PREDICTING THE VOLATILITY OF STOCK PRICE MOVEMENTS

Xiangmin Shen, Michael Zhu

Northwestern University

1. INTRODUCTION AND PRACTICAL UTILITY

Market making at market-close is a popular trading strategy. A market maker is someone who quotes both a buy (bid) and a sell (ask) price in a financial instrument, such as stocks, hoping to make a profit on the bid-ask spread. To quote a price is to post a price, and the bid-ask spread is the difference between the prices quoted for an immediate sale and an immediate purchase of a financial instrument. A market maker tends to be most active during market-close because there is a spike in liquidity during market-close. Market liquidity of a financial instrument refers to how rapidly the instrument can be bought or sold, which is usually measured by its trading volume. A liquid market would actively move back-and-forth and thus provide more opportunities for a market maker to capitalize.

A key decision for a market maker to decide is how far they would like to quote above and below the market. Suppose that the current price of a financial instrument is x. To quote a above the market is to quote a sell (ask) price of $x+a$. Similarly, to quote b below the market is to quote a buy (bid) price of $x-b$. In other words, by quoting a above the market and b below the market, we hope to buy low at $x-b$ and sell high at $x+a$. For each successful trade, we would make a profit of $(x+a)-(x-b)=a+b$. The further we quote above and below the market (the larger the values of a and b), the more profit we would capitalize on each successful trade. However, the likelihood of success decreases as we quote further away from the market. For example, Google currently trades at \$1,750 per share. If we quote \$500 both above and below the market, surely we would make \$1,000 profit on each successful trade, but in reality, we would not make any successful trades at all because no one is willing to buy Google at \$1,250 per share and sell Google at \$2,250 per share. In contrast, if we quote \$1 both above and below the market, then our trades are likely to be successful, but we would only make a small profit of \$2 from each successful trade. In other words, we need to keep a balance between profit per trade and likely hood of success when we make markets.

When we make markets, it would be ideal if we could set our quoting range according to the then-current market condition. If we believe that prices are going to move rapidly, then we could quote far away from the market to maximize our gains per trade and still achieve a decent success rate. On the contrary, if we believe that prices are going to move slowly, then we could quote near the market to make sure that we could still achieve a decent success rate. Therefore, it is very helpful for a market maker to be able to forecast the market condition, at least to a certain degree.

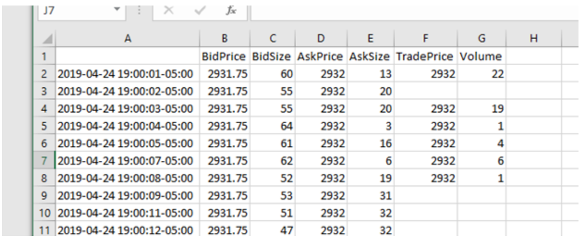
2. PROBLEM STATEMENT

For our final project, we would like to predict the volatility of the S&P 500 index during market-close. The ticker for the S&P 500 index is ES, so we shall refer to it as ES from this point onward. A ticker is an abbreviation of the name of a financial instrument. We chose to study ES because it is one of the most actively traded financial instruments globally, and in real life, many market makers participate during its market close.

The closing price for ES is released at 15:00:00 CST every trading day. The closing price is used as the baseline for market making. For example, to quote \$5 above the market is to quote \$5 above the closing price. The closing price is calculated based on the volume-weighted average price of all transactions during the interval 14:59:30 CST and 15:00:00 CST. Consequently, the market price for ES at 15:00:00 CST is almost always different from the closing price. For our final project, the volatility of ES during market-close shall be defined as the absolute difference between the closing price and the market price at 15:00:00 CST.

3. DATA SOURCE

We have already obtained S&P 500 trading data from 04/24/2019 to 05/08/2019. The data set is stored in a Microsoft Excel file, which includes bid prices, bid sizes, ask prices, ask sizes, trade prices and volume on every second. We choose this data set because a two-week period is about the right size to start with. If the data set spans a long period of time, it may introduce unnecessary noise to the data set. On the other hand, the the data set spans a short period of time, it may not contain enough information to let the model learn the trend. We can certainly acquire more data if necessary. An example of our data set is shown below.



	BidPrice	BidSize	AskPrice	AskSize	TradePrice	Volume
2019-04-24 19:00:01-05:00	2931.75	60	2932	13	2932	22
2019-04-24 19:00:02-05:00	2931.75	55	2932	20		
2019-04-24 19:00:03-05:00	2931.75	55	2932	20	2932	19
2019-04-24 19:00:04-05:00	2931.75	64	2932	3	2932	1
2019-04-24 19:00:05-05:00	2931.75	61	2932	16	2932	4
2019-04-24 19:00:07-05:00	2931.75	62	2932	6	2932	6
2019-04-24 19:00:08-05:00	2931.75	52	2932	19	2932	1
2019-04-24 19:00:09-05:00	2931.75	53	2932	31		
2019-04-24 19:00:11-05:00	2931.75	51	2932	32		
2019-04-24 19:00:12-05:00	2931.75	47	2932	32		

Fig. 1. Dataset

4. METHODS AND EVALUATION

We will use PyTorch to build and train our neural network. The neural network we will develop is like the ones in the assignments, but more complicated. Although we haven't finalized the structure of our models, the process will be similar to what we did in the assignments, so we are confident we can use PyTorch.

We will first train a fully-connected neural network that uses the volatility calculated at the end of each minute for the 30 minutes before the market close. In other words, the inputs of this fully-connected neural network are volatility at 14:30:00 CST, 14:31:00 CST, ..., 14:59:00 CST. This neural network serves as our baseline. This is a reasonable baseline because this is the most straightforward neural network we can build. If the other models we build have better performance than this neural network, we will feel confident that the complicated components we add to the models help the models learn better.

For each trading day we have in the data set, a model shall predict the volatility at market close. We will compare the predicted volatility to the actual observed volatility. Therefore, for each data entry, the error is calculated by the absolute difference between the predicted volatility and the actual observed volatility. We will average this over the entire data set to get a mean error. The mean error shall be used to evaluate the performance of our models.

We only have two team members, so we will each develop a neural network to predict the volatility of market price and have a competition between our models to see whose model has better performance. Before that, we will read a few related papers and discuss what techniques and neural network structures are appropriate for predicting the volatility.

The table below outlines our milestones and expected completion dates.

Milestone	Group Member(s)	Due Date
Data Collection	Michael	11/12
Building the Baseline Model	Xiangmin	11/20
Testing the Baseline Model	Michael	11/21
Finalized Model Structure	Michael and Xiangmin	11/27
Building Our Models	Michael and Xiangmin	12/04
Testing Our Models	Michael and Xiangmin	12/05
Project Report	Michael	12/10
Project Website	Xiangmin	12/10

5. REFERENCES

[1] Can Yang, Junjie Zhai, and Guihua Tao, "Deep learning for price movement prediction using convolutional neural network and long short-term memory," *Mathematical Problems in Engineering*, vol. 2020, pp. 1–13, 07 2020, Available at https://doi.org/10.1155/2020/2746845.

[2] Yakup Kara, Melek Acar Boyacioglu, and Ömer Kaan Baykan, "Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the istanbul stock exchange," *Expert Systems with Applications*, vol. 38, no. 5, pp. 5311 – 5319, 2011, Available at https://doi.org/10.1016/j.eswa.2010.10.027.

[3] J. Ke and X. Liu, "Empirical analysis of optimal hidden neurons in neural network modeling for stock prediction," in *2008 IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application*, 2008, vol. 2, pp. 828–832, Available at https://ieeexplore.ieee.org/abstract/document/4756892.

[4] Erkam Guresen, Gulgun Kayakutlu, and Tugrul U. Daim, "Using artificial neural network models in stock market index prediction," *Expert Systems with Applications*, vol. 38, no. 8, pp. 10389 – 10397, 2011, Available at https://doi.org/10.1016/j.eswa.2011.02.068.

[5] Mustafa Göçken, Mehmet Özçalıcı, Aslı Boru, and Ayşe Tuğba Dosdoğru, "Integrating metaheuristics and artificial neural networks for improved stock price prediction," *Expert Systems with Applications*, vol. 44, pp. 320 – 331, 2016, Available at https://doi.org/10.1016/j.eswa.2015.09.029.