Ministry of Education and Science of Ukraine
National University of Kyiv-Mohyla Academy
Faculty of Computer Sciences
Department of Mathematics

Qualification work

Bachelor's degree

on the theme: "Non linear stochastic models for time series analysis of stock volatilities"

By the student of 4th year
Degree program
"Applied mathematics"
Speciality
113 "Applied mathematics"
Fisun Yelyzaveta

Supervisor:
PhD, Docent, *Shchestyuk N.Yu.*

Reviewer:
Dr.Sc., *Aryasova O.V.*

Qualification work is defended
with a grade of:

____

Secretary of EC:

____
(signature)

"____"________________ 2022

Kyiv – 2022

Ministry of Education and Science of Ukraine
National University of Kyiv-Mohyla Academy
Faculty of Computer Sciences
Department of Mathematics

APPROVED
Head of Department of Mathematics,
Prof., PhD, *Oliynyk B.V.*

____
(signature)

"____"________________ 2021

INDIVIDUAL TASK
for Qualification work
of the student of 4^{th} year of Computer Sciences faculty
Fisun Yelyzaveta

Theme: Non linear stochastic models for time series analysis of stock volatilities.
Output data: Studied non linear stochastic models for financial time series
Text part content of the thesis:

Date of issue "____"__________ 2021 Supervisor ____________________
(signature)
Task resieved ____________________
(signature)

Topic: "Non linear stochastic models for time series analysis of stock volatilities".

Timestamps:

	Name of the step	Deadline	Signature
1.	Discussion of the topic.	01.10.2021	
2.	Review of literature.	01.11.2021	
3.	Study important concepts.	01.12.2021	
4.	Writing the main chapter.	01.04.2022	
5.	Typography work.	01.05.2022	
6.	Preliminary analysis. Corrections.	01.06.2022	
8.	Predefence of the thesis.	15.06.2022	
9.	Final corrections.	25.06.2022	
10.	Defence of the thesis.	05.07.2022	

Contents

Анотація

У фінансовій математиці ми дуже часто стикаємось з такими поняттями, як невизначеність та мінливість, а отже спостерігаємо деякий ризик у кожному припущенні. Стохастичне моделювання є важливим інструментом для аналізу, прогнозу реальних фінансових даних. Вибір “кращої” моделі, яка адекватно описуватиме дані є досить непростою задачею в аналізі часових рядів. Більшість класичних та детально досліджених економетричних моделей є лінійними за параметрами, наприклад такі як AR(p), MA(q), ARMA(p,q). Їх перевагою є простота та невелика кількість параметрів, що дозволяють достатньо добре апроксимувати часові ряди, які за припущенням вважаються стаціонарними. Однак, у реальному житті фінансові показники змінюються дуже непередбачувано, тому постає проблема змінної варіації, яка не може бути описана простими лінійними моделями. Загалом моделювання та прогнозування волатильності на фондовому ринку стало пріоритетним завданням прикладних досліджень останніх років.

Кваліфікаційну роботу присвячено застосуванню теоретичних основ нелінійних стохастичних моделей, а саме ARCH(p), GARCH(p,q) на реальних фінансових даних. У роботі проведена оцінка методом моментів та методом максимальної вірогідності, їх порівняння та симуляція моделей. Спрогнозована поведінка волатильності акцій на певний період.

Ключові слова: умовна варіація, волатильність, стохастична величина, ARCH(p), GARCH(p,q)

Introduction

Common methods of modelling time series (e.g. ARIMA) operate under an assumption of constant variance. However, a lot of financial processes have a change in variance or volatility that can cause problems while modelling. If the variance has an explicit increasing trend, this property of series is called heteroskedasticity. In this case linear models explain and forecast certain economic behavior or economic performance in a relatively poor way, therefore we need another approach to model data.

Nonlinear models are applied in order to explain several phenomena in financial statistics and economics such as "cluster property" of prices, their "disastrous" jumps and downfalls, "heavy tails" of the distributions of the variables $h_n = \ln \frac{S_n}{S_{n-1}}$ that cannot be covered in the scope of linear models. Many macroeconomic indicators (the volumes of production, investment, the general level of prices) and also microeconomic indexes (current prices, the volume of traded stocks) fluctuate with a very high frequency or can be extremely irregular, so nonlinear stochastic models describe recessions and expansions, catastrophic behavior[1].

The ARCH(p) process (Autoregressive Conditional Heteroscedastic) introduced by Engle (1982) allows the conditional variance to change over time as a function of past errors leaving the unconditional variance constant[2]. For such processes, the recent past gives information about the one-period forecast variance[3]. A lag parameter p defines the number of prior residual errors that should be included in the conditional variance equation in order to take account of the long memory effect that is observed in empirical work. Estimating of high-order model leads to the violation of of the non-negativity constraints.

As an extension of ARCH model, GARCH(p,q) (Generalized Autoregressive Conditional Heteroskedastic), was introduced by Bollerslev (1986). It remains modelling not only with a long-memory property but also have more flexible lag structure. The GARCH model combines both the moving average together with autoregressive component.

The main objective of this thesis is the application of theoretical background to real financial time series. This is achieved through the following steps that are implemented manually and in Python:

1. Transforming data in order to follow property of stationarity (unconditional variance).
2. Testing if my data has ARCH/GARCH effects.
3. Determination of order the model.
4. Applying different methods of estimation and comparing results.
5. Simulation of volatility and log returns.
6. Forecast of volatility.

The work consists of two main chapters, a summary, a bibliography and an appendix. The first chapter describes definitions and properties of these models, the second one focuses on numerical results on the real dataset.

1 Nonlinear stochastic models: ARCH, GARCH

1.1 Stochastic processes, model and examples

Analysing time series of the evolution of financial (economic, social) indexes, we strive to find "best" model that fits our data, which actually can be rather a complicated task. Let identify the following main components of statistical data:[1]

- a slowly changing trend component;
- periodic or aperiodic cycles;
- a fluctuating ("stochastic" or "chaotic") component.

The main objectives of financial indexes analysis are to predict the "future dynamics of prices" or to come up with a right investment decision. Facing with uncertainty and variability we definitely observe a risk in every assumption we make. Therefore, in this chapter we introduce essential theoretical background that allows us to design nonlinear stochastic models. (see [4])

Definition 1.1

Let T be the index set, $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space and $(E, \mathcal{G})$ a measurable space. *A stochastic process* is a sequence of random variables $X = X_t; \quad t \in T$ where for each fixed $t \in T, \quad X_t$ is a random variable from $(\Omega, \mathcal{F}, \mathbb{P})$ to $(E, \mathcal{G})$. Ω is known as the sample space, where E is the state space of the stochastic process X_t.
The index set T of the stochastic process is often some subset of the real line, such as natural numbers or an interval, giving the set T the interpretation of time.
The state space E defines the values that the stochastic process can take (integers, real line etc.).
A stochastic process can also be written as $X(t, \omega)\colon t \in T$ to emphasize that it is actually function of two variables, $t \in T$ and $\omega \in \Omega$.

Definition 1.2

If $X(t, w) : t \in T$ is a stochastic process, then for any point $\omega \in \Omega$, the mapping

$$X(\cdot, \omega)\colon T \to S,$$

is called a *sample function (path)* of the stochastic process $X(t, w) : t \in T$.
This type of modelling introduces and predicts outcomes that take to consideration certain levels of unpredictability or randomness. To understand this concept we can compare it to its opposite, deterministic modelling, which in turn gives the same exact results for a particular set of inputs, no matter how many times you re-calculate the model.[6]
The application of stochastic models is commonly used in financial sector, one of the famous examples is Monte Carlo simulation. It models how a portfolio may perform based on the probability distribution of individual stock returns.

Examples of stochastic processes

1. *White noise.* We call ε a **white-noise** process if

$$\begin{aligned} \mathbb{E}(\varepsilon_t) =& 0, \forall t, \\ var(\varepsilon_t) =& \sigma^2, \forall t, \\ cov(\varepsilon_t, \varepsilon_{t-s}) =& 0, \forall s \neq 0 \end{aligned} \tag{1}$$

A time series is a white noise if the variables are independent and identically distributed with a mean of zero and the same variance (σ^2) and each value has a zero correlation with all other values in the series.[24]

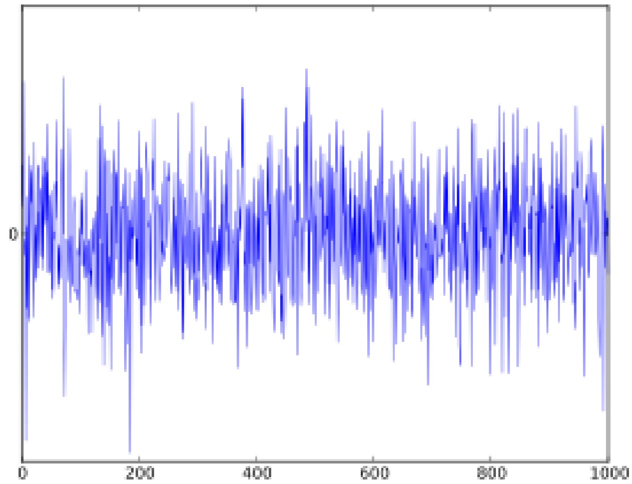

Figure 1: White noise

In fact, white noise is a sequence of random variables and cannot be predicted. In addition, if series follows Gaussian distribution $\varepsilon_t \sim \mathcal{N}(0, 1)$, we call it Gaussian white noise.
2. *Martingale*. A discrete-time stochastic process (sequence of random variables) $x_1, x_2, ..., x_t$ is called **martingale** if it satisfies the following conditions:

$$\begin{aligned} \mathbb{E}(|X_t|) &< \infty, \\ \mathbb{E}(X_{t+1}|X_1, ..., X_t) &= X_t \end{aligned} \tag{2}$$

Conditional expected value of the next observation, given all the past observations, is equal to the most recent observation.
Furthermore, it is essential to define martingale difference sequence (MDS). A stochastic series X_t is MDS if it satisfies the following conditions:

$$\begin{aligned} \mathbb{E}(|X_t|) &< \infty, \\ \mathbb{E}[X_t|\mathcal{F}_{t-1}] &= 0, \forall t \end{aligned} \tag{3}$$

This denotes if Y_t is a martingale, then $X_t = Y_t - Y_{t-1}$ will be martingale difference sequence [23].

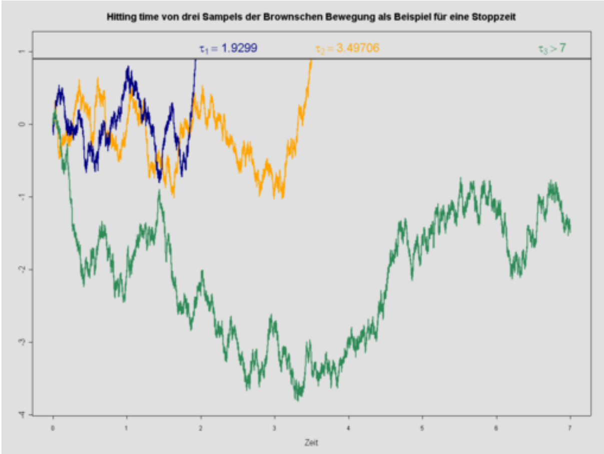

Figure 2: Martingale

Definition 1.3

A time series r_t is *weakly stationary* if $E(r_t) = \mu$, which is constant and $Cov(r_t, r_{t-l}) = \gamma_l$ which only depends on l, (time-invariant). So, the time plot of the data should show that values fluctuate with constant variance. Implicitly in the condition of weak stationarity, we assume that the first two moments of r_t are finite.
The covariance $\gamma_l = Cov(r_t, r_{t-l})$ is called the lag-l autocovariance of r_t.

Correlation and autocorrelation function

Definition 1.4

The correlation coefficient between two random variables X and Y is defined as

$$\rho_{x,y} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{\mathbb{E}[(X-\mu_x)(Y-\mu_y)]}{\sqrt{\mathbb{E}(X-\mu_x)^2\mathbb{E}(Y-\mu_y)^2}},$$

where μ_x and μ_y are the mean of X and Y, respectively, and it is assumed that the variance exists.
This coefficient measures the strength of linear dependence between X and Y, and it can be shown that $-1 \leq \rho_{x,y} \leq 1$ and $\rho_{x,y} = \rho_{y,x}$. The two random variables are uncorrelated if $\rho_{x,y} = 0$. Furthermore, if both X and Y are normal random variables, then $\rho_{x,y} = 0$ iff X and Y are independent.

Definition 1.5

The correlation coefficient between r_t and r_{t-l} is called the lag-l autorrelation of r_t, which under the weak stationarity assumption is a function of l only. Specifically, we define

$$\rho_l = \frac{Cov(r_t, r_{t-l})}{\sqrt{Var(r_t)Var(r_{t-l})}} = \frac{cov(r_t, r_{t-l})}{Var(r_t)} = \frac{\gamma_l}{\gamma_0}$$

Autoregressive model

Forecasting stock prices, investor makes an assumption if new buyers or sellers of that stock are impacted by recent market transactions.
Autoregressive models based on the assumption that past values have an effect on current values.
A sequence $r_t = (r_t)_{t\geq 1}$ called autoregressive model $AR(p)$ of order p if:

$$r_t = \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-2} + ... + \alpha_r y_{t-p} + \varepsilon_t$$

where ε_t is a white noise. We consider lagged values of h_{n-i} as predictors.
For effective use of AR models it is essential to study its basic properties. Consider case of AR(1) the contribution in r_t is made only by the closest in time variable r_{t-1}.

$$r_t = \alpha_0 + \alpha_1 r_{t-1} + \varepsilon_t \tag{4}$$

AR(1) model implies that, conditional on the past observation h_{n-1}, we have:

$$\mathbb{E}(r_t \mid r_{t-1}) = \alpha_0 + \alpha_1 r_{t-1}, \quad Var(r_t \mid r_{t-1}) = Var(\varepsilon_t) = \sigma_e^2.$$

That is, given the past observation r_{t-1}, the current observation is centered around $\alpha_0 + \alpha_1 r_{t-1}$ with variability σ_e^2. This is Markov property such that conditional on r_{t-1}, the value of r_t is not correlated with r_{t-i} for $i > 1$.
We usually restrict AR models with the sufficient and necessary condition for weak stationarity.
1. Expectation of AR(1)
Assuming that we have:

$$\mathbb{E}(r_t) = \mu,\ Var(r_t) = \gamma_0, \text{ and } \quad Cov(r_t, r_{t-j}) = \gamma_j,$$

where μ and γ_0 are constant and γ_j is a function of j, not t. $-1 < \alpha_1 < 1$
Taking the expectation of Eq.(4) and because $\mathbb{E}(\varepsilon_t = 0)$, we obtain

$$\mathbb{E}(r_t) = \alpha_0 + \alpha_1 \mathbb{E}(r_{t-1}). \tag{5}$$

Under assumption of stationarity, $\mathbb{E}(r_t) = \mathbb{E}(r_{t-1}) = \mu$. Substituting this into the previous equation, we get

$$\mathbb{E}(r_t) = \mu = \frac{\alpha_0}{1 - \alpha_1}. \tag{6}$$

Note, the mean of r_t esists if $\alpha_1 \neq 1$ and the mean of r_t is zero if and only if $\alpha_0 = 0$.
2. Variance of AR(1)
Using Eq(6) we have $\alpha_0 = (1 - \alpha_1)\mu$, the AR(1) can be rewritten as

$$r_t - \mu = \alpha_1(r_{t-1} - \mu) + \varepsilon_t. \tag{7}$$

By repeated substitutions, the previous equation implies that

$$\begin{aligned} r_t - \mu &= \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_1^2 \varepsilon_{t-2} + ... \\ &= \sum_{i=0}^{\infty} \alpha_1^1 \varepsilon_{t-1}. \end{aligned} \tag{8}$$

Thus, $r_t - \mu$ is a linear function of ε_{t-1} for $i \geq 0$. With help of linearity and the independence of the series ε_t, we obtain $\mathbb{E}[(r_t - \mu)\varepsilon_{t+1}] = 0$. With the assumption of stationarity we have $Cov(r_{t-1}, \varepsilon_t) = \mathbb{E}[(r_{t-1} - \mu)\varepsilon_t] = 0$.
Taking square and expectation of Eq.(7)

$$Var(r_t) = \alpha_1^2 Var(r_{t-1}) + \sigma_\varepsilon^2$$

Under the stationary assumption, $Var(r_t = Var(r_{t-1}))$, so that

$$Var(r_t) = \frac{\sigma_\varepsilon^2}{1-\alpha_1^2}. \tag{9}$$

In order the variance to be finite and satisfy non-negativity, $\alpha_1^2 < 1$. Hence, weak stationary of AR(1) implies that $-1 < \alpha_1 < 1$ and guarantees the mean and variance of r_t to be finite.

Autocorrelation function of AR(1) model

Multiplying Eq.(7) by ε_t consider independence between ε_t and r_{t-1} and taking expectation, we obtain

$$\mathbb{E}[\varepsilon_t(r-t-\mu)] = \mathbb{E}[\varepsilon_t(r_{t-1}-\mu)] + \mathbb{E}(\varepsilon_t^2) = \sigma_\varepsilon^2,$$

where σ_ε^2 is the variance of ε_t.

Multiplying Eq.(7) by $(r_{t-l}-\mu)$, taking expectation and applying the prior result, we have

$$\gamma_l = \begin{cases} \alpha_1\gamma_1 + \sigma_\varepsilon^2 & \text{if } l = 0 \\ \alpha_1\gamma_{l-1} & \text{if } l > 0, \end{cases}$$

where we use $\gamma_l = \gamma_{-l}$. Consequently, for a weakly stationary AR(1) model, we have

$$Var(r_t) = \gamma_0 = \frac{\sigma^2}{1-\alpha_1^2} \quad \text{and} \quad \gamma_l = \alpha_1\gamma_{l-1}, \quad \text{for} \quad l > 0$$

From the last equation, the ACF of r_t satisfies

$$\rho_l = \alpha_1\rho_{l-1}, \quad \text{for} \quad l \geq 0$$

$\rho_0 = 1$ then $\rho_l = \alpha_1^l$. This result says that ACF of a weakly stationary AR(1) [22] series decays exponentially with rate α_1 and starting value $\rho_0 = 1$.

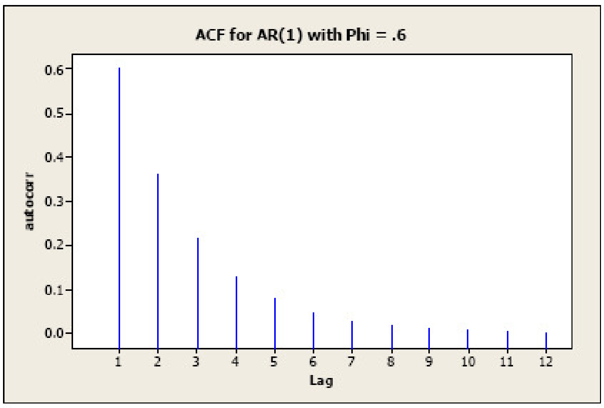

Figure 3: ACF of AR(1) for $\alpha_1 = 0.6$

Why do we model log returns?

Representing financial data in log difference is very common [9]. Calculating simple returns is done using:

$$R = \frac{P_i - P_j}{P_j}. \tag{10}$$

R-is market return,P_i - is ending price, P_j- isstarting price.
Mostly total value of a return consists of n sub-periods, so we need to compound the growth of each period:

$$P_f = P_0(1 + R_1)(1 + R_2)...(1 + R_n).$$

where P_f- is final price, P_0- is initial price, R_x- is return for each sub-period.
We could write this formula in a simplified exponential way if R_x were evenly distributed, but in reality it never happens. Because of this we should take to account a logarithmic form of returns. In this case we are focusing more on relative value rather than absolute one, logarithmic returns show the rate of exponential growth.
Therefore if you earn $r\%$ interest that is compounded continuously, at the end of the year your money will be:

$$P_2 = P_1 \lim_{x\to\infty}(1 + \frac{r}{n})^n = P_1 e^r$$

or grown by a factor:

$$\frac{P_2}{P_1} = e^r$$

Taking natural logarithm we obtain

$$r = ln(\frac{P_2}{P_1})$$

The following equation is derived from Eq.(10)

$$ln(R_i + 1) = ln(\frac{P_i}{P_j})$$

Such representation of an asset return ensures weak-stationary property.

1.2 ARCH model and its properties

The nature of financial data is often as follows:[25]

- The distribution of average and variance over time is not followed under normal law and density function, is characterized by a more elevated critical zone (the presence of so-called thick tails). In other words, most financial time series has such distributions of values in which extreme values of the indicator occur more often than it is provided by the normal distribution.

- In financial processes there is a clustering of volatility: spontaneous strong shocks of an indicator do not subside, and proceed still some time.

Further we declare several nonlinear stochastic models that are popular in financial mathematics and statistics.
The ARCH(p) process (Autoregressive Conditional Heteroscedastic) was introduced by Engle (1982).[1,3]

- AR (Autoregressive) means past financial data influences future data
- CH (Conditional Heteroscedastic) identifies nonconstant volatility related to prior period's high or low volatility.

Let $(\Omega, \mathcal{F}, P)$ be the original probability space and let $\varepsilon = (\varepsilon_n)_{n\geq 1}$ be a sequence of independent, normally distributed random variables ($\varepsilon_n \sim \mathcal{N}(0,1)$) simulating "randomness", "uncertainty" in the models that we consider below.
By $\mathcal{F}_n$ we shall mean the σ-algebra $\sigma(\varepsilon_1, .., \varepsilon_n)$; we set $\mathcal{F}_0 = \{\emptyset, \Omega\}$.
We shall interpret $S_n = S_n(w)$ as the price (or an exchange rate) at time $n = 0, 1....$ Here time can be measured in years, months so on.
To describe the evolution of the variables

$$h_n = ln\frac{S_n}{S_{n-1}}. \tag{11}$$

R.Engle considered the conditionally Gaussian model with

$$h_n = \sigma_n \varepsilon_n. \tag{12}$$

The volatilities σ_n are defined as follows:

$$\sigma_n^2 = \alpha_0 + \sum_{i=0}^{p} \alpha_i h_{n-i}^2. \tag{13}$$

where $\alpha_0 > 0, \alpha_i \geq 0$, in order to guarantee positive variance and $h_0 = h_0(w)$ is a random variable independent of $\varepsilon = (\varepsilon_n)_{n\geq 1}$.
We see that σ_n are predictable functions of (past) functions $h_{n-1}^2, h_{n-2}^2, ...,$ so we can consider that large (small) values of h_{n-i}^2 imply large (small) respectively values of σ_n^2.[4]
Consequently, h_n tends to assume high value, it means that probability of obtaining large shocks is greater.[16]
Shown above nonlinear models can give an interpretation of such phenomena as "cluster property", namely periods of large movements in prices alternate with periods during which prices hardly change. Because of this conditions, the assumption of a constant variance (homoscedasticity) is inappropriate.[8]
For simplicity, we consider $p = 1$:

$$\sigma_n^2 = \alpha_0 + \alpha_1 h_{n-1}^2. \tag{14}$$

The following properties of the $h_n = \sigma_n \varepsilon_n$ are:
1) Expectation of log-returns (First moment)
By applying Theorem of iterated expectations, conditional expectation of measurable random variable and mean of white noise, (see Appendix) we derive

$$\mathbb{E}h_n = \mathbb{E}[\mathbb{E}[(\sigma_n \varepsilon_n) \mid \mathcal{F}_{n-1}]] = \mathbb{E}[\sigma_n \mathbb{E}(\varepsilon_n)] = 0 \tag{15}$$

2) Variance (Second moment)

$$\begin{aligned} Var(h_n) = \mathbb{E}h_n^2 - \mathbb{E}(h_n)^2 = \mathbb{E}[\mathbb{E}[(\sigma_n^2 \varepsilon_n^2 \mid \mathcal{F}_{n-1})]] = \mathbb{E}(\alpha_0 + \alpha_1 h_{n-1}^2) = \\ = \alpha_0 + \alpha_1 \mathbb{E}h_{n-1}^2, \end{aligned} \tag{16}$$

$$\mathbb{E}(h_n^2) \mid \mathcal{F}_{n-1}) = \sigma_n^2 = \alpha_0 + \alpha_1 h_{n-1}^2. \tag{17}$$

If

$$0 < \alpha_1 < 1$$

and assume that equation (16) has unique stationary solution, namely $\mathbb{E}h_{n-1}^2 = \mathbb{E}h_n^2$ we derive

$$\mathbb{E}h_n^2 \equiv \frac{\alpha_0}{1-\alpha_1}, n \geq 0. \tag{18}$$

3) Fourth moment

$$\begin{aligned} \mathbb{E}h_n^4 &= \mathbb{E}\sigma_n^4 \mathbb{E}\varepsilon_n^4 = 3\mathbb{E}\sigma_n^4 = 3\mathbb{E}(\alpha_0 + \alpha_1 h_{n-1}^2)^2 = 3(\alpha_0^2 + 2\alpha_0\alpha_1 \mathbb{E}h_{n-1}^2 + \alpha_1^2 \mathbb{E}h_{n-1}^4) = \\ &= \frac{3\alpha_0^2(1+\alpha_1)}{1-\alpha_1} + 3\alpha_1^2 \mathbb{E}h_{n-1}^4. \end{aligned} \tag{19}$$

Assuming that $0 < \alpha_1 < 1$ and $3\alpha_1^2 < 1$ we can obtain following solution in case ($\mathbb{E}h_n^4 = const$)

$$\mathbb{E}h_n^4 = \frac{3\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)} \tag{20}$$

4) Excess kurtosis
The value of excess kurtosis is

$$K \equiv \frac{\mathbb{E}h_n^4}{(\mathbb{E}h_n^2)^2} - 3 = \frac{6\alpha_1^2}{1-3\alpha_1^2}. \tag{21}$$

Kurtosis measures how fat a distribution's tail is when compared to the center of the distribution. Excess kurtosis helps determine how much risk is involved in a specific investment. We compare excess kurtosis regarding to normal distribution, that's why we substract 3 in (21) (as excess kurtosis of normal distribution equals 0). The values of excess kurtosis can be either negative or positive. When the value of an excess kurtosis is negative, the distribution is called platykurtic.This kind of distribution has a tail that's thinner than a normal distribution. When excess kurtosis is positive, it has a leptokurtic distribution. The tails on this distribution is heavier than that of a normal distribution, indicating a heavy degree of risk. The returns on an investment with a leptokurtic distribution or positive excess kurtosis will likely have extreme values. Excess kurtosis can be at or near zero as well, so the chance of an extreme outcome is rare. This is known as a mesokurtic distribution.[10]
So, in our case kurtosis is positive, which means that we observe heavy tails and distribution of the variables h_n has a peak near the mean value.(see [20])

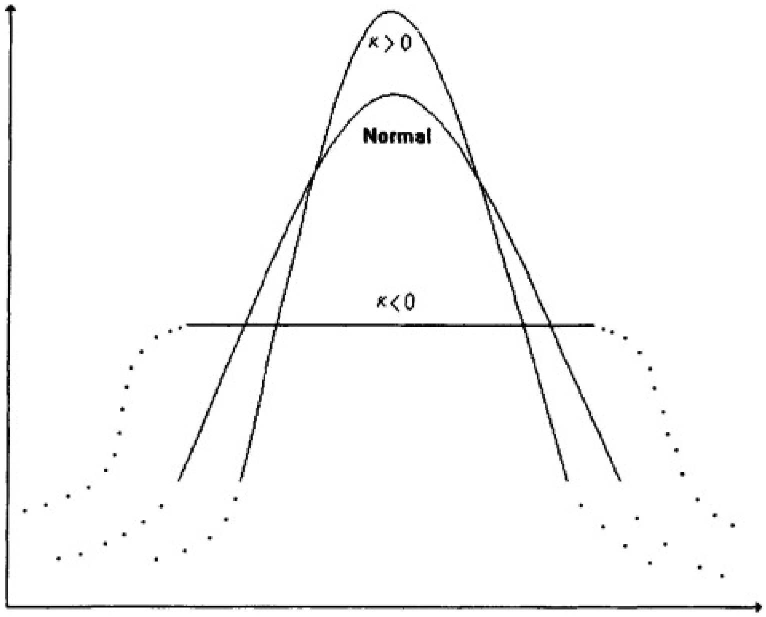

Figure 4: Kurtosis

5) Structure of correlation relationship in h_n
The sequence $h = h_n, h_n = \sigma_n \varepsilon_n$ is serially uncorrelated process.
$\mathbb{E}(h_n h_{n-1}) = \mathbb{E}[\mathbb{E}(h_n h_{n-1} \mid h_{n-1}, h_{n-2}, ...)] = \mathbb{E}[h_{n-1}\mathbb{E}(h_n \mid h_{n-1}, h_{n-2}, ...)] =$
$= \mathbb{E}[h_{n-1} \cdot 0] = 0.$
$Cov(h_n, h_{n-1}) = \mathbb{E}(h_n h_{n-1}) - \mathbb{E}(h_n)\mathbb{E}(h_{n-1}) = 0.$
If two variables are uncorrelated, there is no linear relationship between them, but it doesn't mean that they are independent.
We can prove it considering correlation relationship between squares of h_n^2, h_{n-1}^2 or their abs $|h_n, |h_{n-1}||$.

$$Dh_n^2 = \frac{2}{1-3\alpha_1^2}(\frac{\alpha_0}{1-\alpha_1}^2). \tag{22}$$

$$\mathbb{E}h_n^2 h_{n-1}^2 = \frac{1+3\alpha_1}{1-3\alpha_1^2} \cdot \frac{\alpha_0^2}{1-\alpha_1}. \tag{23}$$

Therefore,

$$p(1) = Corr(h_n^2, h_{n-1}^2) = \frac{Cov(h_n^2, h_{n-1}^2)}{\sqrt{Dh_n^2 Dh_{n-1}^2}} = \alpha_1.$$

Estimation of model using Method of Moments [1]
Firstly, we need to find empirical values of second and fourth moments assuming that they are finite and h_n follows stationary condition.
$\mathbb{E}h_n^2 = \mathbb{E}h_n^2 - (\mathbb{E}h_n)^2$
$\mathbb{E}h_n^4 = \mathbb{E}(h_n^4) - 4\mathbb{E}(h_n)\mathbb{E}(h_n^3) + 6\mathbb{E}(h_n)^2\mathbb{E}(h_n^2) - 3\mathbb{E}(h_n)^4$
Consider the following system of equations:

$$\begin{cases} \mathbb{E}h_n^2 = \frac{\alpha_0}{1-\alpha_1}, \\ \mathbb{E}h_n^4 = \frac{3\alpha_0^2(1=\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}. \end{cases} \tag{24}$$

Then α_0, α_1 can be derived as:

$$\alpha_0 = \mathbb{E}h_n^2 \cdot (1 - \alpha_1), \tag{25}$$

$$\frac{3(Eh_n^2)^2}{\mathbb{E}h_n^4} = \frac{1 - 3\alpha_1^2}{1 - \alpha_1^2}. \tag{26}$$

Forecasting [1]
Financial analysts are interested in predicting future behavior of prices. As I have already mentioned sequence $h = (h_n)$ is martingale difference, so it can not be predicted by prior values $\mathbb{E}(h_{n+m} \mid \mathcal{F}_n^h) = 0$.
So, we will forecast nonlinear function of h_{n+m}, namely h_{n+m}^2

$$\begin{aligned} h_{n+m}^{\hat{2}} = \mathbb{E}h_{n+m}^2 \mid \mathcal{F}_n^h = \mathbb{E}\sigma_{n+m}^2\varepsilon_{n+m}^2 \mid \mathcal{F}_n^h = \\ = \mathbb{E}[\mathbb{E}(\sigma_{n+m}^2\varepsilon_{n+m}^2 \mid \mathcal{F}_{n+m-1}^{\varepsilon}) \mid \mathcal{F}_n^h] = \mathbb{E}\sigma_{n+m}^2 \mid \mathcal{F}_n^h = \sigma_{n+m}^{\hat{2}} \end{aligned} \tag{27}$$

$$h_{n+m}^{\hat{2}} = \alpha_0 \frac{1 - \alpha_1^m}{1 - \alpha_1} + \alpha_1^m h_n^2$$

if $m \to \infty, \quad h_{n+m}^{\hat{2}} \to \mathbb{E}h_n^2 = \frac{\alpha_0}{1-\alpha_1}$
Although ARCH model can model different economic phenomena it has a number of disadvantages.[16]
1. Firstly, analysing structure of the model, it is assumed that positive and negative shocks have the same effects on volatility, because of the squares of previous shocks. However, in practice price of an asset reacts differently to positive and negative shocks.
2. The ARCH model is rather limited. For example, α_1^2 of ARCH(1) should be in interval $[0, \frac{1}{3}]$ provided that series has finite fourth moment.
3. The ARCH model doesn't explain reasons why conditional variance has such behavior.
4.ARCH models are likely to overpredict the volatility because they respond slowly to large isolated shocks to the return series.[16]

1.3 GARCH model and its properties

Since ARCH model was introduced, it became an incentive to generate different variation of itself. One extension known as generalized ARCH was proposed by Bollerslev (1986). Very often fitting financial data for ARCH model requires many lags to adequately describe the volatility process, for instance, consider returns of SP 500 index, ARCH(9) should be applied to model volatility. More complex structure of GARCH model is a good alternative to solve this problem.[1,16]
Then h_n follows a GARCH(p,q) model if

$$h_n = \sigma_n\varepsilon_n, \quad \sigma_n^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i h_{n-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{n-j}^2, \tag{28}$$

where again ε_n is a sequence of i.i.d. random variables with mean 0 and variance 1, $\alpha_0 > 0,\ \alpha_i \geq 0,\ \beta_j \geq 0,$ and $\sum_{i=1}^{max(p,q)}(\alpha_i + \beta_j) < 1$. Here it is understood that $\alpha_i = 0$ for $i > p$ and $\beta_j \geq 0,$ for $j > q$. The last constraint on $\alpha_i + \beta_j$ is needed for unconditional variance of h_n to be finite, whereas its conditional variance σ_n^2 evolves over time.
For $q = 0$ the process reduces to the ARCH(p), and for $p = q = 0$, is simply white noise. In the ARCH(p) process the conditional variance is specified as a function of past sample variances only, whereas the GARCH(p,q) process allows lagged conditional variances to enter as well. This corresponds to some sort of adaptive learning mechanism.

Structure and properties of GARCH(1,1)

$$h_n = \sigma_n \varepsilon_n, \quad \sigma_n^2 = \alpha_0 + \alpha_1 h_{n-1}^2 \beta_1 \sigma_{n-1}^2, \tag{29}$$

where $\alpha_0 > 0,\ \alpha_1 \geq 0,\ \beta_1 \geq 0$

$$\mathbb{E}h_n^2 = \alpha_0 + (\alpha_1 + \beta_1)_{n-1}^2$$

First, a large h_{n-1}^2 gives rise to large σ_t^2. This means that a large h_{n-1}^2 tends to be followed by another large h_n^2, generating again, the well-known behavior of volatility clustering in financial time series. provided that $\alpha_1 + \beta_1 < 1$ then stationary solution of $\mathbb{E}h_n^2$ is

$$\mathbb{E}h_n^2 = \frac{\alpha_0}{1 - \alpha_1 - \beta 1}.$$

If $3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$ then stationary solution for fourth moment is

$$\mathbb{E}h_n^4 = \frac{3\alpha_0^2(1 + \alpha_1 + \beta_1)}{(1 - \alpha_1 - \beta_1)(1 - \beta_1^2) - 2\alpha_1\beta 1 - 3\alpha_1^2}$$

Excess kurtosis is

$$K = \frac{\mathbb{E}h_n^4}{(Eh_n^2)^2} - 3 = \frac{6\alpha_1^2}{1 - \beta_1^2 - 2\alpha_1\beta_1 - \beta_1^2}$$

Consequently, similar to ARCH models, the tail distribution of a GARCH(1,1) process is heavier than that of a normal distribution.
Estimation of GARCH(1,1) based on ARMA representation. [17]
We can assume, that $x_t \equiv h_n^2$ as

$$x_t = \omega + \phi x_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$$

where $\varepsilon_t = x_t - \sigma_t^2$ is a martingale difference sequence with respect to $\mathcal{F}_t$, $\phi = \alpha + \beta > 0$ and $\theta = -\beta < 0$. We shall also assume that $\phi < 1$ in order to guarantee $\mathbb{E}[x_t] < \infty$.
The covariance function is defined as:

$$\gamma(k) = \mathbb{E}[(x_{t+k} - \mathbb{E}[x_t])(x_t - \mathbb{E}[x_t])]$$

Suppose, x_t is stationary with 2nd moment. For stationary process autocorrelation function is $p(k) = \gamma(k)/\gamma(0)$
According to the following set of Yule-walker equations Harvey (1993).

$$p(k) = \phi p(k-1), \quad k = 2, 3..., \tag{30}$$

$$p(1) = \frac{(1+\phi\theta)(\phi+\theta)}{1+\theta^2+2\phi\theta} \tag{31}$$

Let $b = \frac{\phi^2+1-2p(1)\phi}{\phi-p(1)}$, $\phi \neq p(1)$ and express eq.20 in θ, $\theta^2 + b\theta + 1 = 0$
The solution to quadratic equation is

$$\theta = \frac{-b+\sqrt{b^2-4}}{2}$$

We observe that

$$\omega = \sigma^2(1-\phi), \quad \sigma^2 = \mathbb{E}(y_t^2).$$

First, we can estimate ϕ by $\hat{\phi} = \frac{\hat{p(2)}}{\hat{p(1)}}$ Substitute and obtain estimator of θ

$$\hat{\theta} = \frac{-\hat{b}+\sqrt{\hat{b}^2-4}}{2}, \hat{b} = \frac{\hat{\phi}^2+1-2\hat{p}(1)\hat{\phi}}{\hat{\phi}-\hat{p}(1)}$$

This leads to the following estimators of $\lambda = (\alpha, \beta, \omega)$

$$\hat{\alpha} = \hat{\theta} + \hat{\phi}, \quad \hat{\beta} = -\hat{\theta}, \quad \hat{\omega} = \hat{\sigma}^2(1-\hat{\phi}) \tag{32}$$

The literature on GARCH models is enormous; see Bollerslev, Chou, and Kroner (1992), Bollerslev, Engle, and Nelson (1994), and the references therein. The model encounters the same weaknesses as the ARCH model. For instance, it responds equally to positive and negative shocks.

2 Numerical results for ARCH and GARCH models

In this chapter I will illustrate an application of ARCH model on real dataset of Toronto Stock Exchange (TSX).[12] It is canadian stock exchange located in Toronto, Ontario, founded in 1861, the TSX is Canada's premier stock exchange with more than 1,500 listed companies, including those from the energy, mining, technology, and real estate sectors. It is the 11th largest exchange in the world and the third largest in North America based on market capitalization.[13]
We will consider Adjusted Close price in Canadian dollars for the last 5 years from 02/08/2016-30/07/2021. The adjusted closing price factors is anything that might affect the stock price after the market closes. A stock's price is typically affected by supply and demand of market participants. However, some corporate actions, such as stock splits, dividends, and rights offerings, affect a stock's price. Adjustments allow investors to obtain an accurate record of the stock's performance.[14]

Firstly, we plot our raw data.

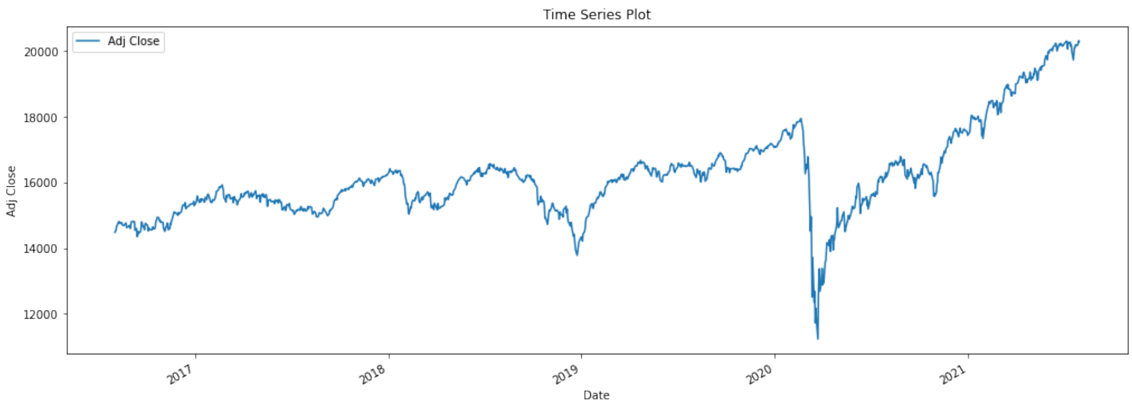

Figure 5: Adj Close price of TSX

Analysing this plot we can assume that non stationarity is present in our data. Firstly, we are tracking positive trend, but in the beginning of 2020 due to covid pandemic we observe a really big drop and afterwards the evolution of stock proceeds with its growth. To make sure that it is true we will apply statistical Augmented Dickey-Fuller. The ADF test is a type of statistical test called a unit root test, because it technically checks if value of α in equation below equals 1 or not.

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

where u_t - is a noise
The null hypothesis (H0) of the test is that the time series is **not stationary**, $\alpha = 1$ (has some time-dependent structure).
The alternate hypothesis (H1) is that the time series is **stationary**, doesn't have a unit root.
We interpret this result using the p-value from the test.
If **p-value** is below a threshold (such as 5% or 1%), we reject the null hypothesis (data is stationary), otherwise a p-value above the threshold suggests we fail to reject the null hypothesis (data is non-stationary). We can also compare ADF Statistic with critical

values, if ADF Statistic is less than critical one we are likely to reject null hypothesis (so stationary data), otherwise we come up with nonstationary data.

```
ADF Statistic: -1.634808
p-value: 0.464975
Critical Values:
        1%: -3.436
        5%: -2.864
        10%: -2.568
```

p-value: 0.46 > 0.05, we fail to reject null hypothesis, so our data is definitely nonstationary.
ADF Statistic: -1.63 > -3.41 - non stationary.
The results obtained by using non-stationary time series may be spurious in that they may indicate a relationship between two variables where one does not exist.
In order to receive consistent, reliable results, the non-stationary data needs to be transformed into stationary data. [15]
In our case we will use log difference approach and represent data as:

$$h_n = ln\frac{S_n}{S_{n-1}}.$$

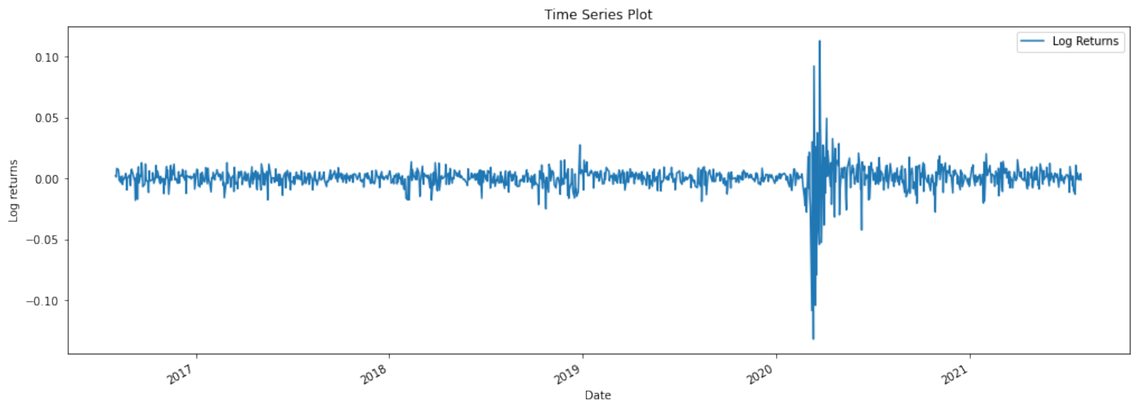

Figure 6: Log-returns

We can see that now our data is more likely to be stationary, except non constant variance at the begining of 2020.
Results of ADF test:

```
ADF Statistic: -8.337829
p-value: 0.000000
Critical Values:
        1%: -3.436
        5%: -2.864
        10%: -2.568
```

p-value: 0.0 < 0.05, we reject null hypothesis, so our data is definetely **stationary**.

ADF Statistic: -8.34 < -2.86 - stationary.

Testing for ARCH effects

The squared series h_n^2 is used to check for conditional heteroscedasticity, where $h_n = r_n - \mu_n$ is the residual of ARMA model. F-statistic can be used to find the joint significance of multiple independent variables. So for the given regression equation:

$$h_n^2 = \alpha_0 + \alpha_1 h_{n-1}^2 + ... + \alpha_m h_{n-m}^2 + \varepsilon_n, \quad t = m+1, ..., T,$$

where ε_n denotes the error term, m is a prespecified positive integer, and T is the sample size.

The null hypothesis $H0$ would be:$\alpha_1 = \alpha_2 = ... - \alpha_m = 0$

Alternative hypothesis $H1$ would be: $\alpha_i \neq 0$

So, if even one of the coefficients is significant, then there is a high possibility of rejecting the null hypothesis as the coefficients are not jointly insignificant anymore. Here the two models can be an unrestricted model which contains all the predictor variables or a restricted model in which we are restricting the number of predictor (for example intercept-only).

Let $SSR_0 = \sum_{t=m+1}^{T}(h_n^2 - \bar{\omega})^2$, where $\bar{\omega}$ is the sample mean of h_n^2 -sum square of residuals of the restricted model

and $SSR_1 = \sum_{t=m+1}^{T} \hat{\varepsilon_n^2}$, where ε_n^2 is the least squares residual of the prior linear regression. It is sum square of residuals of the unrestricted model. Then we have

$$F = \frac{(SSR_0 - SSR_1)/m}{SSR_1/(T - 2m - 1)}$$

Denote:

$df1 = m$: degree of freedom 1,

$df2 = T - 2m - 1$: degree of freedom 2.

The distribution we are gonna compare it with is called the F-distribution. We usually take a confidence interval of 95% which translates to an alpha value of 0.05. Based on the values of the two degrees of freedom and the alpha value we can find the F-critical value on the F-distribution. See table of F-critical values in appendix. If the F-statistic value is greater than the F-critical, we reject the null hypothesis.

So, given model got a F-statistic score of 288.473, $df1 = 1, df2 = 1.25e + 03$. F-critical value for $\alpha = 0.05$ is 3.85. Since, $F - critical$ is much lower than our $F - statistic$, we reject the null hypothesis, which means that the independent variables are jointly significant in explaining the variance of the dependent variable. We can also check the $p - value$ in the summary to determine whether to reject or accept the null hypothesis or not. In our case, $p - value = 2.232 \cdot 10^{-58}$ is much lower than $\alpha = 0.05$, so we reject the null hypothesis.

Order determination

In order to determine order of the model, we need build partial autocorrelation plot, which is a summary of the relationship between an observation in a time series with observations at prior time steps, but only the direct effect is shown (all intermediary effects are omitted). As I have already mentioned variables h_n and h_{n-1} are uncorrelated, but their squares are correlated and can be predicted.

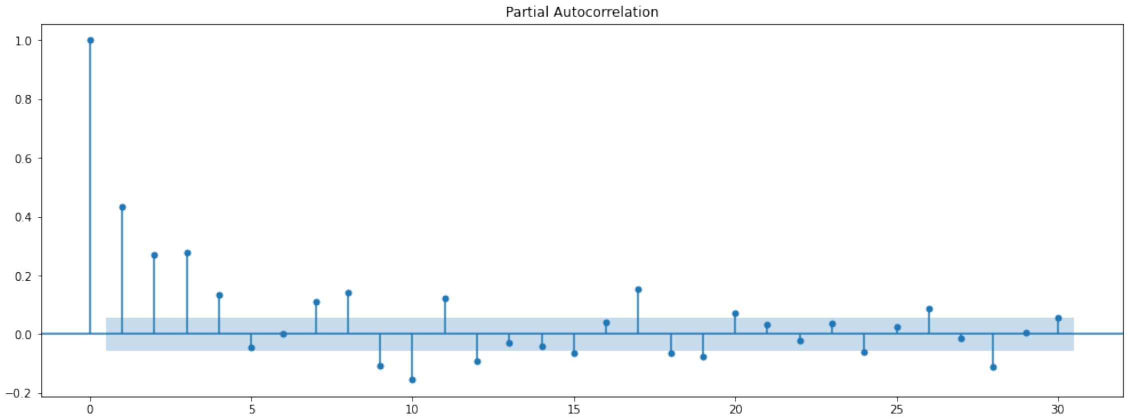

Figure 7: PACF of squared log-returns

We can observe that the first three lags are very significant. So, it is better to use ARCH(3) for this financial data.

2.1 Application of ARCH model

In this paper are covered two different method of estimation the model. First method is the method of moments, but it is difficult to apply it for ARCH model of higher order, so for simplicity I will demonstrate this approach on ARCH(1) model.
Empirical values of moments are:

$$\begin{aligned}\mathbb{E}h_n^2 =&0.000113,\\ \mathbb{E}h_n^4 =&7.0885 \cdot 10^{-5}.\end{aligned}$$

Applying Equations (17-19) from Chapter 1 we obtain the following estimations for coefficients:

$$\alpha_0 = 0.000049.$$

Solving Eq.(19) we obtain two roots:

$$\begin{aligned}\alpha_{11} =& -0.56655,\\ \alpha_{12} =&0.56655.\end{aligned}$$

First one we don't consider, because it is less than zero and the coefficients α_i ought to be positive. Simulated returns of ARCH(1) are shown in the following plot.

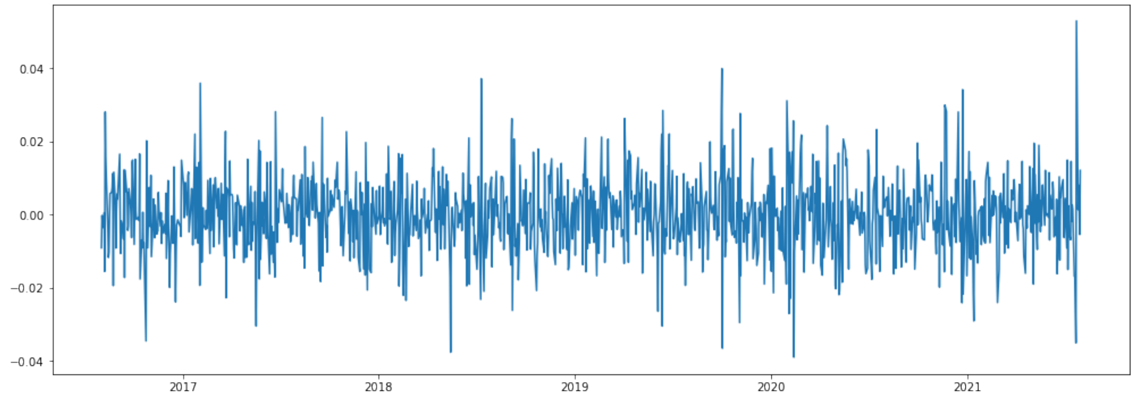

Figure 8: $h_n = \sqrt{0.000049 + 0.56655 h_{n-1}^2} \varepsilon_n$

Also was simulated volatility for ARCH(1).

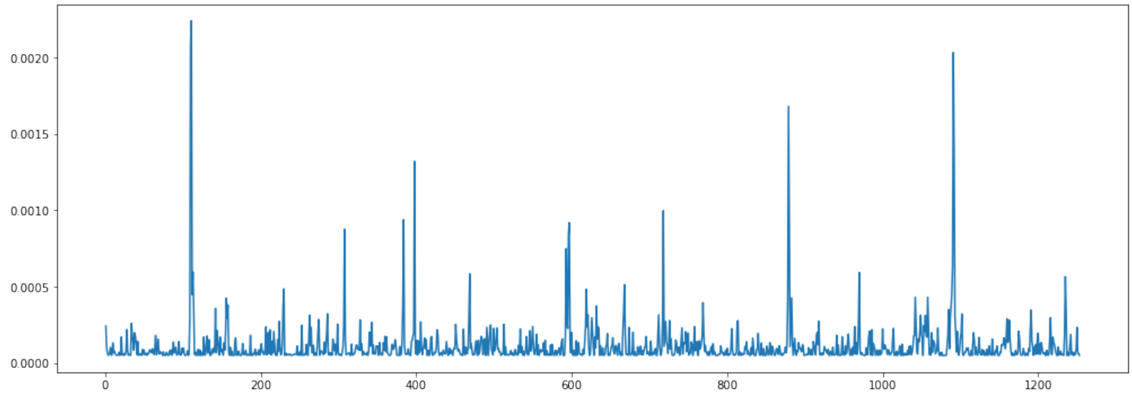

Figure 9: $\sigma_n^2 = 0.000049 + 0.56655 h_{n-1}^2$

Empirical value of excess kurtosis is $K = 51.973$ that confirms presence of heavy-tails in the data.

Maximum of Likelihood function estimation (MLE) [16]

MLE is a probabilistic framework for solving the problem of density estimation. It involves maximizing a likelihood function (optimization problem) in order to find the probability distribution and parameters that best explain the observed data. What we want to calculate is the total probability of observing all of the data, i.e. the joint probability distribution of all observed data points. To do this we would need to calculate some conditional probabilities, which can get very difficult. So, we assume that each data point is generated independently of the others. Thus if the events are independent, then the total probability of observing all of data is the product of observing each data point individually. Here, I am briefly explaining main concepts of algorithm.

Let $\varepsilon_1, ..., \varepsilon_T$ be an independent and identically distributed sample with probability density function (pdf) $f(\varepsilon_t; \Theta)$, where Θ is a $k \times 1$ vector of parameters that characterize $f(\varepsilon_t; \Theta)$. In our case we consider that, $\varepsilon_t \sim \mathcal{N}(\mu, \sigma^2)$ then pdf is defined as:

$$f(\varepsilon_t \mid \Theta) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp^{-\frac{1}{2\sigma^2}(\varepsilon_t - \mu)^2} \; and \, \Theta = (\mu, \sigma^2).$$

The joint probability density of the sample is:

$$f(\varepsilon_1, ..., \varepsilon_t \mid \theta) = f(\varepsilon_1 \mid \Theta)...f(\varepsilon_t \mid \theta) = \prod_{t=1}^{T} f(\varepsilon_t; \Theta).$$

The common approach to find maxima (minima) of the function is to apply differentiation. But the above equation is quite difficult to differentiate, so we can simplify it by taking the natural logarithm of the expression. This is absolutely fine because the natural logarithm is a monotonically increasing function. This is important because it ensures that the maximum value of the log of the probability occurs at the same point as the original probability function. The log-likelihood function $l(\alpha_0, \alpha_1)$ can be written as a function of the parameters α_0 and α_1:

$$\begin{aligned} l(\varepsilon_t \mid \alpha_0, \alpha_1) &= \sum_{t=2}^{n} l_t(\alpha_0, \alpha_1) + lnf_\varepsilon(\varepsilon_1) \\ &= \sum_{t=2}^{n} logf(\varepsilon_t \mid \mathcal{F}_{t-1}) + lnf_\varepsilon(\varepsilon_1) \\ &= -\frac{n-1}{2} log(2\pi) - \frac{1}{2}\sum_{t=2}^{n} log(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2) \\ &\quad - \frac{1}{2}\sum_{t=2}^{n} \frac{\varepsilon_t^2}{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2} + logp_\varepsilon(\varepsilon_1), \end{aligned} \tag{33}$$

where f_ε is the stationary marginal density of ε_t. A problem is that the analytical expression for f_ε is unknown in ARCH models thus can not be calculated. In the conditional likelihood function $l^b = lnf(\varepsilon_n, ..., \varepsilon_2 \mid)\varepsilon_1$ the expression $lnf_\varepsilon(\varepsilon_1)$ dissapears:

$$\begin{aligned} l^b(\alpha_0, \alpha_1) &= \sum_{t=2}^{n} l_t(\alpha_0, \alpha_1) \\ &= \sum_{t=2}^{n} lnf(\varepsilon_t \mid \mathcal{F}_{t-1}) \\ &= -\frac{n-1}{2} ln(2\pi) - \frac{1}{2}\sum_{t=2}^{n} ln(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2) - \frac{1}{2}\sum_{t=2}^{n} \frac{\epsilon_t^2}{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2} \end{aligned} \tag{34}$$

Applying this algorithm, we obtain the following coefficients for ARCH(1) model

Volatility Model

	coef	std err	t	P>|t|	95.0% Conf. Int.
omega	2.6725e-05	4.423e-07	60.425	0.000	[2.586e-05,2.759e-05]
alpha[1]	0.6743	0.111	6.094	1.100e-09	[0.457, 0.891]

Figure 10: Estimation of ARCH(1) with MLE

Estimations for parameters of model obtained via two different methods on simulated values appeared to be similar.
It is shown plot of realised and conditional volatility for ARCH(1).

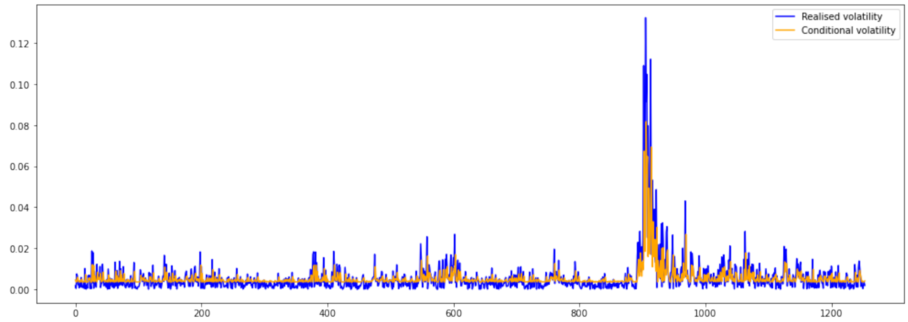

Figure 11: Realised vs conditional volatility

Then it is shown estimation for ARCH(3).

Volatility Model

	coef	std err	t	P>|t|	95.0% Conf. Int.
omega	1.6444e-05	4.864e-12	3.381e+06	0.000	[1.644e-05,1.644e-05]
alpha[1]	0.2557	6.995e-02	3.656	2.563e-04	[0.119, 0.393]
alpha[2]	0.2547	5.858e-02	4.349	1.368e-05	[0.140, 0.370]
alpha[3]	0.2547	8.706e-02	2.926	3.433e-03	[8.411e-02, 0.425]

Figure 12: Estimation of ARCH(3) with MLE

Simulation for log-returns and volatility of ARCH(3) model.

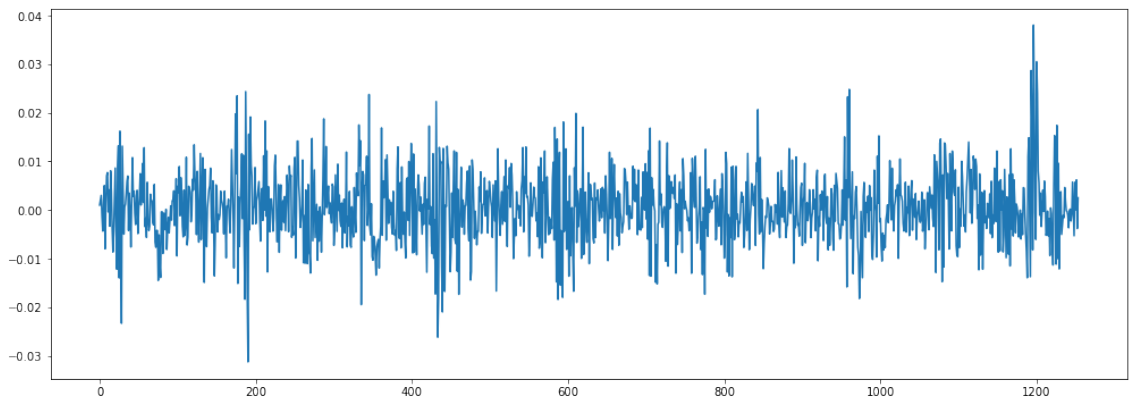

Figure 13: $h_n = \sqrt{1.6444 \cdot 10^{-5} + 0.2557h_{n-1}^2 + 0.2547h_{n-2}^2 + 0.2547h_{n-3}^2}\varepsilon_n$

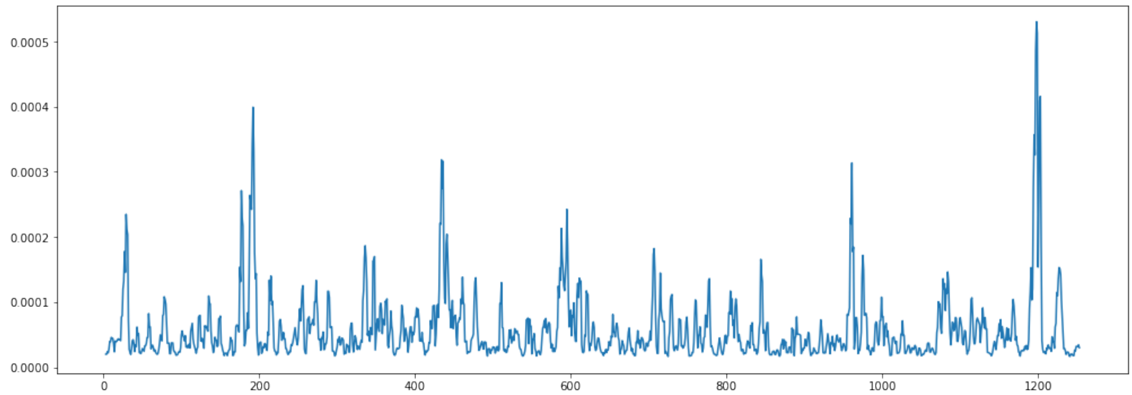

Figure 14: $\sigma_n^2 = 1.6444 \cdot 10^{-5} + 0.2557 h_{n-1}^2 + 0.2547 h_{n-2}^2 + 0.2547 h_{n-3}^2$

Forecast

Applying Eq.(26) from Chapter 1 to $m = 100$ I obtain that $\hat{h_{n+m}^2} = 0.000113$ that is literally equal to $\mathbb{E}h_n^2 = 0.000113$

Forecast with python (test/train sample)

I divided our sample into train and test one in order to predict the last 7 values of volatility of returns. Firstly, I fitted ARCH(1) model on train sample, then applied forecast method on it and added plot of actual and predicted volatility.

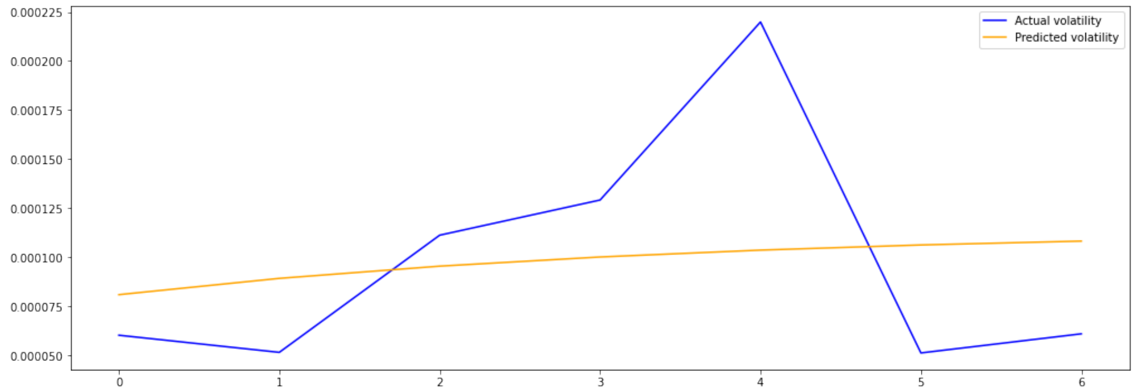

Figure 15: Forecast of volatility for ARCH(1)

2.2 Application of GARCH model

Firstly, I calculated coefficients manually, using Eq.(30) and obtained such results: $\hat{p}(1) = 0.55210036, \quad \hat{p}(2) = 0.57086614$, so $\hat{\phi} = 0.9671, \quad \hat{b} = 2.097, \quad \hat{\theta} = -0.733$

$$\hat{\omega} = 3.715 \cdot 10^{-6}, \quad \hat{\alpha} = 0.234, \quad \hat{\beta} = 0.733$$

$$\sigma_n^2 = 3.715 \cdot 10^{-6} + 0.234 h_{n-1}^2 + 0.733 \sigma_{n-1}^2$$

Estimation with Python (MLE)

Volatility Model

	coef	std err	t	P>|t|	95.0% Conf. Int.
omega	2.2725e-06	1.117e-09	2035.133	0.000	[2.270e-06,2.275e-06]
alpha[1]	0.2000	1.372e-02	14.577	3.962e-48	[0.173, 0.227]
beta[1]	0.7800	1.476e-02	52.847	0.000	[0.751, 0.809]

Simulated returns and volatility for GARCH(1)

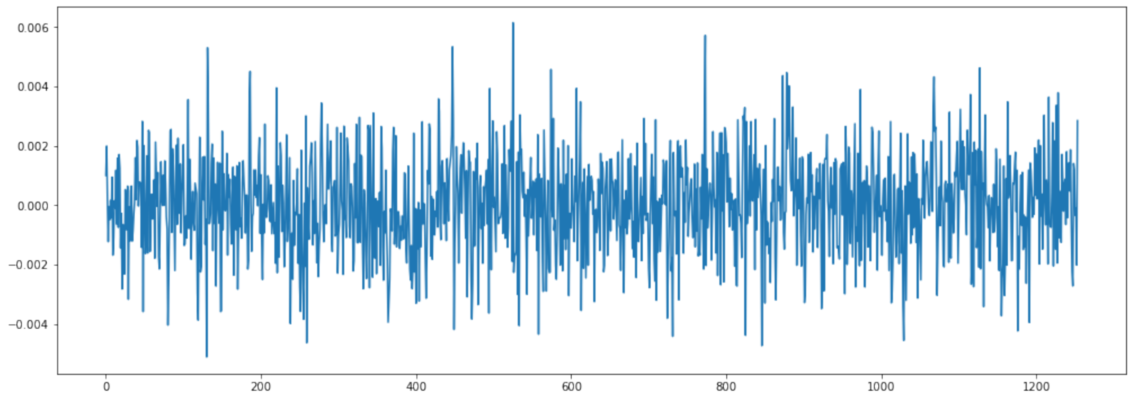

Figure 16: $h_n = \sqrt{2.2711 \cdot 10^{-6} + 0.2h_{n-1}^2 + 0.78\sigma_{n-1}^2}\varepsilon_n$

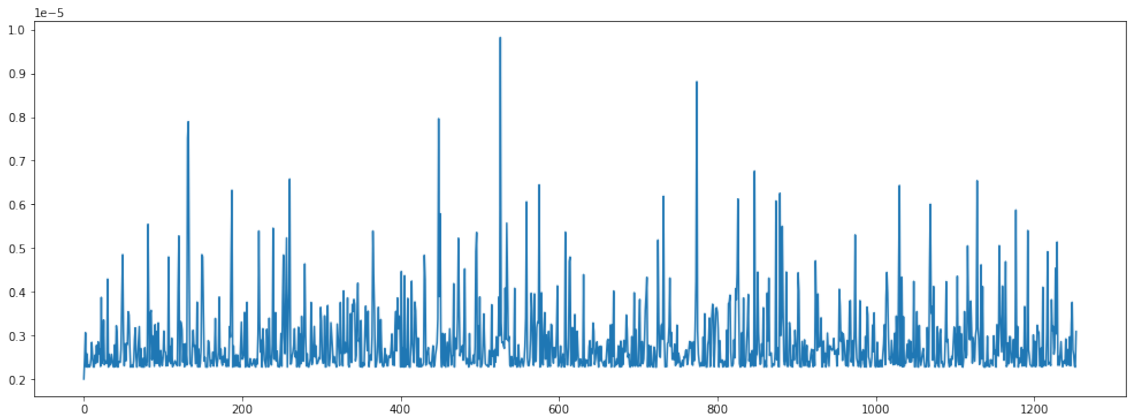

Figure 17: $\sigma_n^2 = 2.2711 \cdot 10^{-6} + 0.2h_{n-1}^2 + 0.78\sigma_{n-1}^2$

2.3 Model diagnostic

In ARCH model standardized shocks $\tilde{a}_t = \frac{a_t}{\sigma_t}$ are independent identically distributed random variables following either normal or standardized Student-t distribution. It is common way to apply the Ljung-Box statistics of $\tilde{a}_t$ to check adequacy of mean equation and $\tilde{a_t^2}$ in order to check if the volatility equation is valid. In addition, it can be useful to analyse QQ-plot, ACF plot and value of skewness, kurtosis.[16]
The Ljung-Box test is a statistical test that checks if autocorrelation exists in a time series. The null hypothesis defines that residuals independently distributed.

The alternative hypothesis is that residuals are not independently distributed.
If the p-value is less than some threshold, you can reject the null hypothesis and conclude that the residuals are not independently distributed, otherwise they are iid.
For ARCH(1) we obtain such results:
Ljung-Box statistics $stat = 0.114188, pvalue = 0.735426$, then p-value is quite large from alpha level and we accept null hypothesis.

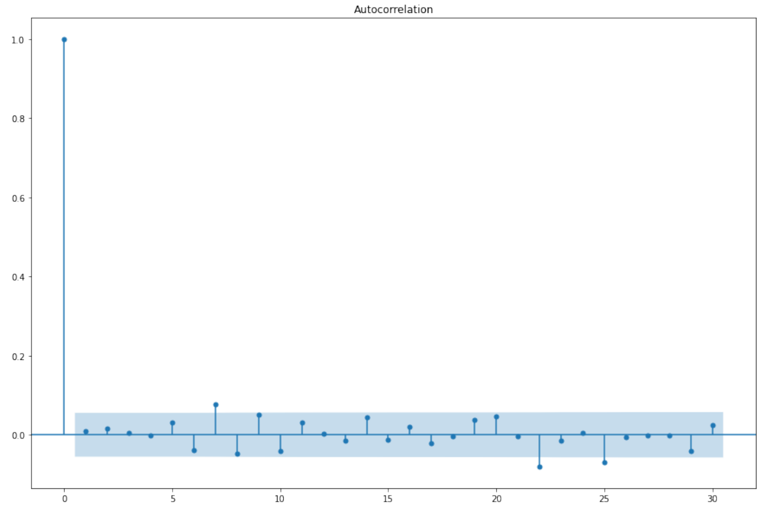

Figure 18: ACF of residuals ARCH(1)

We can observe that residuals are serially uncorrelated, no any significant lags are present.

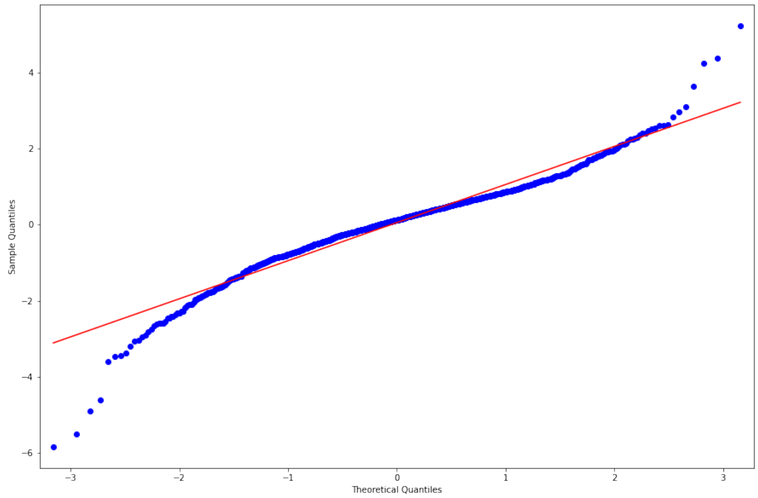

Figure 19: QQ-plot

In general, most values follow law of normal distribution, except on the tails they deviate. Calculated value of excess kurtosis is $K = 0.9015$ it is bigger than kurtosis of Gaussian distribution, hence the distribution may have heavy tails that is shown in the following histogram.

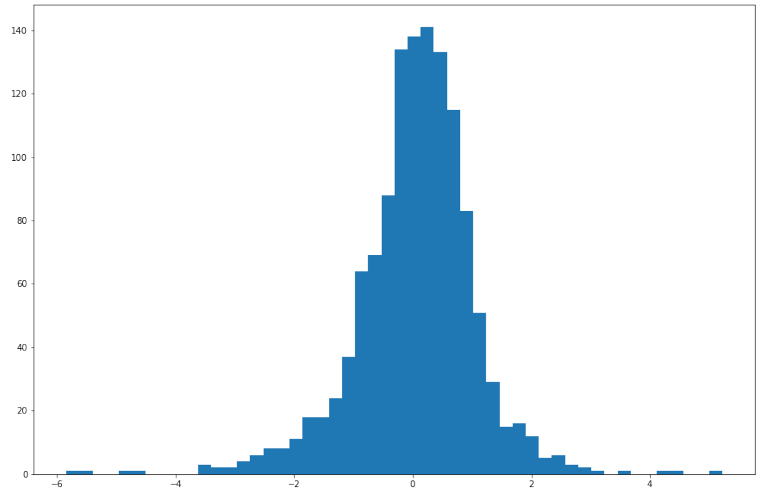

Figure 20: Histogram of residuals ARCH(1)

Analysing GARCH(1,1) we obtain the following results:
Ljung-Box statistics $stat = 0.335288, pvalue = 0.562562$, then p-value is quite large from alpha level and we accept null hypothesis.

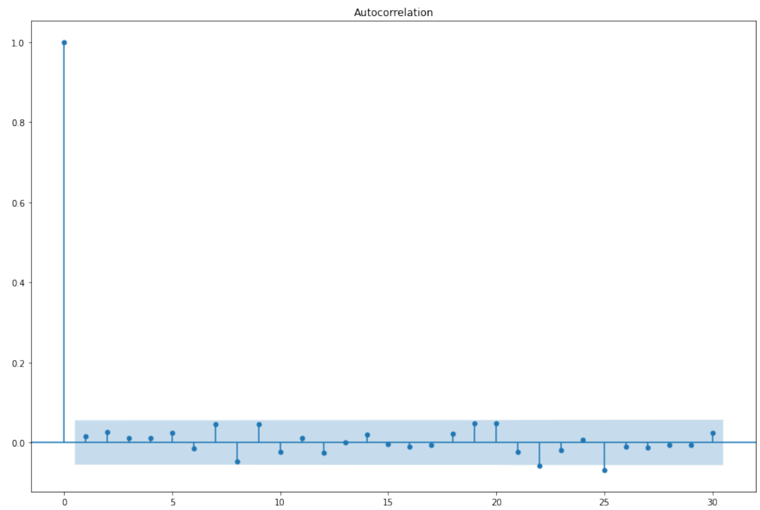

Figure 21: ACF of residuals GARCH(1,1)

Residuals are serially uncorrelated, no any significant lags are present.

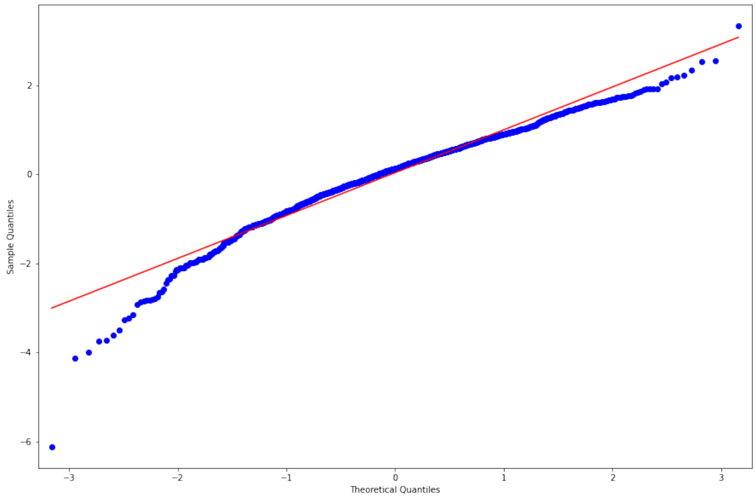

Figure 22: QQ-plot

In general, most values follow law of normal distribution, except on the tails they deviate.

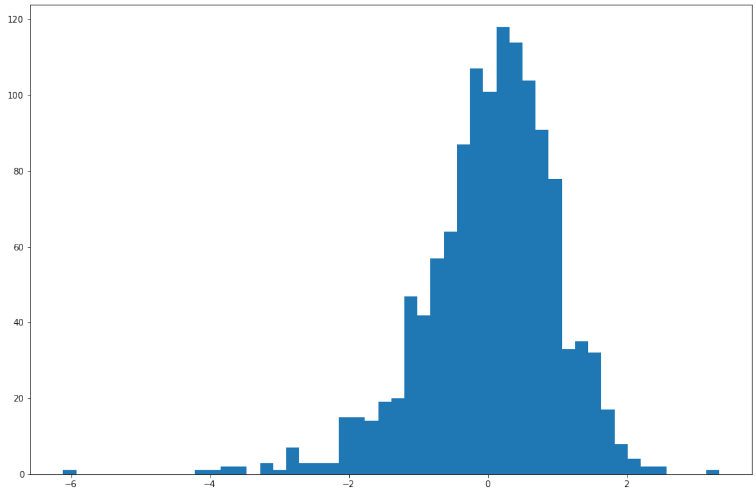

Figure 23: Histogram of residuals GARCH(1,1)

In order to find "best" model we can apply Akaike Information Criterion.[21] The formula for the AIC score is as follows:

$$AIC = 2k - 2ln(\mathcal{L}),$$

where k - is the number of model parameters, $\mathcal{L}$ - is the maximum value of the likelihood function of the model.
The model with the lowest AIC offers the best fit. The absolute value of the AIC value is not important, it can be positive or negative. In our case for ARCH(1) $AIC = -8943.95$, for ARCH(3) - -9113.80, for GARCH(1) - -9176.03 respectively. Therefore, for GARCH(1) model AIC criterion is the lowest, so it is appropriate model.

Conclusion

ARCH / GARCH models belong to the class of nonlinear models with conditional variance.

In this work were implemented the following steps:

- Prices of stocks were transformed into more common representation such as difference of log returns in order to follow weak stationary assumption.
- By the means of F-statistic it was tested and proven that residual series of the model contain ARCH effects, namely conditional heteroscedasticity. The F-statistic was significant, so ARCH model can adequately describe financial time series.
- PACF plot of squared returns h_n^2 was used to determine order of ARCH. It was defined that the first three lags are significant.
- In order to estimate the model was used Method of Moments that provided the following coefficients: $\alpha_0 = 0.000049, \alpha_1 = 0.56655$, thus the obtained equation of returns is $h_n = \sqrt{0.000049 + 0.56655 h_{n-1}^2} \varepsilon_n$. In addition, parameters were evaluated with Maximum Likelihood approach in Python, such that $\alpha_0 = 0.0000267, \alpha_1 = 0.6743$, regarding to quite low level of p-value parameters of model are significant. Consequently, with simulated values estimations of two methods are very similar.
- Volatitility and log-returns were simulated for ARCH(1) and ARCH(3).
- It was confirmed that forecasting log-returns $\hat{h_{n+m}^2}$ for a large step m is striving to second moment under assumption of weak stationarity.
- Ultimately, it was predicted volatility of ARCH(1) for 7 days.

Additionally, I have estimated GARCH(1,1) using both methods, so log-returns are defined with the following equation $h_n = \sqrt{2.2711 \cdot 10^{-6} + 0.2 h_{n-1}^2 + 0.78 \sigma_{n-1}^2} \varepsilon_n$. It was simulated volatility and log returns for GARCH(1,1). Finally, it was predicted volatility of GARCH(1,1) for 7 days. In order to choose the "best" model was applied Akaike Information Criterion which showed that GARCH(1,1) fits data better than other modifications of model.

Bibliography

1. Albert N. Shiryaev Essentials of Stochastic Finance. Facts,Models,Theory, World Scientific, (1999), 152-153
2. Tim Bollerslev, Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics, **31(3)**,(1986), 307-327.
3. Engle, R. F., Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica, **50(4)**, (1982) 987–1007.
4.Albert N. Shiryaev Essentials of Stochastic Finance. Facts,Models,Theory.
5. Christian Gourieroux ARCH Models and Financial Applications.
6.Stephanie Glen. "Stochastic Model / Process: Definition and Examples" From StatisticsHowTo.com.
7.https://en.wikipedia.org/wiki/White_noiseStatistical_properties
8.Mei-Yuan Chen Time Series Analysis: Conditional Volatility Models
9. Lucas L. Why use logarithmic returns in time series modelling
10.Will Kenton. Excess Kurtosis. Investopedia.
11.Kaddour Najim, Enso Ikonen, Ait-Kadi Daoud Stochastic Processes Estimation, Optimization Analysis
12.https://ca.finance.yahoo.com/quote/13. Chen, James. "Toronto Stock Exchange (TSX)". Investopedia.
14. AKHILESH GANTI. "Adjusted Closing Price". Investopedia.
15. Nason, G. P. (2006). Stationary and non-stationary time series. Statistics in volcanology, 60.
16. Ruey S. Tsay, Analysis of Financial Time Series, A Wiley-Interscience Publication JOHN WILEY SONS, INC., (1951)
17. Dennis Kristensen, Oliver Linton. A closed-form-estimator for the GARCH(1,1)-model
18.https://www.reed.edu/economics/parker/312/tschapters/S13_Ch_1.pdf
19.https://en.wikipedia.org/wiki/Martingale_(probability_theory)
20.Kaddour Najim, Enso Ikonen, Ait-Kadi Daoud, Stochastic Processes Estimation, Optimization Analysis, Elsevier, (2004)
21.Nagaraj Naik, Biju R Mohan, Rajat Aayush Jha, GARCH-Model Identification based on Performance of Information Criteria, 171, (2020)
22. https://online.stat.psu.edu/stat510/lesson/1/1.2
23. https://ru.wikipedia.org/wiki/24.https://en.wikipedia.org/wiki/White_noise
25. І. Г. Лук'яненко, В. М. Жук, АНАЛІЗ ЧАСОВИХ РЯДІВ Побудова ARIMA, ARCH/GARCH моделей з використанням пакета E.Views 6.0, (2013)

Appendix

Definition 1.1

$\omega_1, ..., \omega_N$ is called elementary events and finite set $\Omega = \{\omega_1, ..., \omega_N\}$ is called space of elementary events.

Definition 1.2

$Let\ X$ be some set, and let $P(X)$ represent its power set. Then a subset $\mathcal{F} \subseteq P(X)$ is called ***σ-algebra*** if it satisfies the following properties:

1. X is in $\mathcal{F}$.
2. $\mathcal{F}$ is closed under complementation: if $\mathcal{A}$ is in $\mathcal{F}$, then so is its complement X $\setminus\mathcal{A}$ is in $\mathcal{F}$.
3. $\mathcal{F}$ is closed under countable unions: if $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3...$ are in $\mathcal{F}$, then so $\bigcup_{i=1}^{\infty} \mathcal{A}_i$ is in $\mathcal{F}$.

Definition 1.3

An ordered triple $(\Omega, \mathcal{F}, \mathcal{P})$ where Ω is a set of points ω, $\mathcal{F}$ is σ -algebra of subsets of Ω, $\mathcal{P}$ is a probability on $\mathcal{F}$ is called ***probabilistic model or a probability space.***

Definition 1.4

Borel σ ***algebra*** on topological space X is the smallest σ-algebra containing all open sets (or equivalently all closed sets).

Definition 1.5

A random variable is a function X:$\Omega \to \mathbb{R}$. It is said to be ***measurable w.r.t*** $\mathcal{F}$ if for every Borel set $B \in \mathcal{B}(\mathbb{R})$

$$X^{-1}(B) := \{\omega \in \omega : X(\omega) \in B\} \in \mathcal{F}$$

Definition 1.6

Let X be a set. Let $\mathcal{A}, \mathcal{B}$ be σ-algebras on X. Then $\mathcal{B}$ is said to be a ***sub-sigma-algebra*** of $\mathcal{A}$ if and only if $\mathcal{B} \subseteq \mathcal{A}$.

Definition 1.7

Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$, and let $X \in \mathcal{L}^1$ be a random variable. We say that the random variable ξ is the conditional expectation of X with respect to $\mathcal{G}$ - and denote it by $\mathbb{E}[X|\mathcal{G}]$ if

1. $\xi \in \mathcal{L}^1$
2. ξ is $\mathcal{G}$ -measurable,
3. $\mathbb{E}[\xi \mathbb{1}_A] = \mathbb{E}[X \mathbb{1}_A]$, for all A $\in \mathcal{G}$

Definition 1.8

For discrete random variables, the ***conditional probability mass function*** of Y given X=x can be written according to its definition as:

$$p_{Y|X}(y \mid x) = P(Y = y \mid X = x) = \frac{P(\{X = x\} \cap \{Y = y\})}{P(X = x)}$$

Definition 1.9

The ***joint probability mass function*** of two discrete random variables X,Y is:

$$p_{X,Y}(x, y) = P(X = x \text{ and } Y = y)$$

or written in terms of conditional distributions

$$p_{X,Y}(x,y) = P(Y=y \mid X=x)\cdot P(X=x) = P(X=x \mid Y=y)\cdot P(Y=y)$$

$\xi(\omega) = \frac{1}{\mathbb{P}[A]}\mathbb{E}[X\mathbb{1}_A] = \sum\limits_{\omega\in A} X(w)\mathbb{P}[\{\omega\} \mid A],$ or all $\omega \in A$.

Definition 1.11

If X and Y are discrete random variables, the conditional expectation of X given Y is

$$\mathbb{E}(X \mid Y=y) = \sum_x xP(X=x \mid Y=y) = \sum_x x\frac{P(X=x, Y=y)}{P(Y=y)}$$

where $P(X=x, Y=y)$ is joint probability mass function of X and Y.

Statement 1 Law of the unconscious statistician

Let X be a random variable and let $Y=g(X)$ be a function of this random variable. If X is a discrete random variable and $p_X(x)$, the expected value of $g(X)$ is

$$\mathbb{E}[g(X)] = \sum_{x\in\mathcal{X}} g(x)p_X(x).$$

Properties of conditional expectation

Taking out what is known

If X is $\mathcal{H}$-measurable, then $\mathbb{E}(XY \mid \mathcal{H}) = X\mathbb{E}(Y \mid \mathcal{H})$.

Theorem 1(Law of Iterated Expectations, "Adam's Law")

For any random element X $\in \mathcal{X}$ and random variable Y $\in \mathcal{Y} \subset \mathbb{R}$,

$$\mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}X$$

$\mathbb{E}[\mathbb{E}[X \mid Y]] = \sum\limits_y \mathbb{E}[X \mid Y=y]P(Y=y) = \sum\limits_y\sum\limits_x xP(X=x \mid Y=y)P(Y=y)$
$= \sum\limits_y\sum\limits_x xP(Y=y \mid X=x)P(X=x) = \sum\limits_x xP(X=x)\sum\limits_y P(Y=y \mid X=x) = \sum\limits_x P(X=x) = \mathbb{E}[X]$

In the first step let $f(x) = \mathbb{E}[X \mid Y]$ and apply LOTUS (Statement 1).

Projection interpretation

If we consider $\mathbb{E}[X \mid Y]$ as prediction for X given Y then $X - \mathbb{E}[X \mid Y]$ is the residual of that prediction.

Theorem 2 (Projection interpretation)

For any $h: Y \to \mathbb{R}, \mathbb{E}[(Y - \mathbb{E}[Y \mid X])h(Y)] = 0$

Proof

By linearity of conditional expectation we have $\mathbb{E}[Xh(Y)] - \mathbb{E}[\mathbb{E}[X \mid Y]h(Y)] = \mathbb{E}[Xh(Y)] - \mathbb{E}[\mathbb{E}[X \mid Y]h(Y) =$
$\mathbb{E}[Xh(Y)] - \mathbb{E}[\mathbb{E}[Xh(Y) \mid Y]] = \mathbb{E}[Xh(Y)] - \mathbb{E}[Xh(Y)] = 0$

Definition 1.12

The **covariance** of random variable X and Y is defined
$Cov(X,Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y$. If $Cov(X,Y) = 0$, then we say X and Y are **uncorrelated**.

Statement 2

The residual $X - \mathbb{E}X \mid Y$ and $h(Y)$ are uncorrelated for every function $h : \mathcal{Y} \to \mathbb{R}$.

Proof

By linearity and Law of iterated expectation we have

$\mathbb{E}[X - \mathbb{E}[X \mid Y]] = \mathbb{E}X - \mathbb{E}[\mathbb{E}[X \mid Y]] = 0$

$Cov(X - \mathbb{E}[X \mid Y], h(Y)) = \mathbb{E}[(X - \mathbb{E}[X \mid Y])h(Y)] - \mathbb{E}[X - \mathbb{E}[X \mid Y]]\mathbb{E}[h(Y)] = \mathbb{E}[(X - \mathbb{E}[X \mid Y])h(Y)] = 0.$

TABLE A.3

F Distribution: Critical Values of *F* (5% significance level)

v_2 \ v_1	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.36	246.46	247.32	248.01
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.42	19.43	19.44	19.45
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.71	8.69	8.67	8.66
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.87	5.84	5.82	5.80
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.64	4.60	4.58	4.56
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.96	3.92	3.90	3.87
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.53	3.49	3.47	3.44
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.24	3.20	3.17	3.15
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.03	2.99	2.96	2.94
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.86	2.83	2.80	2.77
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.74	2.70	2.67	2.65
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.64	2.60	2.57	2.54
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.55	2.51	2.48	2.46
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.48	2.44	2.41	2.39
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.42	2.38	2.35	2.33
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.37	2.33	2.30	2.28
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.33	2.29	2.26	2.23
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.29	2.25	2.22	2.19
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.26	2.21	2.18	2.16
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.22	2.18	2.15	2.12
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.20	2.16	2.12	2.10
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.17	2.13	2.10	2.07
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.15	2.11	2.08	2.05
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.13	2.09	2.05	2.03
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.11	2.07	2.04	2.01
26	4.22	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.09	2.05	2.02	1.99
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.08	2.04	2.00	1.97
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.06	2.02	1.99	1.96
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.05	2.01	1.97	1.94
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.04	1.99	1.96	1.93
35	4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16	2.11	2.04	1.99	1.94	1.91	1.88
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.95	1.90	1.87	1.84
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.95	1.89	1.85	1.81	1.78
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.86	1.82	1.78	1.75
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02	1.97	1.89	1.84	1.79	1.75	1.72
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95	1.88	1.82	1.77	1.73	1.70
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94	1.86	1.80	1.76	1.72	1.69
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.85	1.79	1.75	1.71	1.68
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.78	1.73	1.69	1.66
150	3.90	3.06	2.66	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.82	1.76	1.71	1.67	1.64
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.80	1.74	1.69	1.66	1.62
250	3.88	3.03	2.64	2.41	2.25	2.13	2.05	1.98	1.92	1.87	1.79	1.73	1.68	1.65	1.61
300	3.87	3.03	2.63	2.40	2.24	2.13	2.04	1.97	1.91	1.86	1.78	1.72	1.68	1.64	1.61
400	3.86	3.02	2.63	2.39	2.24	2.12	2.03	1.96	1.90	1.85	1.78	1.72	1.67	1.63	1.60
500	3.86	3.01	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.77	1.71	1.66	1.62	1.59
600	3.86	3.01	2.62	2.39	2.23	2.11	2.02	1.95	1.90	1.85	1.77	1.71	1.66	1.62	1.59
750	3.85	3.01	2.62	2.38	2.23	2.11	2.02	1.95	1.89	1.84	1.77	1.70	1.66	1.62	1.58
1000	3.85	3.00	2.61	2.38	2.22	2.11	2.02	1.95	1.89	1.84	1.76	1.70	1.65	1.61	1.58

.