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	LogReturns
16330	465.864	468.174	452.422	457.334	457.334	21056800	NA
16331	456.860	456.860	413.104	424.440	424.440	34483200	-7.4643352
16332	424.103	427.835	384.532	394.796	394.796	37919700	-7.2401507
16333	394.673	423.296	389.883	408.904	408.904	36863600	3.5111240
16334	408.085	412.426	393.181	398.821	398.821	26580100	-2.4967660
16335	399.100	406.916	397.130	402.152	402.152	24127600	0.8317417

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 and 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:

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

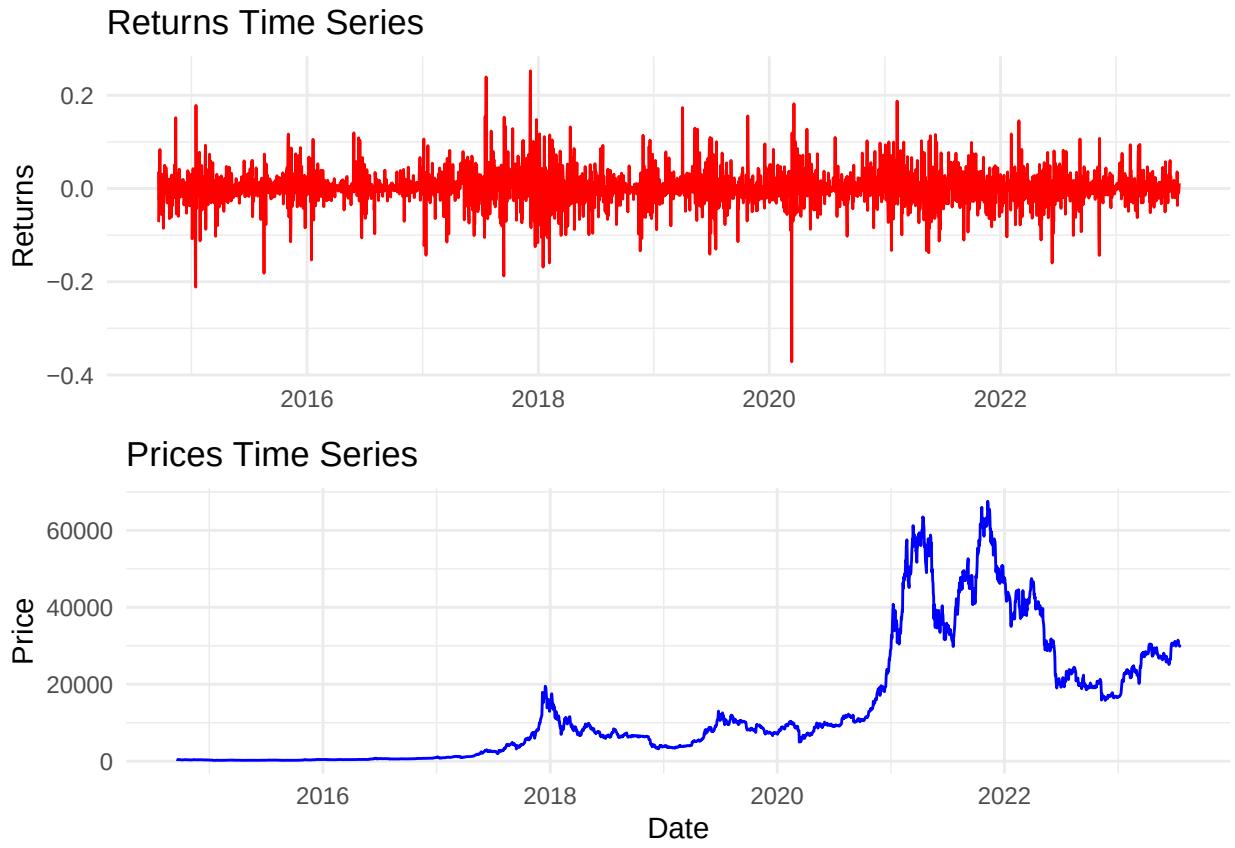
It is a good habit to multiply returns by 100 to express them as a percentage. Some statistical packages are sensitive to the scale of the data. Since log differences can be small, sometimes this creates numerical difficulties.

Thus:

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

Below is a graphical comparison of prices and returns: note some evidence of leverage as descending prices imply higher volatility of returns.

Volatility we will therefore see will be a key issue in choosing our models and will influence our inferential results.



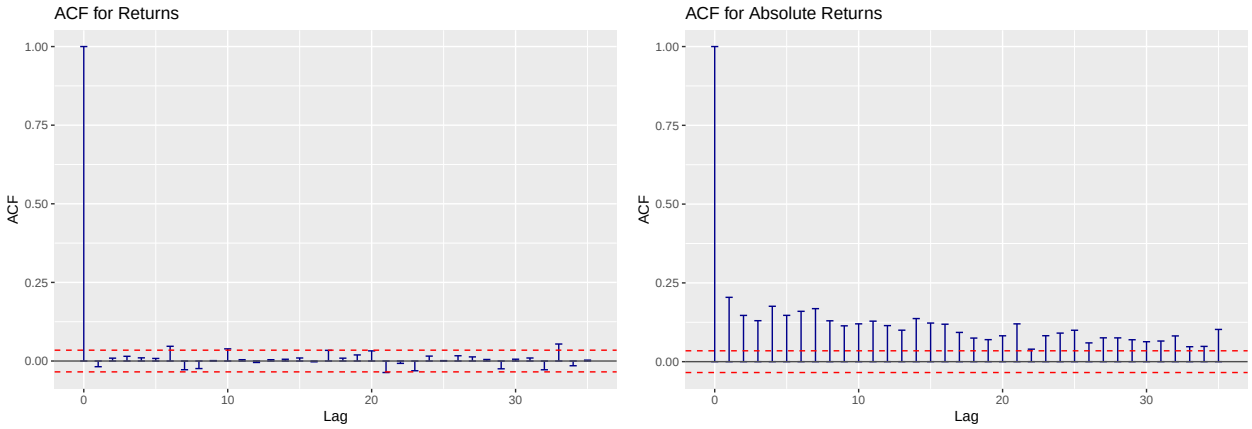
2.2. Volatility and heteroscedasticity

The conditional variance of a time series refers to the variability or dispersion of the data points in the series given the past observations or information. It is a measure of how much the values of the time series fluctuate around their conditional mean, taking into account the historical values of the series. In other words, it quantifies the uncertainty or risk associated with future values of the time series given the available information.

Mathematically, if Y_t represents the value of the time series at time t and the conditional variance of Y_t given the past observations up to time $t-1$, then it can be expressed as:

$$\sigma_t^2 = Var(Y_t|Y_{t-1}, Y_{t-2}, ..., Y_1)$$

Volatility, on the other hand, is a broader concept that generally refers to the degree of variation or dispersion of a financial asset's price (or return) over time. It measures the magnitude of price fluctuations and is often used to assess the risk or uncertainty associated with an investment. In the context of financial markets, volatility is often calculated using various statistical measures, such as standard deviation or variance, to quantify how much an asset's price tends to deviate from its average.



The inspection of the autcorrelograms suggests that:

- returns appear to have weak or no serial dependence
- absolute (and so square) returns appear to have strong serial dependence

In other words we know that the scale of returns changes in time and the (conditional) variance of the process is time varying. In order to capture **volatility clustering**, we need to introduce appropriate time series processes able to model this behavior!

We therefore infer that the volatility of the returns is directly symptomatic of a strong heteroscedasticity of the data.

Heteroscedasticity is the situation in which the variance of the residuals of a regression model is not the same across all values of the predicted variable. In other words, the variability of the residuals (i.e., error term) increases or decreases over the range of predictions.

We therefore infer that the volatility of returns is directly symptomatic of a strong heteroscedasticity of the data. However, it does not make sense to proceed with the visualization of a linear model (lm) fictitious on our data and/or a hypothesis test (e.g., white test) involving the use of an lm. Later we will perform a test to confirm heteroscedasticity and choose an appropriate model (e.g., ARCH model).

2.3. Time series decomposition and stationarity

A given time series is thought to consist of three systematic components including level, trend, seasonality, and one non-systematic component called noise.

These components are defined as follows:

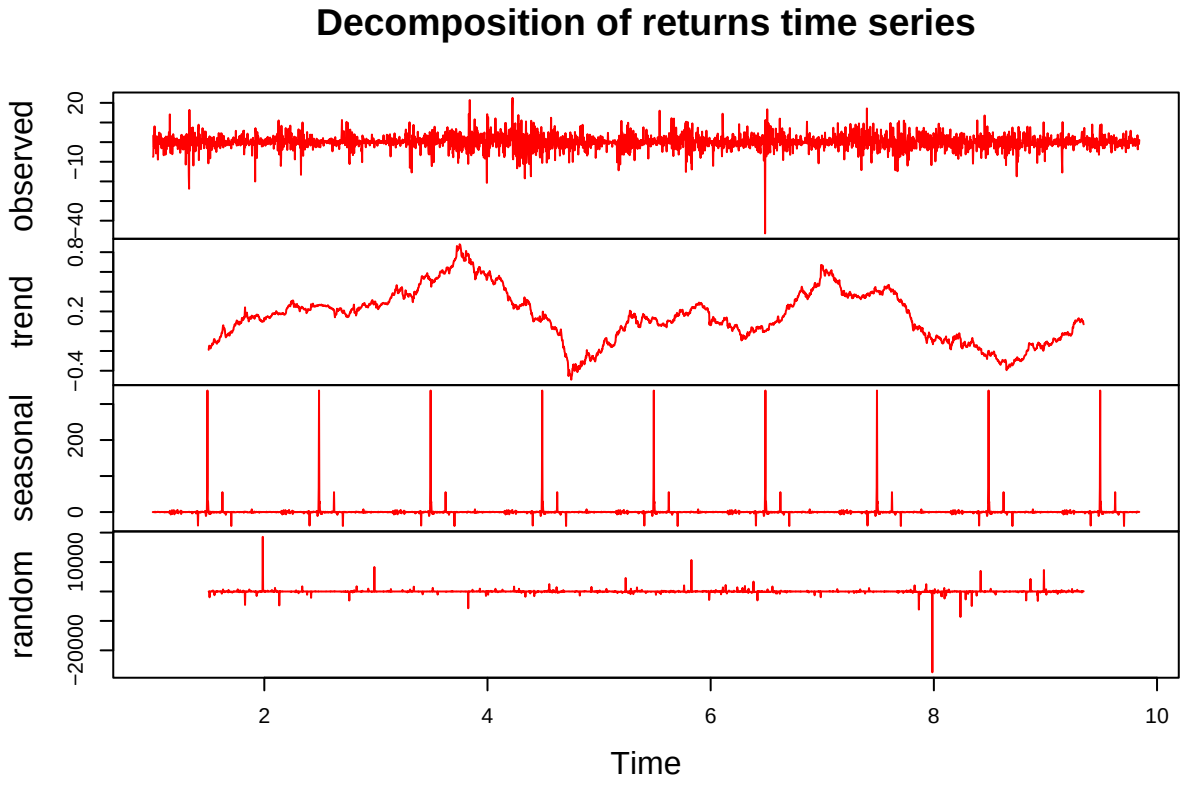
- *Level*: The average value in the series.
- *Trend*: The increasing or decreasing value in the series.
- *Seasonality*: The repeating short-term cycle in the series.
- *Noise (Random)*: The random variation in the series.

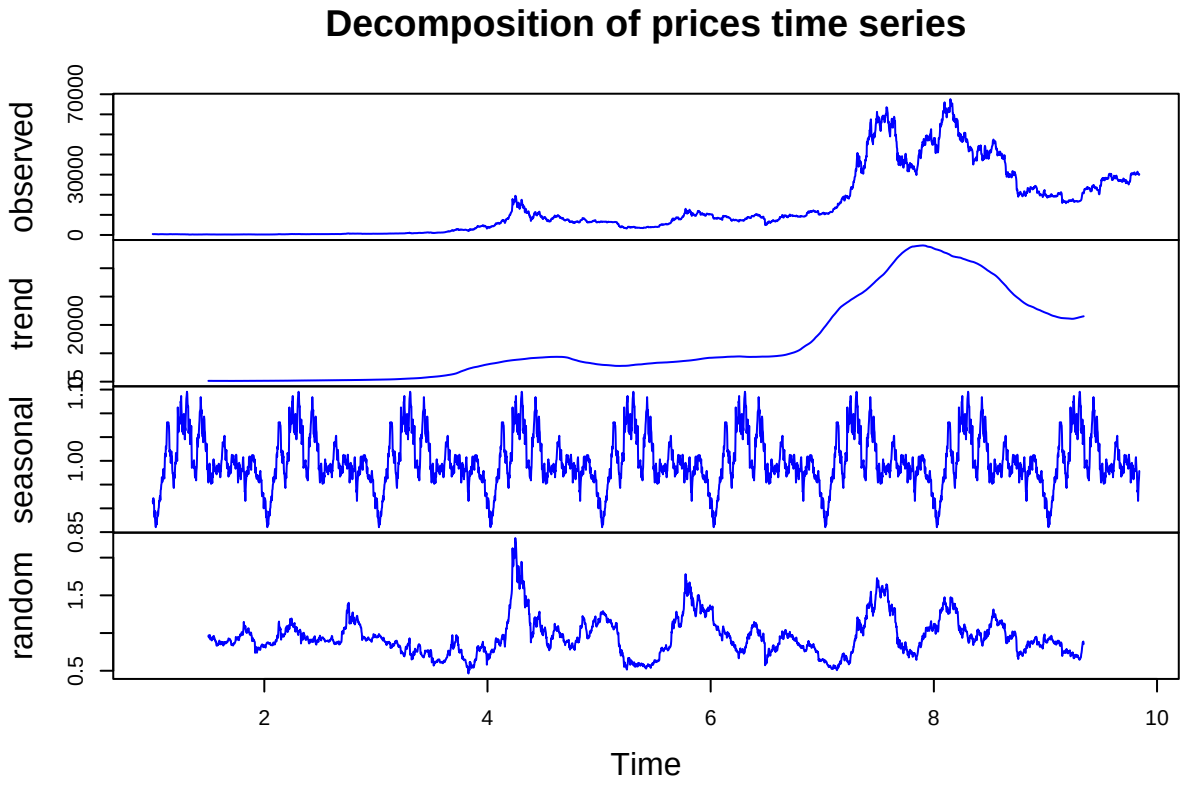
A series is thought to be an aggregate or combination of these four components, all series have a level and noise, while The trend and seasonality components are optional.

In our particular case we decide to consider a **multiplicative model** in which the timeseries is thought as a multiplicative combination of the above components:

$$y(t) = Level \times Trend \times Seasonality \times Noise$$

We report below the decomposition results for *prices* and *returns* considering that we have one observation per day (for 9 years).





2.3.1. Level

Simply observing the **Level** for both our *Price* and *Returns* variables we can propose some obvious findings:

1. our choice of a multiplicative model is justified by obvious nonlinear changes in the values of both variables, i.e. the changes increase and decrease over time;
2. they do not seem to have obvious trends and seasonality;
3. we can guess perhaps some **heteroscedasticity** in the data given large fluctuations in values for adjacent data.

In the following sections we are going to dive into the analysis of the last 2 reported points.

2.3.2. Trend and seasonality

The **Trend** represents the long-term change in the level of a time series. This change can be either upward (increase in level) or downward (decrease in level). If the change is systematic in one direction, then the trend is monotonic. ARIMA-type models for instance are suitable in case of data with an obvious trend, but cannot handle seasonality.

Seasonality refers to periodic fluctuations in certain business areas and cycles that occur regularly based on a particular season. A season may refer to a calendar season such as summer or winter, or it may refer to a commercial season such as the holiday season. An extension of ARIMA models to handle seasonality is the SARIMA model.

In this case, however, we find neither trend nor seasonality

2.3.3 Stationarity

At first glance (with the simple visualization of the observed levels) and based on the previously obtained results we can assume the **stationarity** of our time series. A stationary time series is one whose properties do not depend on the time at which the series is observed.15 Thus, time series with trends, or with seasonality, are not stationary — the trend and seasonality will affect the value of the time series at different times. In general, a stationary time series will have no predictable patterns in the long-term.

the Dickey–Fuller test (Dickey & Fuller, 1979) tests the null hypothesis that a unit root is present in an autoregressive (AR) time series model. The alternative hypothesis is different depending on which version of the test is used, but is usually stationarity or trend-stationarity.

Given that A simple AR model is given by:

$$y_t = \rho y_{t-1} + u_t$$

where:

- y_t is the variable of interest (returns) at time t;
- ρ is a coefficient;
- u_t is an error term.

Write the regression model as:

$$\Delta y_t = (\rho - 1) y_{t-1} + u_t = \delta y_{t-1} + u_t$$

where Δ is the first difference operator and $\delta = \rho - 1$. This model can be estimated, and testing for a unit root is equivalent to testing $\delta = 0$.

The **augmented Dickey–Fuller (ADF) statistic** (augmented since it removes all the structural - autocorrelation - effects in the time series), used in the test, is a negative number. The more negative it is, the stronger the rejection of the hypothesis that there is a unit root at some level of confidence.

```
##
##  Augmented Dickey-Fuller Test
##
```

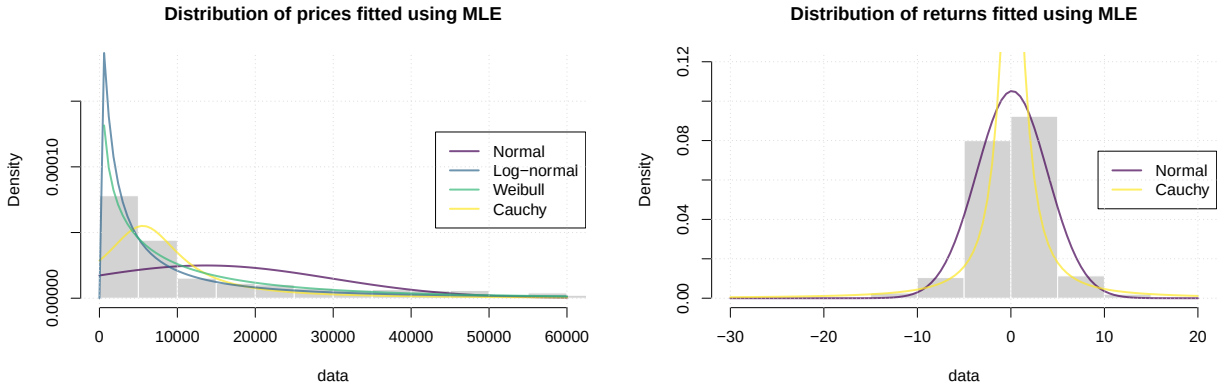
```
## data:  data$LogReturns[-1]
## Dickey-Fuller = -13.988, Lag order = 14, p-value = 0.01
## alternative hypothesis: stationary
```

We got evidence against the the null hypothesis, thus we simply verified the stationarity of our data!

This aspect is going to be a central point since the models we are going to use in order to well represent volatility often imply stationarity of the time series.

2.5. Searching for a prior

We can then begin to analyze the distribution of our data. Let's start with a simple visualization of the distribution (histogram) of our data and then try to fit some possible density functions by selecting a few by intuition and using MLE for parameters.



We note that fortunately for log-returns a normal distribution seems quite consistent for a possible prior assumption in Bayesian framework.

However, we are not satisfied with a simple visualization of this kind and continue the investigation with a **Kolmogorov-Smirnov test**, a *goodnes of fit* test. The resulting p-value, however, indicates that the population does not actually exhibit normal distribution, rejecting the null hypothesis.

```
##
##  Asymptotic one-sample Kolmogorov-Smirnov test
##
## data:  data$LogReturns[-1]
## D = 0.47775, p-value < 2.2e-16
## alternative hypothesis: two-sided
```

We then disprove our initial assumption of a normal prior and via a *QQ plot* the obvious differences between the current data and the normal distribution that we were previously unable to capture.

Traditionally, the errors (and so the returns) have been assumed to be Gaussian, however, it has been widely acknowledged that financial returns display fat tails and are not conditionally Gaussian. Gaussian GARCH models cannot quite capture the key properties of asset returns and to address this, researchers (Bollerslev (1987), He & Teräsvirta (1999) and Bai et al. (2003)) have explored alternative distributions to model asset returns. One approach is to assume that the returns are IID and follow a Student's t-distribution (t-Student) instead of a normal distribution. The t-Student distribution allows for fatter tails, meaning it assigns higher probabilities to extreme events compared to the normal distribution. For example we'll try to use a GARCH model with a t-student prior distribution for returns:

$$y_t = \sigma_t z_t \sim t(d)$$

3. First model

3.1. ARCH vs GARCH

3.1.1. The ARCH model

We veryfied in point _2.1 time series exhibiting conditional heteroscedasticity or autocorrelation in the squared series is said to have autoregressive conditional heteroscedastic -ARCH)

Effects.

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

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

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:

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

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

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

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