Contents

1. Introduction..........3
2. Literature Review..........4
2.1 The mid-point of the stock..........5
2.2 The trading intensity..........5
2.3 Limit orders..........5
2.4 Solution..........6
3. Methodology..........7
3.1 Fit the Empirical Formula..........7
3.2 Theoretical Test Procedure..........7
3.3 Extensions of the model..........8
4. Data Description..........9
5. Theoretical Result..........10
5.1 Comparisons of Three Models..........13
5.2 The Impact of the Choice of Gamma..........15
5.3 The Case of Trend in the Price Dynamics..........17
6. Best-test Results..........19
6.1 Procedure..........19
6.2 Comparison of the three models..........20
7. Conclusions and Future Improvements..........23
7.1 Conclusions..........23
7.2 Future Improvements..........23
8. References..........24

1. Introduction

The object of our project is to investigate optimal limit order placement models as a function of order flow arrival time and depth, order book shape, local volatility, short-term mean reversion, subject to inventory constraints, firm-wide risk-aversion, and real-world transaction costs. The multi-factor model will be considered as many as possible with main effect on inventory and present an approximate solution, with adjustments of other factors, which to achieve the highest out-of-sample profitability compared with a benchmark strategy.

To replicate and verify the market making strategy by real trading data, we utilize the assumed model proposed by Ho and Stoll (1981) to fit the real-world order book data for underlying parameters, which will be as inputs to numerically simulate the performance of our strategy and compare its Profit and Loss to that of a benchmark strategy. If the agent strategy can result in significant positive profit with the naïve trader, we can conclude the model is feasible using timely data.

Rest of the report is structured as follows: Section 2 discusses the literature reviewed for this project, Section 3 elaborates how to get parameters, process of doing simulations and the preliminary results, Section 4 discusses the dataset, Section 5 discusses the theoretical results from simulated performances comparing two models with benchmark, Section 6 elaborates the back-test results using real world bitcoin market data to compare with section 5.

2. Literature Review

To investigate the algorithmic market making, we replicate optimal limit order placement models under a stochastic control scheme described in *Dealing with the Inventory Risk: A solution to the market making problem* by Olivier Guéant, Charles-Albert Lehalle and Joaquin Fernandez-Tapia. The paper studies the optimal submission strategies of bid and ask orders in limit order books, where dealers in securities markets provide liquidity on the exchange by quoting bid and ask prices that they are willing to buy and sell a specific quantity of assets.

Market makers must constantly set bid and ask quotes. Therefore, they face a complex optimization problem with regards to return and at the same time bear several kinds of risk. According to Avellaneda and Stoikov, the two most often addressed risk facing the dealer are (i) the inventory risk arising from uncertainty in the asset's value and (ii) the asymmetric information risk arising from informed traders. In this paper, we will focus on the inventory effect.

The model derives optimal bid and ask quotes around the 'true' price of the asset to account for the effect of the inventory, where the problem of dealers under competition is analyzed and it assumes that our agent is but one player in the market and the 'true' price is given by the market mid-price.

The approach is to combine the utility framework of the Ho and Stoll approach with microstructure of actual limit order books. The optimal bid and ask quotes are derived in an intuition of two steps. First, the dealer computes a personal indifference valuation for the stock given current inventory. Second, we calibrate the bid and ask quotes to the limit order book by considering the probability with which the quotes will be executed as a function of their distance from the mid-price.

The main building blocks of the model describe the dynamics of the mid-market price, the dealer's utility objective and arrival rate of orders as a function of the distance to the mid-price. Based on these, we can be presented an approximate solution and be able to numerically simulate the performance of the strategy and compare it with our benchmark.

2.1 The mid-price of the stock

For simplicity, model assumes money market pays no interest and mid-price of stock evolves as an arithmetic Brownian motion:

$$\mathrm{d}S_t = \sigma dW_t$$

where W_t is a standard one-dimensional Brownian motion and σ is constant.

Obviously, arrival rates depend on the prices S_t^b S_t^a, which are bid and ask price respectively. Inventory q, which is the quantity holds are:

$$\mathrm{q} = N_t^b - N_t^a$$

N_t^b N_t^a are point processes giving the number of shares market maker respectively bought and sold. The difference between the quoted prices and reference price are

$$\delta_t^b = S_t - S_t^b \,; \delta_t^a = S_t^a - S_t$$

2.2 The trading intensity

If we trade continuously, model focuses on Poisson intensity λ with which a limit order will be executed as function of its distance δ. The paper suggests that λ should decay as an exponential or power law function.

$$\lambda^b(\delta^b) = Ae^{-k\delta^b} = Ae^{-k(s-s^b)}$$

$$\lambda^a(\delta^a) = Ae^{-k\delta^a} = Ae^{-k(s^a-s)}$$

Where A and k are positive constants that characterize the liquidity of the stock. Consequently, the market maker has an amount of cash:

$$\mathrm{d}X_t = (S_t + \delta_t^a)\mathrm{d}N_t^a - (S_t - \delta_t^b)\mathrm{d}N_t^b$$

2.3 Limit orders

We add a bound Q to the inventory that a market maker is authorized to have. In other words, we assume that a market maker with inventory Q (will never set a bid quote and symmetrically that a market maker with inventory −Q, that is a short position of Q shares in the stock under consideration, will never set an ask quote. In line with model, we will focus on CARA utility functions and we suppose that the market maker optimizes:

$$\sup_{(\delta_t^a)_t, (\delta_t^b)_t \in \mathrm{A}} E\left[-\exp\left(-\gamma\left(X_T + q_T S_T\right)\right)\right]$$

2.4 Solution

Resulting bid-ask spread quoted by market maker is given as:

2.4.1 Model one:

$$\delta_{\infty}^{b*}(t,q) = \frac{1}{\gamma}\ln\left(1+\frac{\gamma}{k}\right) + \frac{2q+1}{2}\sqrt{\frac{\sigma^2\gamma}{2kA}\left(1+\frac{\gamma}{k}\right)^{1+\frac{k}{\gamma}}}$$

$$\delta_{\infty}^{a*}(t,q) = \frac{1}{\gamma}\ln\left(1+\frac{\gamma}{k}\right) - \frac{2q-1}{2}\sqrt{\frac{\sigma^2\gamma}{2kA}\left(1+\frac{\gamma}{k}\right)^{1+\frac{k}{\gamma}}}$$

$$\psi_{\infty}^{*}(t,q) \simeq \frac{2}{\gamma}\ln\left(1+\frac{\gamma}{k}\right) + \sqrt{\frac{\sigma^2\gamma}{2kA}\left(1+\frac{\gamma}{k}\right)^{1+\frac{k}{\gamma}}}$$

2.4.2 Model two:

$$r(s,t) = s - q\gamma\sigma^2(T-t)$$

$$\psi(t) = \gamma\sigma^2(T-t) + \frac{2}{\gamma}\ln\left(1+\frac{\gamma}{k}\right)$$

$$\delta^a = r + \psi/2 - s$$

$$\delta^b = s - (r - \psi/2)$$

3. Methodology

3.1 Fit the Empirical Formula

The first step of our research is to fit the empirical formula:

$$\lambda^b(\delta^b) = Ae^{-k\delta^b} = Ae^{-k(s - s^b)}$$

and

$$\lambda^a(\delta^a) = Ae^{-k\delta^a} = Ae^{-k(s^a - s)}$$

These two formulas are the most important in our back-test because they connect the theoretical model with practical data. So, the key is to estimate the value of A and k in the empirical formulas.

Here, to estimate the parameters A and k, we take the log-transform of the formulas:

$$ln(\lambda^b) = ln(A) - k \cdot (s - s^b)$$

and

$$ln(\lambda^a) = ln(A) - k \cdot (s^a - s)$$

Next, we just need to run a linear regression to get the estimated value of $ln(A)$ and k. And for simplicity, we assume the log of the estimated A equals the estimated $ln(A)$ (it is not always exact). Then, we get the estimated A and k.

3.2 Theoretical Test Procedure

The solution formula of the bid spread and ask spread for our theoretical model show us how to use the current amount of inventory to set the bid spread and ask spread, while these two empirical formula determine how the bid spread and ask spread influence the order books (namely influence the amount of inventory). It is like an iteration.

To illustrate this in detail, we have the following steps in our simulation:
a) Set our initial assets (amount of inventory) q_0;
b) For t = 0, 1, 2, …, T, input current inventory q_t the into solution formulas to get the optimal bid price s^b and ask price s^a for the next time period;
c) Input the bid price s^b and ask price s^a into the empirical formulas to get the density of the buying order process and selling order process, which are both Poisson processes.
d) Use the densities of the Poisson processes in step c) to generate two random numbers as the selling order book and the buying order book respectively.
e) Repeat step b) to step d) for t = 0, 1, 2, …, T
f) Here we get the order books and bid spreads and ask spreads during the whole time period from time 0 to T. Finally, we can calculate the evolution of the P&L and

the CARA utility curve.

3.3 Extensions of the model

Fitting 2 extensions of the model and back-testing on them.

a) Adding a drift in the price dynamics

$$dS_t = \mu dt + \sigma dW_t$$

We fit the solution of optimal quotes and the bid-ask spread with drift:

$$\delta_\infty^{b*}(q) \simeq \frac{1}{\gamma}\ln\left(1+\frac{\gamma}{k}\right) + \left[-\frac{\mu}{\gamma\sigma^2} + \frac{2q+1}{2}\right]\sqrt{\frac{\sigma^2\gamma}{2kA}\left(1+\frac{\gamma}{k}\right)^{1+\frac{k}{\gamma}}}$$

$$\delta_\infty^{a*}(q) \simeq \frac{1}{\gamma}\ln\left(1+\frac{\gamma}{k}\right) + \left[\frac{\mu}{\gamma\sigma^2} - \frac{2q-1}{2}\right]\sqrt{\frac{\sigma^2\gamma}{2kA}\left(1+\frac{\gamma}{k}\right)^{1+\frac{k}{\gamma}}}$$

$$\psi_\infty^{*}(q) \simeq \frac{2}{\gamma}\ln\left(1+\frac{\gamma}{k}\right) + \sqrt{\frac{\sigma^2\gamma}{2kA}\left(1+\frac{\gamma}{k}\right)^{1+\frac{k}{\gamma}}}$$

b) Adding the market impact that is regarded as adverse selection

$$dS_t = \sigma dW_t + \xi dN_t^a - \xi dN_t^b, \qquad \xi > 0$$

We fit the solution of optimal quotes and the bid-ask spread with market impact:

$$s - s^{b*}(t,q,s) = \delta^{b*}(t,q) = \frac{1}{k}\ln\left(\frac{v_q(t)}{v_{q+1}(t)}\right) + \frac{\xi}{2} + \frac{1}{\gamma}\ln\left(1+\frac{\gamma}{k}\right)$$

$$s^{a*}(t,q,s) - s = \delta^{a*}(t,q) = \frac{1}{k}\ln\left(\frac{v_q(t)}{v_{q-1}(t)}\right) + \frac{\xi}{2} + \frac{1}{\gamma}\ln\left(1+\frac{\gamma}{k}\right)$$

$$\psi^*(t,q) = -\frac{1}{k}\ln\left(\frac{v_{q+1}(t)v_{q-1}(t)}{v_q(t)^2}\right) + \xi + \frac{2}{\gamma}\ln\left(1+\frac{\gamma}{k}\right)$$

4. Data

We are able to access three types of data – public trades market, prices from coinmarketcap and order books. Each of the data is given a time, to the nearest second. For instance, the binance BTC/USDT market, we are able to see the time, the side, buy or sell, amount and price. The following is the sample format:

[Sun May 12 19:04:08 2019]: side: buy, price: 6993.0, amount: 0.2441

For the coinmarketcap data, we have the time and price. The following is the sample format:

[Sun May 12 18:42:48 2019]: 7043.72254532

For the order book, we have the time for the ask/bid prices and the corresponding orders.
The following is the sample format:

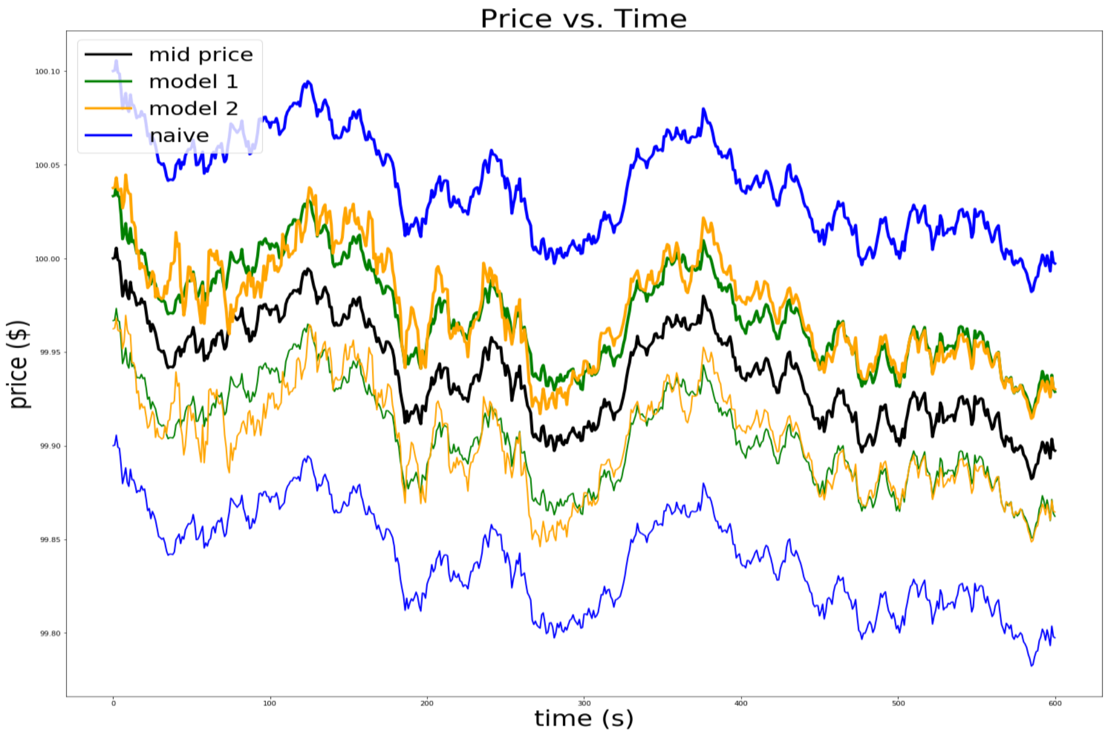
[Sun May 12 18:40:08 2019]:
{ 'asks': { '7059.45': 0.117423,
...},
'bids': { '7007.8': 5.0,
...}}

The order book is the most important data for ask, we are able to know the sequence, amount and prices of the trades, where we can fit the para meters.

5. Theoretical Results

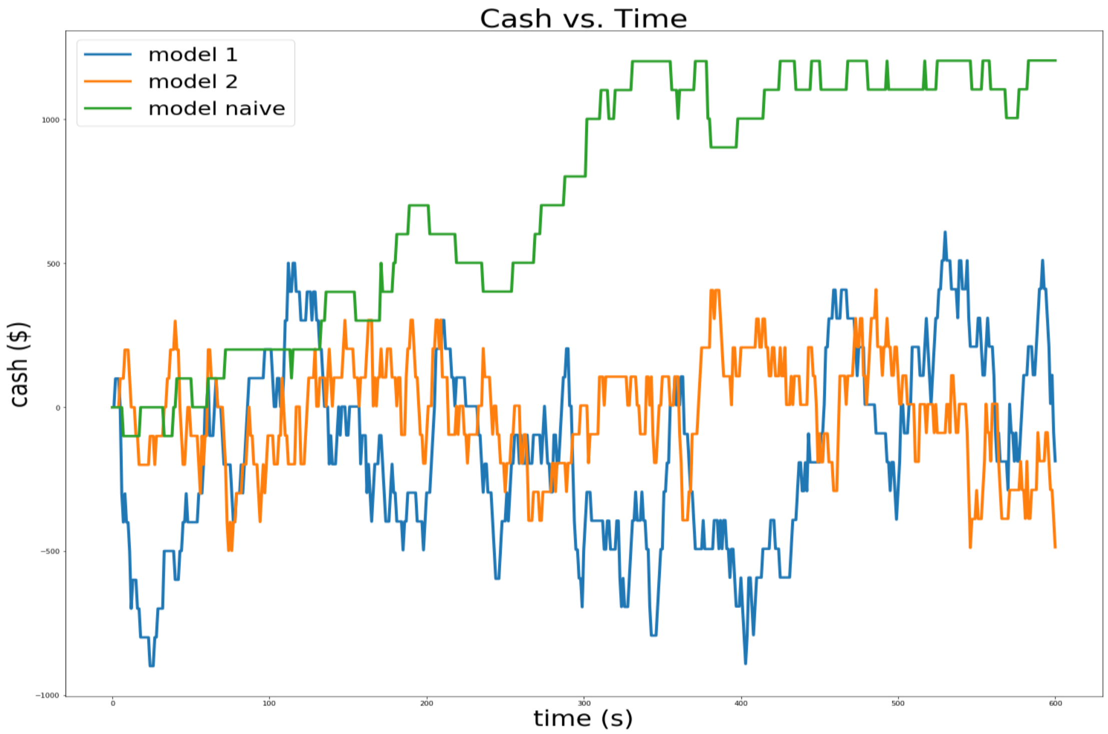
5.1 Comparisons of Three Models

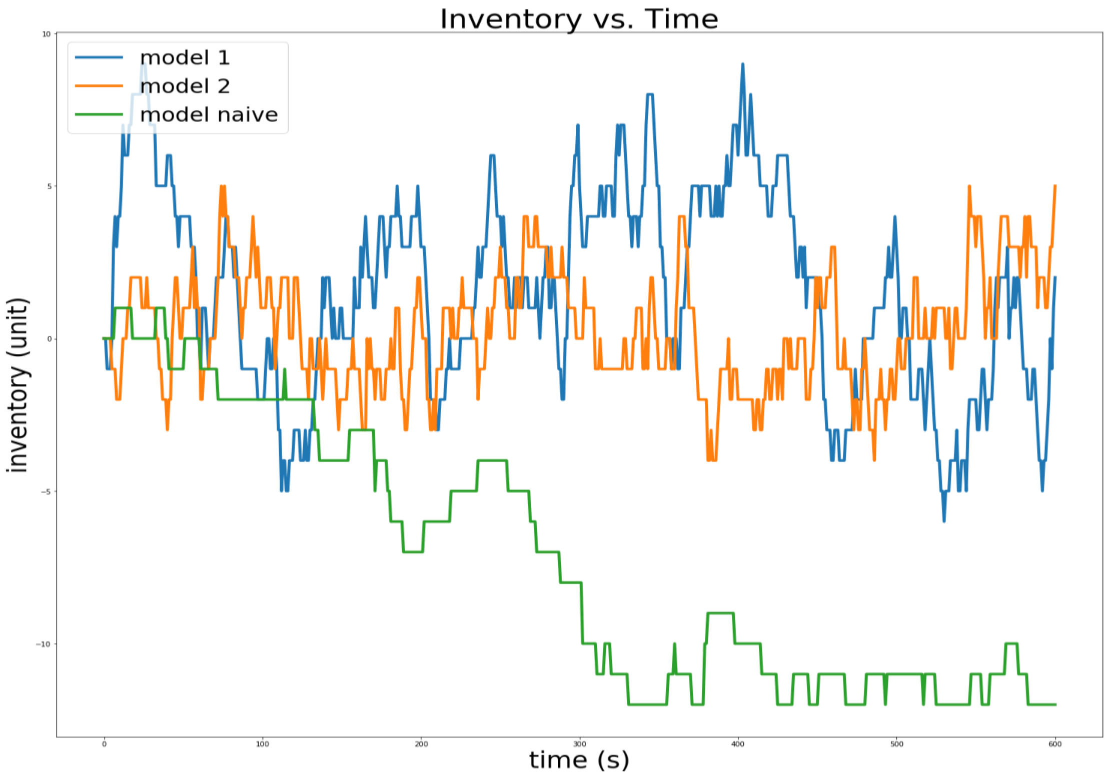
5.1.1 Mid price, bid prices and ask prices



We can see here three models shared the same mid price. The bid and ask prices of model 1 and model 2 are very similar.

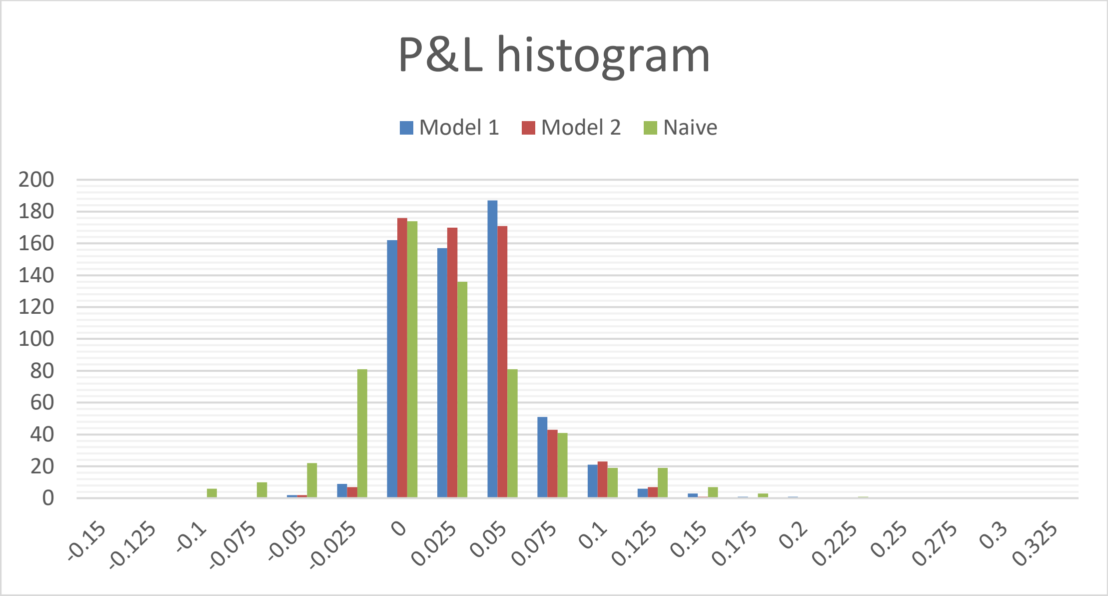
5.1.2 Cash and inventory position





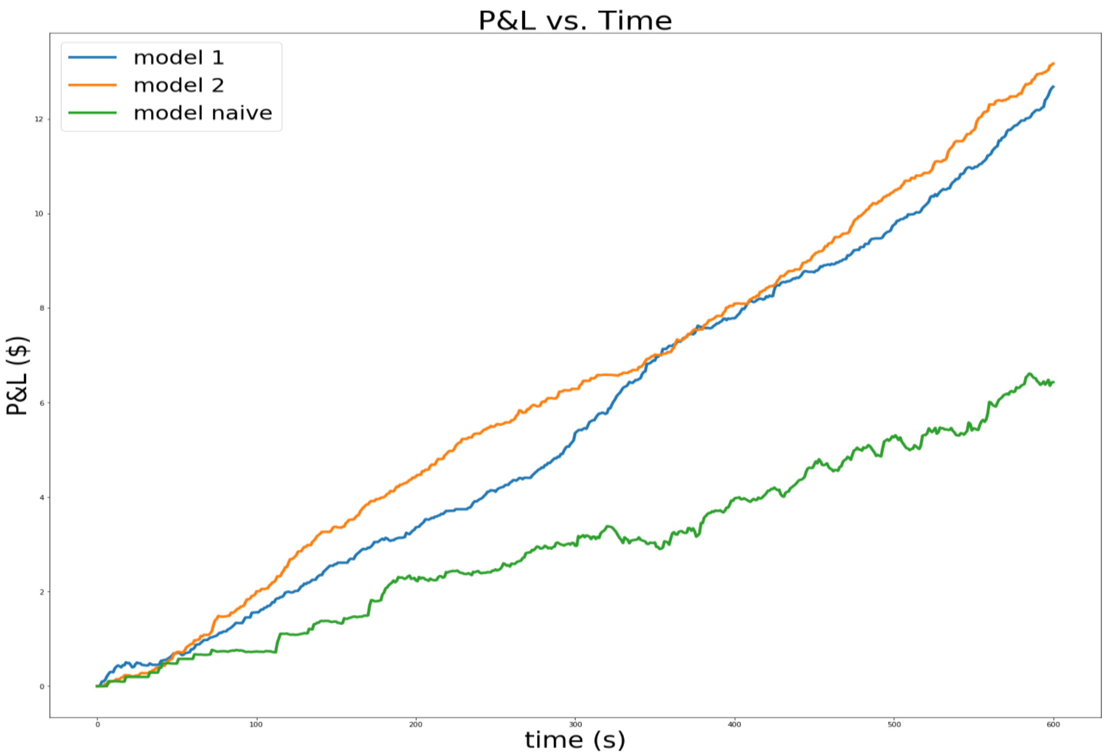
During the trading period, the naïve trader was keep selling inventory while the other two models did not have an obvious trading direction.

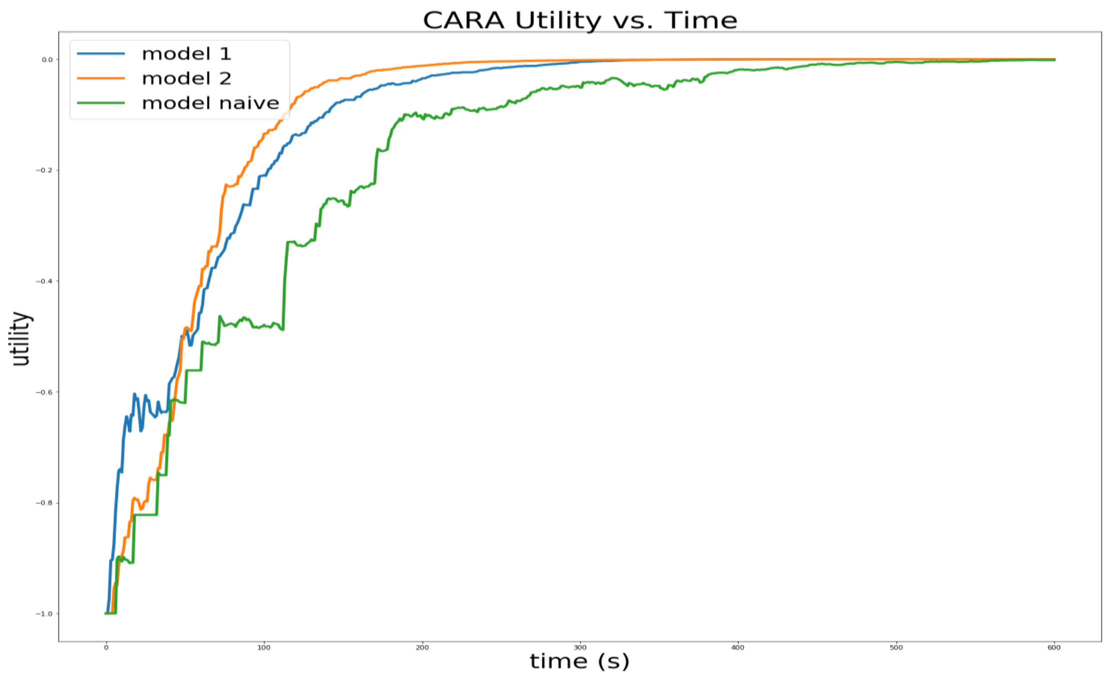
5.1.3 The histogram of P&L



The naïve trader has a much larger tail in the negative part compared to the two models.

5.1.4 Cumulative P&L and utility



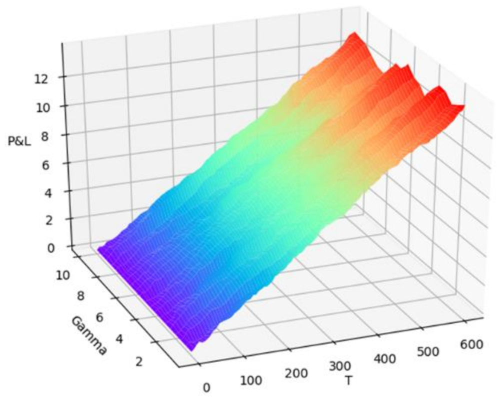


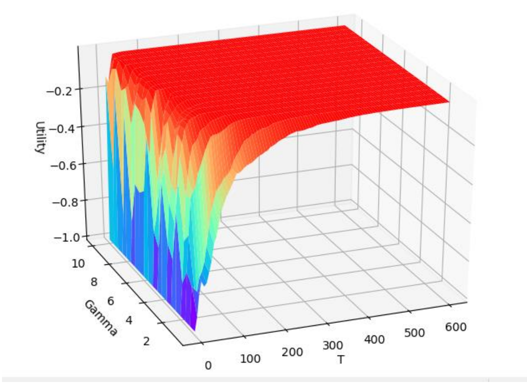
The cumulative profits of the two optimal models are almost twice of the naïve trader's.

5.2 The Impact of the Choice of Gamma

As we mentioned in the previous section, Gamma is a parameter in the formula of the CARA utility function. Since utility curves are kind of subjective, the choice of Gamma is also not clearly determined.

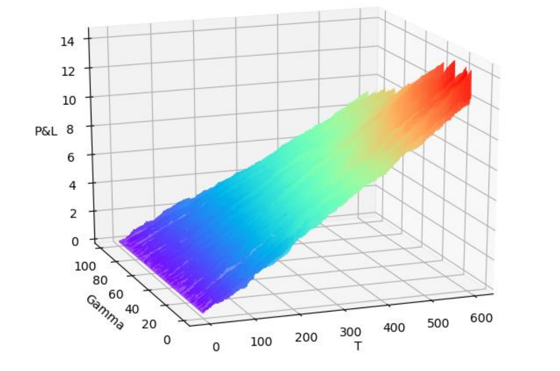
In this section, we tried a range of values of Gamma and explored the related impact of these different values.

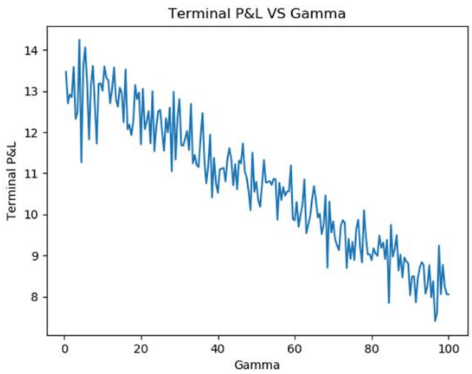




The above two pictures showed that within the range [1, 10], different values of Gamma did not make an obvious difference for any time T.

Then, we took a longer range: [1:100]. It made a difference. The larger the Gamma, the smaller the P&L. Specifically, we took the terminal P&L and showed the graph.

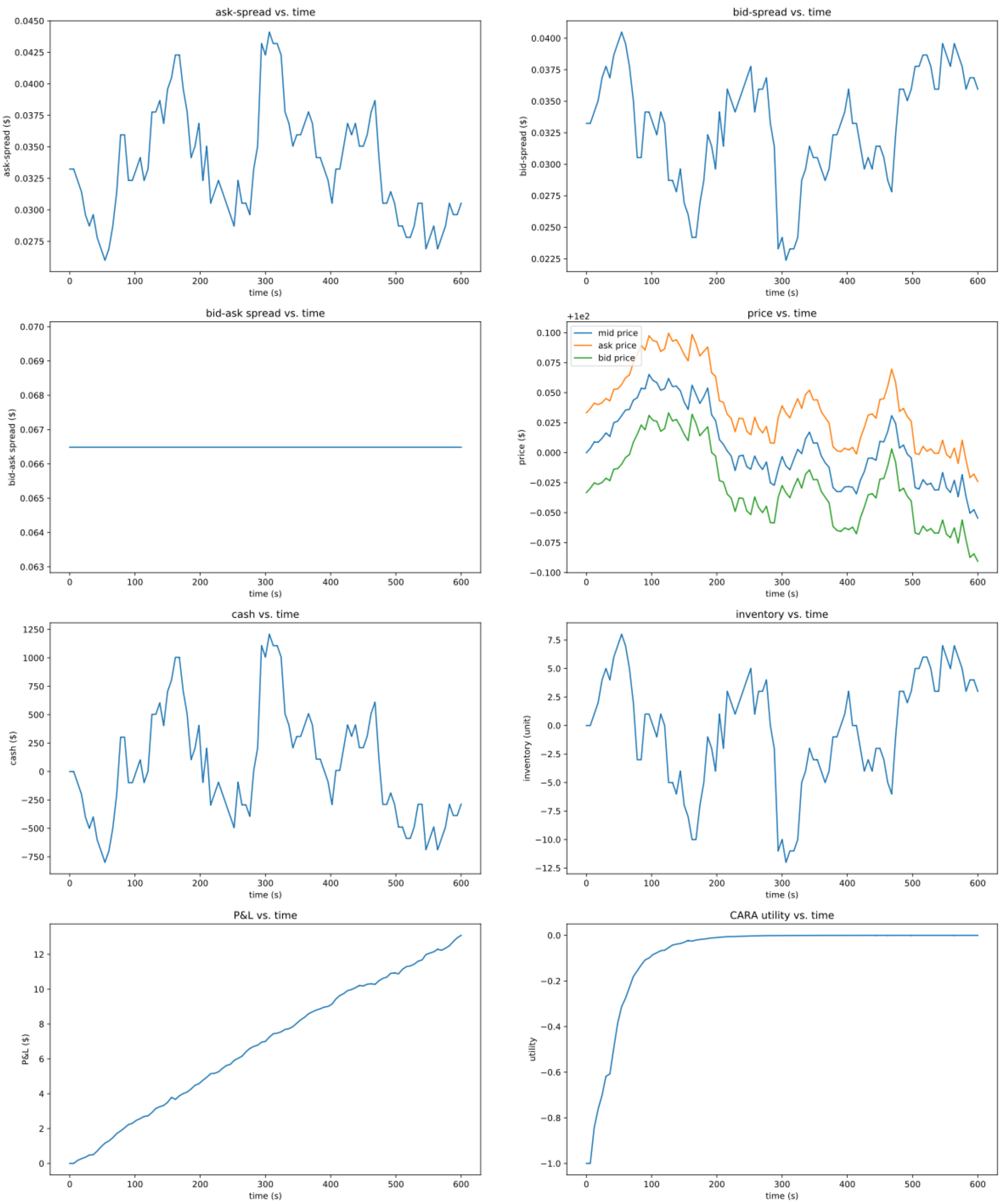




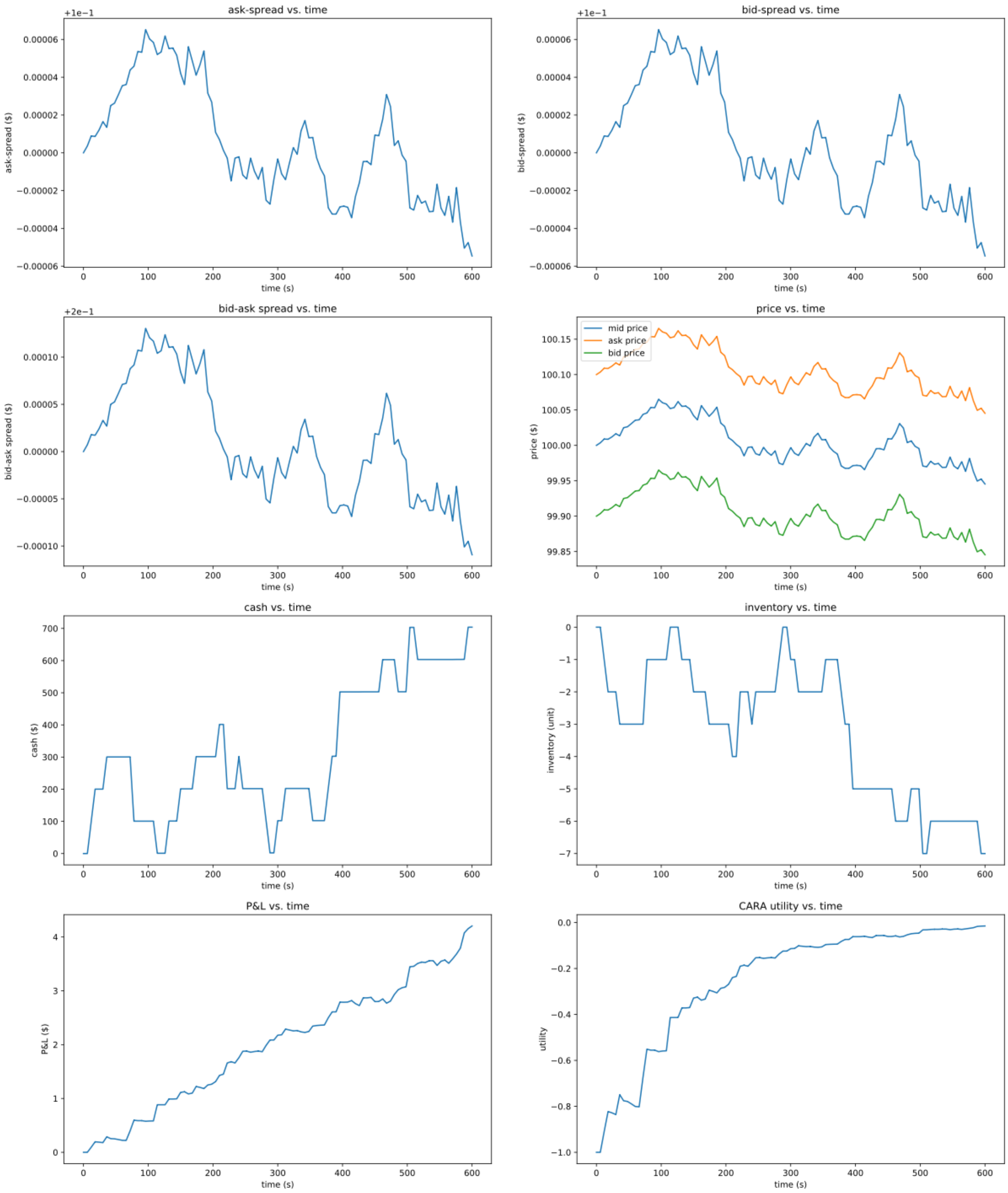
In the 2D graph, it was very clear that the overall correlation of Gamma and P&L should be negative though it had some noise. However, in the interval [1,10], the choice of Gamma was basically not related with P&L. And the practical choice of Gamma is usually in this interval. So, in the latter discussion, we just assumed Gamma = 1 for simplicity.

5.3 The Case of Trend in the Price Dynamics

Take into account the trend in the price dynamics and repeat previous steps.
The results of optimal strategy: μ = 5\$/year



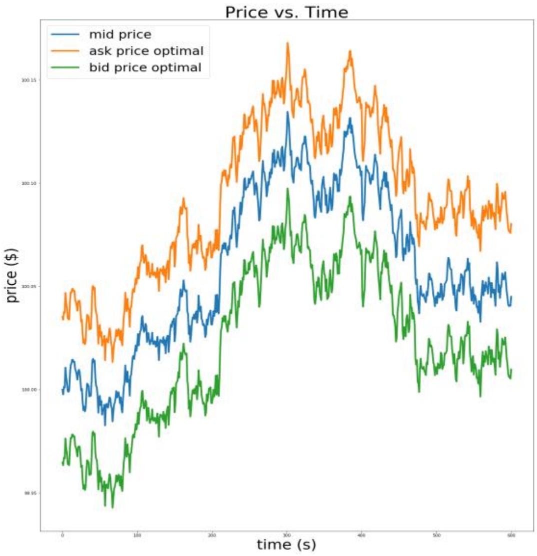
Compare with naïve strategy:

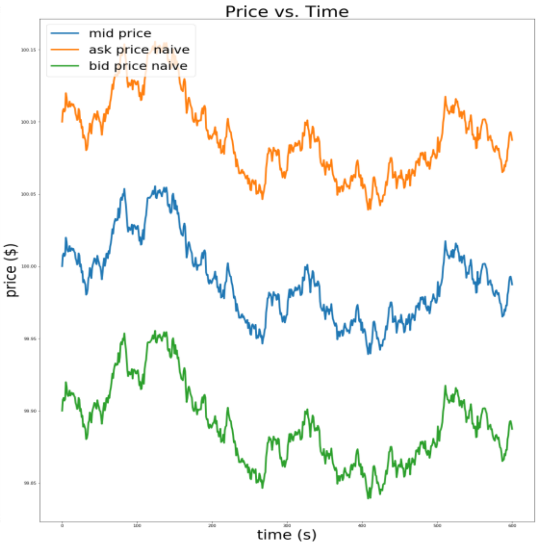


The overall results are very similar to the random walk price model.

5.4 The Case of Market Impact

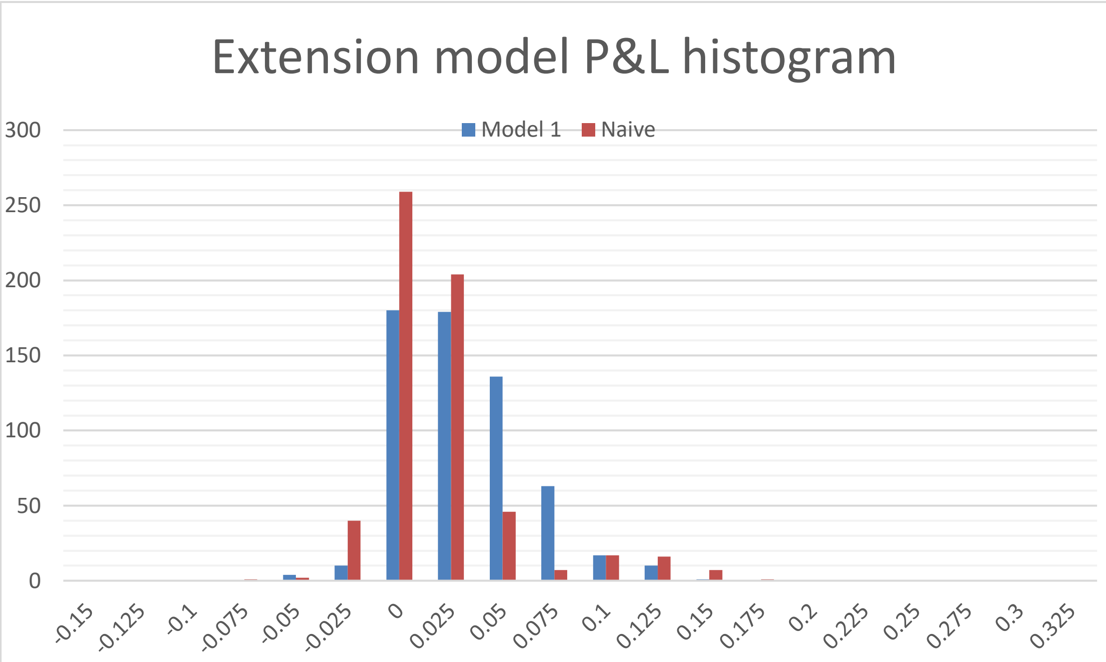
5.4.1 Mid prices, bid prices and ask prices



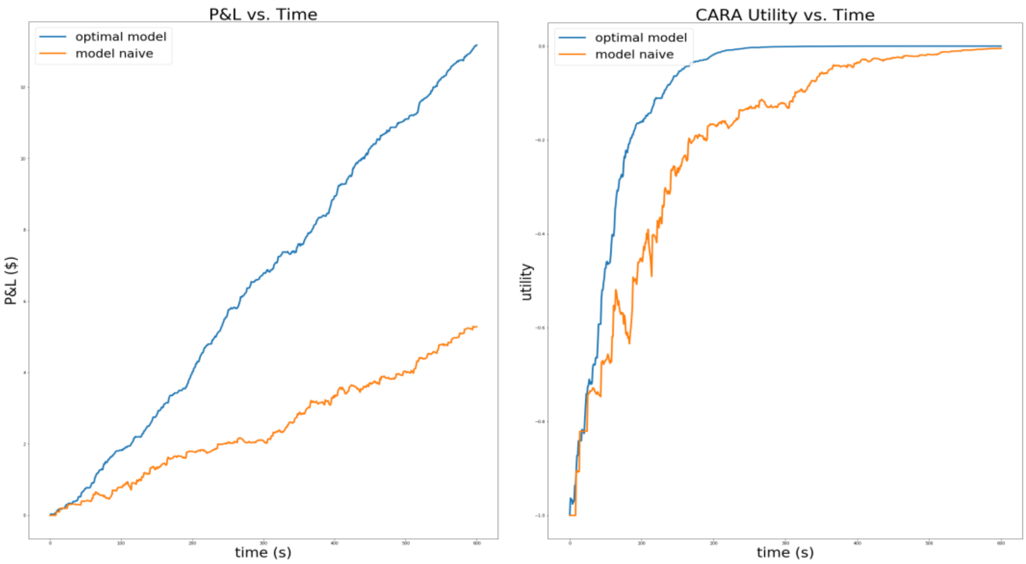


Here, even the mid prices of naïve and optimal trader are different because different strategies influenced the dynamics of the mid-price itself.

5.4.2 Histogram of P&L



5.4.3 Cumulative P&L and utility



The cumulative profits of the optimal model are almost three times of the naïve trader’s.

6. Back-Test Results

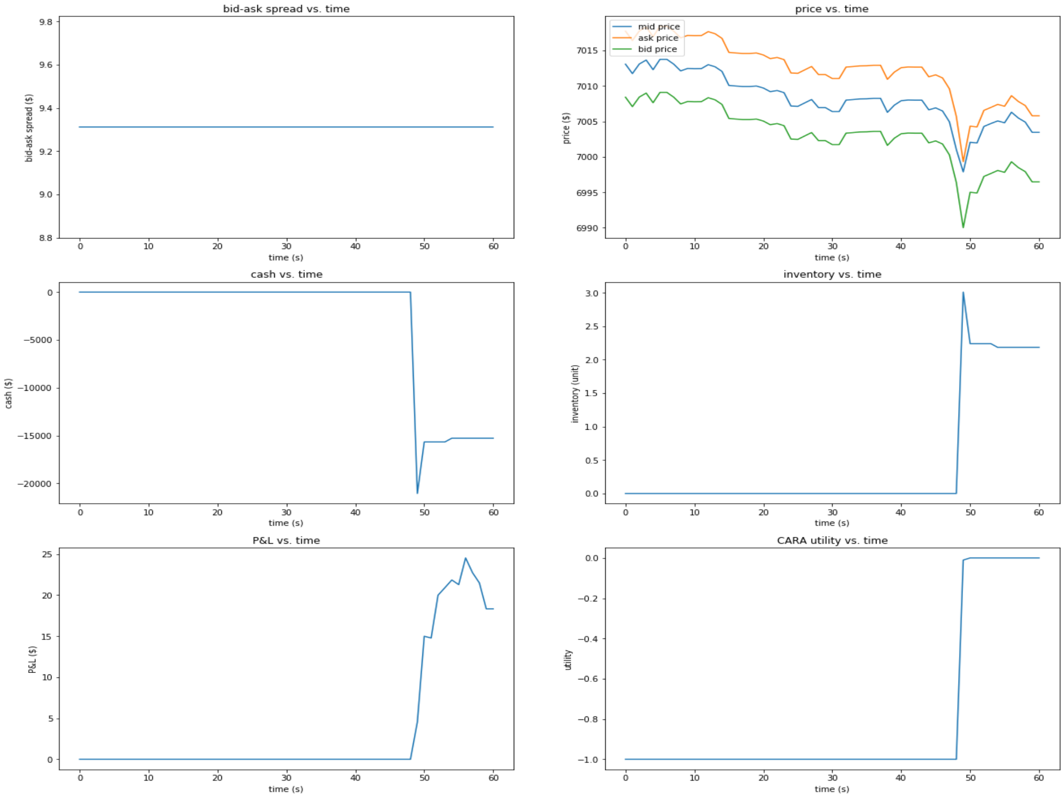
6.1 Procedure

The solution formula of the bid spread and ask spread for our theoretical model show us how to use the current amount of inventory to set the bid spread and ask spread, while these two empirical formula determine how the bid spread and ask spread influence the order books (namely influence the amount of inventory). It is like an iteration. The overall procedure is very similar to the theoretical test. But here, unlike the in theoretical test, we are using the empirical trading data of cryptocurrency to determine whether to accept or reject the order sent to us: If the empirical bid or ask spread is larger than our optimal spread, we accept the order. Otherwise, we reject it.

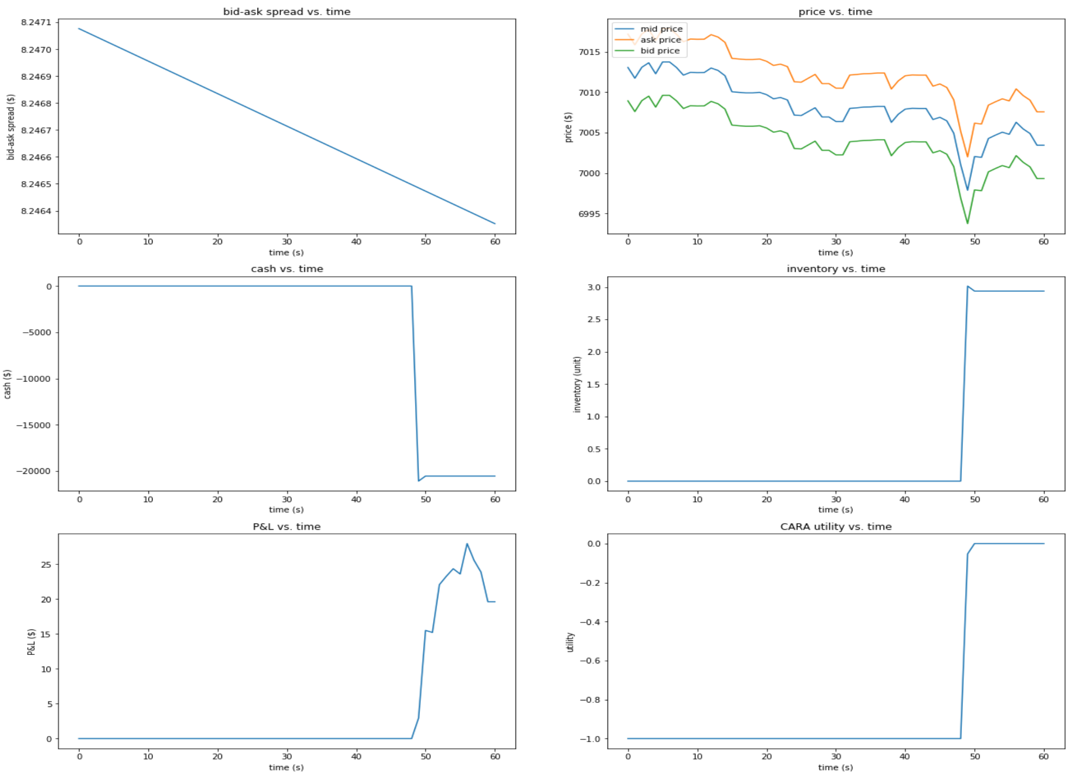
To illustrate this in detail, we have the following steps in our simulation:
a) Set our initial assets (amount of inventory) q_0;
b) For t = 0, 1, 2, …, T, input current inventory q_t the into solution formulas to get the optimal bid price s^b and ask price s^a for the next time period;
c) Compare with the empirical bid & ask price of order book data in the last second to determine whether to accept or reject the orders;
d) Update the cash and the inventory corresponding to the transactions in the last second;
e) Repeat step b) to step d) for t = 0, 1, 2, …, T
f) Here we get the order books and bid spreads and ask spreads during the whole time period from time 0 to T. Finally, we can calculate the evolution of the P&L and the CARA utility curve.

6.2 Comparison of the three models

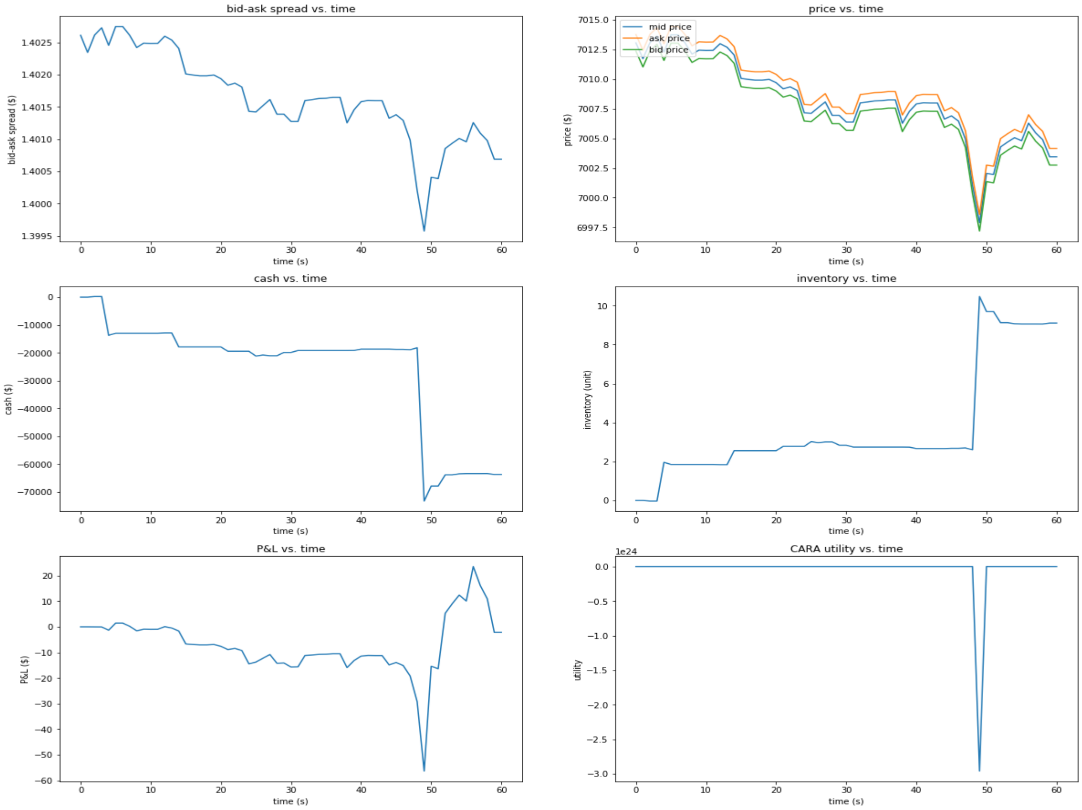
6.2.1 Model one



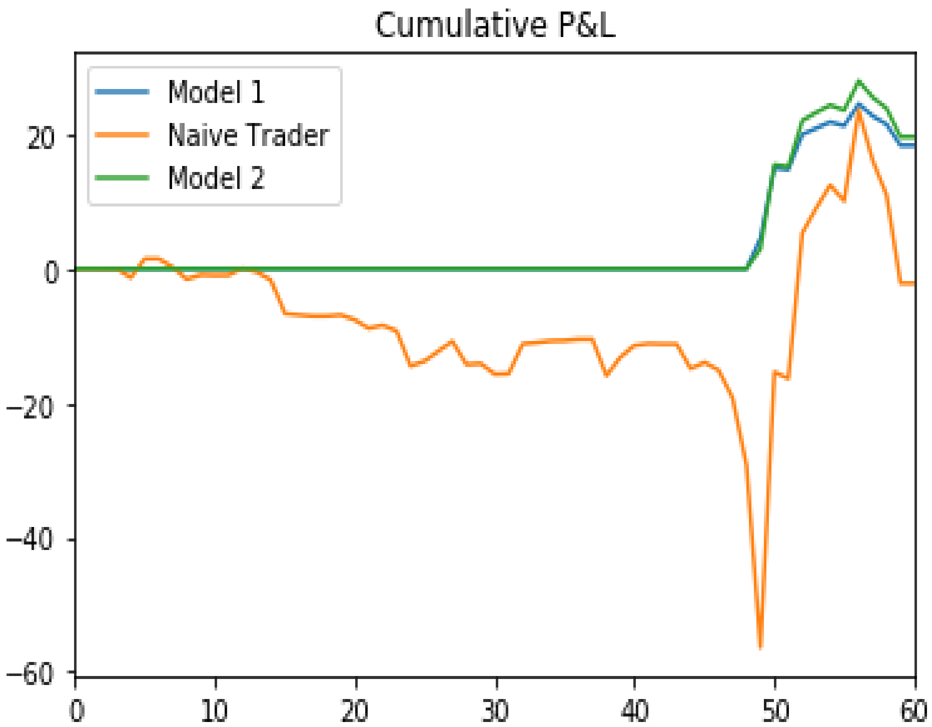
6.2.2 Model two



6.2.3 Naïve model



6.2.4 P&L



We can see there is a price shock at t = 50, which is followed by an immediate recover of price. Thus, there are several kinks in the corresponding parts of the cash graph, inventory graph and P&L graph.

Model one and model two basically did no transactions before the price shock. They rejected all the orders sent to them, which actually protected them from losing money. By contrast, the naïve trader did some transactions and this lead to the huge loss as is shown in the above graph. Although the recover of profit for naïve trader after the price shock was also huge, the profit dropped again in the next cycle. So, we can say our two optimal models are much more stable than the naïve trader in the back test with the empirical high frequency trading data of the cryptocurrency.

7. Conclusions and Future Improvements

7.1 Conclusions

a) The choice of risk-aversion factor Gamma will not influence the test results;
b) Overall, model 1 and model 2 have similar performances in both theoretical results and back test performances;
c) In theoretical test, all of the three models have increasing profits and the profits of model 1 or 2 are twice of the naïve trader;
d) Taking market impact into consideration, the profits of model 1 or 2 are three times of the naïve trader;
e) In back test of cryptocurrency data, model 1 and model 2 still have a larger expected profit than the naïve trader though it is not as obvious as that in the theoretical results. More importantly, model 1 and model 2 have much smaller variances than the naïve trader.

7.2 Future Improvements

a) Take the seasonality into consideration. The empirical formulas we mentioned in 3.1 are derived from econophysics literature. And the model might not always hold. So, to increase the feasibility in the real world, we can use a rolling window to dynamically update the parameters in the empirical formulas;
b) As we mentioned in 3.1, for simplicity, we assume the ln of the estimated A equals the estimated $ln(A)$. But this is only exactly right in some special cases. So, for accuracy, instead of using the log-transform to run a linear regression, we can use the initial formula directly to find the parameter A and k, which can minimize the mean squared error.

8. Reference

1. Olivier Guéant, Charles-Albert Lehalle, Joaquin Fernandez-Tapia: Dealing with the Inventory Risk: A solution to the market making problem. Mathematics and Financial Economics, Springer, **7** (4), 477-507 (2013)
2. Marco Avellaneda, Sasha Stoikov: High-frequency trading in a limit order book. Quantitative Finance, **8** (3), 217–224 (2008)
3. Jean-Philippe Bouchaud, Marc Mézard, Marc Potters: Statistical properties of stock order books: empirical results and models. Quantitative Finance, **2** (4), 251-256 (2002)
4. Bruno Biais, Larry Glosten, Chester Spatt: Market Microstructure: a survey of microfoundations, empirical results, and policy implications. Journal of Financial Markets, **8** (2), 217–264 (2005)
5. Charles-Albert Lehalle, Othmane Mounjid: Limit Order Strategic Placement with Adverse Selection Risk and the Role of Latency. Market Microstructure and Liquidity, **3** (1), 1750009 (2017)
6. Fabien Guilbaud, Huyên Pham: Optimal high frequency trading with limit and market orders. Quantitative Finance, **13** (1), 79-94 (2013)
7. Takahiro Fushimi, Christian González Rojas, Molly Herman: Optimal High-Frequency Market Making. Stanford (2018)
8. Olivier Guéant: Optimal market making. Applied Mathematical Finance, **24** (2), 112-154 (2017)
9. Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. (2008) https://bitcoin.org/bitcoin.pdf?