Stevens Institute of Technology

Glosten-Milgrom Model and PIN model

FE 570: Market Microstructure and Trading Strategies

Professor: Dr. Dan Pirjol

Team 5

Darsh Kachhara

Kinjal Shukla

Run Liu

Yash Kantharia

Introduction

The following report understands and analyses the Coca-Cola Co(KO) stock under two sequential trading models Glosten-Milgrom model and the probability of informed traders (PIN) model.

The GM(Glosten-Milgrom) model analyzes the effects of the process of adverse selection on the bid-ask spread by simulating a market where a risk-neutral specialist interacts with informed and uninformed traders on a single security. Adverse selection refers to when an informed trader makes a trade with a market maker which puts the market maker at a high risk of loss since the trader may know where the stock price is heading. To counter the trading losses from adverse selection, the market makers have to generate a bid-ask spread even when they have zero transaction costs, inventory costs, and profit.

This Probability of informed traders (PIN) model is an extension of the basic Glosten-Milgrom model (Glosten and Milgrom, 1987[2]) where news arrival is dynamic at random times and we try to estimate the probability of informed trading (PIN) defined as the unconditional probability that a randomly chosen trader on a randomly chosen day is informed.

Liquidity

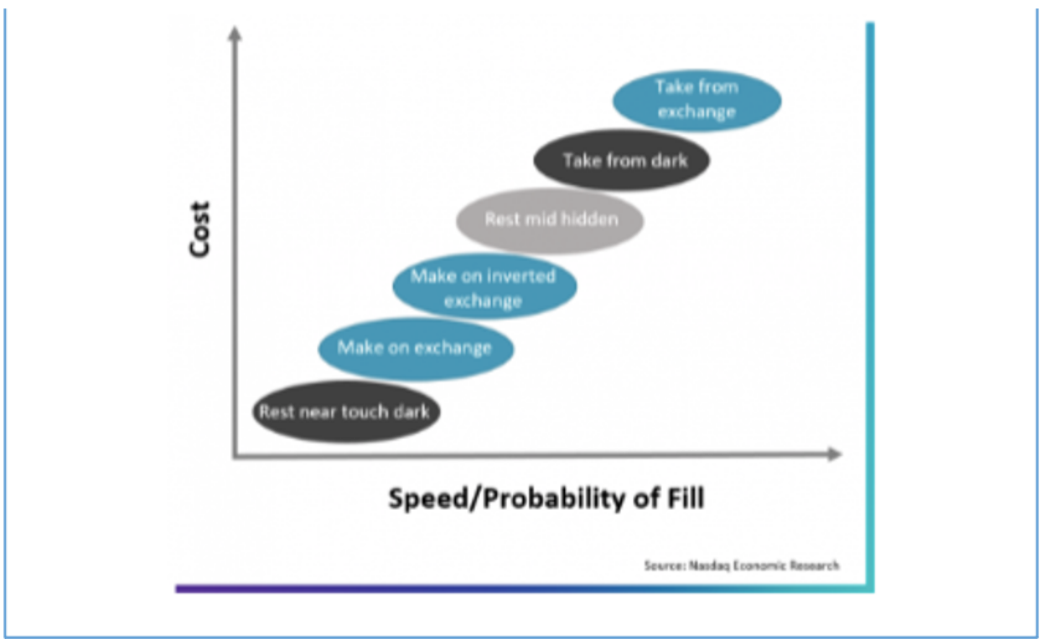
Liquidity is the property of the markets which allows rapid and cheap trade execution. It is the most important characteristic of well-functioning markets.

Dimensions of liquidity

1. Immediacy/time dimension: Execution speed is very important at the microscale. This is called latency, and depends on technology, access to markets (direct or through a broker). Impatient traders use market orders for a quick transaction execution.
2. Depth - trade size dimension Trading a large block of stock introduces market impact.
3. Cost. There are two types of costs associated with trading: Direct costs. Broker fees, exchange, etc. They can also be negative (rebates), which can impact traders' behavior. Implicit cost. Most important in microstructure is the bid-ask spread. For small trades, this is observable in the market, but this can become unknown for large orders, and must be modeled. Called in traders language market breadth.

Orders and liquidity

1. Market orders consume liquidity. (Take liquidity)
2. Limit Orders provide liquidity. (Make liquidity)



Liquidity and Adverse Selection:

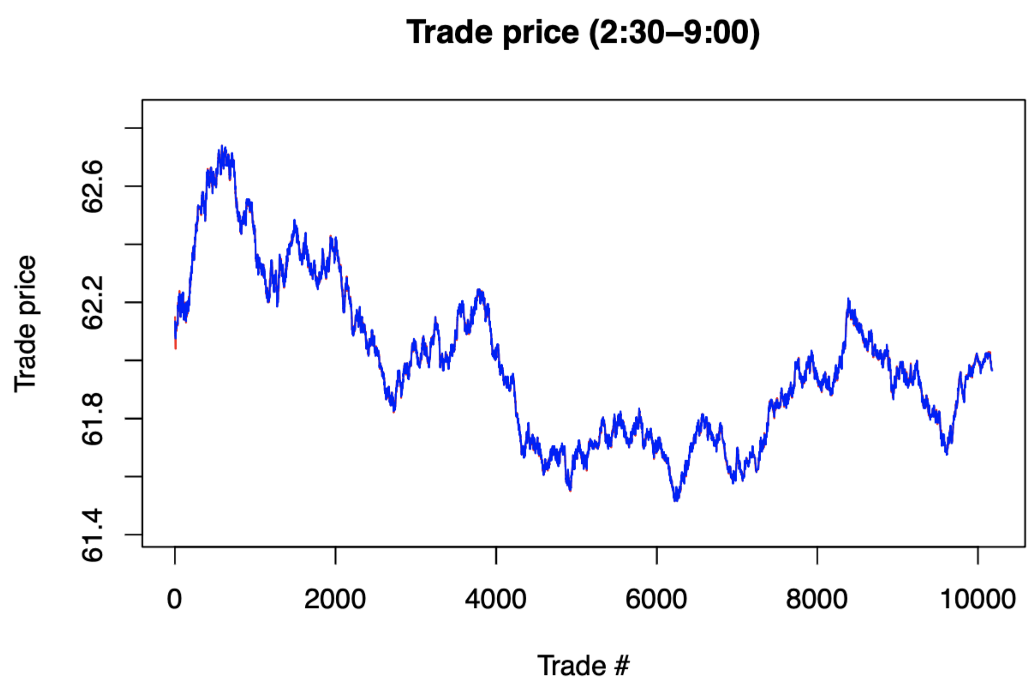
1. The bid-ask spread covers the dealers' risk of trading with counterparts who have superior information about true security value. Informed traders trade at one side of the market and may profit from trading with dealers. This component of the bid/ask spread is called the adverse-selection component since dealers confront one-sided selection of their order flow.

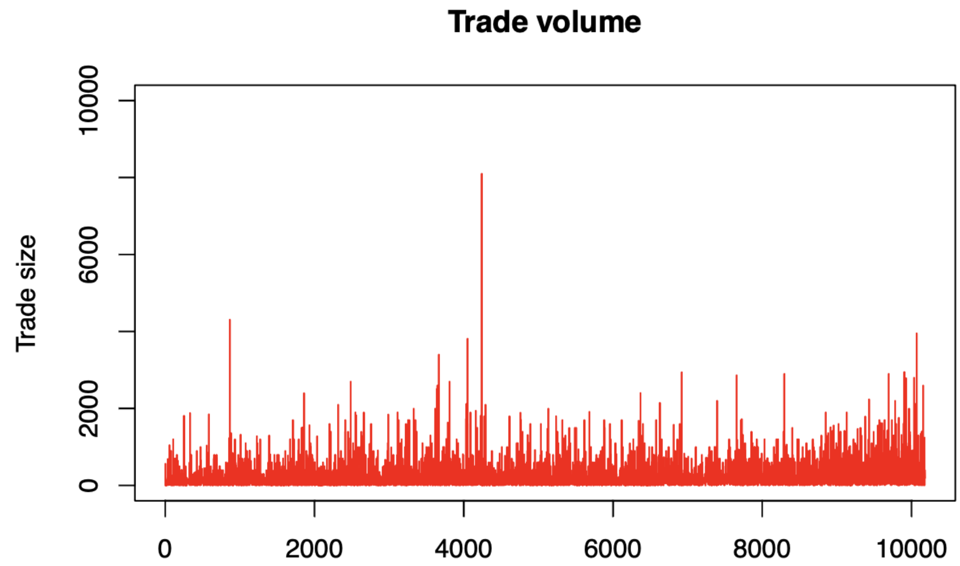
2. Adverse selection plays an important role in market microstructure. Glosten and Milgrom built a model for bid-ask spread arising purely from informational asymmetry between "informed traders" and "dealers."

Data

```
##                         SYMBOL BID     BIDSIZ OFR     OFRSIZ PRICE    SIZE
## 2022-03-01 14:30:01.254 "KO"   "62.10" "  4"  "62.15" " 11"  "62.150" "     1"
## 2022-03-01 14:30:01.255 "KO"   "62.12" "  2"  "62.15" " 11"  "62.150" "     1"
## 2022-03-01 14:30:01.998 "KO"   "62.09" "  1"  "62.13" "  1"  "62.100" "   566"
## 2022-03-01 14:30:02.035 "KO"   "62.07" "  1"  "62.12" "  1"  "62.090" "   200"
## 2022-03-01 14:30:02.107 "KO"   "62.07" "  1"  "62.13" "  2"  "62.120" "     1"
## 2022-03-01 14:30:02.910 "KO"   "62.05" "  1"  "62.12" "  1"  "62.080" "     1"

##                         SYMBOL BID     BIDSIZ OFR     OFRSIZ PRICE    SIZE
## 2022-03-01 20:59:59.011 "KO"   "61.96" " 95"  "61.97" " 27"  "61.970" "  1245"
## 2022-03-01 20:59:59.038 "KO"   "61.96" "117"  "61.97" " 35"  "61.970" "    11"
## 2022-03-01 20:59:59.046 "KO"   "61.96" "117"  "61.97" " 35"  "61.970" "     6"
## 2022-03-01 20:59:59.115 "KO"   "61.96" "117"  "61.97" " 21"  "61.970" "   400"
## 2022-03-01 20:59:59.155 "KO"   "61.96" "115"  "61.97" " 16"  "61.970" "   300"
## 2022-03-01 20:59:59.302 "KO"   "61.96" "125"  "61.97" " 12"  "61.970" "   190"
```





The data used in this report is of the date 1st March 2022, from 2:30 pm to 9 pm (GMT). The trading volume and volatility in the stock after the stock market is closed , explains a lot about the amount of participants trading this stock and thus its liquidity.

The following are the specs of the data:

SR NO.	SPECIFICS	DATA
1	Roll model spread	0.32(cents)
2	EffectiveSpread	0.71(cents)
3	RealizedSpread	0.17(cents)
4	Volume	2150894(shares)
5	Change in price	(0.29%)
6	Volatility	0.0112%

Hence, we could determine that Coca-Cola(KO) is a highly liquid stock.

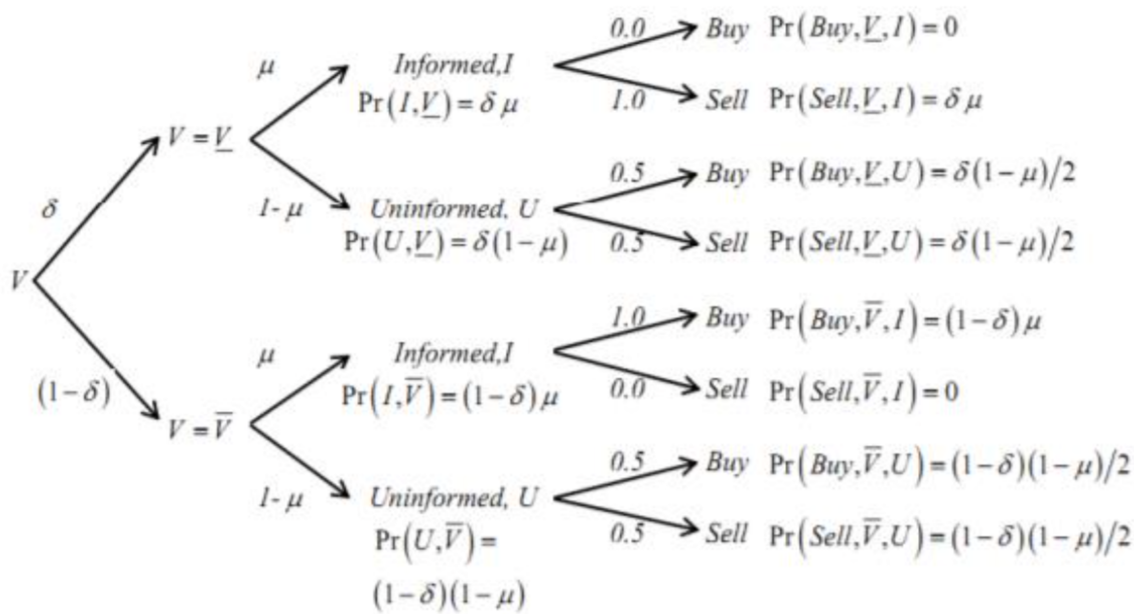
Assumptions made by the GM(Glosten-Milgrom) model

1. The market maker must be prepared for both an informed trader buying or selling at any quoted price. It is assumed the market maker knows the percentage of informed traders in the market so it can use Bayesian reasoning to determine the optimal bid and ask prices to quote.
2. Maker-maker has unlimited cash and securities to trade with zero holding costs of inventory. This assumption is implemented by the zero profit of the market maker for each trade because they do not require a trading profit to offset costs.
3. There is an asymmetry of information between the informed traders and the dealer where the informed traders know the true value of the security and the dealer does not. This is implemented by the informed trader always buying at $\overline{V}$ and selling at $\underline{V}$
4. Noise traders buy and sell randomly with an even probability. This is implemented by the uninformed traders buying or selling with a 50/50 chance no matter the stock price movements.
5. The trade prices form a martingale because the market maker's bid and ask quotes are derived from the unconditional probabilities of a trader buying or selling.

$$\mathbb{E}[P_1] = \mathbb{P}(buy)A(\delta) + \mathbb{P}(sell)B(\delta) = \delta\underline{V} + (1-\delta)\bar{V} = P_0$$

6. The percentage of informed traders in the market is proportional to the bid-ask spread created by the market makers. As the proportion of informed traders in the market increases, the bid price approaches the next high value of the security and the ask price approaches the next low value. At 100% informed traders the spread will have widened to the point that no informed traders can profit from the bid and ask prices equal to the high and low values of the security, respectively.

Event Tree and Model specifications for GM Model



In the event tree we have set the value of the stock to go down as $\overline{V}$ and stock to go down as $\underline{V}$. The value μ describes the number of informed traders in the market and δ denotes the probability that the stock value would go down.

The probabilities of buys and sells are calculated from the event tree by using the formula:

$$\begin{aligned}
P(\text{buy}) &= P(\text{buy}|\underline{V})P(\underline{V}) + P(\text{buy}|\overline{V})P(\overline{V}) \\
&= \frac{(1-\mu)\delta + (1+\mu)(1-\delta)}{2} = \frac{1+\mu(1-2\delta)}{2} \\
P(\text{sell}) &= P(\text{sell}|\underline{V})P(\underline{V}) + P(\text{sell}|\overline{V})P(\overline{V}) \\
&= \frac{(1+\mu)\delta + (1-\mu)(1-\delta)}{2} = \frac{1-\mu(1-2\delta)}{2}
\end{aligned}$$

Now we can calculate the Bid and ask value by the formula: Bid is the Expected value of the stock given buys and Ask is the Expected Value of the stock given sells.

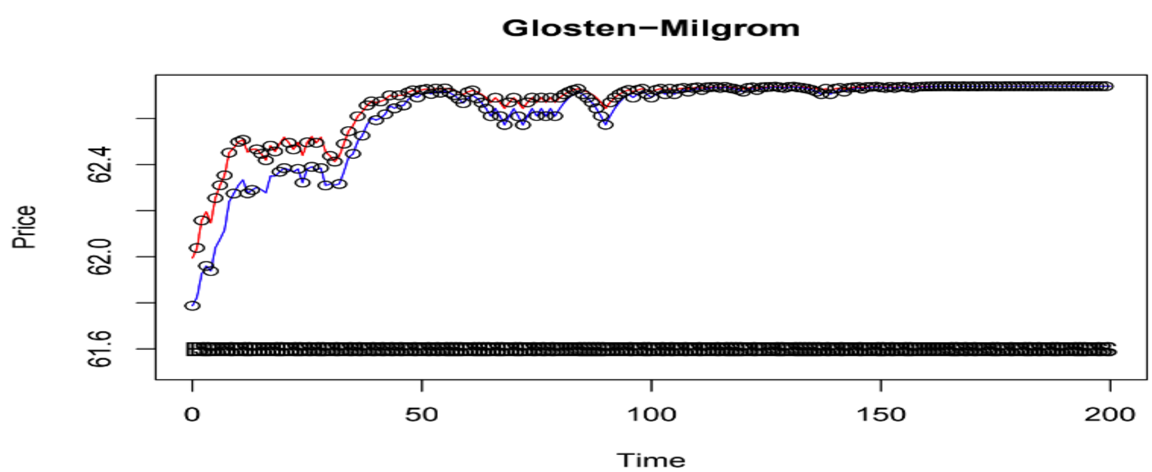
$$Ask^{GM} = \frac{(1+\mu)(1-\delta)\overline{V} + (1-\mu)\delta \underline{V}}{1+\mu(1-2\delta)}$$

$$Bid^{GM} = \frac{(1-\mu)(1-\delta)\overline{V} + (1+\mu)\delta \underline{V}}{1+\mu(2\delta-1)}$$

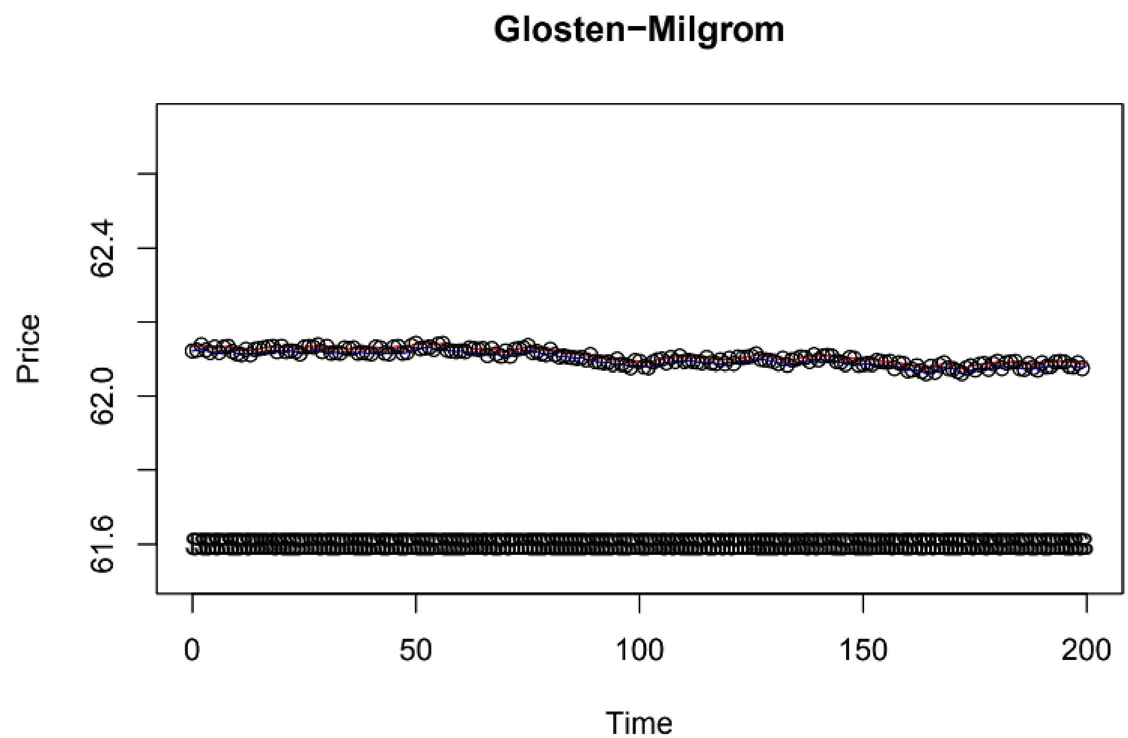
The Bid-Ask spread is calculated from the difference between Ask and Bid

$$\psi^{GM} = Ask^{GM} - Bid^{GM} = \frac{4\mu\delta(\delta-1)(\overline{V}-\underline{V})}{\mu^2(1-2\delta)^2-1}$$

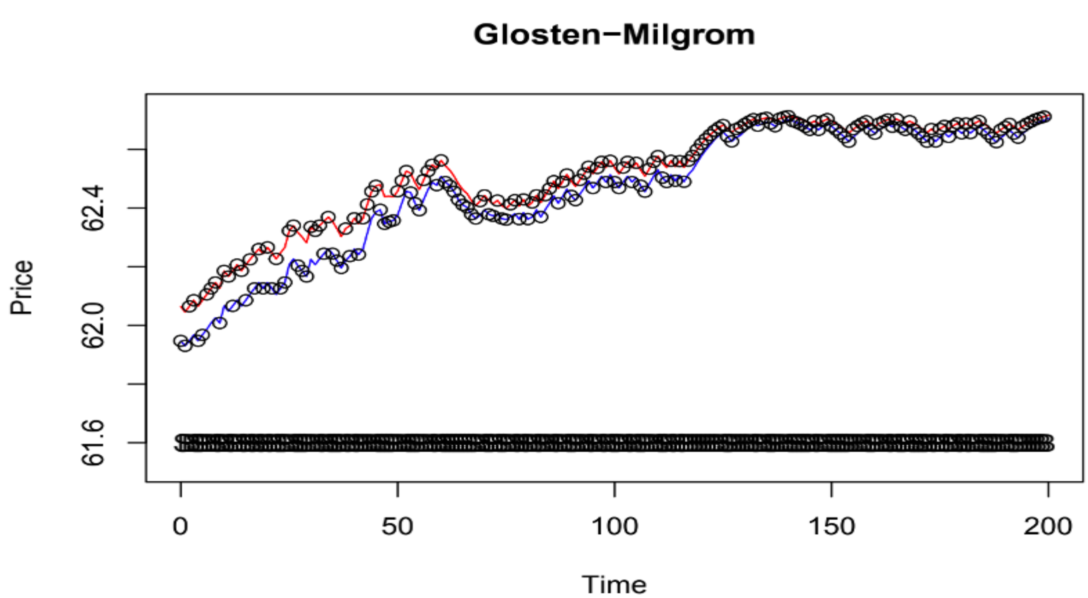
Our observations (code is mentioned in the appendix)



Bids are blue, asks are red. Then we increase the number of uninformed traders from 1% to 20%.



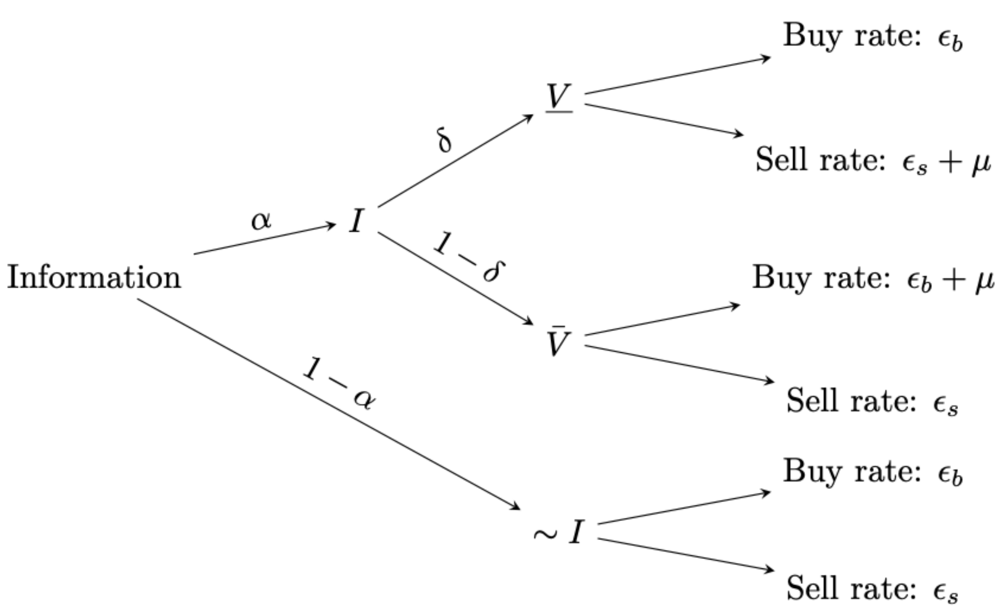
The bid-ask spread becomes much smaller when the number of uninformed traders increases from 1% to 20%. Therefore, altering the percentage of the number of informed traders in the market makes the bid-ask spread vary a lot.



When we increase the mu value which is the percentage of informed traders to 0.5 or more, we can see the spread is more shifted to the buy side indicating if the market contains more informed traders, who know the price of the security value prior, they get to affect the market price drastically.

<u>Event tree and specifications for PIN model</u>

The event tree for the arrival of information and the informed and uninformed buyers and sellers is as shown :



The joint probability distribution with respect to the parameter vector $\Theta \equiv \{\alpha,\ \delta,\ \mu,\ \varepsilon b,\ \varepsilon s\}$ and the number of buys and sells (Bt,St), is specified by

$$f(Bt, St|\Theta) = (1 - \alpha)P(Bt, \varepsilon b)P(St, \varepsilon s) + \alpha[\delta P(Bt, \varepsilon b)P(St, \mu + \varepsilon s) + (1 - \delta)P(Bt, \mu + \varepsilon b)P(St, \varepsilon s)]$$

Where $P(n, \theta) = e - \theta\theta n/n!$ is the Poisson distribution for the random variable n with parameter θ.

The estimates of arrival rates ($\hat{\mu},\ \hat{\varepsilon}\ and\ \hat{\varepsilon}$), along with estimates of the probabilities ($\hat{\alpha}$ and $\hat{\delta}$) can be bs obtained by maximizing the joint log-likelihood function given the order input matrix (Bt,St) over T trading days. The non-linear objective function of this problem can be written as:

$$L(\Theta|T) = \sum_{t=1}^{T} L(\Theta|(B_t, S_t) = \sum_{t=1}^{T} \log[f(B_t, S_t|\Theta)]$$

The maximisation problem is subject to the boundary constraints α, δ ∈ [0, 1] and μ, εb, $\varepsilon s \in [0, \infty)$. The PIN estimate is then given by;

$$\widehat{PIN} = \frac{\hat{\alpha}\hat{\mu}}{\hat{\alpha}\hat{\mu} + \hat{\epsilon}_b + \hat{\epsilon}_s}$$

For our simulation, we use the coding language R and the packages it provides(i.e High frequency, InfoTrade). The data used on this model is the stock data of the Coca-cola co. (KO)

Our observations (code is mentioned in the appendix)

```
##       Parameter       YZ_1d       YZ_5m       YZ_15m       YZ_30m
## 1         Alpha    1.0000000  0.15663255   0.19230592   0.30766950
## 2         Delta    1.0000000  0.15370041   0.00000000   0.00000000
## 3  Epsilon(buy) 4808.9999242 53.23640440 154.28631312 293.89259557
## 4 Epsilon(sell)  870.0430533 66.94439876 206.38466689 413.00002564
## 5            Mu 4500.9568029 65.96328789 160.51298123 247.61532085
## 6           PIN    0.4421372  0.07916463   0.07883667   0.09728769
```

After running the model, for the provided data, the number of buys for the day time period was found to be significantly high. As we can observe from the results obtained from the PIN Model, the number of informed traders for 1-day simulation is fairly high with a probability of 0.442.

The number of informed traders for 5, 15, and 30 minutes time frames is low, with probability of 0.079, 0.078 and 0.097 respectively.

We also look at the Alpha value, which is the rate of arrival of news. For 1-day, α being 1, we can conclude that there is a high chance that there would be some news available with respect to a particular stock under consideration. That news could be either positive or negative. On the other hand when the data under consideration is within 5, 15, or 30 minutes, we can see that the chances of expecting any news is considerably lower. The rate being as low as 0.156.

As we increase the time frame, the model is more confident about the α . Now, coming to the assessment of what would be the trade with respect to the information gathered, we can look at the δ values to be more informed in this regard. δ value is the probability of informed buyers buying given a particular news. The δ value for 1-day is 1, which tells us there is a high likelihood of the news for which the informed seller would unload their position.

On other hand, in all the other time frames , we could see more buying action.

Conclusion

The Glosten-Milgrom model analyzes transaction prices arising from quotes of competing risk-neutral dealers making a market in a single security and facing both privately informed and uninformed traders. The dealers must quote bids and offer prices at which they are willing to, respectively, buy and sell. A trader arrives and decides to buy, sell, or leave. Upon seeing the outcome of the trader's decision, the dealers modify their quotes in preparation for the next arrival.

We analysed this model using an algorithm provided in class, to see the action the stock would have given its specification.

We then used the estimation method for the PIN YZ method in the InfoTrad package by Celik and Tinic, 2018[4] and applied it to live trading data. While it was not easy to get access to real-time TAQ data for many different equities, we used the Coca-Cola (KO) data provided in class. Because of the limited data sets available we split each data set into multiple time periods - 1d, 5m, 15m and 30m - and ran the PIN model estimates on each of these time periods.

PIN model , considers the live trading data and gives the probability of informed traders as an output based on real data, which is more helpful that determining the effective price of the market and making multiple assumptions for doing so.

Thus, we can conclude by saying that the PIN Model is more accurate in terms of the information it gives us, when compared to Glosten Milgrom, helping the dealers and traders to be more informed.

Questions for further research:

1. The use of more complex PIN models that also take into consideration the variability of the arrival of informed traders during different periods of time.

2. The use of fast and accurate optimization routines to estimate the PIN parameters more accurately and consistently.

3. The use of larger datasets with a wide variety of equities to see how PIN parameters vary between different equities or between different classes of investments.

References

1. D. Easley and M. O'Hara. "Price, Trade Size, and Information in Securities Markets". Journal of Financial Economics, 19(1):69-90, 1987.
2. L. R. Glosten and P. R. Milgrom. "Bid Ask and Transaction Price in a Specialist Market with Heterogeneously Informed Traders"". Journal of Financial Economics, 14(1):71-100, 1987.
3. Joel Hasbrouck, Empirical Market Microstructure, Oxford University Press, 2007