HIGH-FREQUENCY TRADING AND PLATFORM FOR FINANCIAL RESEARCH

A PREPRINT

Shiwen An
Department of Particles and Nuclear Studies
The Graduate University for Advanced Studies, SOKENDAI
Oho 1-1, Tsukuba, Japan
sweynan@icloud.com

May 5, 2023

ABSTRACT

High frequency trading (HFT) is a form of algorithmic trading that involves using powerful computer programs and high-speed networks to execute a large number of trades in fractions of a second. The goal of HFT is to profit from very small price movements in financial markets, typically by buying and selling large volumes of securities, such as stocks, futures contracts, or currencies, in a matter of milliseconds.

Keywords High-Frequency Trading · Financial Platform Data · Algorithms

1 Introduction

HFT relies on sophisticated algorithms and advanced technologies, including ultra-low latency data feeds, high-speed trading platforms, and co-location services, which enable traders to access market data and execute trades faster than their competitors. HFT firms use complex mathematical models to identify and exploit market inefficiencies, such as price discrepancies between different exchanges or variations in supply and demand.

HFT has become increasingly popular in recent years, as technological advancements and regulatory changes have made it easier and cheaper for firms to engage in this type of trading. However, HFT has also been controversial, with critics arguing that it can lead to market instability, create unfair advantages for certain traders, and contribute to increased volatility and systemic risk.

There are many different algorithms used in high frequency trading, and the specific ones used by a particular firm may vary depending on their trading strategies and market conditions. However, some of the most commonly used algorithms in HFT include **Market Making, Statistical Arbitrage, Trend Following, and etc.** which will be introduced in the next section.

2 Common Strategies

- Market Making: This algorithm is used to provide liquidity to the market by continuously quoting bid and ask prices for a security, and buying and selling at those prices as orders come in.
- Statistical Arbitrage: This algorithm involves identifying mispricings in related securities and taking advantage of them by simultaneously buying and selling those securities.
- Trend Following: This algorithm uses technical analysis to identify trends in a security's price and then buys or sells based on whether the trend is up or down.
- News-Based Trading: This algorithm uses natural language processing techniques to analyze news and social media feeds for market-moving information, and then executes trades based on the results.

- Scalping: This algorithm involves quickly buying and selling securities to capture small price movements and earn profits on a large number of trades.
- Mean Reversion: This algorithm is based on the idea that prices tend to revert to their mean over time, and involves buying or selling a security when its price deviates significantly from its historical average. Brogaard et al. [2014]

3   Data and Financial Research Platform

The choice of data sources will depend on the specific research question and the type of trading strategy being developed. High-frequency trading research often requires access to large amounts of high-quality data and the ability to process and analyze this data quickly and efficiently. Here I give some brief summary for what kind of data source and simulation is possible for HFT research.

3.1   Brownian Motion

The most in-depth significantly cited paper is from here Campbell et al. [1997]. In some common event, Geometric Brownian Motion is applied.

$$dS_t = \mu S_t dt + \sigma S_t dW_t \tag{1}$$

The equation for geometric Brownian motion is used to generate simulated financial data that exhibits similar statistical properties to real financial data. The equation describes the stochastic evolution of the price of an asset over time, taking into account both the expected mean return and the random fluctuations in price due to volatility.

To generate simulated financial data, we use the equation to calculate the price of an asset at each time step. We begin with an initial price, S_0, and simulate the price at the next time step using the equation:

$$S_t = S_0 e^{(\mu - \frac{\sigma^2}{2})t + \sigma\sqrt{t}Z} \tag{2}$$

where Δt is the time step, μ is the drift rate, σ is the volatility, and Z_1 is a standard normal random variable.

We can continue this process to generate prices at each subsequent time step, using the current price and a new random variable at each step:

$$S_{t+1} = S_t e^{(\mu - \frac{\sigma^2}{2})\Delta t + \sigma\sqrt{\Delta t}Z_{t+1}} \tag{3}$$

By simulating many paths using this equation, we can generate a set of financial data that exhibits similar statistical properties to real financial data, including fat tails, volatility clustering, and autocorrelation. This can be useful for testing and developing trading strategies, risk models, and other financial applications.

```
import numpy as np
def generate_gbm(mu, sigma, S0, T, dt):
  N = int(T / dt)
  prices = np.zeros(N+1)
  prices[0] = S0
  Z = np.random.standard_normal(N)
  for i in range(1, N+1):
    prices[i] = prices[i-1] * np.exp((mu - 0.5*sigma**2)*dt + sigma*np.sqrt(dt)*Z[i
    -1])

  return prices
```

However, simulated data does not always work on the the real market. The data here could just be used for primitive research.

3.2   Market Data

For research on abnormal spike on the market, NYSE Trade and Quote database link is a popular option Campbell et al. [1997]. There is also case study using E-mini S&P 500 Market Value, which drop in short period of time and caused a 'flash crash', where this part of data is part of historical data. Kirilenko et al. [2011]. Some financial firms employed HFT strategies usually collect their own market data and other papers also use Deutsche Boerse DAX index stocks Brogaard et al. [2014], and one section in this research highlight the dataset when considering different task. CRSP data is considered for daily data not included in the HFT dataset, Compustat data are used to incorporate stock characteristics in to the analysis, and Trade and Quote (TAQ) data are used to incorporate additional intraday information Brogaard et al. [2014]. The CBOE S&P 500 Volatility Index (VIX) to capture market-wide volatility Brogaard et al. [2014].
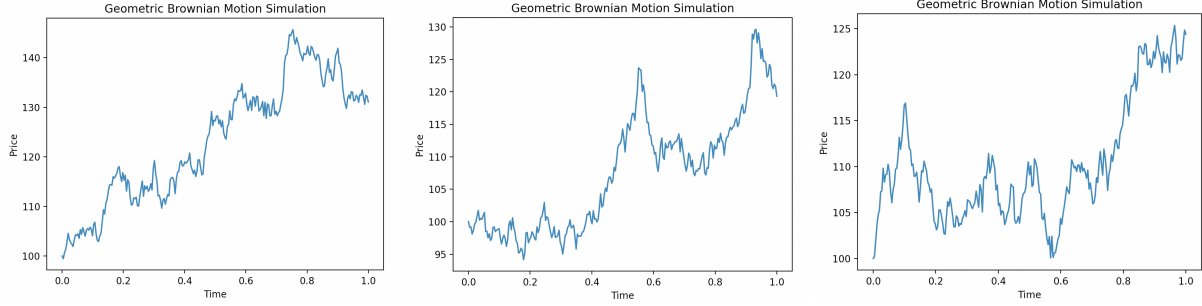
Figure 1: Geometric Brownian Motion Generated Diagram. Frequently used for random in the Financial Research.

3.3 Pros and Cons with Market Data

Market data is one of the most commonly used data sources in high-frequency trading (HFT) research, and it has several advantages and disadvantages that should be considered when conducting research in this area.

Pros of using Market Data for research in HFT:

1. High-quality data: Market data is generally of high quality and is available in real-time or historical form, allowing researchers to analyze market trends and identify trading opportunities.

2. Wide availability: Market data is widely available from a variety of sources, including exchanges and data providers. This makes it easy for researchers to obtain and analyze the data they need.

3. Relevance: Market data is directly related to trading activity and can provide valuable insights into market trends and investor sentiment.

4. Accuracy: Market data is typically very accurate, as it is collected and maintained by highly regulated exchanges and data providers.

5. Standardization: Market data is often standardized, which makes it easy to compare and analyze across different markets and asset classes.

Cons of using Market Data for research in HFT:

1. Expensive: Market data can be expensive to obtain, particularly if real-time data is required. This can be a barrier to entry for smaller research teams or individual traders.

2. Latency: Market data can suffer from latency issues, particularly when real-time data is being used. This can impact the accuracy of trading strategies and make it difficult to compete with faster traders.

3. Limited history: Market data is typically only available for a limited period of time, which can make it difficult to backtest trading strategies over longer timeframes.

4. Noisy: Market data can be noisy and may contain irrelevant information, which can make it difficult to identify useful patterns or trends.

5. Lack of transparency: Market data can lack transparency, particularly when it comes to dark pools and other off-exchange trading venues. This can make it difficult to obtain a complete picture of market activity.

Market data is a valuable resource for high-frequency trading research, but it has some limitations that should be considered. Researchers should be aware of the potential biases and limitations of market data and use it in combination with other data sources to obtain a complete picture of market activity.

For both the strategy and the dataset, Gomber et al. [2011] provides better review for more details. Here I just simply provide a summary for what HFT usually works. In my personal opinion, it is just another algorithm trading with higher processing speed on the market.

References

Jonathan Brogaard, Terrence Hendershott, and Ryan Riordan. High frequency trading and its impact on market quality. *SSRN Electronic Journal*, 2014.

John Y Campbell, Andrew W Lo, and A Craig MacKinlay. Econometric analysis of high frequency financial data. *Handbook of Statistics*, 14:1–101, 1997.

Andrei Kirilenko, Albert S Kyle, Mehrdad Samadi, and Tugkan Tuzun. The microstructure of the 'flash crash': Flow toxicity, liquidity crashes and the probability of informed trading. *Journal of Financial Markets*, 16(2):233–250, 2011.

Peter Gomber, Bj"orn Arndt, Marco Lutat, and Tim Uhle. High-frequency trading. In *The Handbook of High Frequency Trading*, pages 3–24. Elsevier, 2011.