

Minimizing Low-Latency Arbitrage Trading Opportunity for Hft Traders (A Mathematical Model with Hft Simulation)

Dr. C K Sharma, Itishree Swain, Prachismita Samal, Suchismita Sarangi

Department of Mathematics, NM Institute of Engineering and Technology, Bhubaneswar, Odisha

Department of Mathematics, Raajdhani Engineering College, Bhubaneswar, Odisha

Department of Mathematics, Aryan Institute of Engineering and Technology Bhubaneswar, Odisha

Department of Basic Science & Humanities, Capital Engineering College, Bhubaneswar, Odisha

ABSTRACT - All financial markets experts agree HFT creates liquidity in the markets and it is good for the trading community, trading venues and for overall economy of the country [1]. A market maker facilitates trade and supplies liquidity by simultaneously maintaining offers to buy and sell. Because of the market fragmentation securities can be traded across multiple trading venues. As the markets are continuous the price of the traded securities in all the trading venues will not settle instantaneously to NBBO because of the computational and transmission latencies between these trading venues. Basically, this latency between the trading venues involve following components: 1. The intrinsic latency due to geographic distances between information sources; 2. The throughput latency due to local network architecture; and 3. The computational latency in calculating the NBBO. Because of this a new type of HFT strategy in which traders exercise superior speed in order to exploit price disparities between the exchanges. This type of trading is called Low Latency Arbitrage Trading wherein the HFT traders quickly identify such anomalies in the prices between the two trading venues and whichever trading venue the price is less, they buy stock at high speeds and sell in the other market where the price is slightly more and make profits. One more type of arbitrage trading is HFT traders gather all the information in other trading venues market volumes and quickly estimate precisely what will be the next NBBO before the NBBO can be announced and place their orders in queue first and make profits continuously before every announcement of the NBBO. And the HFT traders keep repeating for every announcement of the NBBO. Because of this activity HFT traders are not only liquidity suppliers but also significant liquidity consumers adversely affecting the low-speed algorithmic traders. At the same time low speed algorithmic traders are depending on the HFT for the liquidity and prices. Because of this co-existence of HFT and low speed traders there exists a predator-prey relationship causing dissatisfaction among the low-speed traders. In this project we study and mathematically model a Low latency arbitrage trading eco-system and propose a solution to optimally minimize the low latency arbitrage trading without discouraging the HFT traders as they provide liquidity to the financial markets.

I. INTRODUCTION

High Frequency Trading (HFT) is impacting market dynamics and generated interesting debates. For some, HFT is a relative term. What is called HFT today may be the common form of trading in the future. New technologies such as HFT allow agents to harness an advantage over each other in order to make more profits. HFT is a mode of intervention on financial markets that uses sophisticated software and hardware tools to implement high frequency trading managed by math algorithms. These sophisticated software and hardware tools are developed by individuals and their number is very small when compared to all of the algorithmic traders put together. Interestingly these small no. of HFT traders impact the financial markets in a large way by create liquidity to the markets as the majority of them are market makers. Most of the financial markets researchers and practitioners agree HFT creates liquidity in the markets and it is good for the trading community, trading venues and for overall economy of the country. A market maker facilitates trade and supplies liquidity by simultaneously maintaining offers to buy and sell. Because of the market fragmentation same trading instrument can be traded across multiple trading venues. As the markets are continuous the price of the instruments in all the trading venues will not settle instantaneously to NBBO because of the computational and transmission latencies between these trading venues [2].

Basically, this latency between the trading venues involve following components:

1. The intrinsic latency due to geographic distances between information sources.
2. The throughput latency due to local architecture; and
3. The computational latency

To calculate NBBO the primary exchanges like NASDAQ and NYSE have to procure the volume traded and the prevailing price of an instrument from the connected secondary exchanges and based the overall traded volume of the particular instrument the NBBO is calculated and announced to the traders as part of the tick data. By the time primary exchange collects the secondary exchanges the volume and price traded and calculate the NBBO and to announce there is a latency. This latency depends upon how liquid the asset is and the traded volume. For e.g., stocks like Apple which are highly liquid the NBBO announcement can take up to 1.5ms and some it may even take 60-70ms. During this period there is a price uncertainty for the traders during the period of latency. Because of this a new type of HFT strategy in which traders exercise superior speed in order to exploit price disparities between exchanges. This type of trading is called Low Latency Arbitrage Trading wherein the HFT traders quickly identify such anomalies in the prices between the two trading venues which ever trading venue the price is less, they buy stock at high speeds and sell in the other market where the price is slightly more and make profits.

One more type of arbitrage trading is HFT traders gather all the information in other trading venues market volumes as most of them have Direct Market Access (DMA) to the trading venues. They quickly estimate precisely what will be the next NBBO before the NBBO can be announced and place their orders in queue first and make profits continuously before every announcement of the NBBO. Because of this activity HFT traders are not only liquidity suppliers but also significant liquidity consumers adversely affecting the low-speed algorithmic traders. At the same time low speed algorithmic traders are depending on the HFT for the liquidity and prices. Because of this co-existence of HFT and low speed traders there exists a predator-prey relationship causing dissatisfaction among the low-speed traders. In this project we study and mathematically model a Low latency arbitrage trading eco-system and propose a solution to optimally minimize the low latency arbitrage trading without discouraging the HFT traders as they provide liquidity to the financial markets.

I.1. How the Low Latency Arbitrage Trading works?

Many HFT strategies are designed to exploit advantages in latency – the time it takes to access and respond to market information [3] (Wah and Wellman, 2013) place orders such that they are first in the queue. Some of the industry experts such as Schneider (2012) have estimated that trading on latency advantages account for \$21 billion profit each year. In a way HFT traders are consuming part of the liquidity that they are supplying to the markets. HFT traders are able to obtain such speed advantages over institutional investors by developing sophisticated trading algorithms combined with co-located computer systems, directly linked with trading venues with Direct Market Access.

The Securities and Exchange Commission (SEC) introduced the Regulation National Market System (Reg NMS) in 2007 in order to protect fair access to the best stock price i.e., NBBO for traditional investors. According to Reg NMS rules, trading venues are required to provide trading messages to the primary exchanges such as NASDAQ and NYSE.

These primary exchanges NASDAQ and NYSE collect all relevant data through network of data feeds of securities i.e., their current trading volume, the prevailing price and calculate the respective NBBO. Consequently, stockbrokers are required to execute trading orders at NBBO prices or better [14]. However, considering trading order information from all exchanges, the data feed networks SIP take some finite time, let us say Δt milliseconds, to calculate and for the NBBO to be distributed. Computationally sophisticated traders equipped with front-running high-speed strategies HFT traders can process the order flow in less than Δt milliseconds and out-compute the SIP to calculate the NBBO. With HFT traders sophisticated software and hardware tools quotes within an exchange could update faster than the exchange is able to distribute its new prices to other trading venues for NBBO evaluation.

Thereby low latency arbitrage trading is done by the following two ways:

1. Single Market Trading Model

- Estimates the average latency Δt for NBBO release
- Exploits the latency Δt & computes the NBBO prices in Δt
- Submits orders to be first in the queue exploiting price differential to the two markets simultaneously

In this model for every Δt the HFT traders keep predicting NBBO with sophisticated tools and algorithms before its release and place their orders with the spread between highest Bid and lowest Offer with least impacting trading volumes.

2. Multi-Market Trading Model

- First obtains current price quotes in both markets
- Checks whether an arbitrage latency situation exists
- Submits order exploiting price differential to the two markets simultaneously

In this model the HFT traders quickly identify such anomalies in the price's settlement between the two trading venues. Whichever trading venue's security price is less, they buy stock at high speeds and sell in the same security in other market where its price is slightly more and make profits.

I. 2. Microstructure of the Low Latency Arbitrage Trading

Financial instruments are bought and sold at a financial market. Market makers at a financial market act as intermediaries between the buyer and the seller Semirings, ringoids, algebroids and non-associative algebras play important role in algebra and among them ordered semirings and lattices as well [1,2,3,4]. This is also motivated by idempotent mathematical physics naturally appearing in quantum mechanics and quantum field theory (see, for example, [9] and references therein). They also arise from the consideration of algebroids and ringoids associated with non locally compact groups. Namely, this appears while the studies of representations of non locally compact groups, quasi-invariant measures on them and convolution algebras of functions and measures on them [10,11]. The background for this is A. Weil's theorem asserting that if a topological group has a quasi-invariant σ -additive non trivial measure relative to the entire group, then it is locally compact. Therefore, it appears natural to study inverse mapping systems of non locally compact groups and their dense subgroups. Such spectra lead to structures of algebroids and ringoids. Investigations of such objects are also important for making advances in representation theory of non locally compact groups.

In this paper methods of categorial topology are used This article is devoted to ordered ringoids and semirings with an additional lattice structure. Their continuous morphisms are investigated in [Section 3](#). Preliminaries are given in [Section 2](#). Necessary definitions 2.1–2.4 are recalled. For a topological ringoid K and a completely regular topological space X new ringoids $C(X,K)$ are studied, where $C(X,K)$ consists of all continuous mappings $f:X \rightarrow K$ with point-wise algebraic operations. Their ideals, topological directed structures and idempotent operations are considered in Lemmas 2.6, 2.8, 2.9, 2.12 and Corollary 2.7. There are also given several examples 2.13–2.18 of objects. One of the main examples between them is related to cones in algebras of non locally compact groups. Another example is based on ordinals. Construction of ringoids with the help of inductive limits is also considered.

Then a weak* topology on a family $O(X,K)$ of all order preserving weakly additive morphisms on a Hausdorff topological space X with values in K is taken. The weak* compactness of $O(X,K)$ under definite conditions is proved in Theorem 3.10. Further in Proposition 3.11 there is proved that $I(X,K)$ and $Ih(X,K)$ are closed in $O(X,K)$, where $I(X,K)$ denotes the set of all idempotent K -valued morphisms, also $Ih(X,K)$ denotes its subset of idempotent homogeneous morphisms.

Categories related to morphisms and ringoids are presented in [Subsection 3.2](#). An existence of covariant functors, their ranges and continuity of morphisms are studied in Lemmas 3.14, 3.16, 3.21, 3.34 and Propositions 3.15, 3.22. In Propositions 3.24, 3.26 and 3.29 such properties of functors as being monomorphic and epimorphic are investigated. Supports of functors are studied in Proposition 3.31. Moreover, in Proposition 3.32 it is proved that definite functors preserve intersections of closed subsets. Then functors for inverse systems are described in Proposition 3.33. Bi-functors preserving pre-images are considered in Proposition 3.35. Monads in certain categories are investigated in Theorem 3.38. Exact sequences in categories are considered in Proposition 3.39.

Lattices associated with actions of groupoids on topological spaces are investigated in [subsection 3.3](#). Supports of (T,G) -invariant semi-idempotent continuous morphisms are estimated in Proposition 3.42, where G is a topological groupoid and T is its representation described in Lemma 3.40. Structures of families of all semi-idempotent continuous morphisms associated with a groupoid G and a ringoid K are investigated in Proposition 3.43 and Theorems 3.44, 3.45.

REFERENCES

- [1]. Baez, J.C. The octonions. Bull. Am. Mathem. Soc. 2002, 39, 145–205. [[Google Scholar](#)] [[CrossRef](#)]
- [2]. Birkhoff, G. Lattice Theory; Mathematical Society: Providence, RI, USA, 1967. [[Google Scholar](#)]
- [3]. Bourbaki, N. Algebra; Springer: Berlin, German, 1989. [[Google Scholar](#)]
- [4]. Dickson, L.E. The Collected Mathematical Papers. Chelsea Publishing Co.: New York, NY, USA, 1975; Volumes 1–5. [[Google Scholar](#)]
- [5]. Grätzer, G. General Lattice Theory; Akademie-Verlag: Berlin, German, 1978. [[Google Scholar](#)]
- [6]. Kasch, F. Moduln und Ringe; Teubner: Stuttgart, German, 1977. [[Google Scholar](#)]
- [7]. Kurosh, A.G. Lectures on General Algebra; Nauka: Moscow, Russian, 1973. [[Google Scholar](#)]
- [8]. Schafer, R.D. An Introduction to nonassociative Algebras; Academic Press: New York, NY, USA, 1966. [[Google Scholar](#)]
- [9]. Litvinov, G.L.; Maslov, V.P.; Shpiz, G.B. Idempotent functional analysis: an algebraic approach. Math. Notes 2001, 65, 696–729. [[Google Scholar](#)] [[CrossRef](#)]

- [10]. Ludkovsky, S.V. Topological transformation groups of manifolds over non-Archimedean fields, representations and quasi-invariant measures. J. Mathem. Sci. NY Springer 2008, 147, 6703–6846. [\[Google Scholar\]](#) [\[CrossRef\]](#)
- [11]. Ludkovsky, S.V. Topological transformation groups of manifolds over non-Archimedean fields, representations and quasi-invariant measures, II. J. Mathem. Sci., N. Springer 2008, 150, 2123–2223. [\[Google Scholar\]](#) [\[CrossRef\]](#)