Bitcoin volatility analysis

Author:
Enrico Grimaldi, 1884443

Professor:
Luca Tardella

La Sapienza University of Rome

a.y. 2022/2023

Final project for the course of
Statistical Methods for Data Science 2

Contents

1. Introduction

2. The data set

The Yahoo Finance Bitcoin Historical Data from Kaggle, spanning from 2014 to 2023, capture the evolution of Bitcoin's price over a decade, offering an overview of the following features about Bitcoin value:

Date	Open	High	Low	Close	Adj.Close	Volume
2014-09-17	465.864	468.174	452.422	457.334	457.334	21056800
2014-09-18	456.860	456.860	413.104	424.440	424.440	34483200
2014-09-19	424.103	427.835	384.532	394.796	394.796	37919700
2014-09-20	394.673	423.296	389.883	408.904	408.904	36863600
2014-09-21	408.085	412.426	393.181	398.821	398.821	26580100
2014-09-22	399.100	406.916	397.130	402.152	402.152	24127600

Of the kaggle data set we are solely interested in one of the features: the adjusted closing price of bitcoin (in terms of BTC/USD value).

2.1. Returns rather than prices

In the analysis of financial data, asset equity returns are typically the main variable of interest (rather than prices) There are at least two reasons for this:

1. Returns are easier to interpret
2. Returns have statistical properties which are easier to handle (e.g. stationarity)

Let P_t be the price of an asset at period t ($t = 1, .., T$) the **simple return** is defined as the gross return:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

In other words, R_t is the gross return generated by holding the asset for one period. Simple returns are a natural way of measuring the variation of the value of an asset, however, it is more common to work with **log returns**, defined as:

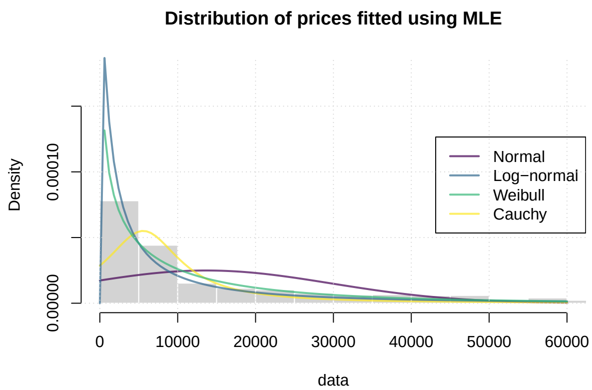
$$\epsilon_t = log(P_t) - log(P_{t-1})$$

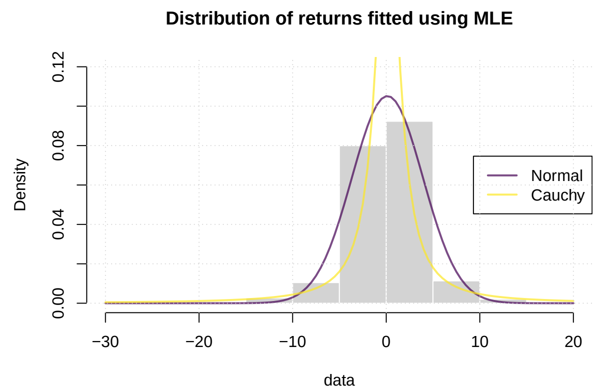
It is a good habit to multiply returns by 100 to express them as a percentage. Some statistical packages are sensitive to the scale of the data. Since log differences can be small, sometimes this creates numerical difficulties. Thus:

$$\epsilon_t = 100 \cdot (log(P_t) - log(P_{t-1}))$$

2.2. Searching for a prior

We can then begin to analyze the distribution of the data in question.





3. First model

3.1. ARCH vs GARCH

3.1.1. The ARCH model

The **ARCH (Autoregressive Conditional Heteroskedasticity)** model is a statistical time series model commonly used in econometrics and finance to capture volatility clustering in financial data. It was introduced by Robert F. Engle in the early 1980s as a way to model the changing volatility observed in financial returns over time. The ARCH model is particularly useful for analyzing financial time series data where the volatility, or the variation in the magnitude of returns, is not constant and can exhibit patterns of clustering or persistence. The **ARCH(1)** model is given by:

$$y_t = \sqrt{\sigma_t^2} \cdot z_t$$
$$z_t \sim D(0, 1)$$

where D is a distribution with mean 0 and variance 1 and in our case $D(0,1) = N(0,1)$. This implies that $y_t = N(\mu, \sigma_t^2)$ and

$$\sigma^2 = \omega + \alpha y_{t-1}^2$$

where $w > 0, \beta > 0, \alpha + \beta < 1$ (in order to have stationarity).

This variables and parameters have a specific meaning in our model:

- y_t is the observed value at time t

- z_t is the white noise (innovation) term at time t
- σ_t^2 is the conditional variance of y_t
- ω is the baseline volatility
- α represents the impact of past squared residuals on the conditional variance

3.1.2. The GARCH model

The **GARCH (Generalized Autoregressive Conditional Heteroskedasticity)** model is an extension of the ARCH (Autoregressive Conditional Heteroskedasticity) model that further captures and models the time-varying volatility in financial and economic time series data. Introduced by Tim Bollerslev in the mid-1980s, the GARCH model addresses some of the limitations of the basic ARCH model by incorporating past values of the conditional variance itself into the volatility modeling process.

Mathematically, a **GARCH(1,1,)** model is given by the following structure:

$$y_t = \sqrt{\sigma_t^2} \cdot z_t$$

$$z_t \sim D(0, 1)$$

$$\sigma^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

As can be seen, the only difference from the previous model is the dependence of volatility on past volatility values (conditional variance) and the introduction of a new β parameter to govern this relationship.

GARCH (Generalized Autoregressive Conditional Heteroskedasticity) and ARCH (Autoregressive Conditional Heteroskedasticity) are both models used to analyze and forecast volatility in financial time series data, such as the volatility of Bitcoin prices. The preference for GARCH over ARCH for modeling Bitcoin volatility is based on several factors:

1. Flexibility and Improved Modeling: GARCH is an extension of the ARCH model that allows for more complex and flexible modeling of volatility dynamics. GARCH models incorporate both lagged conditional variances (as in ARCH) and lagged conditional variances of the squared past returns. This added flexibility often helps capture more intricate volatility patterns observed in financial data like Bitcoin prices.

2. Better Fit to Real Data: Cryptocurrencies like Bitcoin are known for their unique volatility characteristics, including periods of extreme volatility followed by relative stability. GARCH models with their ability to capture changing volatility patterns over time are often better suited to capture these fluctuations and trends in the data.

3. Accommodation of Volatility Clustering: Volatility clustering refers to the phenomenon where periods of high volatility tend to cluster together over time. GARCH models can capture this clustering effect by allowing for the persistence of volatility shocks, making them more suitable for assets like Bitcoin that often exhibit this behavior.

4. More Sophisticated Volatility Forecasting: GARCH models can generate volatility forecasts that are more accurate and reliable compared to ARCH models. This is crucial for risk management and derivative pricing, where accurate volatility forecasts are essential.

5. Statistical Significance and Model Selection: GARCH models often provide more accurate parameter estimates and better model fit, as determined by statistical tests and criteria. This helps in selecting a more appropriate and reliable model for analyzing Bitcoin volatility.

4. Second model

Denote by I_{t-1} the information set observed up to time $t-1$, that is, $I_{t-1} = \{y_{t-1}, i > 0\}$. The general **Markov-switching GARCH** specification can then be expressed as:

$$y_t | (s_t = k, I_{t-1}) \sim D(0, h_{k,t,\xi_k})$$

where $D(0, h_{k,t,\xi_k})$ $D(0, h_{k,t,\xi_k})$ is a continuous distribution with zero mean, time-varying variance h_{k,t,ξ_k} and additional shape parameters gathered in the vector ξ_k. The integer-valued stochastic variable s_t, defined on the discrete space $\{1, ..., K\}$, characterizes the Markov-switching GARCH model.

We define the standardized innovations as $n_{k,t} := y_t / \sqrt{h_{z,t}} \overset{\text{iid}}{\sim} D(0, 1, \xi_k)$