M.184.5453 Advanced Time Series Analysis and Forecasting
winter semester 2023/2024

Project No: 2

Examiner: Prof. Dr. Yuanhua Feng
Assistant: Mr. Dominik Schulz

Project beginning: Februar 18, 2023, 12:00 a.m.
Project due: März 21, 2024, 11:59 p.m.

Submitted by Group No. 14:

Name	Matriculation Number	Contribution
Ahmad-Belal Wali	7125771	Part 1
Mohammad Niaj Uddin Ahmed	6900003	Part 2

Table of Contents

Introduction

In today's financial market landscape, the use of High-Frequency (HF) and Ultra-High-Frequency (UHF) financial data has become increasingly important. These datasets provide detailed insights into the movements of the market, offering traders, analysts, and researchers the information they need to make informed decisions based on intraday activities. Alongside the analysis of such data, we also find the concept of Realized Volatility (RV), a measure which quantifies market volatility by using direct price variations over short periods. This following section is about the core of HF and UHF data, their importance, and the way RV is calculated and applied in financial analysis, incorporating specific formulas that help us better understand these concepts.

PART 1: Description of High-Frequency (HF) and Ultra-High-Frequency (UHF) Financial Data and Realized Volatility (RV)

1.1 High-Frequency and Ultra-High-Frequency Financial Data

High-Frequency (HF) financial data is packed with intraday observations, offering several data points within just one trading day. This kind of data gives us the chance to dig into market behaviour in a way that's far more detailed than just looking at daily closing prices. Going a step further, Ultra-High-Frequency (UHF) financial data, or tick-by-tick data, captures every single transaction happening in the market, giving us an even clearer picture of market activities.

Understanding the difference between HF and UHF data is key for a range of financial analyses, such as studying market microstructures, measuring liquidity, and calculating volatility. UHF data lays out the precise order of trades and quotes, painting a full picture of the market's mood and how traders are acting at any specific time.

However, one of the big problems when working with HF and UHF data is the massive amount of information they contain, which demands a lot of computational power for storing, processing, and making sense of it all. Furthermore, this data is marked by the unpredictability of when observations happen, since transactions are not scheduled at fixed times but rather at dictated times by market activities. This unpredictability adds a layer of complexity to modelling and statistical analysis, helping us to adopt advanced methods to collect valuable insights.

Realized Volatility (RV)

Realized Volatility (RV) stands out as a crucial idea when diving into High-Frequency (HF) and Ultra-High-Frequency (UHF) financial data analysis. It serves as a straightforward and independent way to measure market volatility, determined straight from the squared returns of high-frequency financial data. The core concept of RV is simple: by adding together the squared returns from numerous high-frequency transactions over the course of a day, we can get a solid grasp of the day's overall volatility. Mathematically, this concept can be expressed as follows:

$$RV_1 = \sum_{i=1}^{N_t} y_{t,i}^2$$

where N is the number of high-frequency returns in the day, and y_i denotes the return for the i^{th} interval.

The simplicity and directness of RV make it an invaluable tool for assessing market volatility. Unlike traditional volatility estimators that rely on models with assumptions about return distributions, RV is profiting from the data itself, providing a clear, empirical view of volatility. This feature is particularly beneficial in high-frequency trading environments, where understanding volatility is important for strategy development, decision making and risk management.

1.2 Methodological Considerations

Calculating Realized Volatility (RV) comes with some special adjustments to make sure it is as accurate and useful as possible. For example, sometimes the data can get a bit noisy due to things like the difference between buying and selling prices, prices moving in steps instead of smoothly, and other little interruptions in trading. This noise can mess with our understanding of how volatile the market really is. To fix this, we adjust the RV calculation by using some measurement to ignore the noise and focus on the actual ups and downs of the market.

Also, when we use Ultra-High-Frequency (UHF) data, like transactions that happen every minute or every five minutes, it helps to make our observations regular and easier to compare. This process of making the data suitable for matching regular time slots needs to be done carefully. If not, it could skew our view of how much the market is really moving.

The analysis of High-Frequency and Ultra-High-Frequency financial data, complemented by the calculation of Realized Volatility, offers us profound insights into market dynamics. These methodologies enable the capturing of complex details in market movements, providing a clearer image of volatility and trader behaviour. As financial markets are constantly evolving, the importance of HF and UHF data along with RV will only grow while highlighting the need for robust analytical techniques. These methods and their application in financial analysis indicate better insights in understanding market volatility and market dynamics.

1.3 Description of the ACD Model in High-Frequency Financial Data Analysis

In the world of high-frequency (HF) financial data analysis, the Autoregressive Conditional Duration (ACD) model stands out as a key method for exploring how long it takes for trading events to happen. The following approach gives a solid foundation for anyone looking to understand the timing of trades and how information flows through the markets.

1.4 The ACD Model for Durations

The ACD model is designed to model the time between events in a financial market, typically trades or quotes. It is an extension of the Autoregressive Conditional Heteroskedasticity (ARCH) model to duration data, acknowledging that the time between market events is not uniform but depends on past durations and the information flow in the market. The model can be expressed as follows:

$$Duration_t = \Psi_t \cdot \varepsilon_t$$

where Ψ_t is a deterministic function of past durations, capturing the conditional expectation of the current duration, and ε_t is a residual error term representing the unpredictable part of the duration which can not be estimated.

The ACD model's flexibility allows it to accommodate various specifications for Ψ_t, meaning it can be adjusted in different ways, like linear or exponential, to fit the data one is working with perfectly. This ability to customize the model makes it an excellent tool for looking into how trading happens over time and seeing how different information and what is happening in the market can change trading patterns.

1.5 Estimation of the ACD Model

The estimation process of the ACD model is crucial for its practical application. In general, the estimation is about maximizing a probability of observing the sequence of durations given the model parameters. This process requires computational techniques that can handle the complexities of high-frequency data, including large datasets and the need for precise estimation of parameters.

The successful estimation of the ACD model gives insights into the market mechanisms, such as liquidity dynamics and the impact of news on trading activity. By analyzing the estimated parameters, researchers and practitioners can imply the persistence of durations and how quickly the market assimilates new information.

Implementing the ACD model into the analysis of HF data supports the understanding of market dynamics while offering a differentiated view of how trading activities evolve over time. Through its methodological style and practical applicability, the ACD model stands as an evidence to the ongoing progression in financial econometrics, offering a way to look closely at the structure of financial markets.

In conclusion, high-frequency (HF) data, realized volatility (RV)and methodologies like the ACD model play an indispensable role in modern financial econometrics, offering deep insights into the market's behaviour. While challenges in implementation exist, the ongoing development of analytical tools and software packages promises to bridge the gap, paving the way for further innovations in the field. This exploration into the depths of HF data and its modelling techniques not only enriches our understanding of financial markets but also provides us with the tools to navigate its complexities more effectively.

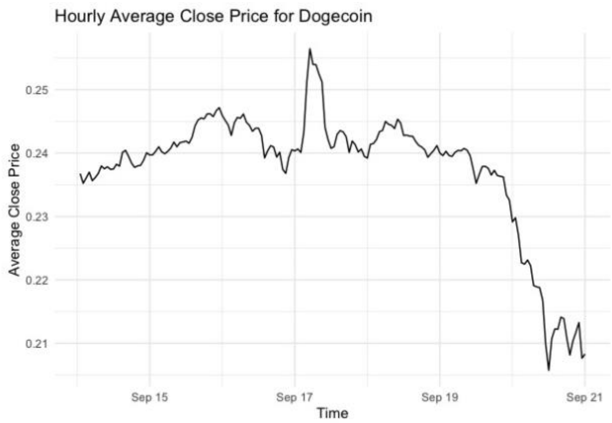
PART 2: Calculating and displaying the UHF-returns, Calculating and displaying the UHF-returns, calculating the RV, detailing & applying of ACD.

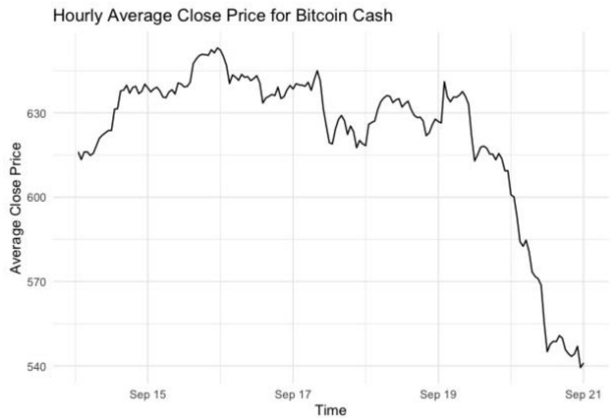
2.1 Data

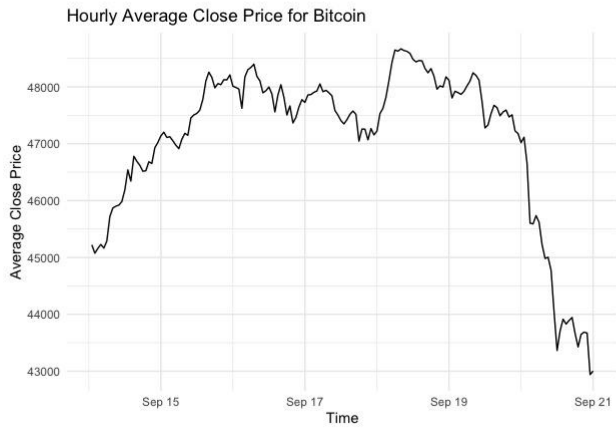
We are using a tick data for crytpto currency. In this data we have 14 different currencies. These are as follows, Bitcoin Cash Binance Coin, Bitcoin, EOS.IO, Ethereum, Classic Ethereum, Litecoin, Monero, TRON, Stellar, Cardano, IOTA, Maker, Dogecoin. We will mostly focus on the currencies which have the most values and which always stay on the trend continuously. They are asset ID = 0, 1, 2, 3, 4, 6 which are Binance coin, Bitcoin, Bitcoin Cash, Cardano, Dogecoin, and Ethereum respectively. The data is about 140 thousand entries for 4 days and has multiple entries in each minutes.

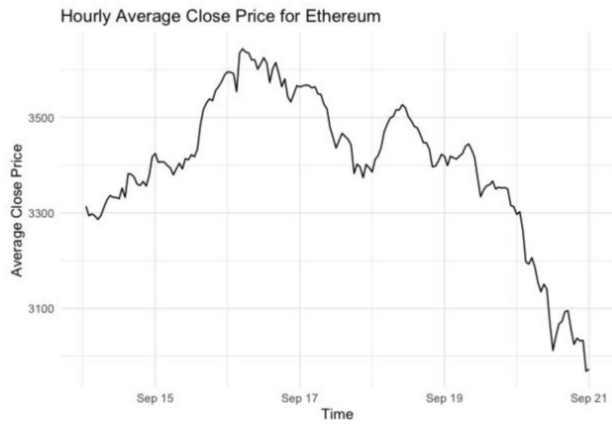
2.2 General hourly overview of the six assets.

Below we have plots for hourly average closing price for each of the six currencies, with that we the intuition of the common followed trend among them.



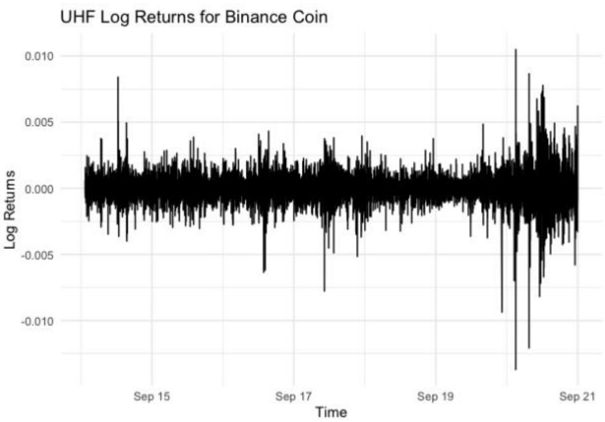


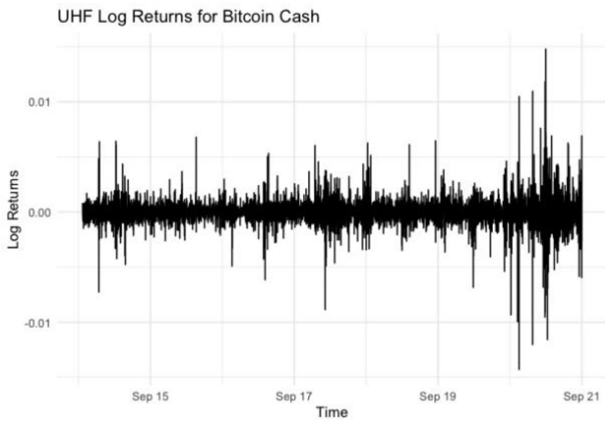


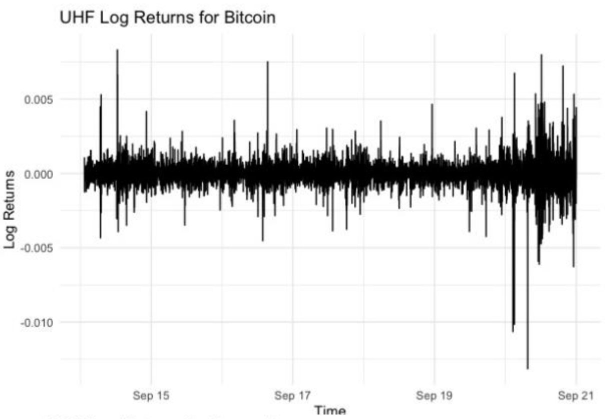


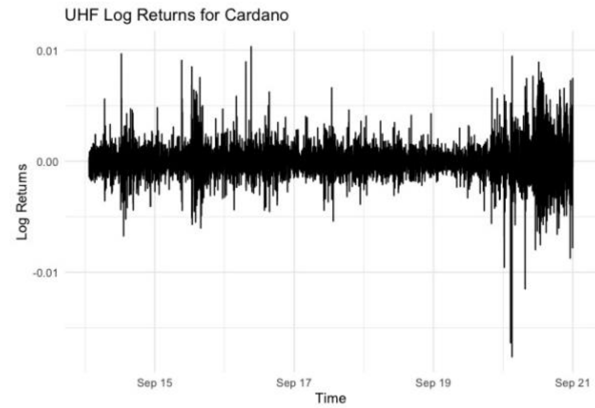
2.3 Calculation and display of UHF.

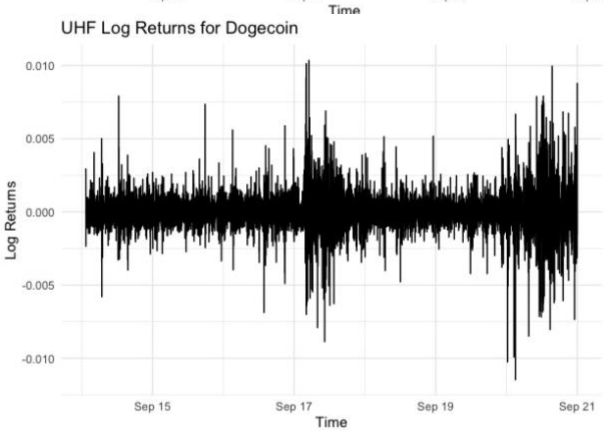
Now we have Ultra High Frequency (UHF) log return plots for these six cryptocurrencies from around September 15 to September 21. Each plot shows the log returns of a cryptocurrency's price, visualizing the fluctuations in returns within the time frame.

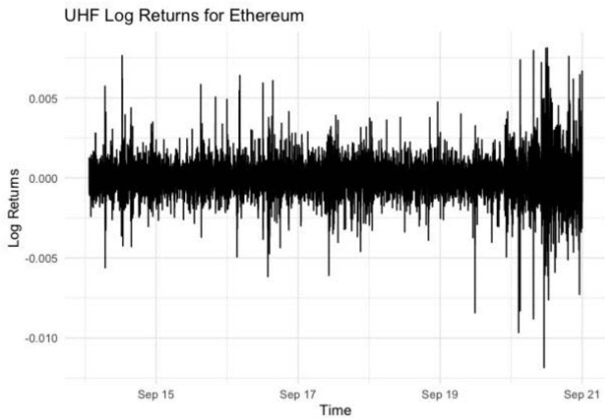










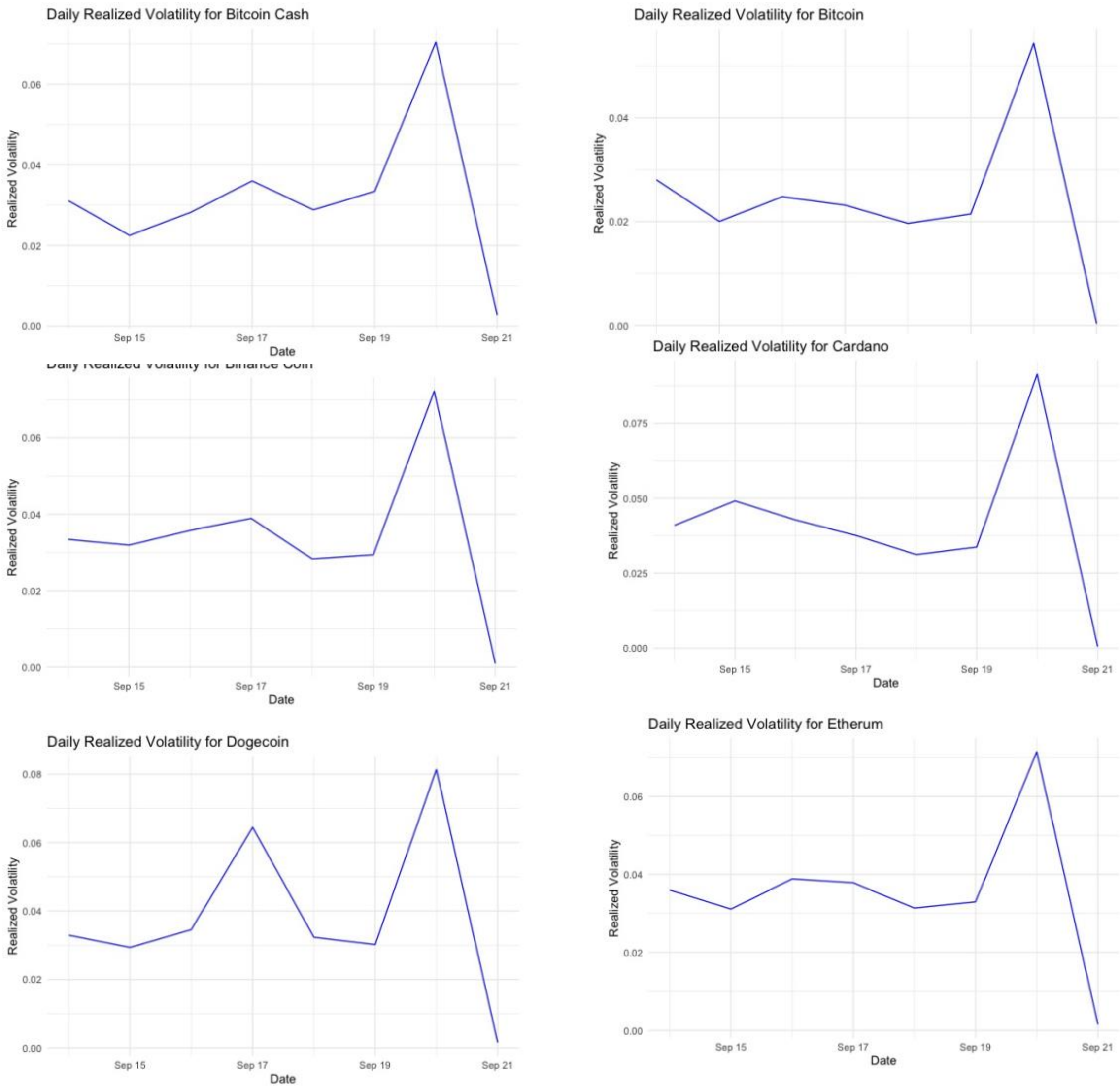


For Binance Coin, the log returns fluctuate around zero, with some instances of higher volatility. There are notable spikes and dips suggesting moments of significant price movement. Bitcoin Cash shows a similar pattern to Binance Coin, with log returns oscillating around the zero line. The volatility spikes seem much more insignificant extreme compared to the assetID zero which is Binance Coin. Bitcoin illustrates volatility around the zero line, with a mix of positive and negative returns. The spread of the returns wide particularly, which tells us that periods of increased trading activity or news events affecting price. Cardano's log returns show a tight clustering around the zero line, with several notable spikes in both positive and negative directions. Dogecoin indicates a volatility pattern as of all other currencies, with a balance of positive and negative returns larger changes can also be seen.

Finally one of the market leaders, Ethereum's UHF log return plot is quite dense with data points closely packed around the zero line. Although similarly moments of pronounced spikes can be observed, reflecting significant price changes.

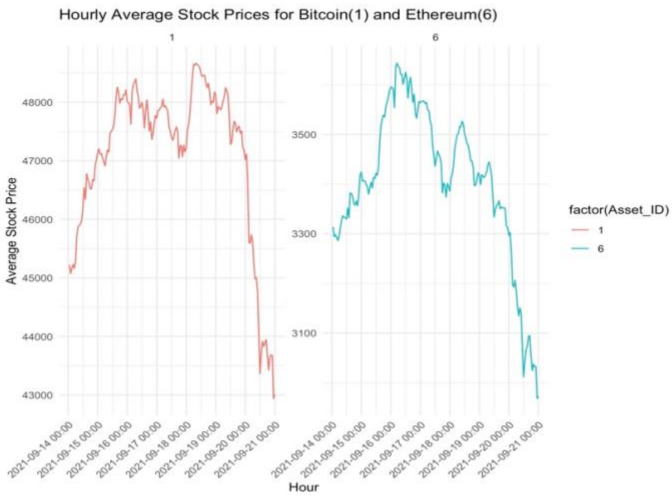
2.4 Realized Volatility Caculation.

Below we can observe realised volatility for each currencies. Where we can see similar trend, with occasional fluctualtions.

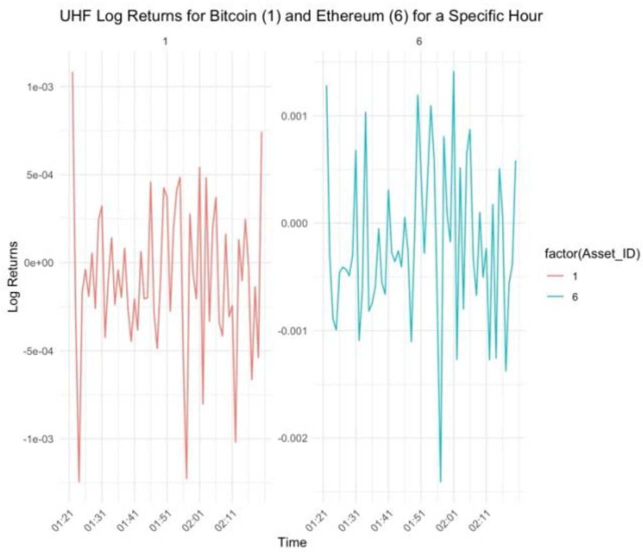


2.5 Realized Volatility and Ultra High Frequency for Ethereum and Bitcoin.

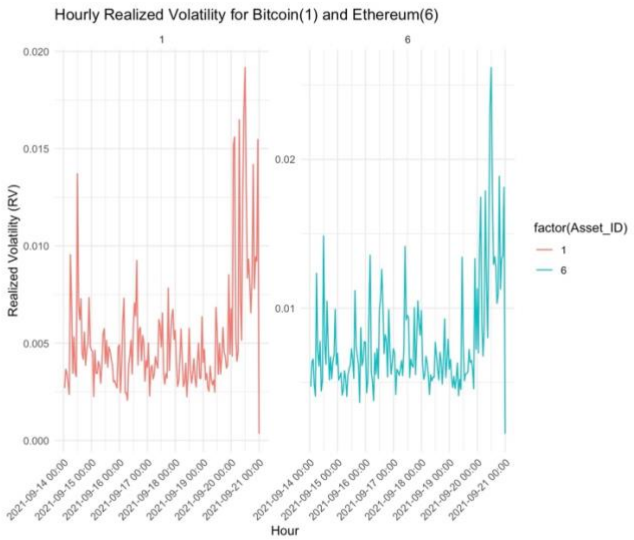
Bitcoin and Ethereum are the most dominant market factor in crypto world. Since these two currencies have bigger influence we will now calculate RV and UHF and compare them with each other. First the stock price will give us some intuitions.



Here both have similar trend that makes them almost identical. Now The UHF is measured over a specific period of time. The duration is within one hour.

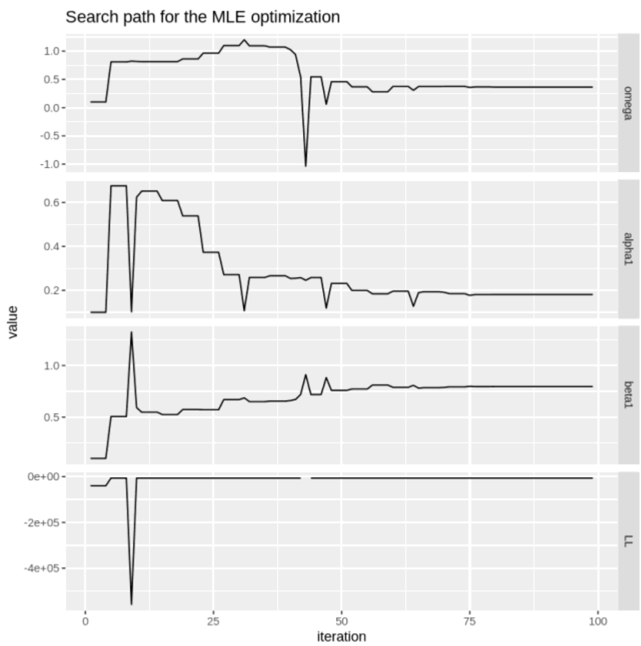


And finally we will compare the two under Realized Volatility calculation in one specifice time.



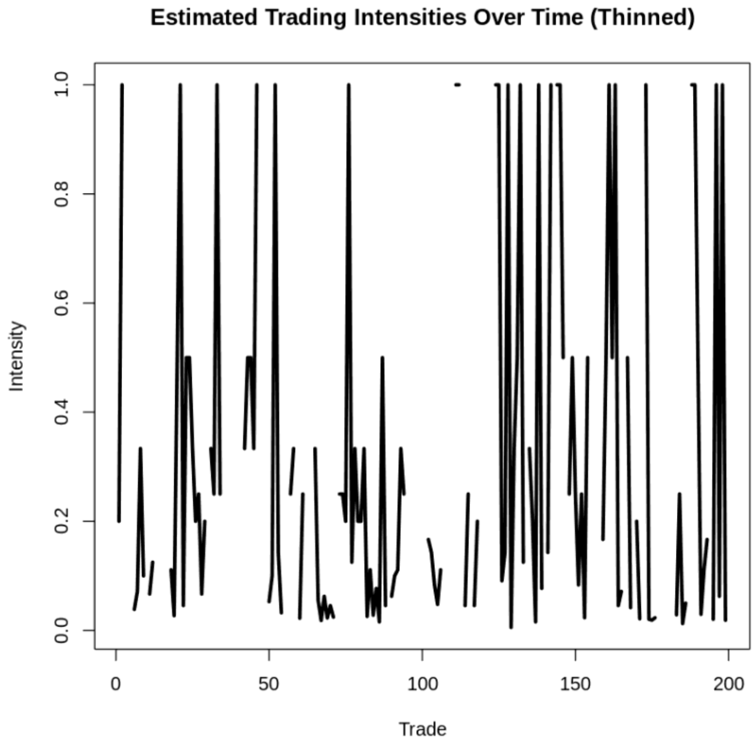
2.6 Autoregressive Conditional Duration.

The ACD model estimation indicates that our exponential ACD(1, 1) model fitting was successful. It shows significant parameters for the baseline intensity (omega), the impact of the previous duration's shock (alpha1), and the past conditional durations (beta1). The goodness of fit metrics like Log-Likelihood, AIC, BIC, and MSE suggest how well the model fits to our durations data, and a convergence value of 0 indicates the optimization algorithm used for estimation found a solution successfully.



2.7 ACD Plot.

we can observe periods of higher trading intensity with tall peaks which is indicating more frequent trading activity, interspersed with periods of lower intensity (shorter spikes or dips), that suggest less frequent trading. The variability in the high peaks suggests fluctuating trading activity within the timeframe. We have about 2 thousand observations in the data. Which represented by the trade and intensity axis, whole intuition is to estimate trading in a duration.



Conclusion:

The significance of High-Frequency (HF) and Ultra-High-Frequency (UHF) data in modern financial analysis is enhanced by the conclusion of the discussed sections. The extensive exploration of these datasets, which includes just under 140,000 entries of tick data from various cryptocurrencies over a span of four days, six currencies explicitly, the and data involves BMW UHF data which have about 2000 entries. It showcases the nature of market dynamics captured in intraday trading activities. Extracting actionable insights from this data has heavily relied on Realized Volatility (RV) calculations and Autoregressive Conditional Duration (ACD) models. Notably, for influencial market leaders like Bitcoin and Ethereum, comparing RV and UHF log returns has provided a detailed perspective on market movements, highlighting their dominant influence in the crypto sphere. This calculations are very significant in todays market. The ACD model used is a generic 1,1 model.

In summary, HF and UHF financial data, in conjunction with RV and ACD models, serve as indispensable tools in financial econometrics, providing deep intuitions into market behavior. Inspite of the challenges associated with their implementation, these tools remain important in ongoing efforts to understand and navigate the complexities of the financial world.