Machine Learning and High Frequency Trading

Abstract

Predictive modeling in high frequency is challenging because the HFT data is huge and there are many predictors to choose from. To tackle this, we take advantage of the advancement in machine learning and big data analysis in recent years. Our goal is to (i) look for trends in higher frequencies using algorithms based on genetic algorithm and (ii) extract features with predictive abilities from market order book using standard methods such as LASSO, Support Vector Machines, and Deep Learning. We believe this research will fill the gap in our understanding of predictive models in higher frequencies relative to low frequency literature.

Predictive models takes the following form:

$$r_{i,t+1} = g(z_{i,t}) + \varepsilon_{i,t+1}$$

Above, $r_{i,t+1}$ is the return of asset i at time t and $g(z_{i,t})$ is a flexible function of a P-dimensional set of predictors. The goal here is to first find a set of predictors, $z_{i,t}$, and later to estimate the form of mixing these predictors, $g(z_{i,t})$. In low frequency, the subscript, t , refers to daily, monthly or yearly observations. However, in high frequency, when t is describing millisecond, second, or minute time intervals, solving the problem becomes challenging: (i) the data becomes large and therefore, there is a large set of predictors to choose from; (ii) the estimation becomes computationally expensive meaning that finding a suitable form of combining predictive variables, $g(.)$, will take time.

This paper aims to take up this challenge using modern-day advancements in machine learning and big data analysis. We examine two sets of predictive models in high frequency. First, we investigate trends. The idea is to see whether there are trends in higher frequencies and if there are, are they profitable? The findings from this analysis sheds light on the weak-form of market efficiency in high frequency. To implement this analysis, we use a machine learning algorithm based on genetic algorithm and employ high-frequency minute-by-minute data from TAQ for the stocks in S&P 500 index in 2011.

Next, we construct several sets of predictors (or features in machine learning terminology) using the information in market order book and then, employ machine learning algorithms to extract true set of predictors. To implement this analysis, we use several supervised and unsupervised machine learning algorithms (such as LASSO, Support Vector Machines, and Deep Learning). The finding from this analysis can clarify the set of features with predicting ability in high frequency. In addition to this, we can use the resulting models to investigate different aspects of high frequency trading such as quote stuffing.

Although strategies based on predictive models or trend realization techniques are a common research topic in low frequencies, they have not been addressed much in higher frequencies. The main reason is the large datasets and the existence of numerous features and predictors that can be extracted from the data. As an example of predictive models in higher frequency, we can refer to the recent paper by Gao et al. (2018). They document that the first half hour return on the market predicts its last half hour return. Considering the large cross section of assets and various other forms of possible relations in higher frequencies, we believe our line of research can tell us much more about prediction in higher frequencies.

On the other hand, this research project can shed light on the decisions made by high frequency traders and the form of predictive-based trading strategies they engage in. The research on high frequency trading is mainly focused on investigating the historical trades made by high-frequency traders (e.g. Brogaard et al., 2014, Baron et al., 2014). Based on these trades, we get to understand what major strategies they engage in. These strategies are mainly categorized as market making, predatory trading, statistical arbitrage, and latency arbitrage. However, when it comes to predictive-based strategies, we know little. With the current research, we aim to overcome this gap.

Reference

Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan. "High-frequency trading and price discovery." The Review of Financial Studies 27.8 (2014): 2267-2306.

Baron, Matthew, et al. "Risk and return in high-frequency trading." Journal of financial and quantitative analysis (2014): 1-57.

Gao, L., Han, Y., Li, S. Z., & Zhou, G. (2018). Market intraday momentum. Journal of Financial Economics, 129(2), 394-414.