Trading Illiquid Goods: Market Making as a Sequence of Sealed-Bid Auctions, with Analytic Results

Peter Cotton and Andrew Papanicolaou

April 7, 2022

Abstract

We provide analytic results for the optimal control problem faced by a market maker who can only obtain and dispose of inventory via a sequence of sealed-bid auctions. Under the assumption that the best competing response is exponentially distributed around a commonly discerned fair market price we examine properties of the market maker's optimal behavior. We show that simple adjustments to skew and width accommodate customer arrival imbalance. We derive a straightforward relationship between the market marker's fill probability and direct holding costs. A simple formula for optimal bidding in terms of (non-myopic) inventory cost is presented. We present the results as a perturbation of an improvement to a "linear skew, constant width" (CWLS) market making heuristic.

1 Introduction

The facilitation of trade by intermediaries often takes on the form of a sequence of sealed-bid auctions. This arises because a customer wishing to sell an item will typically obtain several bids from dealers, and then select the best price. Similarly, a customer wishing to buy an item will perform a similar search then transact with whomever offers the lowest price. The market maker will take on price risk and carrying cost, and must in responding to inquiry take that into account. They are said to mark up their offer or mark down their bid, relative to a fair market price.

The problem addressed in this paper is quite general. We have in mind fungible goods in which the market maker can build up an inventory of two or more items (as with limited edition prints, computers of a certain specification, or even cookie-cutter New York apartments). Moreover we need not sacrifice generality and can allow the inventory be represented by a real number. Negative inventory is allowed to accommodate situations where an intermediary (say on eBay) sells a book before it is written, a computer before it is ordered from elsewhere (as with inter-venue arbitrage schemes), or a security they don't currently possess (short selling).

Almost always the *risk adjusted* running cost of holding a good eventually increases super-linearly with the absolute value of inventory, due to market risk and finiteness of wealth. This super-linear cost provides the market maker with a strong incentive to "lean", or "skew" their prices in a manner that will encourage a return towards a neutral (or alternatively a target) position. In this paper we couch the optimal behavior of the intermediary as a perturbation of a simple heuristic which we believe to be quite widely applied. We refer to this baseline as the Constant Width Linear Skew (CWLS) model for reasons that will shortly be apparent.

There are aspects of the literature that this paper does not address. First, we do not attempt to derive an equilibrium solution in which all market makers are optimal. We instead take what we believe is a more pragmatic approach, assuming that $n-1$ dealers follow a sub-optimal policy that is simply parametrized. Our job is to optimize the behaviour of the n'th dealer only.

Second, we do not *directly* address asymmetry in information amongst the market makers - instead assuming that the fair market price is known to all and agreed upon. Our focus is purely on inventory management and a particularly apt application of this paper's results would be markets where experts have a firm grasp of value, perhaps because of exogenous price discovery. An example is block trading in equities. An individual or entity might posses a large number of shares that they wish to sell on the market, and rather than battle the adverse selection on the exchange they instead solicit bids from a group of large players who are interested in acquiring the position. Thus, the market for the block trade is made through a sealed-bid auction; [9] explains how block trades and market making are done through auction mechanisms; but an extant price reference exists at all times in the form of a central limit order book (such as the New York Stock Exchange) which greatly limits possible heterogeneity in opinions about price.

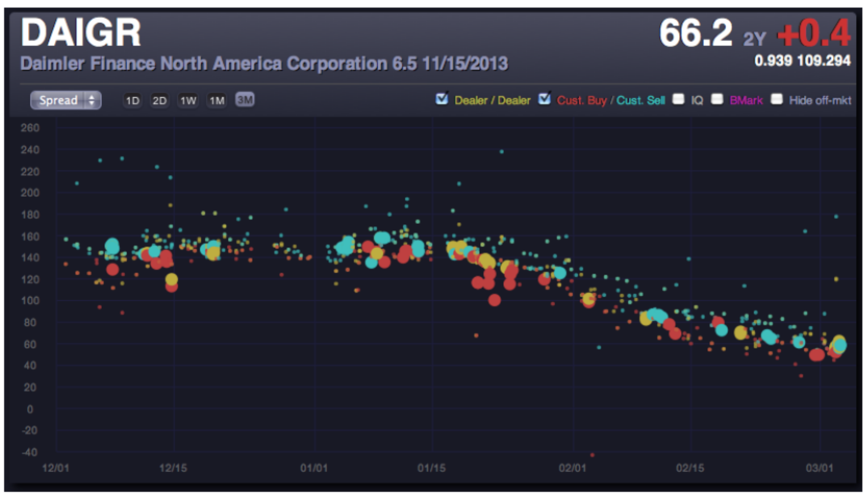

Figure 1: The trading history of Daimler Chrysler (DAIGR) bonds. Trades have different sizes, are either a buy or sell as per the request issued by a costumer, but there may also be dealer to dealer trades too. The statistical phenomenon that we look to capture in our model is clustering, i.e., the phenomenon of trades not occurring for relatively long lengths of time and and at other times occurring in a grouping over time interval. Clustering effectively simulated by letting the arrival times of trade opportunities be a Hawkes process.

1.1 The CWLS Benchmark. Making constant width markets with skew linear in inventory.

To explain the terminology, we define a dealer's width to be the difference between where they are prepared to bid and where they are prepared to offer. We define the skew as the difference between

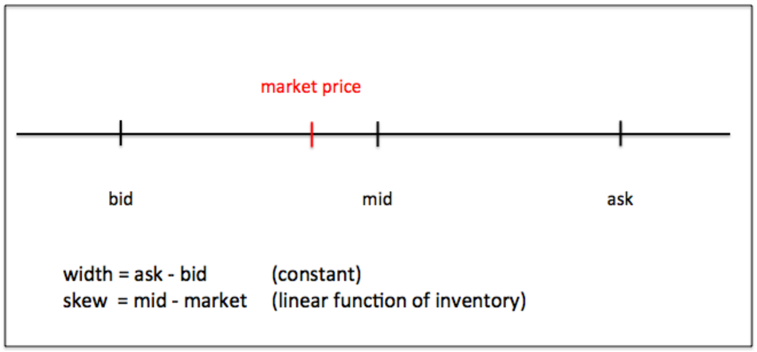

Figure 2: The CWLS heuristic. The width between the market maker's bid and ask is constant, and their mid minus the market price is linear in inventory.

the average of these two and the dealer's estimate of fair value. The CWLS heuristic suggests that a market maker maintain constant width and set skew proportional to inventory - with a negative sign on the skew so that when inventory is high the dealer leans in the direction of reducing it (making their bids more defensive and their offers more aggressive; see Figure 2). Under CWLS the price markup when buying is

$$\text{markup} = \text{const} + bx$$

where b is a parameter and x the signed inventory.

Motivations for CWLS range from its prima facie reasonableness to formal justification as an exact or approximate solution. Anecdotally it is often used without any justification. As an example of an informal argument there is evidently some resemblance to the classic quadratic oscillator problem - although we hasten to add that there are usually implicit assumptions lurking which may or may not be justified in any particular market (for example some linkage must be made there between how far the market maker leans and how quickly the inventory moves back towards zero).

From a practical perspective the CWLS baseline is not a straw man because it is simple. Even in the more economically significant markets, conscious attempts at optimal market making have remained somewhat arcane. Relatively sophisticated mathematical approaches are sometimes employed, but this tends to require equally sophisticated owners of the techniques, non-trivial model maintenance and sometimes, significant technology investment. These are large fixed costs so while the marginal economics may work for hedge funds and banks, the same may not be true of many other situations where items are traded.

Another reason CWLS is not so easy to beat is that it sometimes *is* the solution - or at least can be made to look like the solution after some approximations are introduced - for example see [1] mentioned in Section 1.2. Practitioners who have already stumbled on the baseline method might not be strongly motivated to delve into the details of how and why certain formulations of the problem have approximate solutions of this form.

An analogy might be made to interest rate models whose formulation as a stochastic short rate is simple, but whose justifications as the equilibrium point of an economy are more involved. Typically practitioners find it easier to shoe-horn the solution than to tweak the setup in which this endogenously arises.

1.2 Literature Review

Market making and formation of the bid-ask spread is the sudbject of various papers including [5, 4] where markets are modeled as double auctions. More general auctions literature with some application to finance includes [6, 7, 8, 10], and also [9] where auctions are studied in the setting of negotiated deals and block trades. Turning to more formal justification for CWLS we mention [3, 2], and [1] where market making as a carefully formulated stochastic control problem in which CWLS is provided as a solution. As noted this requires some approximation, however, which more or less ensures linear skew. Those first order considerations may indeed be adequate when the control can be applied in a continuous manner - as with high frequency trading of equities which is the subject of [1] - but less so the trading of rare baseball cards.

2 Model for Illiquid Goods

Let the arrival of trade opportunities be a Hawkes process,

$$d\lambda_t = \kappa(\mu - \lambda_t)dt + \gamma dN_t \ , \tag{1}$$

$$dN_t =_d \text{iid poisson}(\lambda_t dt) \ , \tag{2}$$

and let the i^{th} trade time be denoted as τ_i, i.e.,

$$\tau_i = \inf\{t > 0 | N_t = i\} \ .$$

Each trade is of the same amount, denote by s,

$$s = \text{trade unit size,}$$

and hence the occurrence of large trades are treated as clusters of many s-sized trades; notice in Figure 1 that overall trade arrivals are not necessarily clustered but that occurrences of large trades **are** clustered. Each trade time is a customer arrival met by several quotes from market makers. Effectively, all market makers will submit quotes simultaneously, and the recorded trade price will be the winning quote, i.e., the markup that is best for the customer. At time t, a trade arrival has a chance $q_t \in (0, 1)$ of being a buy opportunity,

$$q_t = \mathbb{P}(\text{buy opportunity}|\text{time } t \text{ is an arrival}) \ .$$

The model for $q(y_t)$ is the logit function driven by an OU process,

$$q_t = q(y_t) = \frac{1}{1 + e^{-y_t}}$$

$$dy_t = -by_t dt + a dB_t \ ,$$

where $b > 0$ and B is a Brownian motion. The average $q(y_t)$ is 1/2.

There is a 'fair' market price that lies within the spread, and which is given by a diffusion,

$$dp_t = \sigma p_t dW_t \ ,$$

where W is another Brownian motion (possible correlated with B). In the event of a sell opportunity a market maker will submit a markup $m^{\uparrow} > 0$ and hence quote the price

$$p_t^{ask} = p_t + m^{\uparrow} \ ,$$

and similarly for a buy opportunity the will submit markup (really a markdown) $m^{\downarrow} > 0$ and hence quote the price

$$p_t^{bid} = p_t - m^{\downarrow} \ .$$

Hence, the challenge for the market maker is to optimally choose markets $m^{\uparrow}$ and $m^{\downarrow}$ at each trade opportunity. This challenge is formulated in the following optimal markups control problem,

$$V(x,y,\ell) = \operatorname*{ess\,inf}_{\mathcal{M}} \left(-\sum_{i=1}^{\infty} \rho^{i-1} \mathbb{E}\left[\text{time } \tau_i \text{ profits} \middle| x_0 = x, y_0 = y, \lambda_0 = \ell \right] \right) , \tag{3}$$

where $\mathcal{M}$ denotes the set of admissible markups, 'ess inf' refers to *essential infimum*, and $\rho \in (0,1)$ is discount parameter.

2.1 Bellman Equation & Exponential Markups

We can formulate a Bellman equation for the optimal markup problem in equation (3). Let $P^{\uparrow}(m,y,\ell)$ and $P^{\downarrow}(m,y,\ell)$ denote the probabilities that a markup for wins for a sell and buy opportunity, respectively. Then for these markups the dynamic programming principle applies and equation (3) is the solution of, where for $\tau_{i-1} \leq t < \tau_i$,

$$V(x,y,\ell) = \mathbb{E}\Big[(\tau_i - t) c(x,y,\ell) \Big| \lambda_t = \ell \Big] + \rho \inf_{\mathcal{M}} \mathbb{E}\left[V(x_{\tau_i}, y_{\tau_i}, \lambda_{\tau_i}) \middle| x_t = x, y_t = y, \lambda_t = \ell \right] .$$

For $t = \tau_i^-$,

$$\begin{aligned} V(x,y,\ell) = \rho \inf_{m^{\uparrow} \geq 0, m^{\downarrow} \geq 0} \Big(& V(x,y,\ell) \Big(1 - q(y_t) P^{\downarrow}(m^{\downarrow},y,\ell) - (1-q(y_t)) P^{\uparrow}(m^{\uparrow},y,\ell) \Big) \\ & + q(y_t) \Big(V(x+s,y,\ell) - (m^{\downarrow} - \epsilon) s \Big) P^{\downarrow}(m^{\downarrow},y,\ell) \\ & + (1-q(y_t)) \Big(V(x-s,y,\ell) - (m^{\uparrow} - \epsilon) s \Big) P^{\uparrow}(m^{\uparrow},y,\ell) \Big) , \end{aligned}$$

where $\epsilon > 0$ is a parameter that accounts for adverse selection. This is solved by first-order conditions,

$$m^{bid}(x,y,\ell) = \max\left(\epsilon + \frac{V(x+s,y,\ell) - V(x,y,\ell)}{s} - \frac{P^{\downarrow}(m^{bid}(x,y,\ell),y,\ell)}{\frac{d}{dm} P^{\downarrow}(m^{bid}(x,y,\ell),y,\ell)}, 0 \right) \tag{4}$$

$$m^{ask}(x,y,\ell) = \max\left(\epsilon + \frac{V(x-s,y,\ell) - V(x,y,\ell)}{s} - \frac{P^{\uparrow}(m^{ask}(x,y,\ell),y,\ell)}{\frac{d}{dm} P^{\uparrow}(m^{ask}(x,y,\ell),y,\ell)}, 0 \right) . \tag{5}$$

The optimal markups given in equations (4) and (5).

2.2 Indifference Liquidation Markup

Suppose now that there is a liquidation markup that the market maker can accept and any time to unload her inventory. Call it $\nu(x,y) > 0$, and suppose that the market maker is indifferent to holding her current inventory, i.e.,

$$V(x,y,\ell) = V(0,y,\ell) + \nu(x,y) \ . \tag{6}$$

The function $\nu(x,y)$ can be thought of as the dealer-to-dealer 'haircut' to which the market maker is indifferent to liquidation or keeping of her inventory. Hence, using the relation of equation (6), equations (4) and (5) simplify,

$$m^{bid}(x,y,\ell) = \max\left(\epsilon + \frac{\nu(x+s,y)-\nu(x,y)}{s} - \frac{P^{\downarrow}(m^{bid}(x,y,\ell),y,\ell)}{\frac{d}{dm}P^{\downarrow}(m^{bid}(x,y,\ell),y,\ell)}, 0\right) \tag{7}$$

$$m^{ask}(x,y,\ell) = \max\left(\epsilon + \frac{\nu(x-s,y)-\nu(x,y)}{s} - \frac{P^{\uparrow}(m^{ask}(x,y,\ell),y,\ell)}{\frac{d}{dm}P^{\uparrow}(m^{ask}(x,y,\ell),y,\ell)}, 0\right) \ . \tag{8}$$

2.3 A Hazard Rate Formulation for the Inside Market

Equations (4) and (5) include the probability of winning the trade $P^{\uparrow}(m^{\uparrow})$. Let $F(z)$ be the distribution of the most competitive markup from another dealer, so that

$$P^{\uparrow}(m^{\uparrow}) = 1 - F(m^{\uparrow}) \ ,$$

and let $f(z)$ denote the corresponding density function. We let

$$h(z) = \frac{f(z)}{1-F(z)} \ ,$$

denote the corresponding hazard rate. Evaluated at $m^{\uparrow}$, the function h is the sum of the hazards of the dealers' markup distributions. As a quick reminder of why let $F_i(z)$ and $f_i(z)$ denote the respective distributions and densities for the markups of the other dealers' markups. Since the probability of winning the trade with a markup of z is $1 - F(z) = \Pi_i(1 - F_i(z))$ we have, differentiating,

$$\begin{aligned} -f(z) &= \sum_i -f_i(z)\Pi_{j\neq i}(1-F_j(z)) \\ &= -\sum_i \frac{f_i(z)}{1-F_i(z)}\Pi_j(1-F_j(z)) \\ &= -(1-F(z))\sum_i \frac{f_i(z)}{1-F_i(z)} \end{aligned}$$

so

$$\frac{f(z)}{1-F(z)} = \overbrace{\sum_i \frac{f_i(z)}{1-F_i(z)}}^{sum\ of\ hazards}$$

The real reason to use hazards is the clean characterization of the *suggested* optimal markups. From the first-order condition in (**??**), it is seen that the best choice of markup is where the increase in profit exactly offsets the possibility of losing the trade. Intuitively, the best choice of markup is where the increase in profit exactly offsets the possibility of losing the trade. For instance, we increased the markup by Δm then:

$$\overbrace{\left(m^{ask} - K^{\downarrow}(x;s)\right)}^{existing\ benefit} \quad \overbrace{h(m^{ask})\Delta m}^{chance\ of\ losing\ it} \quad = \quad \overbrace{\Delta m}^{increase\ in\ profit} \quad .$$

This is consistent with the first-order condition from (**??**), which we can rewrite the optimal markup using the hazard function,

$$m^{ask}(x;s) = \max\left(\frac{1}{h(m^{ask}(s))} + K^{\downarrow}(x;s), 0\right) \ ,$$

which is independent of the distributional assumption for the inside market. Therefore, recalling the definition of $K^{\downarrow}(x;s)$ (given by Definition **??**), the optimal markup when offering to sell breaks down as

$$m^{ask}(x;s) = \max\left(\frac{1}{h(m^{ask}(x;s))} + \epsilon + \frac{\nu(x-s) - \nu(x)}{s}, 0\right) \ .$$

This reads straightforwardly:

"markup = markup width + adverse selection + marginal inventory cost"

though the coupled quantities $\nu(x)$ and $m^{ask}(x;s)$ remain to be found.

2.4 Exponential Markups

Let's now assume the following:

Assumption: The inside hazard rate is constant. In terms of the distribution F, constant hazard rate means

$$\frac{f(z)}{1 - F(z)} = h \ ,$$

where $h > 0$ is the constant hazard rate. This leads to an exponential distribution for best competitor's markup, $F(z) = 1 - e^{-hz}$ for $z > 0$, and the probability of winning the trade is

$$P^{\downarrow}(m^{ask}) = 1 - F(m^{ask}) = e^{-hm^{ask}} \ .$$

By assuming the hazard rate is constant equation (**??**) simplifies to,

$$m^{ask}(x;s) = \frac{1}{h} + \epsilon + \frac{\nu(x-s) - \nu(x)}{s} \ ,$$

and will be globally optimal provided $m^{ask}(x;s) > 0$. To be precise:

$$m^{ask}(x;s) = \max\left(\frac{1}{h} + \epsilon + \frac{\nu(x-s) - \nu(x)}{s}, 0\right)$$

since we can guarantee to trade with zero markup. Similarly our market maker offers a markup of

$$m^{bid}(x;s) = \max\left(\frac{1}{h} + \epsilon + \frac{\nu(x+s) - \nu(x)}{s}, 0\right)$$

when offering to buy. Since we know both the bid and offer, we can now re-express these as skews and widths - the manner in which traders are used to seeing them. To draw an interesting connection we will assume, for the moment, that we are in a small inventory region where neither the optimal bid nor offer require max() in their formulas.

2.5 Reinforcement Learning

Reinforcement learning in the form of Q-learning is used estimate the function $V(x, y, \ell)$. Let Δt be a time step and let,

$$t_n = n\Delta t \ .$$

Let $x_n = x_{t_n}$ be the inventory, $y_n = y_{t_n}$ the OU process for the buy/sell probability, and $\lambda_n = \lambda_{t_n}$ the arrival intensity. After many iterations the Q learning will converge toward the optimal value function, where each iteration is,

$$V(x_n, y_n, \lambda_n) \leftarrow (1-\alpha)V(x_n, y_n, \lambda_n) + \alpha\Bigg(\Delta t c(x_n, y_n, \lambda_n) + \delta(V(x_{n+1}, y_{n+1}, \lambda_{n+1}) - profit_{n+1})\Bigg) \ , \tag{9}$$

where $\alpha \in (0, 1)$ is the learning rate. The Q learning shown in equation (9) is written without minimization because the optimal markups are given by equation (7) and (8).

2.5.1 1-D Example: Constant λ_n and q_n

Take $\Delta t = .1$, $\lambda_n \equiv 5$ and $q_n \equiv 1/2$ for all n, the trade size $s = 1$, the hazard rate $h = .1$, the adverse selection parameter to be $\epsilon = .1$, the discount rate to be $\delta = .9$, the cost to be $c(x) = |x/10|+(x/20)^2$. Let the reinforcement learning run for $n = 1, 2, 3, \dots, 10^7$,with the learning rate starting at $\alpha_0 = .6$ and tuning down to zero as the iterations progress. Figure 3 shows the log of the value function obtained from Q learning. Figure 4 shows the optimal markups along with CWLS, where it is see that having the term $|x/10|$ in the cost function creates a step in the skew.

3 Solving the Exponential Markup Model

Let's consider the limiting problem as $\rho \to 1$. In this limit the control problem in (3) becomes an ergodic control problem,

$$V(x, y, \ell) = \operatorname*{ess\,inf}_{\mathcal{M}} \ \liminf_{N\to\infty} \left(-\frac{1}{N}\sum_{i=1}^{N} \mathbb{E}\left[\text{time } \tau_i \text{ profits}\middle| x_0 = x, y_0 = y, \lambda_0 = \ell\right]\right) \ . \tag{10}$$

The dynamic programming principle, for $\tau_{i-1} \le t < \tau_i$, is

$$\mu + V(x, y, \ell) = \mathbb{E}\Big[(\tau_i - t)c(x, y, \ell)\Big|\lambda_t = \ell\Big] + \inf_{\mathcal{M}} \ \mathbb{E}\left[V(x_{\tau_i}, y_{\tau_i}, \lambda_{\tau_i})\middle| x_t = x, y_t = y, \lambda_t = \ell\right] \ .$$

where μ is a free parameter to make the two sides equal. The solution is found from first-order conditions, which yield the same optimal markups as (4) and (5). If y_t and λ_t are constant, then the dynamic programming simplifies,

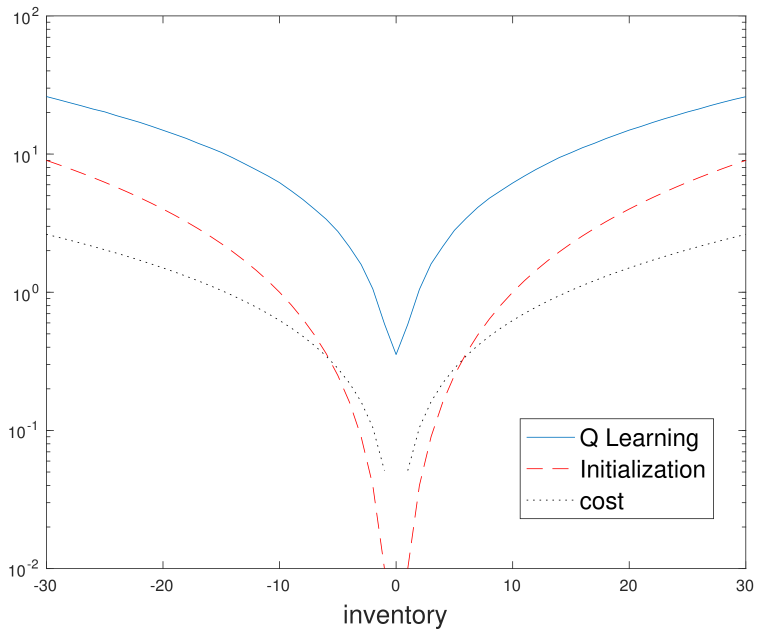

Figure 3: The optimal 1-D value function obtained through Q learning (shown in log scale).

$$\mu + V(x) = \frac{c(x)}{\lambda} + \inf_{m^{\uparrow}\geq 0, m^{\downarrow}\geq 0}\Bigg(V(x)\Big(1 - qP^{\downarrow}(m^{\downarrow}) - (1-q)P^{\uparrow}(m^{\uparrow})\Big) + q\Big(V(x+s) - (m^{\downarrow} - \epsilon)s\Big)P^{\downarrow}(m^{\downarrow}) + (1-q)\Big(V(x-s) - (m^{\uparrow} - \epsilon)s\Big)P^{\uparrow}(m^{\uparrow})\Bigg) . \tag{11}$$

3.1 Implications of a Constant-Width Solution

As a special case we can assume convexity $C(x)$ equals the constant C_0, that is

$$C(x) \equiv C_0 \qquad \forall x \; ;$$

i.e. constant width. This will lead to a linear skew provided that the liquidation and cost functions are consistent with the form implied by the constant-width assumption. In particular, the cost function will need to a hyperbolic cosine function, as seen be inserting a linear $S_\delta(x)$ into equation (**??**).

Since

$$\frac{\nu(x+s) - 2\nu(x) + \nu(x-s)}{2s} = C_0$$

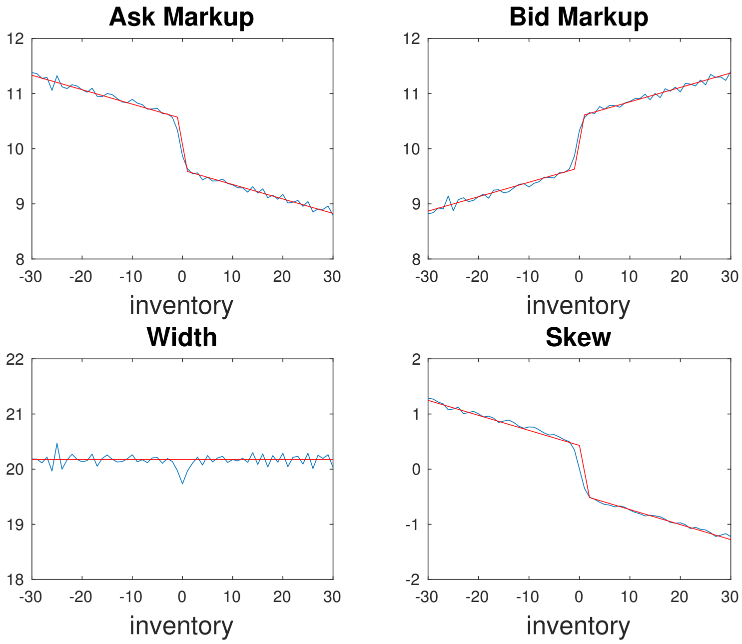

Figure 4: The optimal markups (top plots) and the CWLS heuristic (bottom plots). The Q learning leaves behind some residual noise, but interpolation lets us see the linear behavior. The width is interpolated and seen to be constant, and having the term $|x/10|$ in the cost function creates a step in the skew. If there were no absolute value in the cost (and only a squared penalty), then the skew would be purely linearly.

has solution

$$\nu(x) = \frac{C_0}{s}x^2 + C_1 x + C_2 \ ,$$

$C_2 \geq \frac{sC_1^2}{4C_0}$ so that $\nu(x) \geq 0$. This corresponds to the traditional quadratic choice of inventory cost function. One might think this also corresponds to linear skew, as with the quadratic oscillator and also approximations reached in the literature (see [1]), but such a relationship will depend on the function $c(x)$. Apart from $\nu(x)$, we can solve for skew directly,

$$S_\delta(x) \quad = \quad \frac{1}{h}\cosh^{-1}\left(e^{hC_0}\Omega c(x) + \cosh(hS_\delta(0))\right) \ , \tag{12}$$

and this yields an easily implementable constant width model. However, only in the special case of small x and quadratic cost do we recover linear skew - since the right hand side is approximately proportional to $\sqrt{c(x)}$. This ceases to be the case, however, if funding costs or equity hurdles are material.

If we continue with our assumption $C(x) \equiv C_0$ and using $\nu(x) = \frac{C_0}{s}x^2 + C_1 x + C_2$ as obtained from it, by direct calculation we find,

$$S_\delta(x) = \frac{\nu(x+s) - \nu(x-s)}{2s} - \delta = 2\frac{C_0}{s}x + C_1 - \delta = 2\frac{C_0}{s}x + S_\delta(0) \ , \tag{13}$$

where we've deduced that $C_1 = S_\delta(0) + \delta$. **Equation** (13) **gives the linear skew that we were seeking.** Next, using $C(x) \equiv C_0$ and $S_\delta(x)$ from (13), the cost function can be computed from (**??**):

$$c(x) = \frac{\cosh\left(2\frac{C_0}{s}hx + hS_\delta(0)\right) - \cosh(hS_\delta(0))}{\Omega e^{hC_0}} \ .$$

We assume $c(x) \geq 0$ for all x (positive or negative), and so it must be that

$$S_\delta(0) = 0 \ ,$$

and the expression for cost simplifies to

$$c(x) = \frac{\cosh\left(2\frac{C_0}{s}hx\right) - 1}{\Omega e^{hC_0}} \ . \tag{14}$$

Equation (14) provides an expression for $c(x)$ that indicates the possibility of interval consistency in the CWLS approach; equation (14) will be internally inconsistent with the cost of carrying inventory unless the inventory $c(x)$ was originally calculated to be a hyperbolic cosine function. Even still, (14) may imply parameters that are different from the actual costs faced by a dealer with $\cosh(\cdot)$ costs.

If the cost function of (14) is correct, then for small $2\frac{C_0}{s}hx$ we see that $c(x)$ is approximately the quadratic,

$$c(x) = \frac{2}{\Omega e^{hC_0}}\left(\frac{C_0}{s}hx\right)^2 + \mathcal{O}\left(\frac{C_0}{2s}hx\right)^4 \ .$$

Assuming that $\frac{C_0}{s}x = \mathcal{O}(1)$, a quadratic cost combined with CWLS together imply that h is small.

Finally, assuming liquidation costs at least zero for all x positive or negative, the conclusion that $S_\delta(0) = 0$ and $C_1 = S_\delta(0) + \delta$ (as deduced in (13)) yield the indifference markup function:

$$\nu(x) = \frac{C_0}{s}x^2 + \delta x + C_2 \ ,$$

where $C_2 \geq \frac{s\delta^2}{4C_0}$ so that $\nu(x) \geq 0$.

3.2 Sensitivity of the Solution

Though exponential markups are a stylized assumption, they convey an important intuition about the importance of knowing the other guys' markups versus the importance of knowing where the mid is.

Let π denote the ground truth mid price, h the ground truth hazard rate for the inside market and $w = 1/h$ the corresponding market width. Assume the market maker's estimates for both mid price and market width are normally distributed:

$$\begin{aligned} \hat{\pi} &= \pi + \epsilon_\pi \\ \hat{w} &= w + \epsilon_w \end{aligned}$$

where $\epsilon_h \sim N(0, \sigma_h^2)$ and $\epsilon_w \sim N(0, \sigma_w^2)$ are i.i.d. - an assumption we will return to. For concreteness consider the case $\eta =\downarrow$ when the market maker as an opportunity to sell.

Claim: The location of the best response by competitors, equal to $\hat{\pi} + \hat{w}$, is a sufficient statistic summarizing π and w.
Proof: We have already seen that if the market maker knows π and h exactly then her choice for marked up price $\pi^{\downarrow} = \pi + m^{\downarrow}$ will maximize the gain

$$G(\pi^{\downarrow}) = F(\pi^{\downarrow})\left(m^{\downarrow} - K^{\downarrow}(x;s)\right)$$

where $F(\pi^{\downarrow}) = e^{-h(\pi^{\downarrow}-\pi)}$. Assuming we are in the region where the first order condition is relevant, the best choice $\pi^{ask} = \pi + m^{ask}$ will satisfy

$$\overbrace{\pi^{ask} - \pi}^{m^{\downarrow}} - K^{\downarrow}(x;s) = w$$

with gain

$$\begin{aligned} G(\pi^{ask}) &= e^{-h(\pi^{ask}-\pi)}\left(\pi^{ask} - \pi - K^{\downarrow}(x;s)\right) \\ &= we^{-h(\pi^{ask}-\pi)} \ . \end{aligned}$$

On the other hand if the market maker knows h and π only approximately, with estimates $\hat{h}$ and $\hat{\pi}$ respectively, she will make a different choice π^{ask}_{ϵ} solving the wrong problem. Her choice instead will satisfy

$$\pi^{ask}_{\epsilon} - \hat{\pi} - K^{\downarrow}(x;s) = \hat{w} \ ,$$

and we observe, by subtraction, that this diverges from the optimal choice:

$$\begin{aligned} \pi^{ask}_{\epsilon} - \pi^{ask} &= \hat{w} + \hat{\pi} + K^{\downarrow}(x;s) - \pi - w - K^{\downarrow}(x;s) \\ &= \overbrace{\hat{w} - w}^{\epsilon_w} + \overbrace{\pi - \hat{\pi}}^{\epsilon_\pi} \\ &:= \epsilon_{w+\pi} \\ &\sim N(0,\sigma^2) \ , \end{aligned}$$

showing, incidentally, that the errors in both price and width translates directly into error in the optimal choice of response to the RFQ. The sum of errors is also the error in the market maker's location estimate for the mean of the best price response by his competitors. Thus $\hat{\pi} + \hat{w}$ is a sufficient statistic. ■

From the proof of Claim 3.2 we let $G(\pi)$ denote the expected net benefit to the market maker of the trading opportunity; now we let $E^{\sigma^2}[G(\pi_\epsilon)]$ the same under uncertain knowledge of w and h and define **the relative efficiency:**

$$R := \frac{E^{\epsilon}[G(\pi^{ask}_{\epsilon})]}{G(\pi^{ask})} \ .$$

Claim: Let $\sigma^2 := \sigma^2_\pi + \sigma^2_w$, i.e. the combined variance of $\hat{\pi} + \hat{w}$. If $h = 1/w$ is small (relative to σ^2), then the relative efficiency is approximated by

$$R \approx \bar{R}_0 = e^{\frac{1}{2}\frac{\sigma^2}{w^2}}\left(1 - \frac{\sigma^2}{w^2}\right) \ . \tag{15}$$

Proof: If $\pi_\epsilon > \pi$, then the gain uncertainty markup is

$$\begin{aligned}
G(\pi^{ask}_\epsilon) &= e^{-h(\pi^{ask}_\epsilon-\pi)}\left(\pi^{ask}_\epsilon - \pi - K^{\downarrow}(x;s)\right) \\
&= e^{-h(\pi^{ask}-\pi+\epsilon_{w+\pi})}\left(\overbrace{\pi^{ask}-\pi-K^{\downarrow}(x;s)}^{=w}+\epsilon_{w+\pi}\right) \\
&= we^{-h(\pi^{ask}-\pi)}e^{-h\epsilon_{w+\pi}}\left(1+\frac{\epsilon_{w+\pi}}{w}\right) \\
&= G(\pi^{ask})e^{-h\epsilon_{w+\pi}}\left(1+\frac{\epsilon_{w+\pi}}{w}\right) \ . \qquad (16)
\end{aligned}$$

A somewhat more careful calculation takes into account the possibility that due to errors in estimation the market maker might offer through the mid π and, in that event, always do the trade. In otherwords, if $\pi_\epsilon \leq \pi$ then the dealer automatically makes the trade, and the gain is

$$G(\pi^{ask}_\epsilon) = \pi^{ask}_\epsilon - \pi - K^{\downarrow}(x;s) = w + \epsilon_{w+\pi} \ . \qquad (17)$$

Combining equations (16) and (17), we have the efficiency ratio:

$$\begin{aligned}
R &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-(\pi^{ask}-\pi)}^{\infty} e^{-\frac{1}{2\sigma^2}\epsilon^2}e^{-h\epsilon}(1+h\epsilon)d\epsilon \\
&\quad+\frac{1}{G(\pi^{ask})\sqrt{2\pi\sigma^2}}\int_{-\infty}^{-(\pi^{ask}-\pi)} e^{-\frac{1}{2\sigma^2}\epsilon^2}\left(w+\epsilon\right)d\epsilon \ .
\end{aligned}$$

If h is small then w is large, and assuming σ^2 is not too large we have

$$\pi^{ask} - \pi \gg 1 \ ,$$

and then the chance of offering below the mid is small, which prompts us to define an approximate efficiency ratio

$$\bar{R}(\epsilon_\pi, \epsilon_w) := e^{-\frac{\epsilon_{w+\pi}}{w}}\left(1+\frac{\epsilon_{w+\pi}}{w}\right) \ , \qquad (18)$$

where again, $\epsilon_{w+\pi}$ is the difference between the estimated location of the mean of the competitors' best ask and the ground truth. Integrating directly we have, approximately:

$$\begin{aligned}
\bar{R}_0 &= E^\epsilon R(\epsilon_\pi, \epsilon_w) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(\epsilon)^2}e^{-h\epsilon}(1+h\epsilon)d\epsilon \\
&= e^{\frac{1}{2}\frac{\sigma^2}{w^2}}\left(1-\frac{\sigma^2}{w^2}\right) \ .
\end{aligned}$$

as claimed. ■

3.3 A Humble Market Maker

As (18) warns, assuming the correctness of point estimates might not be the best way for the market maker to choose m^{ask}. Instead, she should acknowledges her errors in $\hat{\pi}$ and $\hat{w}$ and act defensively.

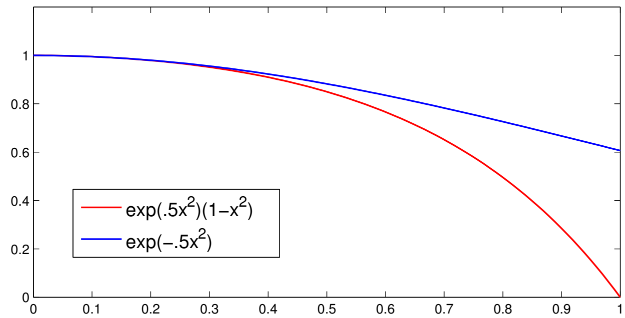

Figure 5: The relative efficiency approximation $\bar{R}_0$ given by equation (15) and the maximized efficiency $\bar{R}_\varphi$ given by (19), both expressed as functions of $x = \sigma/w$ where σ is the standard error in the market maker's estimate of $\pi + w$.

Claim: Expected efficiency is maximized by replacing the point estimate $\hat{\pi} + \hat{w}$ with a cautious estimate $\hat{\pi} + \hat{w} + h\sigma^2$ whereupon the mean efficiency is approximated by

$$\bar{R}_{h\sigma^2} = e^{-\frac{1}{2}\frac{\sigma^2}{w^2}} \ . \tag{19}$$

Proof: If the market maker biases her estimate of $\hat{\pi} + \hat{w}$ by an amount φ then $\epsilon_{\pi+w}$ will be normally distributed around φ with variance equal to her error σ^2. As with (16) her mean efficiency must be

$$\begin{aligned} \bar{R}_\varphi &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(\epsilon-\varphi)^2} e^{-h\epsilon}(1 + h\epsilon) d\epsilon \\ &= e^{\frac{h^2\sigma^2}{2}} e^{-h\varphi}(-h^2\sigma^2 + h\varphi + 1) \end{aligned} \tag{20}$$

Let $\upsilon(\varphi) := e^{-h\varphi}(-h^2\sigma^2 + h\varphi + 1)$. Since $\upsilon'(\varphi) = \left(-h + \frac{h}{-h^2\sigma^2 + h\varphi + 1}\right)\upsilon(\varphi)$ we set the term in the denominator equal to 1 to maximize efficiency whereupon $\varphi = h\sigma^2$ and

$$\bar{R}_\varphi = e^{-\frac{h^2\sigma^2}{2}} \ ,$$

as claimed. ■

There are two flaws with this approximation. First, we can only apply an offset of $\hat{h}\sigma^2$ not $h\sigma^2$ since the latter is unknown. Second, our approximation assumes no chance of offering through the true mid. However if $\pi_\epsilon < \pi$ the gain (in fact loss) is actually

$$G(\pi_\epsilon^{ask}) \quad = \quad \pi_\epsilon^{ask} - \pi - K^{\downarrow}(x; s)$$

because we always trade so (16) is not valid for all π_ϵ^{ask} and either is (18) for all $\epsilon_{\pi+w}$. In practice then, a market maker might want to deviate from (20) especially when given an opportunity to get out of a large inventory. The optimal efficiency would maximize the exact formula,

$$R_\varphi \quad = \quad \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(\epsilon-\varphi)^2} \min\left(e^{-h\epsilon}, 1\right)(1 + h\epsilon) d\epsilon \ ,$$

but Claim 3.3 is merely a heuristic and does not an exact solution.

4 Conclusions

We have presented properties of the optimal choice of bid and offers in a repeated sequence of sealed-bid auctions, and demonstrated that under the assumption of exponentially distributed adversary markups these take on a simple functional form. Within a region of validity the amount by which a market maker marks up the price is given by

$$markup = \text{const} + \frac{\nu(x-s)+\nu(x)}{s}$$

and thus is essentially synonymous with marginal inventory cost $\nu(x)$. We have quantified the extent to which this marginal cost differs from a linear function of inventory x, thus providing a simple improvement to a ubiquitous "constant width linear skew" trading heuristic.

References

[1] M. Avellaneda and S. Stoikov. High-frequency trading in a limit order book. *Quantitative Finance*, 8(3):217–224, 2008.

[2] T. Ho and R. Macris. Dealer bid-ask quotes and transaction prices. *The Journal of Finance*, 39(1):23–45, 1984.

[3] T. Ho and H. R. Stoll. On dealer markets under competition. *The Journal of Finance*, 35(2):259–267, May 1980.

[4] T. Hubbard, H. Paarsch, and W. Wright. Hui: A case study of a sequential double auction of capital. 2014. U. of Chicago Working Paper.

[5] H. Luckock. A steady-state model of a continuous double auction. *Quantitative Finance*, 3(1):385–404, 2003.

[6] P. Milgrom and R. Weber. A theory of auctions and competitive bidding. *Econometrica*, 50(5):1089–1122, 1982.

[7] R. Myerson. Optimal auction design. *Math. Oper. Res.*, 6(1):58–73, February 1981.

[8] T. Roughgarden, V. Syrgkanis, and E. Tardos. The price of anarchy in auctions, 2016. available on arXiv, arXiv:1607.07684.

[9] Seymour Smidt. Continuous versus intermittent trading on auction markets. *The Journal of Financial and Quantitative Analysis*, 14(4):837–866, 1979.

[10] W. Vickrey. Counterspeculation auctions, and competitive sealed tenders. *Journal of Finance*, 16(1):8–37, 1961.