M_{\text{sell}})}^m y\,u(y)\,dy\,.$$

The percentage slippage from mid is

$$S_{\text{sell}}(M_{\text{sell}}) = \frac{P_{\text{achieved}}(M_{\text{sell}})}{m} - 1\,.$$

Definition. Sell volume at percentage p is defined as the biggest sell market order that can be made suffering slippage at most $p\%$ from the mid price S. This is

$$\Lambda^p_{\text{sell}} := \max\left(M : S_{\text{sell}}(M) \leq p\right) = \frac{1}{S}\left(\frac{\frac{1}{x_{\min}(M_{\text{sell}})-S}\int_{x_{\min}(M_{\text{sell}})}^{S} y\, u(y)\, dy}{S} - 1\right). \tag{5}$$

Example Consider a limit order book that looks as follows:

	Price p	Volume V
	110	2
Offers	105	1
Mids	100	0
Bids	95	-1
	90	-2
	85	-3
	80	-4

Then we have the following

$$\Lambda^{1\%}_{\text{sell}} = 0\,,\ \ \Lambda^{5\%}_{\text{sell}} = 1\,,\ \ \Lambda^{10\%}_{\text{sell}} = 4\,.$$

Once we fix a percentage p we define the order book liquidity (or the liquidity provided by a single participant as appropriate) as

$$\lambda(u) := \min\left(\Lambda^p_{\text{sell}}, \Lambda^p_{\text{buy}}\right), \tag{6}$$

where Λ^p_{sell} and Λ^p_{buy} are given by (5) and (4) respectively.

3.4 Liquidity across time

Since market-makers are to be rewarded for providing liquidity we must have a way of considering how the liquidity available exists over time. There are two basic ways to see this. First is to consider the average liquidity provided by a participant over a time interval $[0, T]$:

$$\frac{1}{T}\int_0^T \lambda(u_t)\, dt\,.$$

This allows market makers to completely withdraw liquidity for brief periods and compensate by providing more during calm periods. This will generally be undesirable from the point of view of the exchange. The other is the minimum liquidity provided:

$$\min_{t\in[0,T]} \lambda(u_t)\,.$$

4 Agent based models

We now consider two different agent-based models. One model, in Section 4.1, models a mechanism where the trading fee increases if both agents want to decrease their commitment and conversely the trading fee decreases if both agents are willing to increase their bonded commitment. The other model, in Section 4.2, models liquidity demand.

It is worth noting that for either of the models we cannot compute Nash equilibrium explicitly and so we resort to numerical approximation. Broadly, there are two approaches one may use to solve such problems numerically: one based on the Bellman / HJB equation, the other on Pontryiagin's optimality principle. The first approach consists of writing the Bellman / HJB equations for the value functions each of the agents need to solve. These can then be approximated using finite difference methods, see e.g. [6] and then solved iteratively, or first linearised, see e.g. [9] and then approximated using finite differences. However, in this paper we employ the "Method of Successive Approximations" to solve the control problem each agent is solving, see [2].

It is not a priori clear that as a whole differential game there is Nash equilibrium within the space of pure strategies. This can, most likely, be fixed by permitting relaxed (mixed) strategies, but this would be done as part of future work. Moreover we have not analysed the numerical algorithms for convergence and so while they appear to work correctly, we don't have full mathematical proof.

4.1 Two competing market makers in a single market - not modelling liquidity demand

We have two agents and each has different beliefs about the market, which is captured by the volume response function V^i, $i = 1, 2$. We have $(f, S) \mapsto V^i(f, S)$ where $f \in (0, 1)$ denotes the trading fee and $S \in (0, \infty)$ is the total market making stake committed to the market. We assume that $S \mapsto V^i(f, S)$ is increasing for every f (the more stake on a market the more volume it can support) while $f \mapsto V^i(f, S)$ is monotone decreasing (higher fee leads to less trading volume).

The market makers share the income from trading proportionally to their stake size. For a period of time dt this is given by $fV(f, S)\, dt$. Using S^i to denote the stake of MM i she obtains $\frac{S^i}{S^1+S^2} fV(f, S)\, dt$ in that time period. On the other hand, they have to maintain liquidity above a certain level and if they fail they are penalised by the amount $\sigma^i_\lambda S^i$. The other cost would be the cost of capital, which can easily be included, but we omit it for brevity. Finally the market maker is penalised (with $\delta > 0$ small) by $\delta|\alpha^i|^2$ with α^i denoting how quickly they change their stake. At first sight this might look strange but it promotes predictability for other participants, since it prevents the MM from pulling out liquidity immediately.

The MM i adjusts their stake at rate α^i_t and the trading fee is determined by the enthusiasm of MMs to increase their stake - the more they increase the stake the more the fee decreases:

$$dS^1_t = \gamma_S \alpha^1_t\, dt\,, \quad dS^2_t = \gamma_S \alpha^2_t\, dt \;\text{ and }\; df_t = -\gamma_f(\alpha^1_t + \alpha^2_t)\, dt\,. \tag{7}$$

The optimisation problem agent i is solving is to maximise

$$J^i(f, S^1, S^2, \alpha^i, \alpha^{j,*}) = \int_0^T \left[\frac{S^i}{S^1_t + S^2_t} f_t\, V^i(f_t, S^1_t + S^2_t) - \sigma^i_\lambda\, S^i_t \right] dt\,.$$

over all (admissible) strategies $\alpha^i = (\alpha^i_t)$ with the strategy of the other agent assumed to be fixed (and optimal for the other agent).

The resulting game can be solved by an iterative scheme based on the Pontryagin optimality principle.[3] To formulate the scheme let us define $x^i = (f, S^i)$, $x^i_t = (f_t, S^i_t)^\top$,

$$b^i(\alpha^1_t, \alpha^2_t) = (-\gamma_f(\alpha^1_t + \alpha^2_t), \gamma_S \alpha^i_t)^\top$$

and

$$H^i(x^i, S^j, y^i, a^i, a^j) = H^i(f, S^i, S^j, y^i, a^i, a^j) = b^i(a^i, a^j) \cdot y^i + \left(\frac{S^i}{S^i + S^j} f\, V(f, S^i + S^j) - \sigma^i_\lambda S^i \right)\,.$$

[3] This effectively extends the usual first order condition from finite dimensional case, which tells us that a concave function is maximised when its gradient is zero, to the infinite dimensional strategy case.

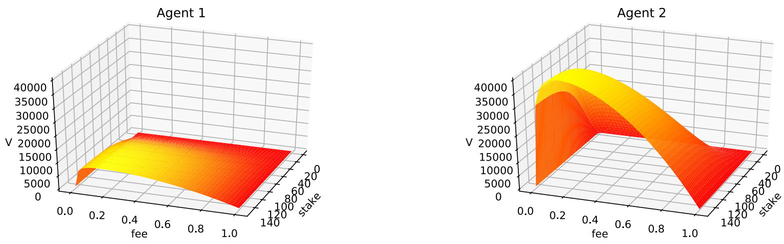

Figure 2: The function $(f,S) \mapsto f\,V^1(f,S)$ (left) and $(f,S) \mapsto f\,V^2(f,S)$ (right).

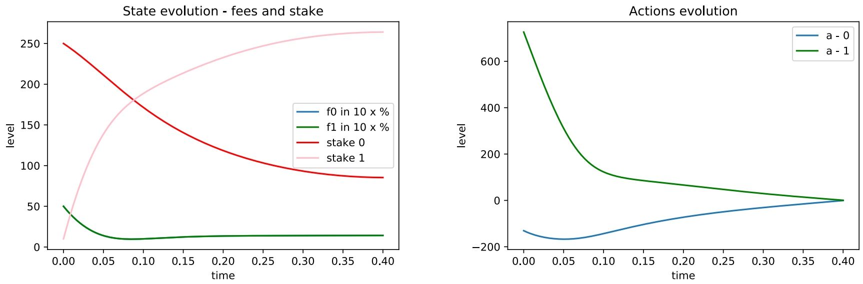

Figure 3: Fees and stake evolution (left) and agents' actions evolution (right).

The optimal control is given by solving

$$
\begin{aligned}
dx_t^{i,n} &= b^i(\alpha_t^{i,n}, \alpha_t^{j,n-1})\,dt\,,\;\; t \in [0,T]\,,\;\; x_0^i = (f, S^i)\,,\\
dy_t^{i,n} &= -(\nabla_x H^i)(x_t^{i,n}, y_t^{i,n}, \alpha_t^{i,n}, \alpha_t^{j,n-1})\,dt\,,\;\; t \in [0,T]\,,\;\; y_T = 0\,,\\
\alpha_t^{i,n} &= \arg\max_a H^i(x_t^{i,n}, y_t^{i,n}, a, \alpha_t^{j,n-1})\,,\;\; t \in [0,T]
\end{aligned}
\tag{8}
$$

iteratively, solving for $\alpha^{1,n}$ with $\alpha^{2,n-1}$ as input and solving for $\alpha^{2,n}$ with $\alpha^{1,n-1}$ as input and repeating. In practice we can replace the maximisation in (8) by one step of gradient ascent with learning rate ρ:

$$
\alpha_t^{i,n} = \rho(\nabla_{a^i} H)(x_t^{i,n}, y_t^i, a_t^{i,n-1}, \alpha_t^{j,n-1})\,,\;\; t \in [0,T]\,.
$$

Simulation results: To run a simulation we need to choose the volume response function; the details are in the Python notebook. Figure 2 displays this function for agents 1 and 2. Agent 2 assumes the same response but higher maximum trading volume.

To test, we create a setting where Agent 1 is staking a market but she thinks that there won't be too much trading (for whatever reason) while Agent 2 starts with almost no stake but has much higher belief in the volume. What we see is that Agent 1 reduces her stake (which on its own would lead to an increase in fees) but Agent 2 increases her stake aggressively, so overall the fees fall. The results are in Figure 3.

4.2 Multiple competing market makers - basic liquidity demand model

We will have $i = 1, \dots, N$ market makers. Each market maker uses two controls: $\theta_t^i = \left[\alpha_t^{i,f}, \alpha_t^{i,S}\right]$ which fix the speed of change in their desired fee and stake respectively:

$$
df_t^i = \alpha_t^{i,f}\,dt \;\text{ and }\; dS_t^i = \alpha_t^{i,S}\,dt\,.
$$

We fix constants $\kappa_D > 0$ (volume response to liquidity demand), $\kappa_f > 0$ (volume response to fee level) and $\mathrm{LD}^*_F > 0$ (market liquidity demand). The trading volume in the market evolves as

$$dV_t = \kappa_D \left(\mathrm{LD}^*_F - \frac{V_t}{\bar{S}_t} \right) V_t dt - \kappa_f L(\bar{f}_t) V_t dt, \quad V_0 = v_0\,,$$

where $\bar{S}_t = \sum_{i=1}^N S^i_t$ and where

$$L(f) = \frac{1}{1 + \exp(-(f - \mathrm{FP}^{\mathrm{mid}}))}$$

is a logistic function determining the fee level from $\bar{f}_t$. Let us now look at how $\bar{f}_t$ is calculated. First we sort f^i_t from smallest to largest (we use π to denote the sorting permutation) :

$$f_t^{\pi(1)} \leq f_t^{\pi(2)} \leq \cdots \leq f_t^{\pi(N)}\,.$$

We calculate the cumulative stake corresponding to market makers providing fees, sorted from smallest to largest:

$$C^i_t := \sum_{j=1}^{i} S_t^{\pi(j)}\,.$$

We check the index of the last market maker providing commitment needed to meet current liquidity demand:

$$i^* := \max \left\{ i = 1, \ldots, N : \kappa_C C^i \leq \frac{V_t}{\bar{S}_t} \right\}.$$

Finally we set

$$\bar{f}_t := \sum_{j=1}^{i^*+1} f_t^{\pi(i)} \frac{S_t^{\pi(i)}}{C^{i^*+1}}\,.$$

This matches the mechanism described in Section 2.3 .

The system dynamics for each agent is

$$d\boldsymbol{X}^i_t = b^i \left(\boldsymbol{X}^{-i}_t, \boldsymbol{X}^i_t, \boldsymbol{\theta}^{-i}_t, \boldsymbol{\theta}^i_t \right) dt$$

where $\boldsymbol{\theta}^{-i}$, $\boldsymbol{X}^{-i}$ denotes the strategy and state of all the other market makers / agents and

$$\boldsymbol{X}^i_t = \left[V_t, f^i_t, S^i_t \right], \quad \boldsymbol{\theta}^i_t = \left[\alpha^f_t, \alpha^S_t \right]$$

with

$$b^i \left(\boldsymbol{X}^{-i}, \boldsymbol{X}^i, \boldsymbol{\theta}^{-i}, \boldsymbol{\theta}^i \right) = \begin{bmatrix} \kappa_D \left(\mathrm{LD}^*_F - \frac{V}{\kappa_S S} \right) V - \kappa_f L(\bar{f}_t) V \\ \alpha^{i,f} \\ \alpha^{i,S} \end{bmatrix}.$$

The optimisation criteria for market maker / agent i is to maximise

$$J^i(\boldsymbol{\theta}^{-i}, \boldsymbol{\theta}^i) = \int_0^T F^i(\boldsymbol{X}^{-i}, \boldsymbol{X}^i_t, \boldsymbol{\theta}^{-i}, \boldsymbol{\theta}^i_t)\, dt + U^i_{\mathrm{volume}}(V_T)\,,$$

over strategies $\boldsymbol{\theta}^i$, where $\boldsymbol{\theta}^{-i}$, $\boldsymbol{X}^{-i}$ denotes the strategy and state of all the other market makers / agents and where

$$F^i(\boldsymbol{X}^{-i}_t, \boldsymbol{X}^i_t, \boldsymbol{\theta}^{-i}_t, \boldsymbol{\theta}^i_t) = U^i_{\mathrm{cash}} \left(L(\bar{f}^\theta_t) V^\theta_t \frac{S^i_t}{\bar{S}^\theta_t} - \gamma^i S^i_t \right) - \frac{\delta_f}{2} (\alpha^{i,f}_t)^2 - \frac{\delta_S}{2} (\alpha^{i,S}_t)^2\,.$$

Notice that each agent has their own financial utility and terminal volume utility. We are aiming to find the Nash equilibrium i.e. strategies $\hat{\theta}$ such that for any $i = 1, \ldots, N$ using some other strategy θ^i would mean that

$$J^i(\hat{\boldsymbol{\theta}}^{-i}, \hat{\boldsymbol{\theta}}^i) \geq J^i(\hat{\boldsymbol{\theta}}^{-i}, \boldsymbol{\theta}^i)\,.$$

Let

$$H^i(\boldsymbol{X}_t^{-i}, \boldsymbol{X}_t^i, \boldsymbol{\theta}_t^{-i}, \boldsymbol{\theta}_t^i, \boldsymbol{P}_t^i) = b(\boldsymbol{X}_t^{-i}, \boldsymbol{X}_t^i, \boldsymbol{\theta}_t^{-i}, \boldsymbol{\theta}_t^i) \cdot \boldsymbol{P}_t^i + F^i\left(\boldsymbol{X}_t^{-i}, \boldsymbol{X}_t^i, \boldsymbol{\theta}_t^{-i}, \boldsymbol{\theta}_t^i\right).$$

The Nash equilibrium is the approximated by recursively. We denote by $\boldsymbol{X}_t^{-i,n}, \boldsymbol{X}_t^{i,n}, \boldsymbol{\theta}_t^{-i,n}, \boldsymbol{\theta}_t^{i,n}, \boldsymbol{P}_t^{i,n}$ the state at iteration n. Starting from an initial guess of a strategy for each player we then repeatedly solve

$$\begin{aligned}
d\boldsymbol{X}_t^{i,n} &= b^i(\boldsymbol{X}_t^{-i,n}, \boldsymbol{X}_t^{i,n}, \boldsymbol{\theta}_t^{-i,n}, \boldsymbol{\theta}_t^{i,n})\, dt\,, \;\; t \in [0,T]\,, \;\; \boldsymbol{X}_0^{i,n} = \boldsymbol{x}^i\,, \\
d\boldsymbol{P}_t^{i,n} &= -(\nabla_x H^i)\left(\boldsymbol{X}_t^{-i,n}, \boldsymbol{X}_t^{i,n}, \boldsymbol{\theta}_t^{-i,n}, \boldsymbol{\theta}_t^{i,n}, \boldsymbol{P}_t^{i,n}\right) dt\,, \;\; t \in [0,T]\,, \;\; \boldsymbol{P}_{T,1}^{i,n} = (\nabla_V U^i_{\text{volume}})(V_T^n)\,, \\
\boldsymbol{\theta}_t^{i,n+1} &= \arg\max_{\boldsymbol{a}} H^i\left(\boldsymbol{X}_t^{-i,n}, \boldsymbol{X}_t^{i,n}, \boldsymbol{\theta}_t^{-i,n}, \boldsymbol{a}, \boldsymbol{P}_t^{i,n}\right)\,, \;\; t \in [0,T]
\end{aligned} \tag{9}$$

iteratively, solving for $\alpha^{1,n}$ with $\alpha^{2,n-1}$ as input and solving for $\alpha^{2,n}$ with $\alpha^{1,n-1}$ as input and repeating. In practice we can replace the maximisation in (9) by one step of gradient ascent with learning rate $\rho > 0$. To improve stability of the algorithm it may be useful to add some random perturbations. With $\sigma \geq 0$ and $(Z_n)_{n\in\mathbb{N}}$ i.i.d. random variables (e.g. $N(0,1)$) we would then use

$$\boldsymbol{\theta}_t^{i,n+1} = \rho \cdot (\nabla_{\boldsymbol{a}^i} H^i)\left(\boldsymbol{X}_t^{-i}, \boldsymbol{X}_t^{i,n}, \boldsymbol{\theta}_t^{-i,n}, \boldsymbol{\theta}_t^{i,n}, \boldsymbol{P}_t^{i,n}\right) + \sigma\sqrt{\rho} Z_n\,, \;\; t \in [0,T]\,.$$

The results are in Figure 4.

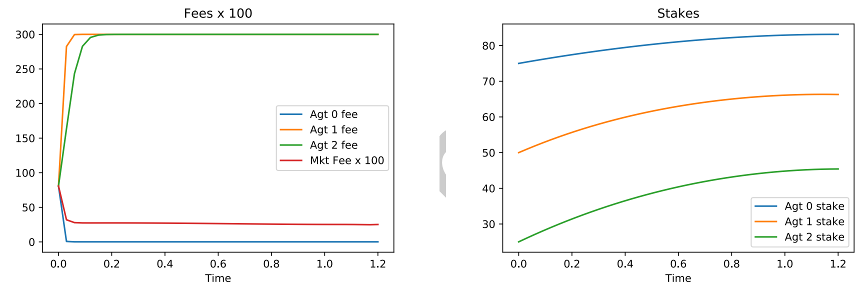

Figure 4: Fees and stake evolution (left) and agents' actions evolution (right).

References

[1] F. Black. Toward a Fully Automated Stock Exchange. *Financial Analysts Journal*, 27(4), 28–35+44, 1971.

[2] F. L. Chernousko and A. A. Lyubushin. Method of Successive Approximations for Solution of Optimal Control Problems. *Optimal Control Applications & Methods*, 3, 101–114, 1982.

[3] S. J. Grossman and J. E. Stiglitz. On the Impossibility of Informationally Efficient Markets. *The American Economic Review*, 70(3), 393–408, 1980.

[4] A. S. Kyle. Continuous Auctions and Insider Trading. *Econometrica*, 53(6), 1315–1335, 1985.

[5] L. R. Glosten and P. R. Milgrom. Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders. *Journal of Financial Economics*, 14, 71–100, 1985.

[6] I. Gyöngy and D. Šiška. On Finite-Difference Approximations for Normalized Bellman Equations. *Applied Mathematics & Optimization*, 60(3), 297–339, 2009.

[7] G. Danezis, D. Hrycyszyn, B. Mannerings, T. Rudolph, and D. Šiška. Vega Protocol: A liquidity incentivising trading protocol for smart financial products. *Vega research paper*, 2018.

[8] M. Feng, R. Bhat, and C. P. Las Marias. Liquidity Mining: A marketplace-based approach to market maker compensation, *Hummingbot research paper*, 2019.

[9] B. Kerimkulov, D. Šiška, and L. Szpruch. Exponential Convergence and Stability of Howards's Policy Improvement Algorithm for Controlled Diffusions. To appear in *SIAM Journal on Control and Optimization*, 2020.