# 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

## 1.1 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. The data set

# 3. First model

## 3.1. ARCH vs GARCH

### 3.1.1. The ARCH model

In order to capture volatility clustering, in 1982 Robert Engle proposed the *AutoRegressive Conditional Heteroskedasticity* (ARCH) model. 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

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

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

# 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 $h_{k,t,\xi_k}$ and additional shape parameters gathered in the vector $\xi_k$. 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 $n_{k,t} := y_t/\sqrt{h_{z,t}} \overset{\text{iid}}{\sim} D(0, 1, \xi_k)$