High Frequency Trading Strategies

Stanford MSE448

Joachim Sasson
ICME
Stanford University
jsasson@stanford.edu

Wei Hong Ho
ICME
Stanford University
whho@stanford.edu

Finsam Samson
Computer Science
Stanford University
finsam@stanford.edu

Abstract

In recent years, significant work in Quantitative Finance has been done to build efficient High Frequency Trading strategies. These strategies take advantage of high frequency data and attempt to trade intraday from frequencies ranging from microseconds to minutes. In this paper we focus our interest on two strategies. The first one involves the microprice: a fair estimator of a stock's price that is an adjustment of the mid price. The second strategy takes into account the inventory risk and aims at deriving an optimal trading strategy by solving a Hamilton-Jacobi-Bellman equation. While understanding and implementing the microprice strategy was insightful, we fail to make significant profit by using it as an estimator of the fair price. The second strategy however proves to be much more profitable than a classic symmetric strategy consisting of quoting symmetrically at both sides of the midprice.

1 Introduction

The problem of optimal market making trading strategy is well known in classical trading. The agent has views or opinions on the price of an asset at a given time horizon and trades in consequence. In high-frequency however, the trading strategy is very dynamic and trades can occur every nanosecond. Therefore, a high frequency trader will want to have an estimation of the fair price of the stock at any time. Many approaches can be used to estimate the fair price. This work explores two of them: the **microprice** and the **Avellaneda-Stoikov model** which takes into account the inventory risk. Before diving deeply into the above two papers, we will try to motivate the above models by considering two examples and the pitfall of both examples.

1.1 Pitfall for mid-price

The mid-price is defined as

$$M = \frac{1}{2}(P^a + P^b) \tag{1}$$

where P^a is the best (lowest) ask price and P^b is the best (highest) bid price. The mid-price is widely used in the market. However, it does suffer from a few drawbacks. First, changes in mid-prices are highly auto-correlated, this is something known as the bid-ask bounce. Second, changes in the mid-prices are infrequent compared to the changes of quote updates, making it a low frequency signal. Finally, the mid price does not take into account the volumes that are quoted for the bid and ask.

1.2 Pitfall for weighted mid-price

We will try to take into account the volume at the top of the order book: the best bid's volume Q_t^b and best ask's volume Q_t^a. For that purpose we define the weighted mid price as follows:

$$W_t = I_t P_t^a + (1 - I_t) P_t^b \tag{2}$$

Where $I_t := \frac{Q_t^b}{Q_t^a + Q_t^b}$ quantifies the predominance of the best bid with respect to the best ask. However, if it takes into account the volumes, the weighted mid price suffers from at least three flaws. First, because the weighted mid-price changes on every update of the imbalance, it is a quite noisy signal. Second, there is no theoretical justification for the weighted mid-price to be considered a 'fair' price, since it is not necessarily a martingale. Finally, in practice there exist some cases in which the behavior of W_t is counter intuitive. For instance assume $P^b = 10.00$, $Q^b = 9$, $P^a = 10.02$ and $Q^a = 27$. At this moment the weighted mid-price is 10.005. However, if a sell order of size one arrives at 10.01, the new weighted mid-price will be updated upward to 10.009. This is counter intuitive since the impact of a new sell order should surely be to update the price downward.

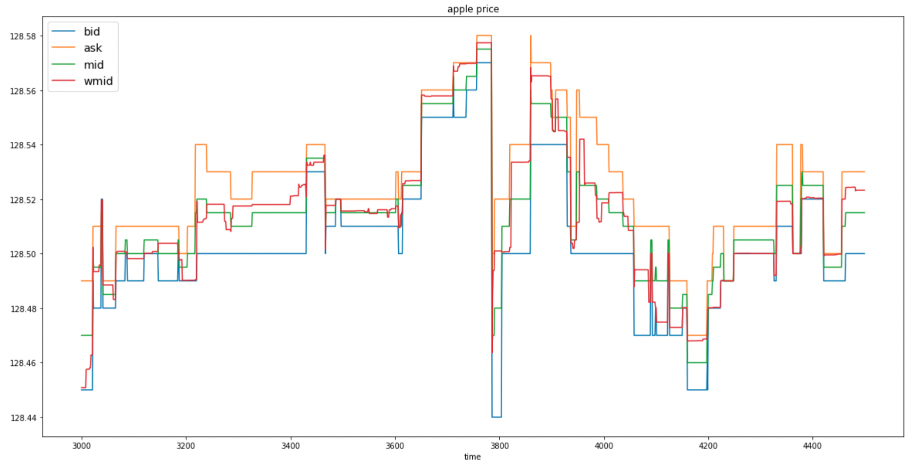

Figure 1: AAPL dataset with Naive pricing

2 MicroPrice

Now that we have presented the two mainly used ways of estimating the fair price, we focus our interest on a construction called **the microprice** defined by Sasha Stoikov in [1]. This construction aims at correcting the few drawbacks of the weighted mid price and especially it gives a theoretical foundation to the price construction.

2.1 Definitions

We define the microprice as follows:

$$P_t^{micro} := \lim_{i \to \infty} \mathbf{E}\left[M_{\tau_i} | \mathcal{F}_t\right] \tag{3}$$

where the τ_i's are the times when the mid-price changes:

$$\tau_1 = \inf\{u > t, M_u - M_{u^-} \neq 0\} \tag{4}$$

$$\tau_{i+1} = \inf\{u > \tau_i, M_u - M_{u^-} \neq 0\} \tag{5}$$

This construction is pretty natural and will remind the reader of the Black-Scholes theory in which we define the price of a derivative as the expectation of its discounted payoff at maturity. Now that we have provided the definition of the microprice, we need to make some assumptions to derive a close form formula.

2.2 Assumptions

Following the paper, we will make two main assumptions:

- The information in the order book is determined by the processes of the mid, the imbalance and the spread:
$$\mathcal{F}_t = \sigma(M_s, I_s, S_s; s \leq t)$$
- Mid price increments are independent from mid-price level:
$$\mathbf{E}[M_{\tau_{i+1}} - M_{\tau_i} | M_t = M, I_t = I, S_t = S] = \mathbf{E}\left[M_{\tau_{i+1}} - M_{\tau_i} | I_t = I, S_t = S\right]$$

The first assumption leads us to think of the order book as a Markov chain defined by $X_t := (M_t, I_t, S_t)$. The second one is very natural and will remind the reader of the property of independence of the increments for Brownian motions. This assumption leads us to think of the order book only as $X_t := (I_t, S_t)$

Given those two assumptions we can simplify the prediction of the ith micro-price as:

$$\mathbf{E}\left[M_{\tau_i}|\mathcal{F}_t\right] = M_t + \sum_{k=1}^{i} g^k(I_t, S_t) \tag{6}$$

where

$$g^1(I, S) = \mathbf{E}\left[M_{\tau_1} - M_t | I_t = I, S_t = S\right] \tag{7}$$

is the first order micro-price adjustment. The $i+1$-th order micro-price adjustment

$$g^{i+1}(I, S) = \mathbf{E}\left[g^i(I_{\tau_1}, S_{\tau_1}) | I_t = I, S_t = S\right] \tag{8}$$

Under further assumptions that we will describe in the next section, we will be able to derive a closed form for the microprice and to evaluate the quality of this estimator.

2.3 Finite-State space model

We denote by $X_t := (I_t, S_t)$ the state of the order book. If we assume there is a a finite number of possible values of the imbalance I_t and spread S_t, we can express the micro-price as:

$$P_t^{micro} = M_t + \sum_{k=1}^{\infty} B^k G^1 \tag{9}$$

where:

$$\begin{aligned}
B &:= (I - Q)^{-1} T && (10)\\
G^1 &:= (I - Q)^{-1} R K && (11)\\
Q_{xy} &:= \mathbf{P}(M_{t+1} - M_t = 0 \cap X_{t+1} = y | X_t = x) && (12)\\
T_{xy} &:= \mathbf{P}(M_{t+1} - M_t \neq 0 \cap X_{t+1} = y | X_t = x) && (13)\\
R_{xk} &:= \mathbf{P}(M_{t+1} - M_t = k | X_t = x) && (14)
\end{aligned}$$

2.4 Implementation

2.4.1 Estimation

The first step approximation to the micro-price is:

$$\begin{aligned}
g^1(i) =& \mathbb{E}\left[M_{\tau_1} - M_t | X_t = i\right]\\
=& \left(1 - Q\right)^{-1} R^1 \underline{k}
\end{aligned}$$

Where $Q_{ij} := \mathbb{P}(M_{t+1} - M_t = 0 \wedge X_{t+1} = j | X_t = i)$ are the transition probabilities for transient states (in which the mid price does not move) and $R^1_{ik} := \mathbb{P}(M_{t+1} - M_t = k | X_t = i)$ are the transition probabilities into absorbing states (mid price does move) and $\underline{k} = [-0.01, -0.005, 0.005, 0.01]^T$.

Now, we can compute recursively:

$$g^{n+1} = B g^n$$

Where $B := \left(1-Q\right)^{-1} R^2$ and R^2 is a new matrix of absorbing states:

$$R^2_{ik} := \mathbb{P}(M_{t+1} - M_t \neq 0 \wedge I_{t+1} = k | I_t = i)$$

We observe experimentally that the series will converge very well within 6 iterations.

2.4.2 Data Manipulation

- On every quote, compute $I_t, S_t, (M_{t+1} - M_t)$, after having discretized the state space
- Symmetrize the data, by making a copy where $I^2_t = n - I_t, S^2_t = S_t, (M^2_{t+1} - M^2_t) = -(M_{t+1} - M_t)$
- Estimate transition probability matrices Q, R^1, R^2
- compute the first micro-price adjustment

$$p^1 - M = g^1 = \left(1-Q\right)^{-1} R^1 \underline{k}$$

- compute recursively the p^6 adjustment

$$p^6 - M = g^1 + g^2 + \ldots + g^6 = g^1 + Bg^1 + \ldots + B^5 g^1$$

2.4.3 Results

For our project, we trained the model for AAPL and CVX. We trained the model on T-1 day and run the testing on T. We pick stocks from 2 completely different industry hoping to see different properties of micro-prices from it. For AAPL, we trained the data on 2021/01/05, getting data at 0.1s interval for a total of 288000 data points. On figure 2.4.3 we can see how the dataset is organized.

```
[2]: df= twx.bookquery('20210105', 'aapl', 'time: 9:00am 5:00pm 0.1s', 'direct.bid, direct.ask, direct.mid, direct.bsize, direct.asize')

[3]: len(df)

[3]: 288000

[3]: df.head()

[3]:
```

datetime	direct.bid	direct.ask	direct.mid	direct.bsize	direct.asize
2021-01-05 09:00:00.000	128.4	128.42	128.41	988	1561
2021-01-05 09:00:00.100	128.4	128.42	128.41	1088	1561
2021-01-05 09:00:00.200	128.4	128.42	128.41	1088	1561
2021-01-05 09:00:00.300	128.4	128.42	128.41	1088	1561
2021-01-05 09:00:00.400	128.4	128.42	128.41	1088	1561

```
[5]: df.columns = ['bid', 'ask', 'mid', 'bs','as']
     df['imb']= df.bs/(df.bs + df['as'])
     df['wmid']=df.imb*df.ask+(1-df.imb)*df.bid
     df['spread']=df.ask-df.bid
```

Figure 2: AAPL dataset

As mentioned in section 2.3, the state-space is completely defined by the imbalance and the spread. We will now take a look at how imbalance and spread affects the micro-prices for different equities.

Imbalance vs Different type of prices Since the mid price is independent of volume (therefore imbalance), we can see it is just a horizontal line. We can see for the results are quite similar for both stocks. The curves are smoother for the AAPL compared to Cheveron(CVX). This is probably due to the fact that AAPL is more liquid than CVX. As expected, the micro-prices should be between the mid-prices and weighted mid prices.

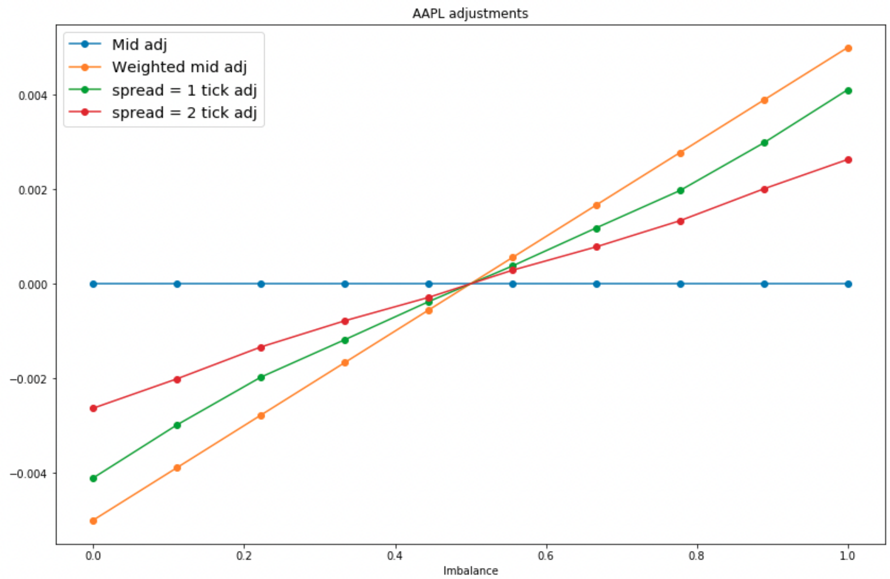

Figure 3: AAPL imbalance vs adjustments with respect to the mid

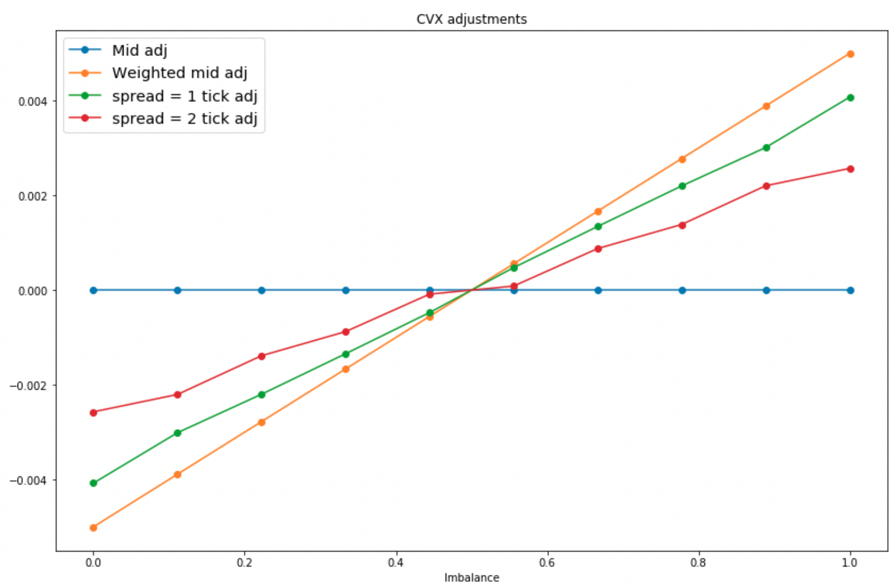

Figure 4: CVX imbalance vs adjustments with respect to the mid

Stationary distribution We can also plot frequency of imbalances with different spread. In this example, the stationary distribution is quite different for both stocks. For 1-tick spread in both AAPL and CVX, we can see it has an N-shaped curve. This implies that most of the time, the imbalance is not extreme. It is more likely for the best bid and best ask volume to be similar than lopsided. However, for the 2-tick spread, we can see a difference in behavior. It's a U-shaped curve for AAPL but it's a N-shaped curve for CVX. Our guess is that, for a very liquid stock, it is rare to have a 2-tick spread. But when such a scenario exists, it would imply there is a strong movement(strong buying or selling), causing the order book to be lopsided.

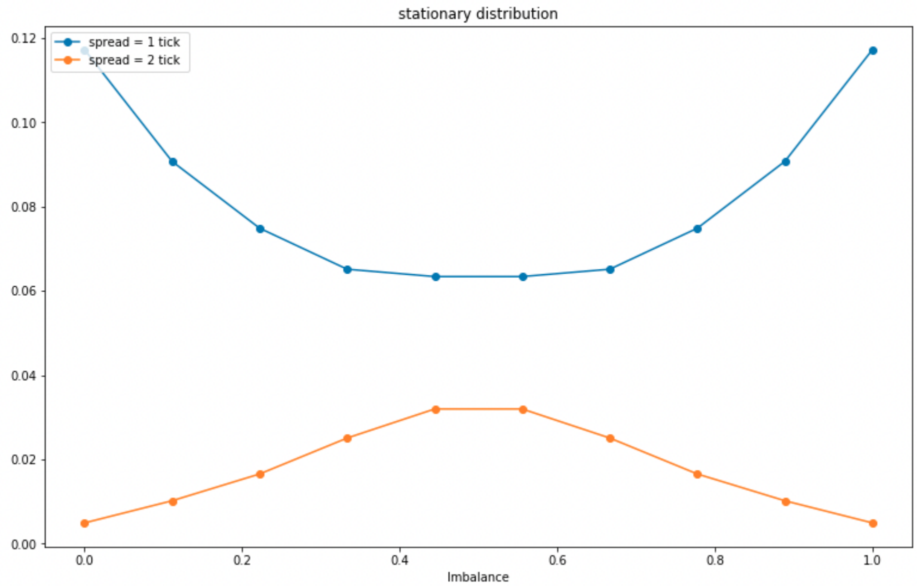

Figure 5: AAPL stationary distribution

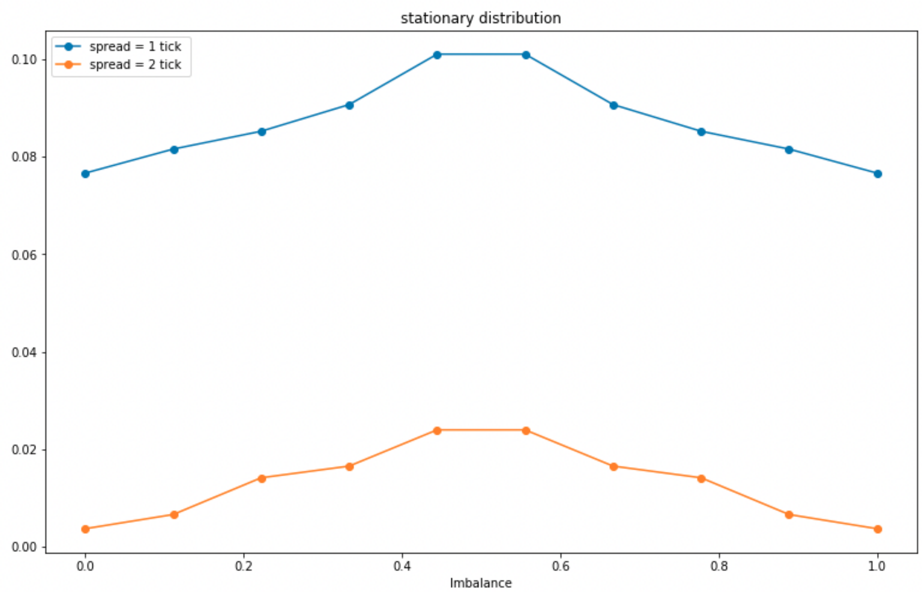

Figure 6: CVX stationary distribution

Out of Sample Prediction In this subsection, we will analyze the plot of the microprice over time. We can see in 2.4.3 and 2.4.3 that the micro-prices in the graph has a lot more movement as compared to the bid and ask. This might give us the chance to create more trading signals.

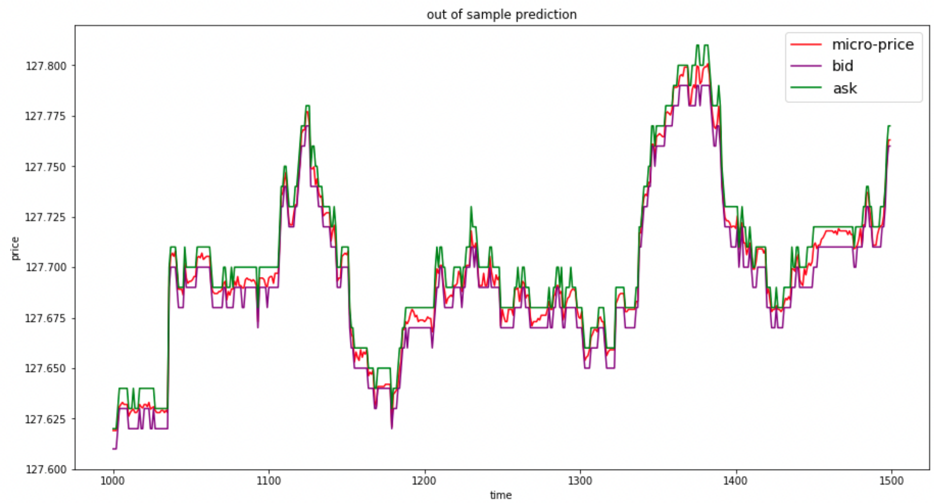

Figure 7: AAPL Out of Sample microprice

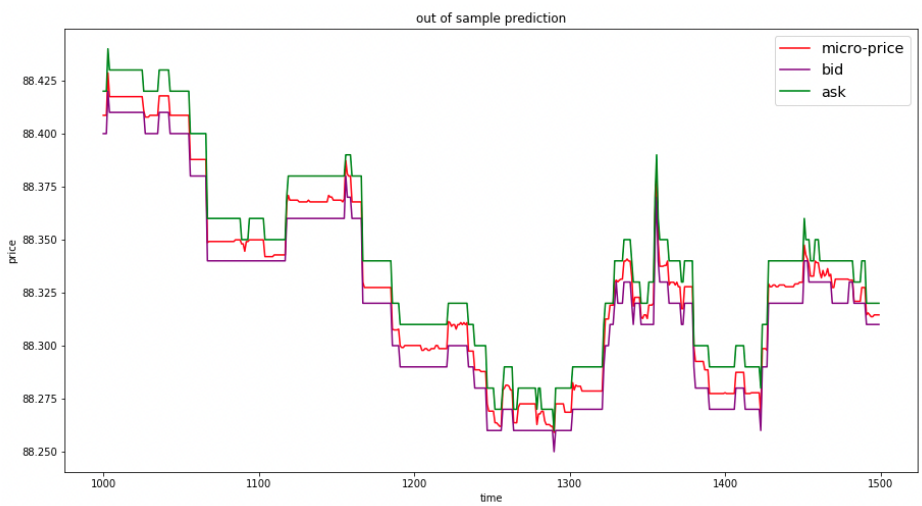

Figure 8: CVX Out of Sample micro-price

Conclusion for micro-price Let us briefly summarize the uses and benefits of this model. We started off with the weighted mid-price which is a lot better than the mid-price as it takes into account the volume of orders. However, we realized the model can be improved by including the spread to the model. Below are the advantages of using the microprice compared to the mid and weighted mid price.

- The micro-price adjustment is only dependent on the spread and imbalance
- The micro-price is the expected mid-price in the distant future
- In practice, the micro-price converges rapidly and is adequately approximated when we compute the series until τ^6 (time of the 6th price move)
- The micro-price can fit very different microstructures
- The micro-price seems to live between the bid and the ask

Even though the micro-price takes the spread and imbalance into account. The main risk of a market maker is often inventory risk. The profit and rebates we can make through market making trades can easily be dwindled when the price only moves in one direction. In what follows, we develop a method that takes into account the inventory risk.

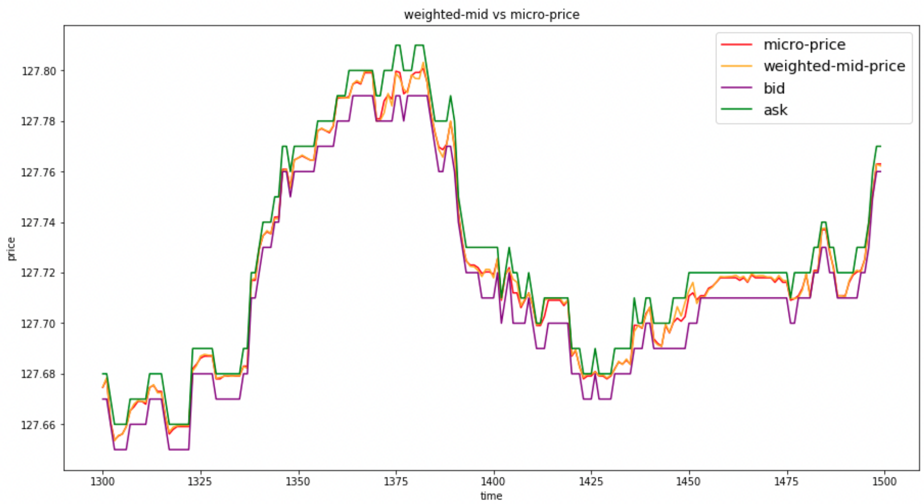

Figure 9: AAPL Out of Sample micro-price vs Weighted mid

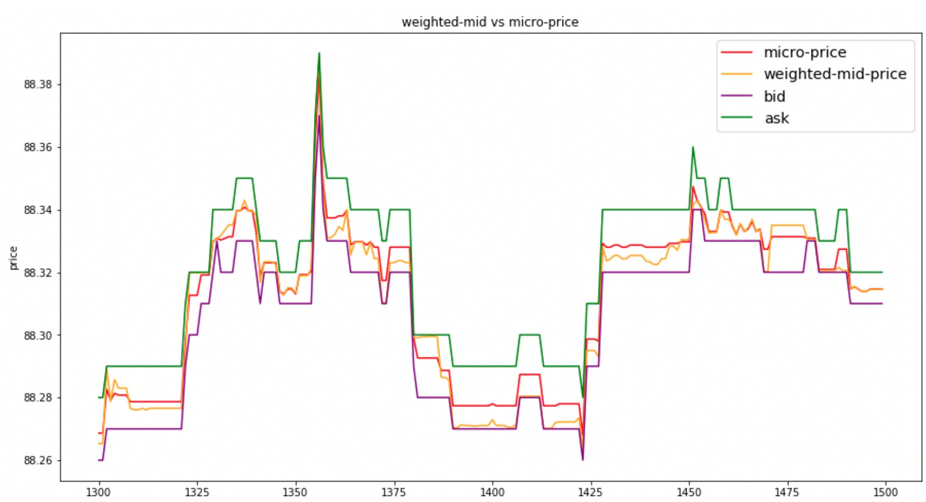

Figure 10: CVX Out of Sample micro-price vs Weighted mid

3 The Avellaneda-Stoikov model

The main idea of this section comes from the paper [2]. As said before, this paper takes the inventory risk into account. Using utility functions and arrival rates of buy and sell orders, we are able to come up with a close form solution for the fair price. We will refer to this fair price as the reservation price.

3.1 Construction of the model

3.1.1 Assumptions

The Avellaneda-Stoikov model relies on two major assumptions:

- The mid price is a geometric brownian motion with volatility σ, the agent has no opinion on the drift or any autocorrellation structure of the stock

$$dS_u = \sigma dW_u \tag{15}$$

- The agent's value function is:

$$v(x, s, q, t) = \mathbf{E}\left[-\exp(-\gamma(x + qS_T))\right] \tag{16}$$

  It can be written as $-\exp(-\gamma x)\exp(-\gamma qs)\exp\left(\frac{\gamma^2 q^2 \sigma^2 (T-t)}{2}\right)$ where x is the initial wealth in dollars.

We can notice that the agent's value function is an increasing function of the final wealth and the product qS_T. Indeed, if the inventory is large then if S_T is large the agent maximizes his utility.

3.1.2 Reservation bid and ask

In this framework, the two quantities of interest are called the reservation bid and ask.

- The reservation bid r^b is defined by:

$$v(x - r^b(s, q, t), s, q + 1, t) = v(x, s, q, t) \tag{17}$$

  It is the price that would make the agent indifferent between his current portfolio and his current portfolio plus one stock.

- The reservation ask r^a is defined by:

$$v(x + r^a(s, q, t), s, q - 1, t) = v(x, s, q, t) \tag{18}$$

  It is the price that would make the agent indifferent between his current portfolio and his current portfolio minus one stock.

Our goal in general will be to estimate r^a and r^b. Under the assumptions described in 3.1.1 we can solve analytically the implicit equations that define r^a and r^b.

$$r^b(s, q, t) = s - (1 + 2q)\frac{\gamma\sigma^2(T-t)}{2} \tag{19}$$

$$r^a(s, q, t) = s + (1 - 2q)\frac{\gamma\sigma^2(T-t)}{2} \tag{20}$$

If the agent is long stock ($q \geq 0$), the reservation price is below the mid-price, indicating a desire to liquidate the inventory by selling stock. On the other hand, if the agent is short stock ($q \leq 0$), the reservation price is above the mid-price, since the agent is willing to buy stock at a higher price.

In this framework, we define the **reservation price** $r(s, q, t) = s - q\gamma\sigma^2(T - t)$ as the mid of these two prices. It is an adjustment of the midprice which accounts for the inventory held by the agent.

3.1.3 Market impact

Now we allow the agent to trade the stock through limit orders that he sets around the mid-price.

- p^b denotes the bid quote ($\delta^b = s - p^b$)
- p^a denotes the ask quote ($\delta^a = s - p^a$)

Let us assume the agent places an order to buy Q stocks. Let p^Q be the price of the highest limit order executed in the trade. We define $\Delta p = p^Q - s$ to be the **temporary market impact** of the trade.

From these definition we can derive a Poisson model for the trading intensity:

- The orders to sell stock will hit the agent's buy limit order at Poisson rate $\lambda^b(\delta^b)$, a decreasing function of δ^b.
- The orders to buy stock will hit the agent's sell limit order at Poisson rate $\lambda^a(\delta^a)$, a decreasing function of δ^a.

Intuitively, the further away from the mid- price the agent positions his quotes, the less often he will receive buy and sell orders. In this framework, the wealth and inventory are now stochastic and depend on the arrival of market sell and buy orders.

$$dX_t = p^a dN_t^a - p^b dN_t^b \tag{21}$$

where N_t^b is the amount of stocks bought by the agent, N_t^a is the amount of stocks sold by the agent. N_t^b and N_t^a are Poisson process of intensity $\lambda^b(\delta^b)$ and $\lambda^a(\delta^a)$. The inventory is thus defined as:

$$q_t = N_t^b - N_t^a \tag{22}$$

3.1.4 Final optimization problem

The objective of the agent who can set limit orders is:

$$\max_{\delta^a, \delta^b} \mathbf{E}_t \left[-\exp(-\gamma(X_T + q_T S_T)) \right] \tag{23}$$

We will assume from now on that we have symmetric, exponential arrival rates:

$$\lambda^a(\delta) = \lambda^b(\delta) = Ae^{-k\delta} \tag{24}$$

which corresponds to $\Delta p \simeq ln(Q)$.

3.2 Implementation of the model

We now attempt to test the effectiveness of this strategy. To do this, we need a control strategy to be used as a benchmark. The control strategy places symmetrical trade around the "mid-price" rather than the reservation price.

The Avellaneda-Stoikov (AS) strategy, will be placing trades at $r^a = r + \delta^a$ and $r^b = r - \delta^b$. The implementation will be as follows:

- Estimate σ thanks to the sample variance of the stock, and κ the arrival rate of orders as the change in volume per second. We try two values (0.1 and 0.5) for the risk-averson parameter γ.
- Calculate the reservation price based on current inventory
- Calculate the optimal bid and ask price using δ^a, δ^b
- Create market orders based on reservation price
- If market orders are competitive(at the best price or better), it will have a probability of being executed with a probability related to κ

3.3 AS strategy results

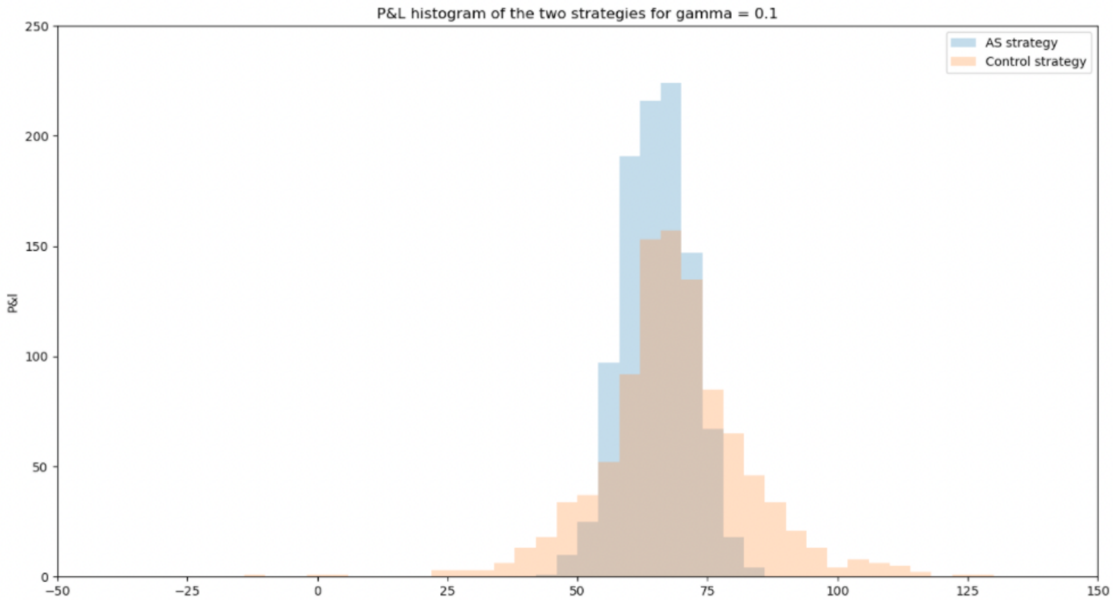

Figure 11: AS vs control strategy, gamma=0.1

Strategy	Profit	Std(Profit)	average inv	Std(inv)
AS	65.288	6.45	-0.333	2.99
Control	68.932	14.24	0.531	8.49

Table 1: AS vs control table gamma=0.1

We first take a deeper dive into the figure with parameter $\gamma = 0.1$. In this case, we are only slightly inventory risk adverse. The profit of the control strategy is greater than the AS strategy, this is expected as there are more constraints in the control strategy (less inventory). We can truly see shine by comparing the inventory. For the AS strategy, the average inventory is -0.333 while the control strategy has an average inventory holding of 0.531. The standard deviation of the AS strategy is 3 to 4 times lower than the control strategy.

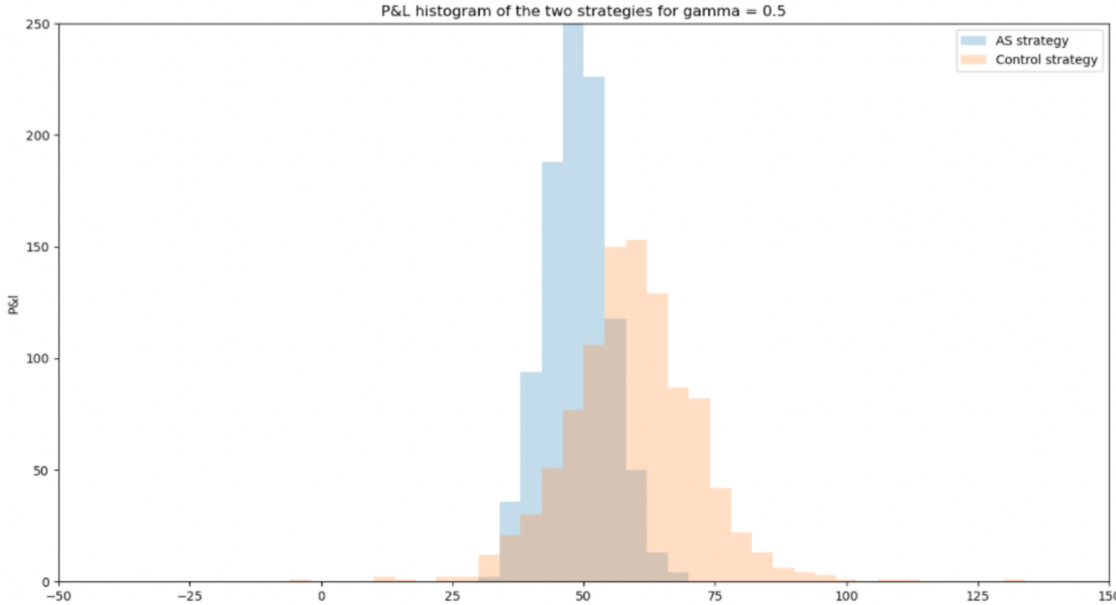

Figure 12: AS vs control strategy, gamma=0.5

Strategy	Profit	Std(Profit)	average inv	Std(inv)
AS	48.95	5.92	-0.055	2.10
Control	59.97	12.27	-0.296	6.75

Table 2: AS vs control table gamma=0.5

For the scenario $\gamma = 0.5$, we can see that the AS strategy has significantly lower profit than the control strategy (20% lesser). From the graph, it is obvious that the AS strategy is skewed more to the left compared to the control strategy. However, the average inventory is significantly lower as well.

3.4 Conclusion and improvement for AS strategy

We are able to see the impact of AS strategy in reducing inventory sizing, which is a great risk for a market maker. However, there are many strong assumptions in the model and some of it might not be realistic. We will point out some of them here:

- Optimal order quantity is not considered, we only place limit order of 1 bid and 1 ask
- Treating volatility as a constant might be too strong of an assumption
- There is strong evidence that book order liquidity varies with time, however it is also treated as a constant in this paper.

References

[1] Sasha Stoikov. The micro-price: A high frequency estimator of future prices. 2017.

[2] Marco Avellaneda and Sasha Stoikov. High frequency trading in a limit order book. *Quantitative Finance*, 8:217–224, 04 2008.