Chapter 8
High-Frequency Data and Machine Learning

Eli Hadad Junior
https://orcid.org/0000-0003-2985-9241
Universidade Presbiteriana Mackenzie, Brazil

ABSTRACT

Forecasting financial asset prices has always been a challenging task due to a vast array of factor. Forecasting stock prices using econometric models is satisfactory when the data exhibits a certain statistical regularity, however, this regularity is not standard for this type of data. Considering that the advancement of computer technology, making them increasingly faster, the popularization of high-speed internet, the development of secure trading platforms, and the interconnection of markets have significantly increased the volume of transactions. With the increase in the number of transactions per unit of time, it became impossible for a human trader to keep up with this volume of buying and selling transactions of financial assets, thus giving rise to trading robots. In this chapter is made a comparision between traditional methods and machine learning methods to forecast time series

INTRODUCTION

Predicting how financial assets, like stocks and bonds, will move in the future has always been a difficult task. There are many different factors that can influence these prices, which we will explain later on. When using econometric models to forecast stock prices, the predictions can work well if the data shows certain statis-

DOI: 10.4018/979-8-3693-5777-4.ch008

tical patterns. However, such patterns do not always occur in financial data, making it tricky to rely solely on these models.

As technology continues to improve, with computers becoming faster and more efficient, alongside the spread of high-speed internet and the development of secure trading platforms, the number of financial transactions has skyrocketed. The growing connection between markets worldwide has also played a role in this increase. As more and more transactions happen in shorter amounts of time, it became impossible for human traders to keep track of all the buying and selling. This gave rise to the use of trading robots—automated systems designed to carry out trades at high speeds without the need for human intervention.

These trading robots have had a major impact. While they have increased the volume of trades significantly, they've also created more financial noise. This is due to the rapid placement and removal of buy and sell offers, which happens constantly. Back in the 1980s and 1990s, most traders used data that was recorded monthly or daily for building trading strategies and managing portfolios. This was because of the slower nature of financial markets at that time.

However, as trading speeds increased, transactions started happening within the same day, creating what is now known as intraday trading. This shift changed the entire way the financial markets operated and made price forecasting even more complicated. As a result, traditional econometric models became less effective and less reliable for making accurate predictions.

At this point, the use of high-frequency data—data collected at much shorter intervals than daily data—became more common. Alongside this development, machine learning techniques gained significant attention from researchers and practitioners alike. The main reason is that machine learning can potentially increase both the speed and accuracy of financial predictions, working in tandem with econometric models. The fast-moving, unpredictable, and often volatile nature of financial markets has reduced the effectiveness of traditional models, causing them to frequently fail in capturing real-time market dynamics.

Therefore, applying machine learning models that work with high-frequency data has started to produce promising results. These models represent an important step forward in the quest to improve price predictions, offering new opportunities for making more accurate and timely financial forecasts.

WHAT IS HIGH-FREQUENCY DATA (HFD)?

High-frequency data (HFD) represents a type of data collection characterized by its granularity and temporal precision. Unlike traditional datasets, which may aggregate information over daily, weekly, monthly, or even yearly intervals, HFD is recorded at extremely short intervals, often ranging from seconds to milliseconds.