NATIONAL RESEARCH UNIVERSITY HIGHER SCHOOL OF ECONOMICS

FACULTY OF ECONOMIC SCIENCES

BACHELOR'S PROGRAMME ECONOMICS

PROJECT PROPOSAL

MARKET MAKING EFFICIENCY EVALUATION WITH ML

Lev Koteyev, BEC1710

Advisor:
Lukyanchenko Peter P., Teacher
Department of Big data and information search
Language Consultant:
Anikudimova Evgeniya N., Teacher
Department of Foreign Languages

MOSCOW

2021

This research is devoted to the search for an effective model of market making for a professional market participant. The uniqueness of this study is the application of the latest methods of artificial intelligence using old proven pricing models for the trader. We will check how effective it is to use the machine learning capabilities for placing quotes. We will test from simple and clear methods - linear regression, trees to simple neural networks. Similar developments are already underway by the world's largest hedge funds and major liquidity providers. Also unique is a new approach that is extremely rare to find in scientific articles - testing hypotheses by means of a backtest on the full list of orders coming to the exchange, which will allow us to fully reproduce trading at the lowest level. Data in this format is very expensive and only the largest market representatives have it, but it is easier to find it on the Russian stock market. This approach will allow us to get results as close to reality as possible and applicable in practice.

Introduction

Background. For quite a long time in the traditional stock markets, major players have been providing liquidity by engaging in market making. It is their activities that allow private investors to participate in investing in companies with the lowest spread costs - because the greater the competition between market makers, the smaller the spread between buying and selling shares. And accordingly, any private investor can trade on more favorable terms.

Since the end of the 80s, when electronic trading methods began to spread actively in trading and the entire exchange became a large IT hub, serving a large number of requests. There was a problem of constant liquidity for not the most popular shares. If you look at it more deeply, this problem is not just for traders and investors - if the company is not liquid, it will not be interesting for investment, which will limit its access to capital and to growth opportunities. Taking this into account, the exchanges of different countries began to actively encourage market makers and liquidity providers, giving them various privileges. Thanks to this, the traditional capital markets-stocks and bonds-received the necessary liquidity. So many organizations and states have gained access to capital.

However, recently, due to the excess capital and low interest rates, competition in the stock market and in particular in the market making market has increased, many liquidity providers are looking for new improved methods of quoting. Moreover, the emergence of new inefficient markets, in particular the cryptocurrency market and ICO, has set new challenges for market makers. In many places, traditional methods of technical analysis were no longer effective and often lagged behind. New statistical studies in the field of quantitative finance, in particular the Avellaneda-Stoikov model, corrected the situation for a while.

However, the rapid development of computing power and advances in machine learning have presented traditional statistics and quantitative finance with new, non-obvious challenges. There are already quite a few articles on the use of machine learning in trading practice, but very few of them are focused on market making. Moreover, many of them use candlesticks for coding-with a frequency of at best a minute, which is as far from practice as possible. The task of a market maker is largely a task for High frequency trading, for which it is necessary to use more professional tools to make it effective.

In this paper, I will consider the general approaches applied from the point of view of traditional science and used by major players of the Russian stock market with using hft.

Problem Statement. The main research issue is the possibility of using machine learning to solve market maker problems. The purpose of the study is to develop a ready-made model of pricing and bidding for a market maker using machine learning methods.

The main method of validating the result will be a backtest on the full order log of the exchange and comparing it with traditional models.

To achieve these goals, the following action plan is expected:

1) Evaluation and analysis of current research for their practicality and applicability to the market

2) Formation of basic models based on them and their programming for the backtest

3) Creation of a backtest for any models based on Russian market data

4) Testing of hypotheses and various models with artificial intelligence methods for quoting prices and their software implementation

5) Comparison of the baseline and the resulting ML model on the backtest

6) Interpretation of the obtained results and formation of recommendations for ML HFT market making.

Professional significance of this study is that the large players that provide liquidity to the market and perform the functions of a market maker will be able to apply the ideas tested here in their models and more effectively set quotes.

Literature Review

High Frequency Trading in a Limit Order Book Avellaneda-Stoikov

One of the most famous articles related to the tasks of the market maker, which is very much referenced. The authors use the statistical apparatus to suggest the utility function of the market maker, taking into account the risks it faces. the proposed model is focused on minimizing these risks and shows a high stable P&L

The traditional job for a stock dealer is to optimize the submission of bid and ask prices to the order book. The market maker faces inventory risk due to the diffusive nature of the average stock price and operational risk due to the Poisson arrival of market buy and sell orders. After setting the agent's problem within the maximum expected utility, we get the solution in a two-step procedure. First, the dealer calculates a personal indifference score for the stock, given its current holdings. Second, it calibrates its bid and requests quotes in the market's limit order book. We compare this "stock-based" strategy with the "naive" best bid/best ask strategy by simulating stock price trajectories and displaying the P&L profiles of both strategies. We found that our strategy has a P&L profile that has both higher returns and lower variance than the reference strategy. (Avellaneda & Stoikov, 2006).

On Optimal Pricing Model for Multiple Dealers in a Competitive Market

The article is based on previous developments in the framework of modeling for market making, including the work of Avellaneda & Stoikov. However, the authors began to consider the market maker problem from the point of view of a problem from game theory. They looked at this problem not from the point of view of an individual dealer, but from the point of view of the fact that there are many of them and they make decisions, including taking into account what their competitors do.

The main idea of the article is to optimize and build models for all dealers in the market, and not for one specific one. In this way, the author tries to approach a more market situation, when all the players react quite actively to the actions of the other and try to bypass each other. For each of the approaches is whether the model Avellaneda & Stoikov or Ho Stoll only or someone else were held on 1000 simulations for 7 or more dealers. Market players gradually adjusted to each other. One way or another, they came to a general equilibrium state, which is very different from their original models.

Market Making with basic Machine Learning Methods

An article from Stanford that describes the use of basic machine learning techniques to predict prices and trade based on these predictions. The authors of the article did not manage to achieve a stable profit for the market maker with the methods they considered, however, they told about their implementation in the code of this. They also described important basic concepts if you try to reproduce them. It is also worth noting that the results were tested in the most efficient blue chip market in America and possibly in a less efficient market for say cryptocurrencies these methods may become relevant.

Their strategy uses machine learning models to predict the evolution of the stock price they are considering, based on ML. They periodically test the state of the market and use these models to output a signal indicating whether the price will rise, decline, or remain unchanged on a global scale. Taking this evolution into account, we place ask and bid orders at a given price (which may be, for example, the best ask and bid quote or the second best bid and ask quote, as they do in our strategy), adjusted for evolution to better capture the market trend

High Frequency Trade Execution Model for Supervised Learning

Newer articles try to use neural networks to predict price movements. I am not considering RL methods as they are quite heavy for HFT at the moment. The authors propose a light neural network and test its effectiveness on a snapshot of the glass, which brings this study quite close to the real market.

This paper uses an algorithmic approach to predict the next price flip event based on a short sequence of observations of the depth of the book of limit orders and market orders. As predictors, we choose the space-time representation (Sirignano, 2016; Dixon, 2017) of the limit order book in combination with the history of market orders. Our approach solves the problem of sequence classification - a short sequence of observations of the depth of the book and market orders can be classified into a directional movement of the average price. The sequence classifier provides a potential significant benefit to market participants. For example, a market maker may use a classifier to continuously adjust quotes, potentially reducing the likelihood of an unfavorable price choice. Sequence classification has been considered in other literature sources to predict lower frequencies of price movements at historical prices (Leung et al., 2000; Dixon et al., 2016). Thus, the novelty of our approach lies in the application of a recurrent neural network classifier to the space-time representation of the book of limit orders in combination with the history of market orders to predict price reversals.

Learning the recurrent architecture of a neural network can be done using stochastic gradient descent (SGD), which learns the weights and offsets in the architecture between layers. Drop-out (DO) performs variable selection (Srivastava et al., 2014). RNNs rely on a moderate amount of training time series data together with a flexible architecture to "match" the sampling performance and beyond, measured by the mean error, area under the curve (AUC), or F1 score, which is the harmonic mean of accuracy and recall.

Methods

The main hypothesis of my work is to use machine learning methods to predict the price movement and as a consequence of the fair price based on the snapshot of the glass and the trades provided by the exchange. This approach, combined with the ideas already proposed by Avellaneda Stoikov and their followers, seems to me to be able to give a high and stable P&L for a market maker.

I'm going to start by repeating the experience of the Stanford researchers and try various artificial intelligence methods such as SVM, floating trees, and random forest to predict price spikes up or down. If this doesn't work, I would also try to double-check Sirignano's research and use simple neural networks.

What is unique and new for me is that I, unlike all researchers, will try to do all this using the lowest-level data available on the exchange. Most researchers abroad did not use this approach because they basically checked everything on the NYSE, which sells data of this level for very expensive. Unlike them, in the Russian market, such data is much easier to get and I have them. However, even with such data, a very solid IT work is required, I already have the necessary IT blanks for such solutions and I have the opportunity to test any hypotheses with a ready-made backtest framework.

In addition to the fact that I will try to build a state of the art model, I will also compare them with clear and proven benchmarks. The benchmarks that I would like to consider and the backtest of which I already have are the basic models using technical analysis tools-market making based on envelopes band and moving average (one of the most popular models of market making used in the Russian market, with the difference that base price is considered in different markets in its own way), and market maker based on bollinger bands. In addition, I would also like to collect and test the models of avellaneda stoikov and their followers and also use them as benchmarks. All of the above models need to be assembled so that they can trade in increments and rewrite this into a well-tested code, which I will give at the end of my work.

It is assumed that the ML model will predict the movement of the market in this format, you can compare it with the problem of logregression

$$Sig_t = \begin{cases} 1 & \text{for} \quad Midprice_t > \theta \\ 0 & \text{for} \quad -\theta < Midprice_t < \theta \\ -1 & \text{for} \quad Midprice_t < \theta \end{cases}$$

Results Anticipated

At the very beginning of the work, we studied the classic methods of market making that were relevant 10 years ago. In addition to the simple methods of technical analysis that appeared with the advent of e-commerce, the well-known methods of Avellaneda Stoikov models were also considered, which from the point of view of theoretical validity often win even the latest state of the art ML models. In the process of reviewing this part, it turned out that the data and accuracy of the backtest are extremely important for obtaining good results in modeling, which is important not only for market making but also for any HFT trading algorithm.

Based on this, I found the most in-depth data on the Russian securities market for 2015-2016, which is the Full order log of the exchange. Many professional participants usually try to use such data, since they can more accurately take into account the order queue and any kind of "fast" orders(orders for placing and orders for withdrawing go almost instantly). Many market participants do not provide such a level of backtest for free, at best, you can find solutions for backtest on candles, which are also very limited in functionality. In total, at this stage of work, I put together a ready-made simulation of the exchange and conducted a backtest of the benchmark models that I wrote about above. This stage is already ready.

The next stage I see is the backtest and optimization of the ML strategies described in the articles, taking into account this level of backtest. I will make an assumption here that all algorithms are able to fit into 100 ms for making a trading decision, which is possible very strong doubling, especially for some neural network architectures, however, I will assume that due to the use of quantization and low code, this can be achieved in the real market.

The final stage of my work is to combine these models. Most likely, this will be in the format that first the basic simple benchmark algorithms tell the ML model their quotes and the ML model, as in the stacking prototypes, will either choose one of these options, or take them into account as a feature for its prediction and thus will be able to achieve a high sharp ratio for a long time. I will present all the code used in the work.

Conclusion

The stock market has changed beyond recognition in the last few years, and many trading models used by dealers are becoming increasingly irrelevant with the development of computing power. All market players understand this and try to try all new trends in pricing and modeling, in particular, the latest trend has long been the use of ML methods. Many market participants, especially not the largest ones, have been using gradient boosters and other lightweight algorithms for their trading strategies for a long time, and some of them show good results. But technologies do not stand still and more and more research is emerging with models using various neural network architectures, including RL models.

Nevertheless, despite this, many of the ideas used earlier may still be relevant and applicable for the purposes of market making. Moreover, for large players with entire risk and security departments, agreeing on incomprehensible and complex ML models for trading can often be an unrealistic task. Therefore, many large market players are interested in something as simple as possible, but also containing proven ML concepts and implemented within the framework of a large slow-moving player. This approach can certainly have a lost benefit, but it is more stable and it is such solutions that many are interested in.

References

1) S. Stoikov&Marc Avellaneda, “High-frequency trading in a limit order book. Quant Finance”; 213–254, 2006.
2) Dixon M., “A High Frequency Trade Execution Model for Supervised Learning”, 2017
3) Ching Wai-Ki-“On Optimal Pricing Model for Multiple Dealers in a Competitive Market”, 2015
4) Hans R. Stoll&Thomas Ho, - “Optimal Dealer Pricing Under Transactions And Return Uncertainty”, Finance Journal(35), 248 – 256,1979
5) M. Mezar, J. Bouchaud, M. Potters – “Statistical properties of stock order books: empirical results and models”, Quantitative Finance, 2, 251–256, 2002