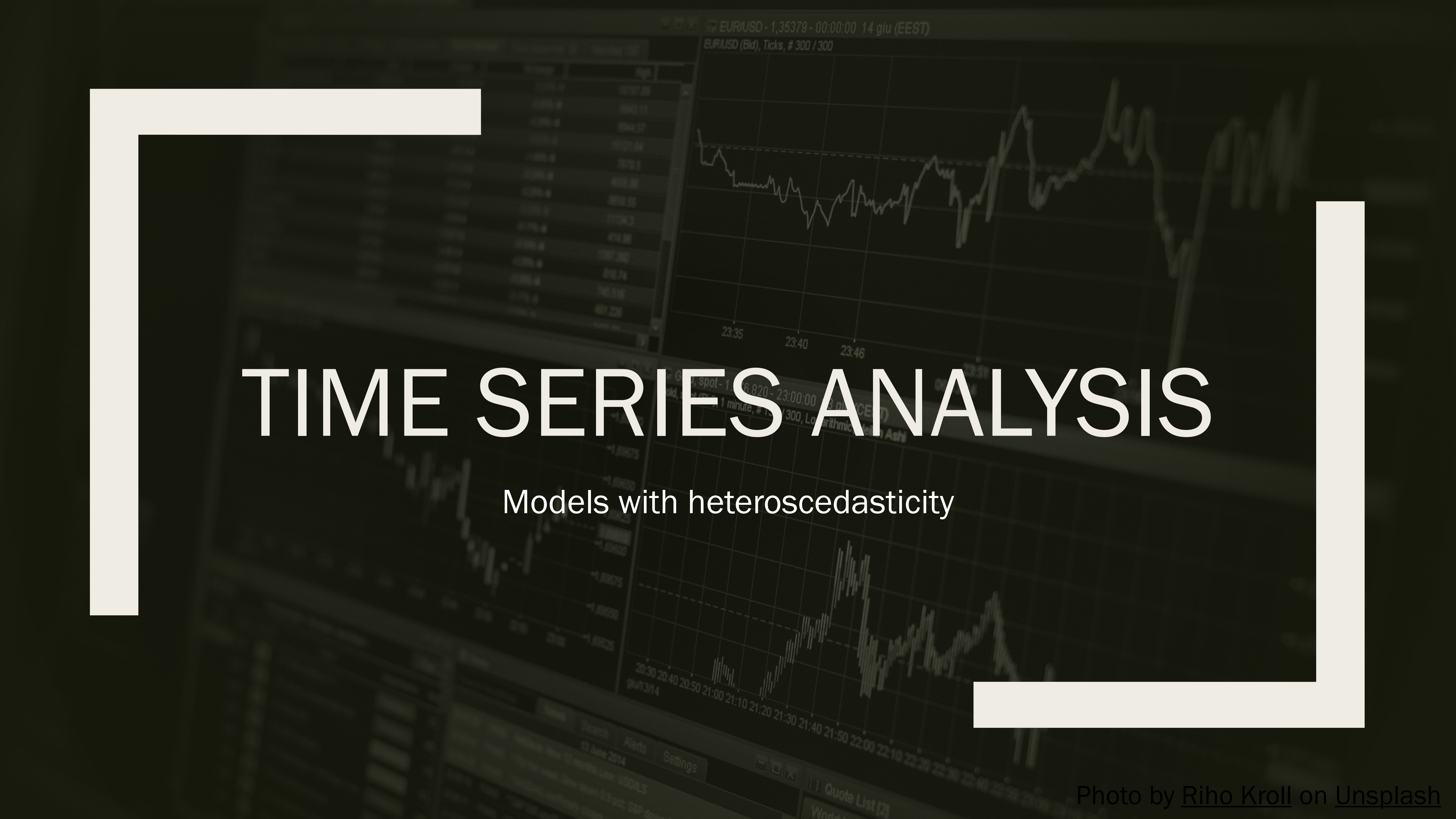


TIME SERIES ANALYSIS

Models with heteroscedasticity

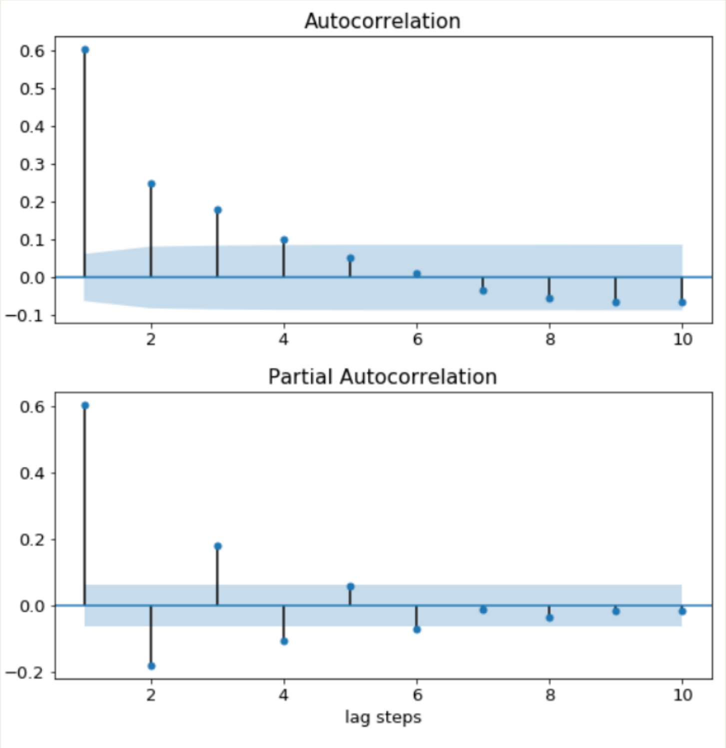
Photo by Riho Kroll on Unsplash

Time series analysis with ARIMA models (reminder)

- A time series is a series of data points indexed (or listed or graphed) in time order (usually at successive equally spaced points in time):

$$y_1, y_2, \dots, y_t, \dots, \qquad y_t \in R$$

- ARIMA (p,d,q) = we turn a non-stationary series into a stationary one by differentiating it d times in a row and apply the model ARMA (p,q)
- PACF shows the order of AR part (the last significant lag defines parameter p)
- ACF shows the order of MA part (the last significant lag defines parameter q)
- Choose the best model using AIC and BIC criteria

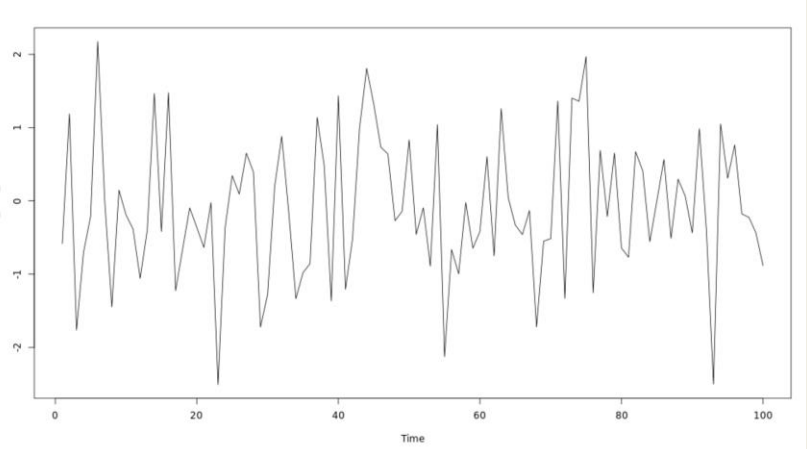


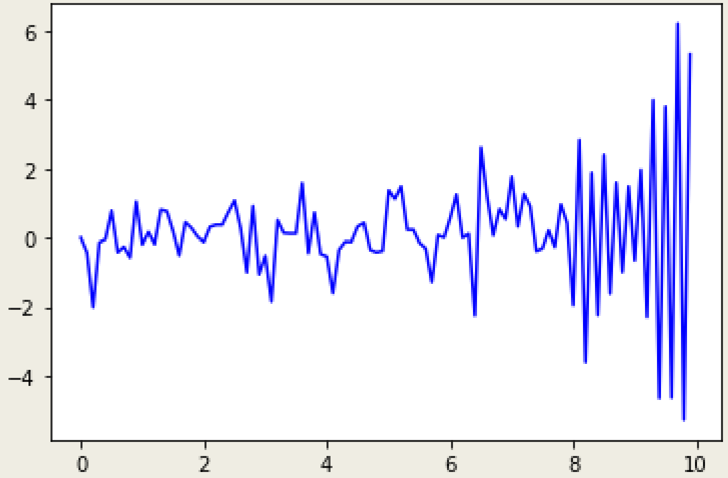
Time series analysis with ARIMA models (reminder)

- A time series is a series of data points indexed (or listed or graphed) in time order (usually at successive equally spaced points in time):

$$y_1, y_2, \dots, y_t, \dots, \qquad y_t \in R$$

- ARIMA (p,d,q) = we turn a non-stationary series into a stationary one by differentiating it d times in a row and apply the model ARMA (p,q)
- One of the assumptions is the homoscedasticity of the errors, i.e. the constant variance



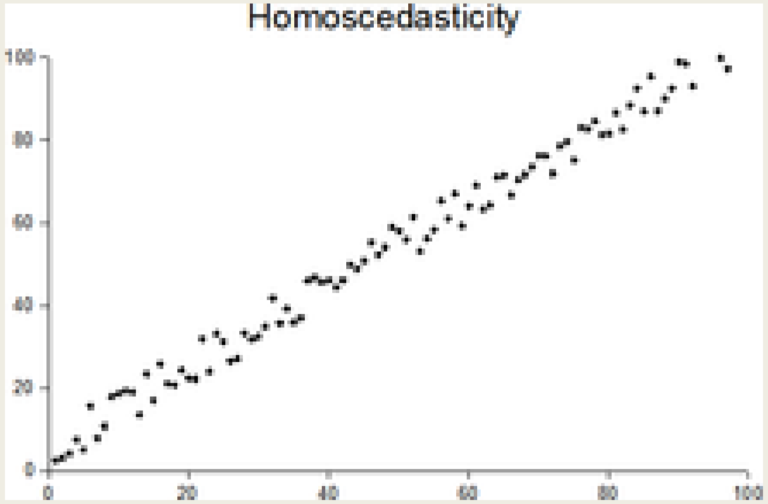


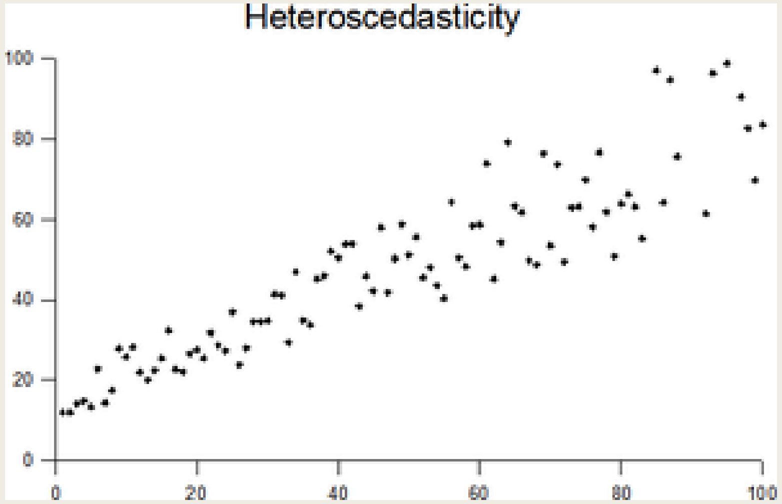
Time series analysis with ARIMA models (reminder)

- A time series is a series of data points indexed (or listed or graphed) in time order (usually at successive equally spaced points in time):

$$y_1, y_2, \dots, y_t, \dots, \qquad y_t \in R$$

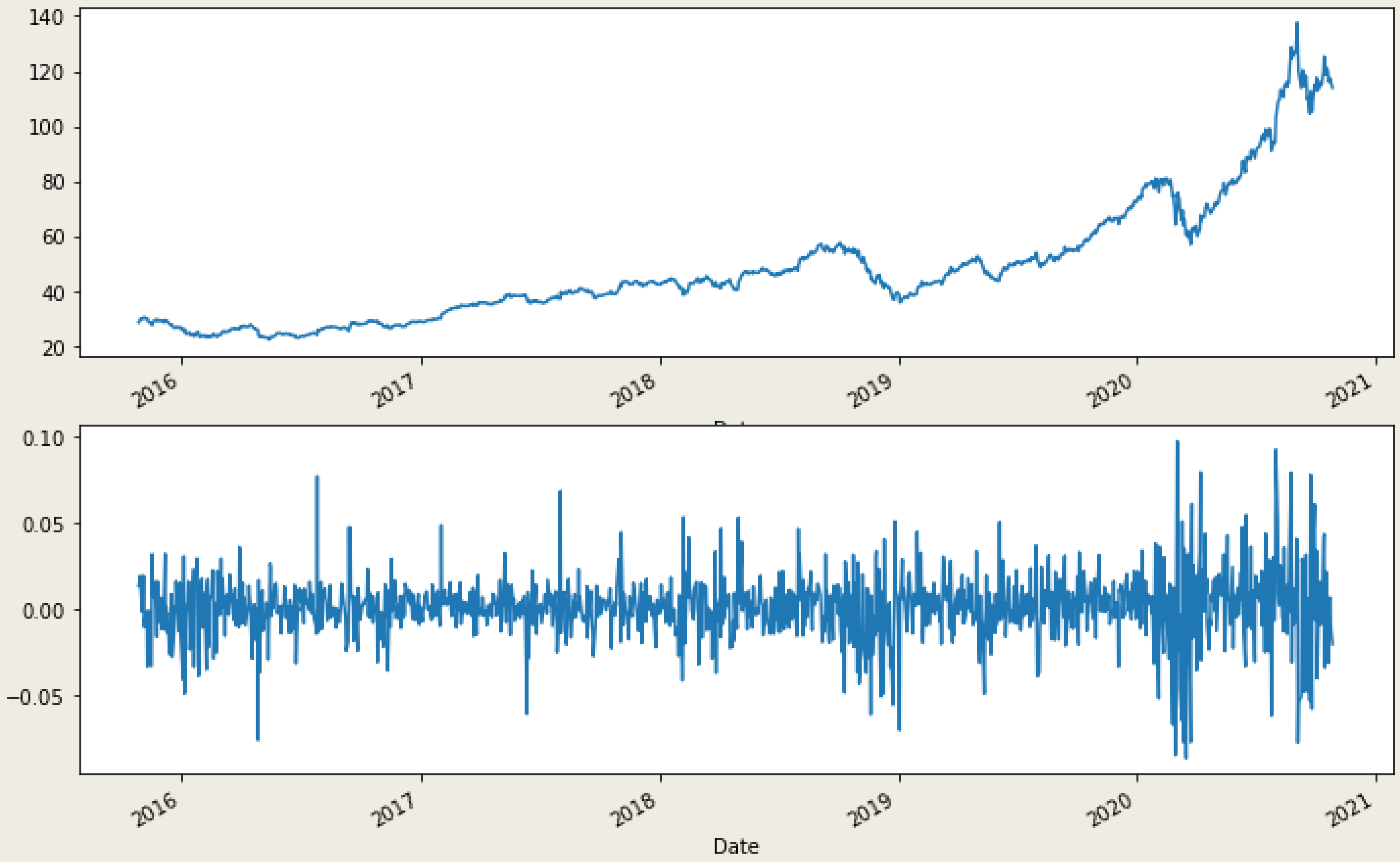
- ARIMA (p,d,q) = we turn a non-stationary series into a stationary one by differentiating it d times in a row and apply the model ARMA (p,q)
- One of the assumptions is the homoscedasticity, i.e. the constant variance.





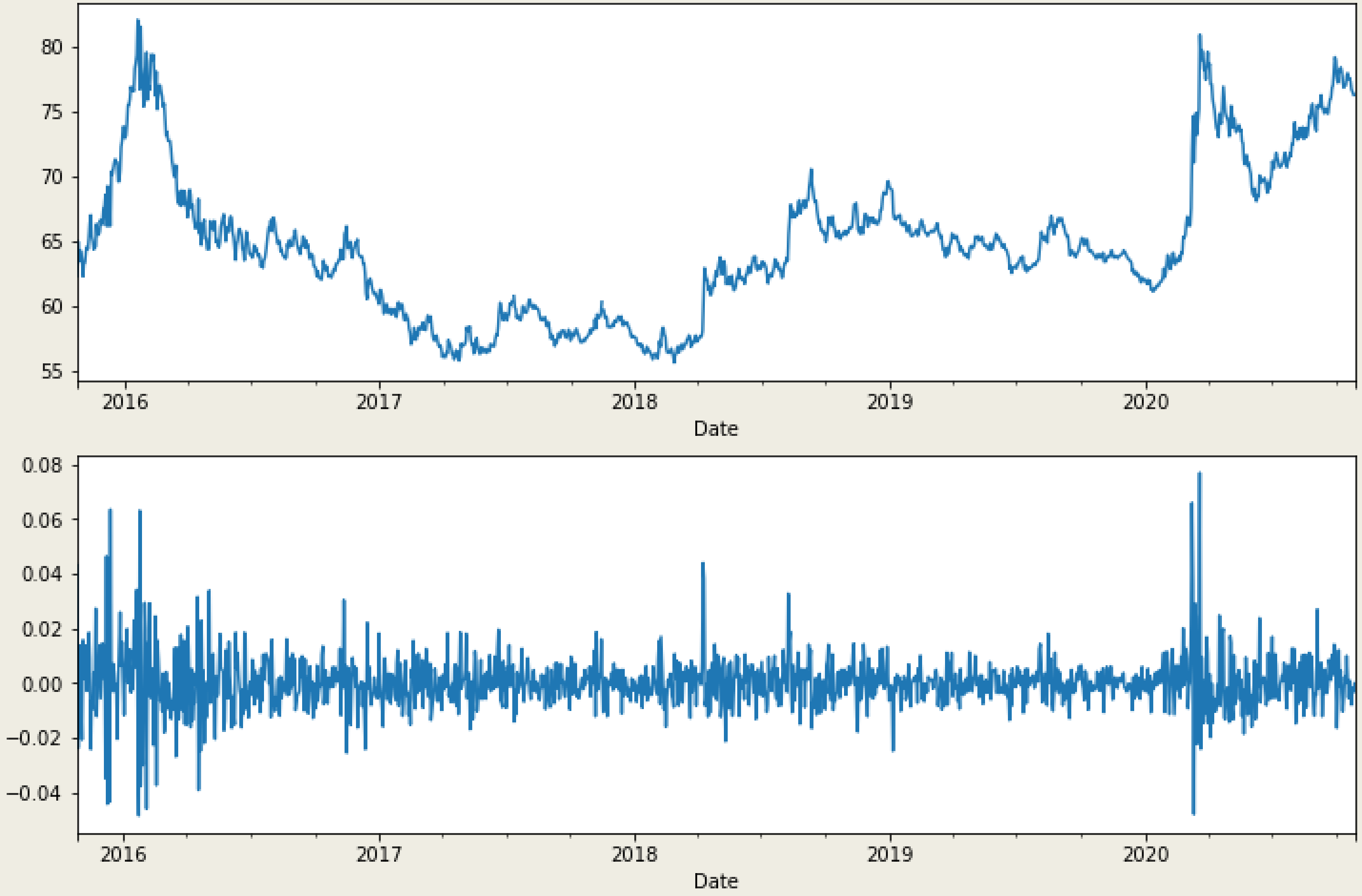
ARIMA cannot catch the clustering of the volatility

- Unfortunately, in many real-life problems time series has **heteroscedasticity** (when variance might change over time), especially in finance.
- Apple stock prices for Oct 27, 2015 - Oct 27, 2020 and its return $r_i = \frac{S_i - S_{i-1}}{S_{i-1}}$
- Average return is 0, but its variability changes over time



ARIMA cannot catch the clustering of the volatility

- Unfortunately, in many real-life problems time series has **heteroscedasticity** (when variance might change over time), especially in finance.
- USD/RUB for Oct 27, 2015 - Oct 27, 2020 and its % change of the currency pair USD / RUB

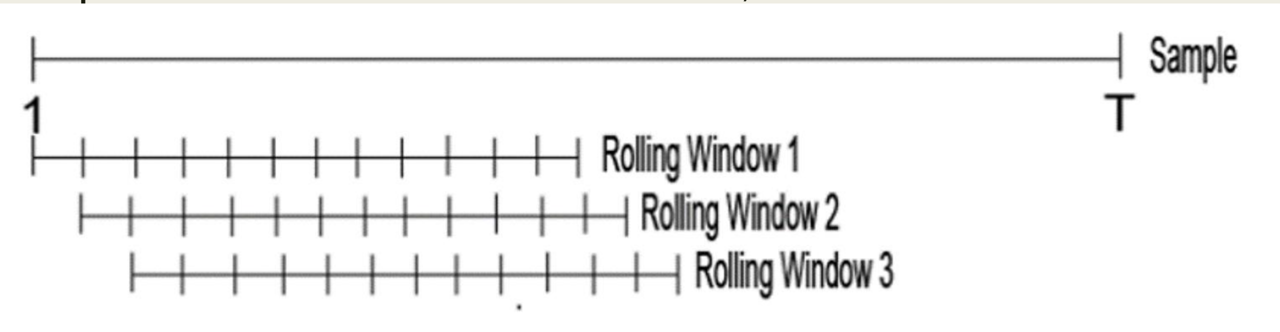


ARIMA cannot catch the clustering of the volatility (i.e. heteroscedasticity)

- Volatility is the degree of variation (dispersion) of financial asset prices over time, usually measured by the standard deviation or variance of price returns:
- Returns $r_i = \frac{S_i - S_{i-1}}{S_{i-1}}$
- Mean return $m = \frac{\sum_{i=1}^{n} r_i}{n}$ and standard deviation $\sigma = \sqrt{\frac{\sum_{i=1}^{n}(r_i - m)^\wedge 2}{n-1}}$
- The higher the volatility, the riskier a financial asset

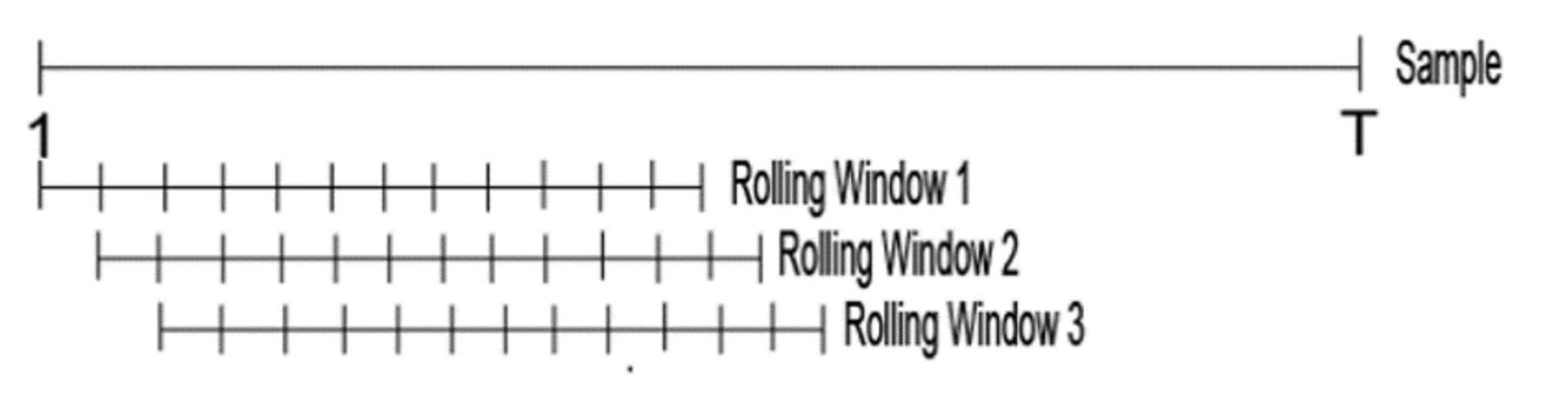
ARIMA cannot catch the clustering of the volatility

- Volatility is the degree of variation (dispersion) of financial asset prices over time, usually measured by the standard deviation or variance of price returns:
- Returns $r_t = \frac{S_t - S_{t-1}}{S_{t-1}}$
- (unconditional) mean return $m = \frac{\sum_{t=1}^{n} r_t}{n}$ and (unconditional) standard deviation $\sigma = \sqrt{\frac{\sum_{t=1}^{n} t-m)\hat{}2}{n-1}}$
- The higher the volatility, the riskier a financial asset
- If we use the rolling window to compute the standard deviation σ, then we find out that σ depends on time σ_t



ARIMA cannot catch the clustering of the volatility

- Returns $r_t = \frac{S_t - S_{t-1}}{S_{t-1}}$
- (unconditional) mean return $m = \frac{\sum_{t=1}^{n} r_t}{n}$ and (unconditional) standard deviation σ $= \sqrt{\frac{\sum_{t=1}^{n} t-m)\hat{}2}{n-1}}$
- If we use the rolling window to compute the standard deviation σ, then we find out that σ depends on time σ_t



ARIMA cannot catch the clustering of the volatility

Idea of GARCH models

- The estimation of σ_t requires modelling. We will first use GARCH models (Generalized AutoRegressive models with Conditional Heteroscedasticity)
- ARCH model was invented by R.Engle and their generalized version GARCH was invented by T.Bollerslev.
- Our goal is to predict future return r_t using the information available at time t: $I_t = \{r_{t-1}, r_{t-2}, \dots\}$
- We can compute the (conditional) mean return $\mu_t = E(r_t|I_t)$, but it is not perfect and subject to error $\varepsilon_t = r_t - \mu_t$
- The volatility of return is $\sigma_t = \sqrt{Var(r_t|I_t)}$, where $\sigma_t^2 = Var(r_t|I_t) = E((r_t - \mu_t)^2|I_t) = E(\varepsilon_t^2|I_t)$ is a (conditional) variance

Idea of ARCH models

Auto Regressive:
future values depends on past values

Conditional Heteroscedasticity:
Volatility as a weighted average of past information

- The estimation of σ_t requires modelling. We will first use GARCH models (Generalized Auto Regressive models with Conditional Heteroscedasticity)
- ARCH model was invented by R.Engle and their generalized version GARCH was invented by T.Bollerslev.
- Our goal is to predict future return r_t using the information available at time t: $I_t = \{r_{t-1}, r_{t-2}, \dots\}$
- We can compute the (conditional) mean return $\mu_t = E(r_t|I_t)$, but it is not perfect and subject to error $\varepsilon_t = r_t - \mu_t$
- The volatility of return is $\sigma_t = \sqrt{Var(r_t|I_t)}$, where $\sigma_t^2 = Var(r_t|I_t) = E((r_t - \mu_t)^2|I_t) = E(\varepsilon_t^2|I_t)$ is a (conditional) variance.
- Volatility is related to the residuals $\varepsilon_t = \sigma_t \cdot u_t$, where u_t is a white noise (uncorrelated random variables with zero mean and a finite variance)

ARCH model

- The expected value μ_t can be found as sample mean for the last M values: mean with rolling window $\mu_t = \frac{\sum_{i=1}^{M} r_t - i}{M}$, or use time series model like ARMA
- The same idea for the variance: variance with rolling window $\sigma_t^2 = \frac{\sum_{i=1}^{M} \varepsilon_{t-i}^2}{M}$ or ARCH(p) model $\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2$ (here we have weights)
- ARCH(1): $\sigma_t^2 = \omega + \alpha\ \varepsilon_{t-1}^2$

ARCH model

- The expected value μ_t can be found as sample mean for the last M values: mean with rolling window $\mu_t = \frac{\sum_{i=1}^{M} r_t - i}{M}$, or use time series model like ARMA
- The same idea for the variance: variance with rolling window $\sigma_t^2 = \frac{\sum_{i=1}^{M} \varepsilon_{t-i}^2}{M}$ or ARCH(p) model $\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2$ (here we have weights)
- ARCH(1): $\sigma_t^2 = \omega + \alpha\ \varepsilon_{t-1}^2$
- GARCH(p,q): $\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1q} \beta_i \sigma_{t-i}^2$(here the variance depends not only on previous squared prediction error, but also on the previous variance prediction)
- GARCH(1,1): $\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$

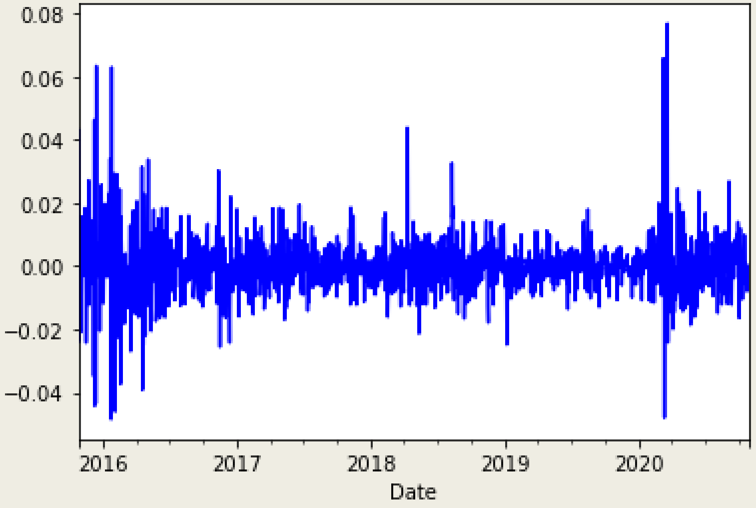
GARCH model

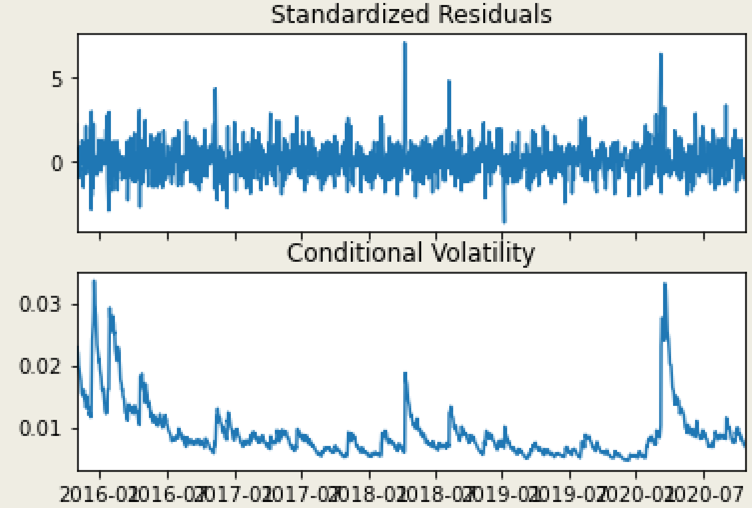
- GARCH(1,1) model: $\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$
- The larger the α, the bigger the immediate impact of the shock
- The larger the β, the longer the duration of the impact
- There are several restrictions on the parameters:
  - *ω, α and β must be positive to have positive variance σ_t^2*
  - *$\alpha + \beta < 1$ to ensure that σ_t^2 will return to the ω over time (the long run variance will be $\frac{\omega}{1-\alpha-\beta}$)*
- In the normal GARCH model we assume $r_t = \mu_t + \varepsilon_t$, where $\varepsilon_t \sim \mathrm{N}(0, \sigma_t^2)$, so the standardized returns $\frac{r_t-\mu_t}{\sigma_t} \sim N(0,1)$

Python "arch" package

- https://bashtage.github.io/arch/univariate/introduction.html
- Kevin Sheppard. (2020, June 24). bashtage/arch: Release 4.15 (Version 4.15). Zenodo. https://doi.org/10.5281/zenodo.593254
- A complete ARCH model is divided into three components:
  - *a mean model: "constant" (default), "zero" , "AR";*
  - *a volatility model: "GARCH" (default), "ARCH" , "EGARCH"*
  - *a distribution for the standardized residuals: "normal" (default), "t" , "skewt".*
- Estimation is done by maximum likelihood method

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






```python
garch_result = garch_model.fit()

print(garch_result.summary())
```

```
Iteration:      1,   Func. Count:      6,   Neg. LLF: 25588303783.447334
Iteration:      2,   Func. Count:     20,   Neg. LLF: 9.287725591613866e+18
Iteration:      3,   Func. Count:     35,   Neg. LLF: 26797269377.987156
Iteration:      4,   Func. Count:     48,   Neg. LLF: -4413.431043823735
Optimization terminated successfully    (Exit mode 0)
            Current function value: -4413.431048171028
            Iterations: 8
            Function evaluations: 48
            Gradient evaluations: 4
                     Constant Mean - GARCH Model Results
==============================================================================
Dep. Variable:                 return   R-squared:                      -0.001
Mean Model:             Constant Mean   Adj. R-squared:                 -0.001
Vol Model:                      GARCH   Log-Likelihood:                4413.43
Distribution:                  Normal   AIC:                          -8818.86
Method:            Maximum Likelihood   BIC:                          -8798.17
                                        No. Observations:                 1305
Date:                Wed, Oct 28 2020   Df Residuals:                     1301
Time:                        21:44:50   Df Model:                            4
                                 Mean Model
===============================================================================
                 coef    std err          t      P>|t|        95.0% Conf. Int.
-------------------------------------------------------------------------------
mu         -1.5870e-04  1.151e-05    -13.783  3.226e-43 [-1.813e-04,-1.361e-04]
                              Volatility Model
============================================================================
                 coef    std err          t      P>|t|       95.0% Conf. Int.
----------------------------------------------------------------------------
omega      2.0179e-06  1.375e-11  1.467e+05      0.000 [2.018e-06,2.018e-06]
alpha[1]       0.1000  3.391e-02      2.949  3.189e-03   [3.354e-02,  0.166]
beta[1]        0.8800  2.473e-02     35.587 2.225e-277   [  0.832,  0.928]
============================================================================
```

Mean Models

arch_model(data, p = 1, q = 1, mean = ‘zero', vol = 'GARCH', dist = 'normal')

- zero - Model with zero conditional mean
- constant - Constant mean model
- AR – Model with the mean as an autoregressive (AR) process: $\mu_t = \mu + \theta * (r_{t-1} - \mu) + \varepsilon_t$

A task for you!!

- Simulate the GARCH model
- Send me the csv file with the simulated data (index is arange, please) and the python notebook with the code also
- You can send the initial data (like stock prices from example) s.t. their differences or returns will be modelled by the GARCH model
- lyude@inbox.ru, legorova@hse.ru

Significance of the parameters

- Null hypothesis (H0): parameter value = 0 (check whether we can have the results by chance)
- If H0 cannot be rejected, you should leave out the parameter. Common threshold is 1% or 5%
- Reject the null hypothesis if p-value < signicance level

Mean Model

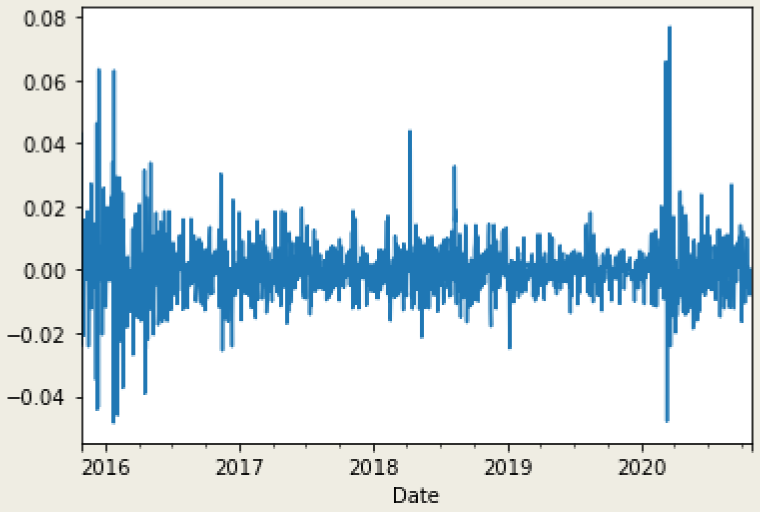
	coef	std err	t	P>|t|	95.0% Conf. Int.
mu	-1.5870e-04	1.151e-05	-13.783	3.226e-43	[-1.813e-04,-1.361e-04]

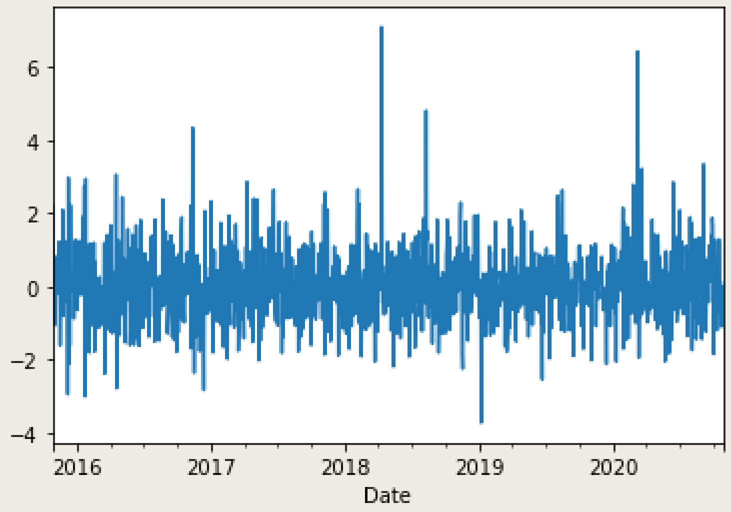
Volatility Model

	coef	std err	t	P>|t|	95.0% Conf. Int.
omega	2.0179e-06	1.375e-11	1.467e+05	0.000	[2.018e-06,2.018e-06]
alpha[1]	0.1000	3.391e-02	2.949	3.189e-03	[3.354e-02, 0.166]
beta[1]	0.8800	2.473e-02	35.587	2.225e-277	[0.832, 0.928]

Goodness-of-fit

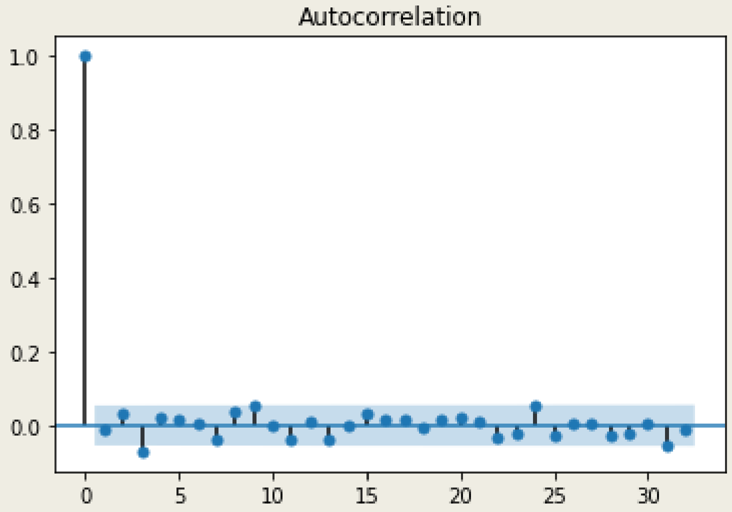
- Visual check: compare the initial return data and the standardized residuals – they should be like white noise and do not have clear clustering





Goodness-of-fit

- Check autocorrelation in the standardized residuals: existence of autocorrelation in the standardized residuals indicates the model may not be sound
- Or use Ljung-Box test, which test whether any of a group of autocorrelations of a time series are different from zero
- H0: the data is independently distributed
- And if P-value < 5%: the model is not sound

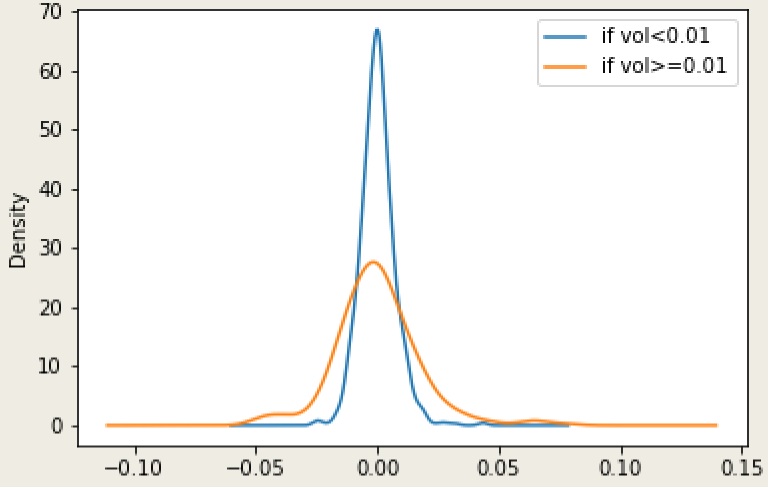


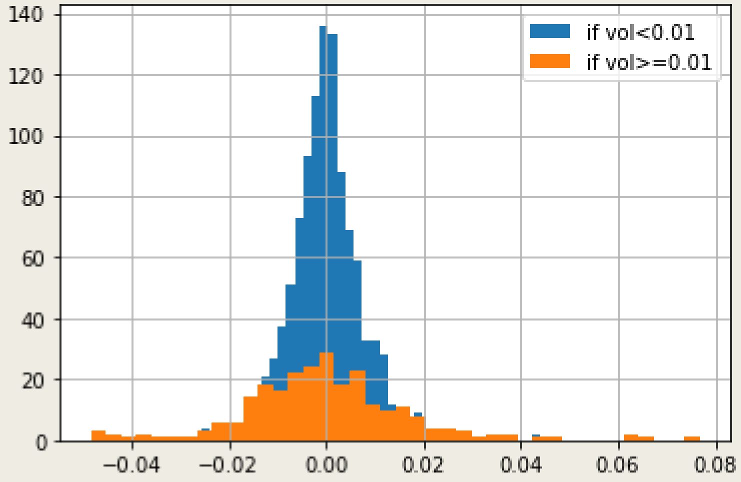
Goodness-of-fit

- To choose between different models, you may use Likelihood function and information criteria
- Likelihood function = the probability of getting the data observed under the assumed model and larger likelihood values must be preferred
- AIC (Akaike's Information Criterion) and BIC (Bayesian Information Criterion) penalizes for the more complexity of the models, so the lower information criterion is better.

Further improvement of GARCH models

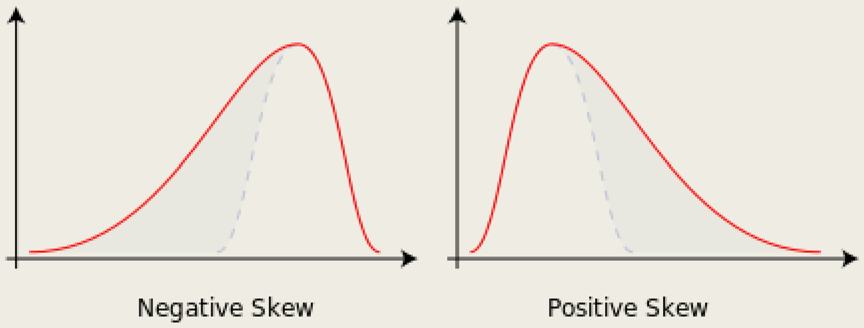
- Markets take the stairs up, but the elevators down!
- In the normal GARCH model we assume $r_t = \mu_t + \varepsilon_t$, where $\varepsilon_t \sim \mathrm{N}(0, \sigma_t^2)$, so the standardized returns $\frac{r_t - \mu_t}{\sigma_t} = \frac{\varepsilon_t}{\sigma_t} \sim N(0,1)$
- But the distribution might be skewed and have "fat tails"



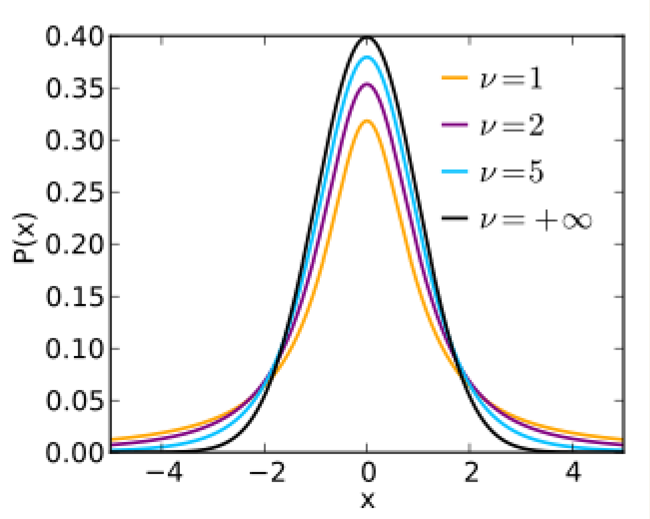


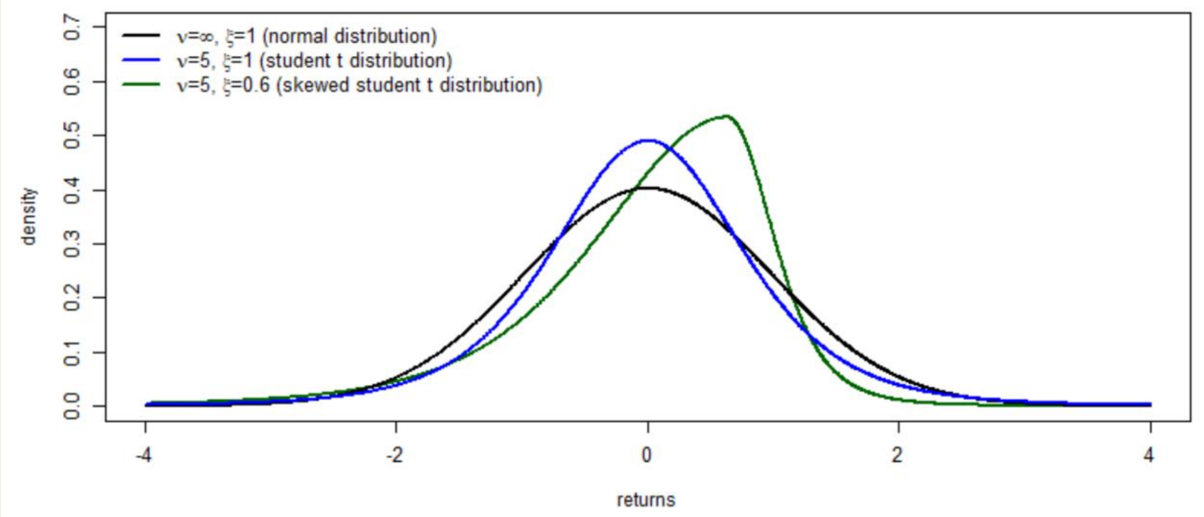
Fat tails and skewness

- Skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean



Student's *t* distribution with $\nu = n - 1$ degrees of freedom and skewed Student's distibution





Asymmetric shocks

- The asymmetry of news: bad news (negative shocks) usually have a greater impact on volatility than good news (positive shocks), that is, volatility is higher in a falling market than in a growing one.
- This effect is sometimes called the effect of leverage, which is associated with one of the explanations for this phenomenon that stock prices decline, increasing the financial leverage of companies, and hence the level of risk (which corresponds to greater volatility).
- Within the framework of classical GARCH models, this effect cannot be explained, since the conditional variance depends on the squares of the past values of the series and does not depend on the signs.

Asymmetric shocks

- The asymmetry of news: bad news (negative shocks) usually have a greater impact on volatility than good news (positive shocks), that is, volatility is higher in a falling market than in a growing one.
- GJR-GARCH proposed by Glosten, Jagannathan and Runkle:

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \gamma\varepsilon_{t-1}^2 I_{t-1} + \beta\sigma_{t-1}^2, \text{ where } I_{t-1} = \begin{cases} 0, if\ r_{t-1} \geq \mu, \\ 1, if\ r_{t-1} < \mu. \end{cases}$$

- In python the additional parameter o should be introduced, which includes o lags of an asymmetric shock which transforms a GARCH model into a GJR-GARCH

Asymmetric shocks

- The asymmetry of news: bad news (negative shocks) usually have a greater impact on volatility than good news (positive shocks), that is, volatility is higher in a falling market than in a growing one.
- Another way is to use EGARCH (exponential GARCH) by Nelson & Cao:

$$\log \sigma_t^2 = \omega + \alpha g(z_{t-1}) + \beta \log \sigma_{t-1}^2 \text{, where } g(z_{t-1}) = \theta z_t + \lambda(|z_t| - E(|z_t|))$$

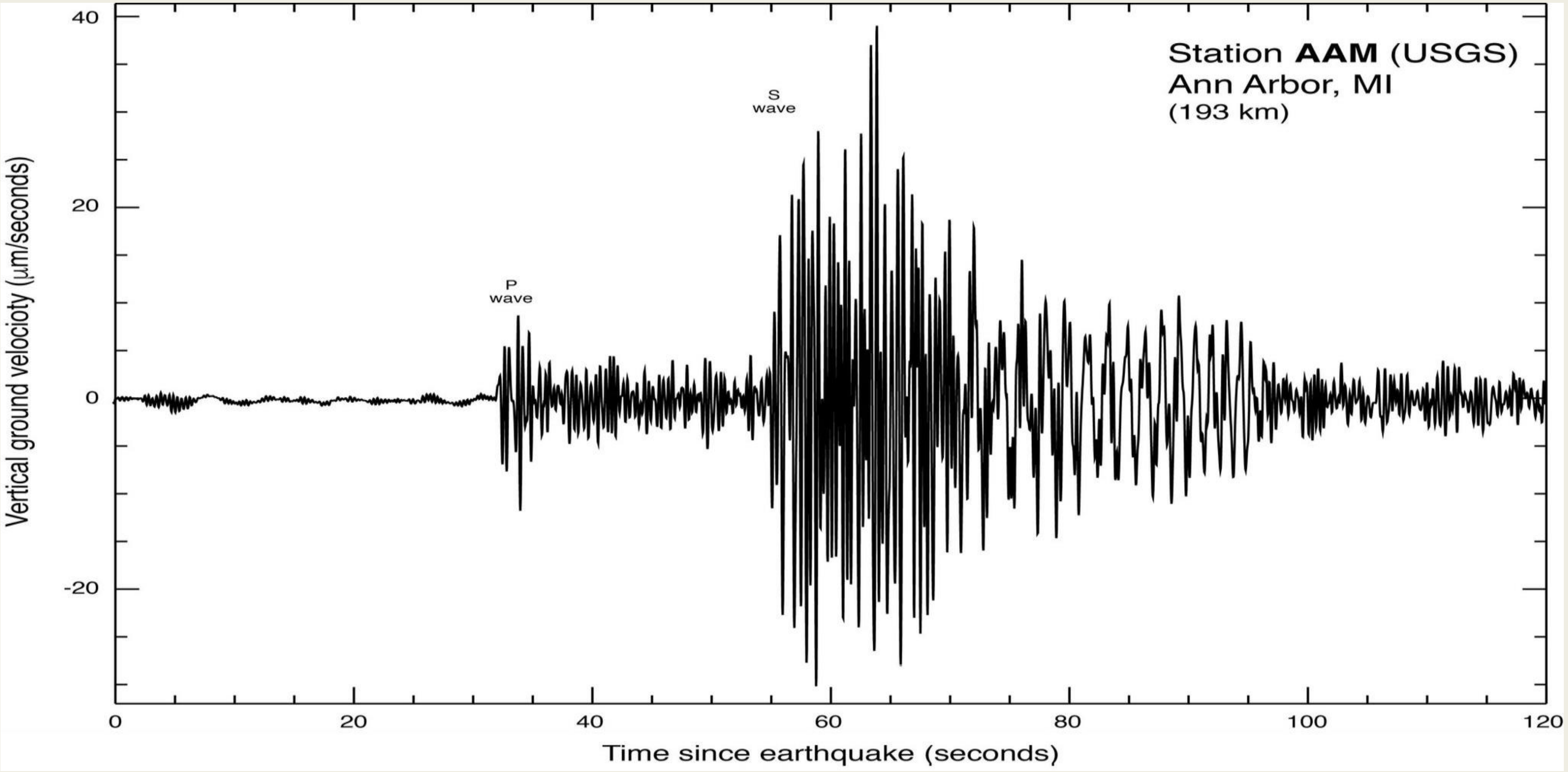
- z_t may be a standard normal variable. The formulation for z_t allows the sign and the magnitude of z_t to have separate effects on the volatility.
- Since z_t may be negative, there are no sign restrictions for the parameters.
- In python the additional parameter o should be introduced, which includes o lags of an asymmetric shock which transforms a GARCH model into a GJR-GARCH

A task for you!!

- Fit the GARCH models to the data sent to you and find the best model.
- Send me the python notebook with the code where I can check all the models you've tried
- Do not delete or replace the models, but copy them if you want to change the parameters.
- Write the answer in the end – what model should be preferred and why.

- You may send me (legorova@hse.ru) your report till October 30, 23:59.

Not only finance!



THANK YOU!

Photo by Riho Kroll on Unsplash