# 241222SOsession AlgorithmicTradingBook Ch2

for Heuristic Life ksh

December 2024

basic market making model that focuses on inventory and inventory risk, as well as the trade-off between execution frequency and profit per trade, basic measures of liquidity.

Market Microstructure is the study of the process and outcomes of exchanging assets under explicit trading rules" (O'Hara,1995)

: exchanging assets: trading strategies, and their outcomes: asset prices, volume, risk transfers, etc

vVho has what information, how does that information affect trading strategies, and how do those trading strategies affect trading outcomes in general, and asset prices in particular.

"market prices are an efficient way of transmitting the information required to arrive at a Pareto optimal allocation of resources" ( Grossman  Stiglitz (1976)).

liquidity and price discovery

Types of Trading : via personal deals settled over a handshake in a club, via decentralised chat rooms where traders engage 1 Abergel, Bouchaud, Foucault, Lehalle  Rosenbaum (2012) provides a general overview of the determinants and effects of liquidity in security markets and related policy issues. each other in bilateral personal transactions, via broker-intermediated over-thecounter (OTC) deals, via specialised broker-dealer networks, on open electronic markets, etc.

Liquidity Trader

passive' market maker (MM), who facilitates trade and profits from making the spread and from her execution skills, and must be quick to adapt to changing market conditions

Ative MM, who exploits her ability to anticipate price movements and must identify the optimal timing for her market intervention.

or example an MM is willing to purchase shares of company XYZ at 99$andwillingtosellat$101 per share.(=Quoted Spread is 2 dollar.) Note that by posting LOs, the MM is providing liquidity to other traders who may be looking to execute a trade quickly, e.g. by entering a market order (MO). Hence, we have the usual dichotomy that separates MMs as liquidity providers from other traders, considered as liquidity takers.

GrassmanMiller Model(1985):describes how MMs obtain a liquidity premium from liquidity traders that exactly compensates MMs for the price risk of holding an inventory of the asset until they can unload it later to another liquidity trader.

- **Time periods:** $t \in \{1, 2, 3\}$.
- **Agents:** $n$ market makers (MMs), liquidity trader LT1 (at $t = 1$), and liquidity trader LT2 (at $t = 2$).
- **Asset's cash value at $t = 3$:** $S_3 = \mu + \epsilon_2 + \epsilon_3$, where:
  - $\mu$: Constant component of asset value.
  - $\epsilon_2, \epsilon_3$: Independent normally distributed shocks with mean 0 and variance $\sigma^2$.

All agents are risk-averse with the expected utility function:

$$U(X) = -\exp(-\gamma X)$$

where $\gamma > 0$ is the risk aversion parameter.
The Grossman  Miller (1988) framework provides insights into how the cost of holding assets, driven by uncertainty faced by risk-averse MMs, impacts liquidity through trading costs ($|S_1 - \mu|$) and the demand for immediacy. At $t = 1$, LT1 only executes $\frac{n}{n+1} i$ instead of the desired $i$. Competition among MMs plays a crucial role in determining trading costs.
To understand what drives $n$ (the number of competing MMs), we consider both participation costs and trading costs:

- **Participation Costs:** These are fixed costs, denoted $c$, required to be actively present in the market, such as time, investments, and opportunity costs.
- **Trading Costs:** These are proportional to the number of shares traded, parameterized by $\eta$ (per-share trading fees).

—

Suppose each trader pays $\eta$ per share traded, whether buying or selling. For simplicity:

- Remaining inventories after $t = 2$ are liquidated at $t = 3$.
- LT1 wants to sell $i$ units ($i > 0$), and LT2 wants to buy the same amount.

At $t = 2$:

- **LT2's Demand:**

$$q_2^{\mathrm{LT2}} = \frac{\mathbb{E}[S_3 - \eta|\epsilon_2] - (S_2 + \eta)}{\gamma\sigma^2}$$

- **Price at $t = 2$:** Since everyone anticipates liquidating positions later, trading fees do not affect $S_2$:

$$S_2 = \mathbb{E}[S_3|\epsilon_2] = \mu + \epsilon_2$$

At $t = 1$:

- **LT1's Supply:**

$$q_1^{\text{LT1}} = \frac{\mathbb{E}[S_2 - \eta] - (S_1 - \eta)}{\gamma\sigma^2}$$

- **MMs' Demand:**

$$q_1^{\text{MM}} = \frac{\mathbb{E}[S_2 - \eta] - (S_1 + \eta)}{\gamma\sigma^2}$$

- **Market Clearing Condition:**

$$i = nq_1^{\text{MM}} + q_1^{\text{LT1}}$$

Substituting the above equations, the equilibrium price at $t = 1$ becomes:

$$S_1 = \mu - \frac{\gamma\sigma^2 i}{2(n+1)} - \eta$$

—

The total liquidity discount with trading fees is given by:

$$\Delta = \mu - S_1 = \frac{\gamma\sigma^2 i}{2(n+1)} + \eta$$

Almost all trading fees are borne by the initiating liquidity trader (LT1). Specifically:

- LT1 pays their own trading fee $\eta$ per share.
- LT1 indirectly pays a fraction of the MMs' trading fees, $\frac{n}{n+1}(2\eta)$, through a lower sale price $S_1$.

—

- **Competition:** Participation costs ($c$) directly reduce $n$, while trading fees ($\eta$) indirectly affect $n$ by lowering MMs' expected profits.
- **Immediacy:** Trading fees also reduce the immediacy for LT1. With fees, LT1's holdings at the end of $t = 1$ are:

$$q_1^{\text{LT1},*} = -\frac{i}{n+1} + \frac{2\eta n}{(n+1)^2\gamma\sigma^2}$$

Thus, trading fees introduce an additional liquidity discount and reduce the immediacy available in the market.
The Grossman  Miller (1988) model demonstrates how trading costs, whether setup costs or trading fees, are primarily borne by liquidity traders. These costs manifest either explicitly (as trading

fees) or implicitly (through a greater liquidity discount when selling or a larger premium when buying). To measure liquidity, we consider divergences from 'efficient' prices and how these can be observed in electronic exchanges.

—

In the Grossman  Miller (1988) model, equilibrium prices are solved in a 'Walrasian auctioneer' framework, where trading happens simultaneously at a single price. However, in electronic asset markets, trading decisions are made sequentially. For example:

- Suppose liquidity traders send market orders (MOs) to an exchange.
- These MOs interact with limit orders (LOs) resting in the limit order book (LOB), posted by patient MMs.
- As LT1's MOs execute against LOs, the price moves, adjusting the LOB. LT1 stops trading after selling $\frac{n}{n+1}i$ shares, as the price has moved too far.

The price $S_1$, therefore, reflects the average price LT1 receives, either from a single large order or multiple smaller orders that walk the LOB. The liquidity discount is the difference between $S_1$ and

the initial mid-price (a proxy for the efficient price, $\mathbb{E}[S_2]$).
The relationship between price and quantity traded, $q_{\text{LT1}}$, can be written as:

$$S_1 = \mu - \lambda q_{\text{LT1}}$$

where:

- $\lambda = \frac{\gamma\sigma^2}{n}$: Price impact parameter, representing the market's price reaction to LT1's order.
- $q_{\text{LT1}} = \frac{n}{n+1} i$: LT1's total quantity traded.

—

A second way to measure liquidity is by observing the autocovariance of asset price changes. This measure is related to the autocovariance of returns but is easier to compute for the Grossman-Miller model. To illustrate:

- Introduce an additional date $t = 0$, prior to LT1's order submission ($t = 1$).
- A public news event $\epsilon_1$ affects the asset's liquidation price:

$$S_3 = \mu + \epsilon_1 + \epsilon_2 + \epsilon_3$$

At $t = 0$, there are no liquidity traders, and the equilibrium price is:

$$S_0 = \mathbb{E}[S_3] = \mu$$

Subsequent equilibrium prices are:

$$S_1 = \mu + \epsilon_1 - \lambda q_{\mathrm{LT1}}$$

$$S_2 = \mu + \epsilon_1 + \epsilon_2$$

Define:

$$\Delta_1 = S_1 - S_0, \quad \Delta_2 = S_2 - S_1$$

The autocovariance of price changes is:

$$\mathrm{Cov}[\Delta_1, \Delta_2] = \mathrm{Cov}[\epsilon_1 - \lambda q_{\mathrm{LT1}}, \epsilon_2 + \lambda q_{\mathrm{LT1}}]$$

Expanding this:

$$\mathrm{Cov}[\Delta_1, \Delta_2] = -\lambda^2 \mathrm{Var}[q_{\mathrm{LT1}}] < 0$$

In this model, the autocovariance of price changes captures liquidity similarly to price impact. As liquidity increases ($\lambda \to 0$), the autocovariance approaches zero, and the price process converges to the martingale process $\mu_t$, representing the 'efficient price'.

—

The two liquidity measures—price impact ($\lambda$) and autocovariance of price changes—capture different dimensions of market liquidity:

- **Price Impact ($\lambda$)**: Reflects the temporary price movement caused by an individual trader's order.
- **Autocovariance**: Reflects the overall reaction of the market to incoming orders and provides insights into the dynamics of price changes over time.

In richer dynamic models, these measures become distinct. For example:

- The price impact ($\lambda$) captures temporary effects visible in execution prices.
- Autocovariance reflects both temporary and permanent effects on the efficient price.

In transitioning from the Walrasian auctioneer framework in the Grossman-Miller model to the measurement of price impact, we propose that market makers (MMs) participate by posting limit orders (LOs). This section examines why an MM would behave in this way and how to determine optimal behavior.

—

We consider a simplified, static version of the Ho Stoll (1981) model, capturing essential elements of the MM's problem. Key assumptions are:

- The MM is a small, risk-neutral trader with costless inventory management and infinite patience.
- The MM liquidates her inventory at the midprice ($S_t$) at no cost.
- Uncertainty comes from the timing and size of large incoming market orders (MOs), and all information is public.
- The MM makes money by posting LOs to the limit order book (LOB), unaffected by other MMs' decisions.

The MM's problem is to choose where to place her LOs to maximize profit per trade. She posts:

- A sell LO at $S_t + \delta_+$,
- A buy LO at $S_t - \delta_-$,

where $\delta_+$ and $\delta_-$ represent the distance of the LOs from the midprice.

—

The uncertainty from MOs comes from:

- The probability that an MO arrives ($P_\pm$).
- The probability that the MO "walks the book" to the MM's LO, represented by the cumulative distribution function $P_\pm(\delta_\pm)$.

Assume the distribution of other LOs in the LOB follows an exponential distribution with parameter $\kappa$. Then, the probability of execution for the MM's buy and sell LOs is:

$$P_-(\delta_-) = Pe^{-\kappa\delta_-}, \quad P_+(\delta_+) = Pe^{-\kappa\delta_+}$$

where $P$ is the probability of an MO arriving.

—

The MM's profit per trade is proportional to the distance of the LO from the midprice ($\delta_\pm$). The MM's objective is to balance the trade-off between:

- Higher profit per trade ($\delta_\pm$) as the LO is placed further from the midprice.
- Lower execution probability as $\delta_\pm$ increases.

The MM's profit per trade is given by:

$$\Pi_\pm = \delta_\pm P_\pm(\delta_\pm) = \delta_\pm P e^{-\kappa\delta_\pm}$$

The MM maximizes this profit by choosing $\delta_\pm$:

$$\delta_\pm^* = \frac{1}{\kappa}$$

—

The optimal depth $\delta_\pm^* = \frac{1}{\kappa}$ is equal to the mean depth of the LOB under the exponential distribution assumption. This result reflects:

- A trade-off between execution probability and profit per trade.
- The simplicity of the MM's decision in this static framework.

—

While this model captures key elements of limit order placement, it makes simplifying assumptions:

- The stochastic components ($P_{\pm}$ and $p_{\pm}$) are exogenous and static.
- The MM's decisions and those of other traders are independent.
- The MM's objective is static and ignores inventory risk or dynamic adjustments.

Future chapters (e.g., Chapter 10) explore these issues in dynamic inventory models, incorporating interactions with informed traders and other complexities.
So far, we have considered trading in contexts where all information is public. However, many trades arise because one party has (or believes they have) better information than what is reflected in current prices. This section explores how traders can exploit an informational advantage while accounting for their price impact, using the model introduced by Kyle (1985).

—

Kyle (1985) examines the decision problem of an informed trader in a market where the price is "efficient." The key features of the model are as follows:

- **Traders:**
  - An *informed trader* who knows the true value of the asset ($v$).
  - *Liquidity traders* with price-insensitive random net demand ($u$), where $u \sim N(0, \sigma_u^2)$ and is independent of $v$.
  - *Market makers (MMs)* who observe and compete for order flow, setting prices based on the total net order flow.
- **Asset Value:** The future cash value of the asset ($v$) is normally distributed, $v \sim N(\mu, \sigma_v^2)$.
- **Market Pricing:** MMs set the price ($S$) as a function of net order flow ($x + u$), where $x$ is the informed trader's demand.

—

The trading process proceeds as follows:

1. The informed trader observes $v$ and chooses their trade size $x(v)$.
2. The liquidity traders' random net demand $u$ is realized.
3. MMs observe the total order flow ($x + u$) and set the price $S(x + u)$.

The price is determined by the zero-profit condition for MMs:

$$S(x+u) = \mathbb{E}[v \mid x+u]$$

This reflects semi-strong efficiency, where prices incorporate all public information, including the order flow.

—

The informed trader aims to maximize their expected profit:

$$\max_x \mathbb{E}[x(v - S(x + u))]$$

Assuming the price function is linear in net order flow:

$$S(x + u) = \mu + \lambda(x + u)$$

where $\lambda$ is the sensitivity of the price to order flow. Substituting this into the objective function and taking expectations with respect to $u$, the problem becomes:

$$\max_x x(v - \mu - \lambda(x + u))$$

The first-order condition yields the optimal trading strategy:

$$x^*(v) = \beta(v - \mu), \quad \beta = \frac{1}{2\lambda}$$

—

To solve for $\lambda$, note that $S(x+u) = \mathbb{E}[v \mid x+u]$. Using the projection theorem for normal random variables:

$$S(x+u) = \mu + \frac{\sigma_v^2}{\sigma_v^2 + \sigma_u^2}(x+u-\mu)$$

Equating this with $S(x+u) = \mu + \lambda(x+u)$, we find:

$$\lambda = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_u^2}$$

—

1. **Price Impact:** The parameter $\lambda$ represents the market's price impact. It increases with $\sigma_v^2$ (greater informational advantage) and decreases with $\sigma_u^2$ (more noise from liquidity traders). 2. **Adverse Selection:** MMs are adversely selected, as informed traders buy when prices are low and sell when prices are high, leading to losses for MMs. This adverse selection is compensated through $\lambda$, which adjusts prices based on order flow. 3. **Informed Trading Profits:** The informed trader camouflages

their trades within the noisy order flow of liquidity traders, generating profits by exploiting their informational advantage.

—

The price impact parameter $\lambda$ reflects the additional premium required to compensate for adverse selection. This differs from the liquidity premium in Grossman Miller (1988), which compensates for inventory risk. The similarity is that in both models, prices move with order flow:

$$S(x+u) = \mu + \lambda(x+u)$$

where $\lambda > 0$, indicating that prices rise with buy orders and fall with sell orders.

—

The Kyle (1985) model illustrates how informed traders strategically exploit their informational advantage while accounting for price impact. The equilibrium reveals a balance where informed traders profit from their advantage, MMs adjust prices to order flow, and liquidity traders bear the cost through price adjustments. This highlights the importance of adverse selection in determining market prices and liquidity.

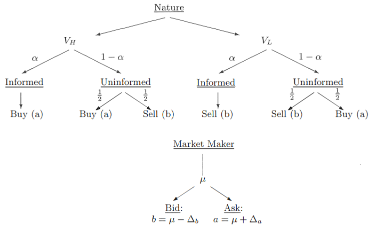


**Figure 2.3** The Glosten-Milgrom model.

Figure: Glosten Milgrom

The Kyle model focuses on the informed trader's problem, characterizing market makers' (MMs) decisions through competition. Glosten Milgrom (1985) shift the focus to MMs, placing them at the center of the problem when trading with better-informed counterparties.

—

In this simplified and static version of the Glosten-Milgrom model, we consider:

- **Asset Value:** The future cash value of the asset $v$ can take one of two values:

$$v \in \{V_H, V_L\}, \quad V_H > V_L$$

with an unconditional probability $p$ of $v = V_H$.

- **Traders:**
  - *Liquidity traders:* Price-insensitive traders, who buy or sell with equal probability $(1/2)$, regardless of the true asset value.
  - *Informed traders:* Traders who know the true value $v$. When $v = V_H$, they buy one unit if $a < V_H$. When $v = V_L$, they sell one unit if $b > V_L$.
  - *Market makers:* Competitive MMs post limit orders (LOs) at prices $a$ (ask) and $b$ (bid), seeking to balance profitability and competitiveness.
- **Trader Population:** A proportion $\alpha$ of the population are informed traders, and $1 - \alpha$ are uninformed liquidity traders.

—

The MM's task is to set $a$ and $b$ optimally, ensuring zero expected profits due to competition. Prices are set as:

$$a = \mu + \Delta_a, \quad b = \mu - \Delta_b$$

where:

- $\mu = \mathbb{E}[v \mid F]$: Expected asset value given all public information ($F$).
- $\Delta_a$: Ask half-spread, and $\Delta_b$: Bid half-spread.

—

The MM's expected profit from posting a price $a$ depends on the source of the buy order:

- If the buyer is an uninformed liquidity trader (probability $\frac{1-\alpha}{2}$):

$$\text{Profit} = a - \mu = \Delta_a$$

- If the buyer is an informed trader (probability $\alpha p$):

$$\text{Loss} = a - V_H = \Delta_a - (V_H - \mu)$$

The total expected profit from the ask price $a$ is:

$$\mathbb{E}[\text{Profit}] = \frac{1-\alpha}{2}\Delta_a + \alpha p(\Delta_a - (V_H - \mu))$$

Setting this to zero (no-profit condition), we solve for $\Delta_a$:

$$\Delta_a = \frac{\alpha p}{\alpha p + \frac{1-\alpha}{2}}(V_H - \mu)$$

Similarly, for the bid price $b$, the expected profit is:

$$\mathbb{E}[\text{Profit}] = \frac{1-\alpha}{2}\Delta_b + \alpha(1-p)(\Delta_b - (\mu - V_L))$$

Setting this to zero, the bid half-spread is:

$$\Delta_b = \frac{\alpha(1-p)}{\alpha(1-p) + \frac{1-\alpha}{2}}(\mu - V_L)$$

—

The spreads ($\Delta_a$ and $\Delta_b$) depend on:

- **Prevalence of informed trading ($\alpha$):** Higher $\alpha$ increases the spreads as the MM faces greater informational disadvantage.
- **Magnitude of informational advantage ($V_H - \mu$, $\mu - V_L$):** Larger deviations from $\mu$ increase the spreads.
- **Buy and Sell "Toxicity":**
  - Buy toxicity: $\alpha p$, associated with informed traders buying at $a$.
  - Sell toxicity: $\alpha(1-p)$, associated with informed traders selling at $b$.

The total spread $(\Delta_a + \Delta_b)$ reflects the MM's compensation for adverse selection risks due to informed trading.

—

The Glosten-Milgrom model illustrates how MMs manage their informational disadvantage by adjusting bid-ask spreads. These spreads compensate for the risk of adverse selection when trading with informed traders. The model provides a foundation for understanding market liquidity and the effects of asymmetric information.

The simple Glosten-Milgrom model can be extended in two complementary ways:

- Incorporating a time dimension.
- Allowing liquidity traders to become price-sensitive.

—

Incorporating a Time Dimension

To include a time dimension, consider the following:

- Set the interest rate to zero to avoid tracking discounting effects.

- Index all variables by time $t$, with $T$ representing the time when the asset's true cash value $v$ is determined.
- Adjust probabilities and expectations to reflect the accumulation of public information from trade, as captured by the filtration $\mathcal{F}_t$, which represents the history of observed trades up to time $t$.

At each time $t$, MMs observe the sequence of buy and sell orders and update their beliefs about the distribution of $v$ using Bayes' rule:

$$P_t = \mathbb{P}(v = V_H \mid \mathcal{F}_t), \quad \mu_t = \mathbb{E}[v \mid \mathcal{F}_t]$$

The bid and ask prices are then adjusted dynamically:

$$a_t = \mu_t + \Delta_a(t), \quad b_t = \mu_t - \Delta_b(t)$$

where $\Delta_a(t)$ and $\Delta_b(t)$ are the ask and bid half-spreads at time $t$, respectively.

—

Dynamic Bid-Ask Price Adjustments
The bid and ask prices evolve in response to the history of trading. This dynamic adjustment reflects the public information embedded

in the observed order flow:

- As buy and sell orders arrive, MMs continuously update $P_t$ and $\mu_t$ using the new information.
- The updated probabilities $P_t$ influence the bid and ask prices, ensuring they reflect the most current information about $v$.

—

Execution Prices as Conditional Expectations

At every execution:

- The execution price equals the conditional expectation of the asset's value based on the history of order flow ($\mathcal{F}_t$) and the type of trade (buy or sell).
- For a market buy order, the execution price is the ask price ($a_t$).
- For a market sell order, the execution price is the bid price ($b_t$).

This ensures the realized price process (the sequence of execution prices) is a martingale under the objective probability measure. Formally:

$$\mathbb{E}[P_{t+1} \mid \mathcal{F}_t] = P_t$$

—

## Implications for Price Dynamics

The model highlights:

- **Order Flow Impact:** The bid-ask spread adjusts dynamically to reflect the updated probabilities derived from observed buy and sell orders.
- **Martingale Property:** Execution prices reflect all available information and evolve as a martingale, ensuring they are unbiased predictors of the asset's future value.
- **Price Discovery:** The order flow acts as a mechanism for incorporating public information into prices, driving the market toward efficiency over time.

An interesting extension to the static model is to allow liquidity traders to avoid trading if the half-spread ($\Delta$) is too high. This can also be incorporated into the dynamic setting discussed earlier.

—

## Utility of Liquidity Traders

Liquidity traders are assumed to derive an additional (exogenous) utility, $C_i$, from executing their desired trade. The trader $i$'s decision depends on the relationship between their utility and the transaction cost:

- If $\Delta > C_i$, trader $i$ avoids the trade.
- If $\Delta \leq C_i$, trader $i$ executes the trade.

The distribution of the utility parameter $C_i$ across the population of liquidity traders is described by a cumulative distribution function $F(c)$, where $F(c)$ gives the proportion of traders with $C_i \leq c$. We refer to $C_i$ as the trader's *urgency parameter*.

—

Expected Profit for Market Makers

The expected profit for the MM from setting an ask price $a = \mu + \Delta_a$ is modified to account for the reduced participation of liquidity traders:

$$\mathbb{E}[\text{Profit}] = (1 - F(\Delta_a))\frac{1-\alpha}{2}\Delta_a + \alpha p\left(\Delta_a - (V_H - \mu)\right)$$

This expression incorporates the fact that only a fraction $1 - F(\Delta_a)$ of liquidity traders will execute their trade when the ask half-spread is $\Delta_a$.

---

Updated Half-Spreads

The new half-spreads are implicitly defined as:

$$\Delta_a = \frac{\alpha p}{\alpha p + \frac{1-\alpha}{2}(1 - F(\Delta_a))}(V_H - \mu)$$

and similarly:

$$\Delta_b = \frac{\alpha(1-p)}{\alpha(1-p) + \frac{1-\alpha}{2}(1 - F(\Delta_b))}(\mu - V_L)$$

---

Market Collapse and Extreme Solutions

A key implication of introducing price-sensitive liquidity traders is that as the MM increases the half-spread ($\Delta_a$ or $\Delta_b$), the number of liquidity traders willing to trade decreases:

- If the urgency parameters $(C_i)$ in the population are small, the MM faces a shrinking pool of liquidity traders.
- At extreme levels of the half-spread, the solutions simplify to:

$$\Delta_a = V_H - \mu, \quad \Delta_b = \mu - V_L$$

These extreme solutions correspond to a market collapse where only informed traders participate, and the order flow fully reveals the asset's true value.

In such a scenario:

- No gains from trade occur for liquidity traders.
- Trades made by informed traders immediately reveal the asset's underlying value, and the market price becomes strongly efficient.

---

Conclusion

Introducing price-sensitive liquidity traders increases the bid-ask half-spreads and introduces the risk of market collapse. When urgency parameters are low, the MM must balance setting profitable spreads with maintaining sufficient participation from liquidity traders.