How to respond to an RFQ

Peter Cotton

September 2, 2021

Overview

Introduction

- Request for quote (RFQ) dominated markets
- Trading heuristics
- An optimal control approach
- Previous academic work
- New results for RFQ markets
- Notation for trading opportunities
- Notation for price markups and markdowns
- Notation for adverse selection
- Notation for carrying cost of inventory
- Reasonable linear cost parameters for flow trading
- Reasonable quadratic cost parameters for flow trading

Non-myopic inventory cost

Properties of the optimal solution

The intuition behind optimal flow trading

A financial argument

Break-even markups (“strikes”)

A hazard rate formulation for the inside market

Optimal markups as a function of zero profit markups

Exponential markups

Skew and Width

Consistent skew and width

Order imbalance

Relating width and skew

Solving the exponential markup model

A better constant width model

Implied cost fitting

Sensitivity of the solution

A humble market maker

Introduction

Request for quote (RFQ) dominated markets

We consider a steady state solution for an optimal market maker who competes with other dealers.

1. Customer inquiry is sent to multiple dealers
2. Dealers compete on price
3. Trade is filled

Of relevance to corporate bonds, but also many other areas.

Trading heuristics

Trading strategy development often takes a less formal path than the econometric forecasting work on which it rests. Examples:

1. Keep width constant, but use skew that is linear in inventory.
2. Trading rules optimized in back-tests

But note that the first method solves the quadratic oscillator problem - and this fact itself helps us understand whether or not linear skew is optimal for our purposes.

(That's a test question for the reader)

An optimal control approach

1. Design suitable penalty for holding inventory.
2. Model the markups of our competitors
3. Solve for the trading strategy that maximizes return
4. Gain intuition
5. Determine which parameters are sufficient statistics
6. Determine sensitivity of profitability to parameter estimates
7. Figure out where to focus efforts

... and this does not preclude lots of ad-hoc back-tests along the way.

Previous academic work

(More research required here)

- Avellaneda and Stoikov. *High-frequency trading in a limit order book*. Authors show that, to an approximation, one should use constant width and linear skew .

Sound familiar? Why can't this be right for corporate bonds?

New results for RFQ markets

1. A financial argument for constraints on inventory cost $\nu(x)$.
2. Clarity around how width and skew relate to inventory cost.

In the special case of exponentially distributed competitor markups:

3. Analytic expression for order imbalance impacting skew and width
4. Analytic expression for width as a function of inventory assuming constant width (and vice versa)
5. A numerical approach improving on both 3 and 4 above.
6. Analytic expression for market making efficiency based on variance in parameter estimates

Notation for trading opportunities

For simplicity all incoming inquiry is for a hypothetical trade of size s (whilst this would, to be pedantic, restrict the possible inventory levels to a lattice, we would want to solve for continuous x as the assumption will be violated in practice).

The probability of the market maker having an opportunity to buy is p, versus $1 - p$ for the probability that the next inquiry is a customer wanting to buy.

The time between opportunities is δt, assumed Poisson.

Notation for price markups and markdowns

Let x denote the current inventory of a market maker. The dollar value of holdings will be this quantity multiplied by a fair price, though the price itself will not feature directly in what follows.

It is sufficient to let m^η and q^η variables denote the *markups* relevant to the next trading opportunity, where η is either $\uparrow$ for a buying opportunity or $\downarrow$ for a selling opportunity

The market maker bids $m^\uparrow$. The competition bids $q^\uparrow$. The market maker wins the trade if $m^\uparrow < q^\uparrow$.

Notation for adverse selection

We don't really assume our market maker knows the fair market price.

Rather, we assume that customers know more: that the true fair value is separated by ϵ from where the market makers believe the mid to be, and always in an adverse direction from her perspective.

Put simply, our market maker's profitability for every trade is ϵ less than the markup would naively suggest.

Notation for carrying cost of inventory

We assume a carrying cost $c(x)$ per unit of time. For concreteness take

$$c(x) = c_+ x^+ + c_- |x^-| + c_2 x^2$$

where x^+ and x^- are the positive and negative parts of inventory x respectively.

1. Piecewise-linear costs include funding and return on equity hurdles.

2. Quadratic costs capture the risk for a mean-variance investor.

Costs here are treated as a meta-parameter set by management: informed by econometric estimates such as volatility, and risk limits.

Reasonable linear cost parameters for flow trading

Recall the direct cost function:

$$c(x) = c_+ x^+ + c_- |x^-| + c_2 x^2$$

Without going into details, a typical funding cost is on the order of 60-80 bps for a long position whereas the linear cost for short positions will depend on repo rates (maybe 50 bps less).

However, the direct funding cost is *not the relevant penalty* to use in determining a flow trading strategy. Why?

Flow businesses expect high return on equity - say $>$30% for concreteness assuming leverage of, say, 8. To encourage turnover we might set:

$$c_{+} = \frac{return\ on\ equity}{leverage} + funding = \frac{0.30}{8} + 0.07 \sim 4.5\%$$

Similarly $c_{-} \sim 4\%$ say for short positions.

Reasonable quadratic cost parameters for flow trading

Quadratic costs are a proxy for all manner of actual constraints on trading intended to capture default and gap risk (in addition to whatever diffusive risk is assumed by a market-risk sensitive variance minimizing investor, which is what motivates the quadratic term).

A practical way to set quadratic risk terms is via one of the following questions:

1. At what inventory level should the risk penalty first surpass the linear costs?
2. At a fixed inventory corresponding (say $\$10$m long), what fraction of linear costs for a long position should the risk penalty comprise?
3. At what inventory level should the risk cost be prohibitive?

In practice I have preferred the second option, setting

$$c_2 x^2 = 0.1 c_+ x$$

and thus

$$c_2 = \frac{0.1 c_+}{x = 10,000,000}$$

for example.

To summarize. In the direct cost function

$$c(x) = c_+ x^+ + c_- |x^-| + c_2 x^2$$

some reasonable parameters are

$$\begin{aligned} c_+ &= 0.045 \\ c_- &= 0.04 \\ c_2 &= \frac{0.1 \times 0.045}{10,000,000} = 4.51e^{-10} \end{aligned}$$

Modify as you will.

Non-myopic inventory cost

We let $\nu(x)$ denote the biggest markdown a market maker will accept if offered the chance to get out of her inventory x immediately.

More precisely, suppose $x > 0$ and the current price of the asset is p. Suppose a third party bids with price $p - \nu(x)$ and size x equal to the market maker's inventory. We assume there is a unique $\nu(x)$ for which the market maker is indifferent as to whether to trade or not.

Properties of the optimal solution

The intuition behind optimal flow trading

In what follows:

- Every incoming inquiry has a potential benefit to the market maker. She has the option not to bid at all, so this is strictly positive.
- In deciding on the true cost of moving from inventory x to $x+s$, she will weigh not only the differential in direct costs proportional to $c(x+s)-c(x)$ but also the differential value of trading opportunities at each inventory level. The larger the inventory $|x|$, generally, the greater the value to her of each trading opportunity.

Naive approaches ignore the second term, overstate $\nu(x)$, and are too defensive.

A financial argument

In order to impose conditions on $\nu(x)$ we consider two courses of action.

1. The market maker liquidates her holding and pays $\nu(x)$ to do so. She then waits until the next trade opportunity, makes a trading decision, and then again liquidates her position if necessary.

2. She defers liquidation until after the trading opportunity.

Both courses of action recombine into the same final inventory $x = 0$. So by definition of $\nu(x)$ as the indifference markup, both paths must incur the same costs.

Optimal RFQ responses

Denote the costs for the first path as Δ_1. Recall that the market maker must decide on a markup ($m^{\uparrow}$ for buying, $m^{\downarrow}$ for selling) and $\iota(q;m) = q > m$ is the indicator function for winning the trade. If a buying opportunity arrives first with probability p then the cost for the best policy is the original liquidation cost plus the net final liquidation cost (after deducting the immediate gain from that trade). Thus

$$
\begin{aligned}
\Delta_1 \quad = \quad & \overbrace{\nu(x)}^{first\ liquidation} \qquad + p \inf_{m^{\uparrow}} \left\{ \overbrace{\left(\nu(s) - (m^{\uparrow} - \epsilon)s\right)}^{net\ cost} \overbrace{E^q\left[\iota(q^{\uparrow}, m^{bid})\right]}^{=P^{\uparrow}(m^{bid})} \right\} \\
& + (1-p) \inf_{m^{\downarrow}} \left\{ \left(\nu(x-s) - (m^{\downarrow} - \epsilon)s\right) P^{\downarrow} \right\}
\end{aligned}
$$

On the other hand if the market maker defers liquidation she also pays a direct cost for carrying the inventory to the next trading opportunity, and must liquidate even if there is no trade at that time. Hence:

$$
\begin{aligned}
\Delta_2 \quad = \quad & \overbrace{E[\delta t] c(x)}^{carry} + \\
& + p \inf_{m^{\uparrow}} \left\{ \left(\nu(x+s) - (m^{\uparrow} - \epsilon) s \right) P^{\uparrow}(m^{\uparrow}) \nu(x) \left(1 - P^{\uparrow}(m^{\uparrow}) \right) \right\} \\
& + (1-p) \inf_{m^{\downarrow}} \left\{ \left(\nu(x-s) - (m^{\downarrow} - \epsilon) s \right) P^{\downarrow}(m^{\downarrow}) + \nu(x) \left(1 - P^{\downarrow} \left(m^{\downarrow} \right) \right) \right\}
\end{aligned}
$$

where again $P^{\uparrow}(m^{\uparrow})$ is the probability of winning the trade when the market maker marks up by $m^{\uparrow}$.

Break-even markups ("strikes")

We equate Δ_1 and Δ_2 but first, define

$$K^{\uparrow}(x;s) = \epsilon + \frac{\nu(x+s)-\nu(x)}{s} \tag{1}$$

Similarly

$$K^{\downarrow}(x;s) = \epsilon + \frac{\nu(x-s)-\nu(x)}{s} \tag{2}$$

as the "strike" when selling, which may well be negative for positive inventory. We shall interpret these as the non-myopic break-even markups, as will be shortly apparent.

Optimal RFQ responses

Indifference to the choice of path means $\Delta_1 = \Delta_2$. This simplifies after we subtract $\nu(x)$ from both sides, write $\tau = E[\delta t]$ for the average waiting time, swap signs and replace $\inf$ with $\sup$:

$$
\begin{aligned}
\frac{\tau c(x)}{s} \quad = \quad & p \sup_{m^{\uparrow}} \left\{ \overbrace{\left(m^{\uparrow} - K^{\uparrow}(x;s)\right)}^{net\ gain} P^{\uparrow}(m^{\uparrow}) \right\} \\
& + (1-p) \sup_{m^{\downarrow}} \left\{ \left(m^{\downarrow} - K^{\downarrow}(x;s)\right) P^{\downarrow}(m^{\downarrow}) \right\} \\
& - p \sup_{m^{\uparrow}} \left\{ \left(m^{\uparrow} - K^{\uparrow}(0;s)\right) P^{\uparrow}(m^{\uparrow}) \right\} \\
& - (1-p) \sup_{m^{\downarrow}} \left\{ \left(m^{\downarrow} - K^{\downarrow}(0;s)\right) P^{\downarrow}(m^{\downarrow}) \right\}
\end{aligned}
$$

then we start to appreciate the intuition. Differential "option" value of the trading opportunities at x versus zero inventory equates to the direct cost of carry.

Optimal RFQ responses

Moreover there was nothing special about $x = 0$ in the argument so

$$
\begin{aligned}
\frac{\tau(c(x) - c(x'))}{s} \quad = \quad & p \sup_{m^\uparrow} \left\{ \left(m^\uparrow - K^\uparrow(x; s) \right) P^\uparrow(m^\uparrow) \right\} \\
& + (1-p) \sup_{m^\downarrow} \left\{ \left(m^\downarrow - K^\downarrow(x; s) \right) P^\downarrow(m^\downarrow) \right\} \\
& - p \sup_{m^\uparrow} \left\{ \left(m^\uparrow - K^\uparrow(x'; s) \right) P^\uparrow(m^\uparrow) \right\} \\
& - (1-p) \sup_{m^\downarrow} \left\{ \left(m^\downarrow - K^\downarrow(x'; s) \right) P^\downarrow(m^\downarrow) \right\} \qquad (3)
\end{aligned}
$$

holds for any two inventories x and x'. Note that the $m^\uparrow$ and $m^\downarrow$'s in respective $\sup$'s are distinct.

Can we get rid of the $\sup$'s?

A hazard rate formulation for the inside market

Inside the $\sup$'s we have the probability of winning the trade $P^{\uparrow}(m^{\uparrow})$.

If $F(q)$ is the distribution of the most competitive markup from another dealer, and $f(q)$ the corresponding density function, we let $h(q) = \frac{f(q)}{1-F(q)}$ denote the corresponding hazard rate.

This 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

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

Optimal markups as a function of zero profit markups

The real reason to use hazards is the clean characterization of the *suggested* optimal markups. We interpret $K^{\downarrow}(x;s)$ as the zero profit markup for a market maker when offering to sell a quantity s to move inventory from x to $x-s$.

Evidently the best choice of markup is where the increase in profit exactly offsets the possibility of losing the trade. For say we increased the markup by Δm then:

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

Optimal RFQ responses

So always have the first order condition

$$m^{ask}(x;s) = \frac{1}{h(m^{ask}(s))} + K^{\downarrow}(x;s)$$

for the optimal markup, independent of the distributional assumption for the inside market. And recalling the definition of $K^{\downarrow}(x;s)$ the optimal markup when offering to sell therefore breaks down as

$$m^{ask}(x;s) = \frac{1}{h(m^{ask}(x;s))} + \epsilon + \frac{\nu(x-s) - \nu(x)}{s} \tag{4}$$

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.

Exponential markups

If the inside hazard rate is assumed constant our life gets easier. The first order optimality condition becomes:

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

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

$$m^{ask}(x;s) = \max\left(\frac{1}{h} + \epsilon + \frac{\nu(x-s) - \nu(x)}{s}, 0\right) \tag{5}$$

since we can guarantee to trade with zero markup.

Similarly our market maker offers a markup of

$$m^{bid}(x;s) = \max\left(\frac{1}{h} + \epsilon + \frac{\nu(x+s) - \nu(x)}{s}, 0\right) \tag{6}$$

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.

Skew and Width

Let p^{market} be the market mid that all markups are defined with respect to. Then our market maker's bid p^{bid} is at $p^{market} - m^{bid}$ and her ask p^{ask} is at $p^{market} + m^{ask}$. The arithmetic midpoint of her bid and offer is

$$p^{mid} = p^{market} + \frac{m^{ask} - m^{bid}}{2}$$

We can re-express the optimal bid and offer as

$$m^{bid}(x) = \frac{1}{h} + K^{\uparrow} = \overbrace{\epsilon + \frac{1}{h}}^{\Delta} + \overbrace{\frac{\nu(x+s) - \nu(x)}{s}}^{K^{\uparrow} - \epsilon} \quad (7)$$

$$m^{ask}(x) = \frac{1}{h} + K^{\downarrow} = \epsilon + \frac{1}{h} + \overbrace{\frac{\nu(x-s) - \nu(x)}{s}}^{K^{\downarrow} - \epsilon} \quad (8)$$

where $\Delta = \frac{1}{h} + \epsilon$. Substituting these into p^{mid} we see that our market maker is skewing the mid point of her quotes as follows:

$$p^{mid} = p^{market} - \frac{\nu(x+s) - \nu(x-s)}{2s}$$

Normally the second term here would be called skew.

Optimal RFQ responses

We notice that the skew term is the secant slope and we are prompted to define slope and convexity of the inventory cost:

$$
\begin{aligned}
S(x) &:= \frac{\nu(x+s)-\nu(x-s)}{2s} = \frac{K^{\uparrow}-K^{\downarrow}}{2} \\
C(x) &:= \frac{\nu(x+s)-2\nu(x)+\nu(x-s)}{2s} = \frac{K^{\uparrow}+K^{\downarrow}}{2} + \epsilon
\end{aligned}
$$

so that slope = skew.

Similarly, to interpret the role of convexity $C(x)$ we note that

$$C(x) + S(x) = K^{\uparrow} - \epsilon$$

and

$$C(x) - S(x) = K^{\downarrow} - \epsilon$$

Since

$$\begin{aligned} m^{ask} &= \frac{1}{h} + K^{\downarrow} = \overbrace{C(x) + \Delta}^{width} + \overbrace{S(x)}^{skew} \\ m^{bid} &= \frac{1}{h} + K^{\uparrow} = C(x) + \Delta - S(x) \end{aligned}$$

it is clear that $C(x)$ plays the role of discretionary "width" over and above a minimum constant width Δ. That is, convexity = discretionary width.

Consistent skew and width

We are now in a position to recast the inventory cost relation (3) in terms of a relationship between optimal skew and width.

$$
\begin{aligned}
\frac{\tau h}{s}\left(c(x)-c(x')\right) &= p\,\overbrace{e^{-hC(x)+hS(x)}}^{P^{buy}(m^{bid})}+(1-p)\,\overbrace{e^{-hC(x)-hS(x)}}^{P^{sell}(m^{ask})} \\
&\quad -pe^{-hC(x')+hS(x')}-(1-p)e^{-hC(x')-hS(x')}
\end{aligned} \tag{9}
$$

for all x and x', provided we are in the region where our assumptions are valid.

Order imbalance

If there are more buying opportunities for our market maker than selling opportunities, or vice versa, we will have the analytic inconvenience $p \neq \frac{1}{2}$. But this turns out to be a superficial difficulty. Indeed we can derive the impact of order imbalance on the optimal skew and width as follows, even before solving $\nu(x)$.

Define

$$\delta = \frac{\log(1-p) - \log(p)}{2h}$$

and

$$\gamma = \frac{\log\left(\frac{1}{2}\right) - \frac{\log(1-p)+\log(p)}{2}}{h}$$

It is clear from concavity of $\log$ that $\gamma > 0$.

Optimal RFQ responses

For reasons that will shortly be clear we allow δ to move the skew, defining

$$S_\delta(x) = S(x) - \delta.$$

In similar fashion we allow γ to increase the adverse selection $\epsilon \to \epsilon + \gamma$ resulting in an increase to the non-discretionary component of width:

$$\Delta_\gamma = \Delta + \gamma.$$

Claim: Order imbalance requires the market maker to skew by δ and widen by γ.

Optimal RFQ responses

<u>Proof:</u> Substitute S_δ into the consistency relation. To remove clutter we multiply by $e^{h\Delta}$:

$$\begin{aligned}
\frac{\tau h}{s}\left(c(x) - c(x')\right) e^{h\Delta} &= pe^{-hC(x)+h(S_\delta(x)+\delta)} + (1-p)e^{-hC(x)-h(S_\delta(x)+\delta)} \\
&\quad -pe^{-hM(x')+h(S_\delta(x')+\delta)} - (1-p)e^{-hC(x')-h(S_\delta(x')+\delta)}
\end{aligned}$$

and more pertinently by $e^{h\gamma}$:

$$\begin{aligned}
\frac{\tau h}{s}\left(c(x) - c(x')\right) e^{h\Delta_\gamma} &= e^{h\gamma}pe^{-hC(x)+h(S_\delta(x)-\delta)} + (1-p)e^{h\gamma}e^{-hC(x)-h(S_\delta(x')} \\
&\quad -pe^{h\gamma}e^{-hC(x')+h(S_\delta(x')+\delta)} - (1-p)e^{h\gamma}e^{-hC(x')-h(S} \\
&= \frac{1}{2}e^{-hC(x)+hS_\delta} + \frac{1}{2}e^{-hC(x)-hS_\delta(x')} \\
&\quad -\frac{1}{2}e^{-hC(x')+hS_\delta(x'))} - \frac{1}{2}e^{-hC(x')-hS_\delta(x')}
\end{aligned}$$

where we have used $e^{h\delta+h\gamma} = \frac{1}{2p}$ and $e^{h\gamma-h\delta} = \frac{1}{2(1-p)}$.

We have recovered the symmetric relation with $S_\delta(x)$ in place of $S(x)$ and $\Delta + \gamma$ in place of Δ. Thus our market maker will have to be more defensive and use a wider market, as claimed. ■

There is another way to look at the impact of order imbalance.

Claim: Order imbalance increases the effective carrying cost by

$$e^{\log\left(\frac{1}{2}\right) - \frac{\log(1-p)+\log(p)}{2}} > 1$$

Proof: The second claim follows from inspection of the left hand side of (11), from which it is clear that the solution could also be achieved by solving the symmetric problem where the direct carrying cost $c(x)$ has been multiplied by

$$e^{h\gamma} = e^{\log\left(\frac{1}{2}\right) - \frac{\log(1-p)+\log(p)}{2}}$$

Of course $e^{h\gamma} > 1$ since $\gamma > 0$.

$\blacksquare$

Relating width and skew

With the distraction of order imbalance behind us we clean up the consistency relation (11) by defining the modified gain

$$G_\delta(x) = e^{-hC(x)} \cosh\left(hS_\delta(x)\right)$$

and cost coefficient

$$\Omega = \frac{\tau h}{s} e^{h\Delta_\gamma} = \frac{\tau h}{s} e^{1+h\epsilon+h\gamma}$$

Then internal consistency can be tersely written

$$G_\delta(x) - G_\delta(x') = \Omega\left(c(x) - c(x')\right) \tag{12}$$

which holds for all x, x'.

The relation (12) defines family of self-consistent solutions for $S(x)$ and $C(x)$. We can relate the former to the latter readily if we know that $C(x)$ and $S_\delta(x')$ are specified, by inverting (12):

$$S_\delta(x) = \frac{1}{h}\cosh^{-1}\left(\Omega e^{hC(x)}c(x) + e^{h(C(x)-C(x'))}\cosh\left(hS_\delta(x')\right) - e^{hC(x)}\Omega c(x')\right)$$

Returning to the special case $x' = 0$ we have some simplification. We have $c(0) = 0$ and by symmetry, $S_\delta(0) = 0$. Consistency implies that the gain is an affine function of carrying cost.

$$G_\delta(x) = \Omega c(x) + e^{-hC_0}$$

If convexity $C(x)$ is known,

$$S_\delta(x) = \frac{1}{h}\cosh^{-1}\left(e^{hC(x)}\left\{\Omega c(x) + e^{-hC_0)}\right\}\right)$$

Solving the exponential markup model

A better constant width model

As a special case we can assume convexity $C(x)$ equals the constant C_0. Since

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

has solution

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

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 (Avellaneda and Stoikov), but that isn't true here.

Instead we solve for skew directly:

$$S_\delta(x) \quad = \quad \frac{1}{h}\cosh^{-1}\left(1+e^{hC_0}\Omega c(x)\right) \tag{13}$$

and this yields an easily implementable constant width model.

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.

We can similarly motivate a linear skew model, noting that for $\nu(x) = \frac{C_0}{s}x^2$ as before we have, by direct calculation,

$$S^*_\delta(x) = \frac{\nu(x+s) - \nu(x-s)}{2s} = \frac{1}{2}\frac{C_0}{s}x \tag{14}$$

This is, of course, a different skew than before and if we seek a consistent width $C(x)$ we'll have

$$C^*(x) = \frac{1}{h}\log\left(\frac{\cosh\left(\frac{C_0}{2s}hx\right)}{\Omega c(x) + e^{-hC_0}}\right) \tag{15}$$

The fact that C, S and C^*, S^* don't coincide exactly merely reflects the fact that $\nu(x) \propto x^2$ is not an internally consistent inventory cost function.

Moral: We need to solve for $\nu(x)$!

Implied cost fitting

One direct approach to finding an approximate $\nu(x)$ is to specify a parametric form and then optimize the fit between the left and right hand sides of (3).

A current implementation uses a functional form suggested by Nicholas West:

$$\nu(x) = \begin{cases} c_{l_0} + c_{l_1}\frac{x}{s} + c_{l_2}\left(\frac{x}{s}\right)^2 + c_{l_3}\left(\frac{x}{s}\right)^3 & x < 0 \\ c_{r_0} + c_{r_1}\frac{x}{s} + c_{r_2}\left(\frac{x}{s}\right)^2 + c_{r_3}\left(\frac{x}{s}\right)^3 & x \geq 0 \end{cases}$$

It is only necessary to calculate $\nu(x)$ once on startup.

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 $\epsilon_h \sim N(0, \sigma_h^2)$ 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 $\pi + w$, is a sufficient statistic summarizing π and w.

Furthermore let $G(x; s)$ denote the expected net benefit to the market maker of the trading opportunity; $E^{\sigma^2}[G(x; s)]$ the same under uncertain knowledge of w and h; and $R_{\sigma^2} := \frac{E^{\epsilon}[G(x;s)]}{G(x;s)}$ the relative efficiency. Then

Claim: The relative efficiency is given by

$$\bar{R}_0 = e^{-\frac{3}{2}h^2\sigma^2}$$

where $\sigma^2 := \sigma^2_{\hat{\pi}} + \sigma^2_{\hat{w}}$ is the combined variance of $\hat{\pi} + \hat{w}$.

Optimal RFQ responses

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

$$\begin{aligned} G(\pi^{ask}) &= e^{-h(\pi_{\epsilon}^{ask} - \pi)} \left(\pi^{ask} - \pi - K^{\downarrow}(x;s) \right) && (16) \\ &= w e^{-h(\pi_{\epsilon}^{ask} - \pi)} && (17) \end{aligned}$$

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_ϵ^{ask} solving the wrong problem. Her choice will satisfy

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

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

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

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 $\pi + w$ is a sufficient statistic.

Provided $\pi_\epsilon > \pi$ the gain for other choices is

$$
\begin{aligned}
G(\pi_\epsilon^{ask}) &= e^{-h(\pi_\epsilon^{ask}-\pi)}\left(\pi_\epsilon^{ask} - \pi - K^{\downarrow}(x;s)\right) \\
&= e^{-h(\pi^{ask}-\pi+\epsilon_{w+\pi})}\left(\overbrace{\pi^{ask} - \pi - K^{\downarrow}(x;s)}^{=w} + \epsilon_{w+\pi}\right) \\
&= w e^{-h(\pi^{ask}-\pi)} e^{-h\epsilon_{w+\pi}}\left(1 + \frac{\epsilon_{w+\pi}}{w}\right) \\
&= G(\pi^{ask}) e^{-h\epsilon_{w+\pi}}\left(1 + \frac{\epsilon_{w+\pi}}{w}\right) \qquad (18)
\end{aligned}
$$

This prompts us to define the efficiency ratio

$$R(\epsilon_\pi, \epsilon_w) := e^{-\frac{\epsilon_{w+\pi}}{w}} \left(1 + \frac{\epsilon_{w+\pi}}{w}\right) \tag{19}$$

where, again, $\epsilon_{w+\pi}$ is the difference between the estimated location of the mean of the competitors' best ask and the ground truth. If the error is small relative to w then the chance of offering through the mid is small, and so integrating directly we have, approximately:

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

as claimed.

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.

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

■

A humble market maker

As (19) 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. But by how much?

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 given by

$$\bar{R}_{h\sigma^2} = e^{-\frac{3}{2}\frac{\sigma^2}{w^2}}$$

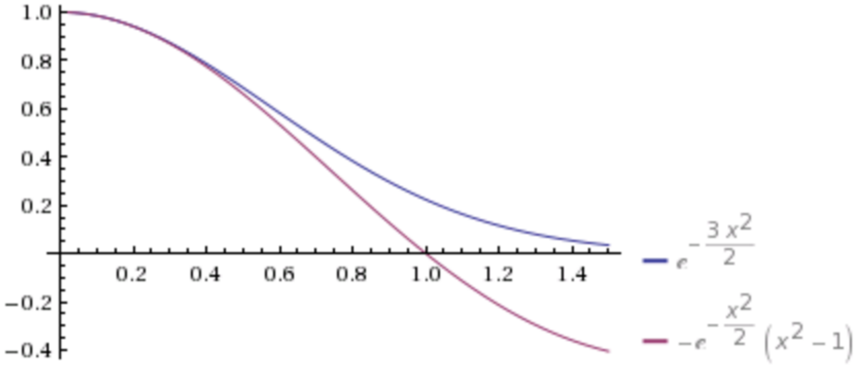

Figure 1: Relative efficiency as a function of the σ/w, where σ is the standard error in the market maker's estimate of $\pi + w$.

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 (18) her mean efficiency must be

$$\bar{R_\varphi} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(\epsilon-\varphi)^2} e^{-h\epsilon}(1+h\epsilon)d\epsilon$$
$$= e^{-\frac{h^2\sigma^2}{2}} e^{-h\varphi}(-h^2\sigma^2 + h\varphi + 1) \quad (21)$$

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

$$\bar{R} \approx e^{-\frac{3}{2}h^2\sigma^2}$$

■

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 (18) is not valid for all π_ϵ^{ask} and either is (19) for all $\epsilon_{\pi+w}$. In practice then, a market maker might want to deviate from (21) especially when given an opportunity to get out of a large inventory. A slightly better approximation might be

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

but this is merely a heuristic and does not follow from the considerations above.

Questions?