

TIME SERIES ANALYSIS: LAB SESSION 1

Hugo S. ROCHA

2025-02-28

Exercise 1-a Studying a Gaussian White Noise

We begin by setting the seed to ensure reproducibility of the results.

```
set.seed(344)
```

1. Generate 1000 observations from a normal distribution with 0 mean and a standard deviation of 1.7 This process is named (ut).

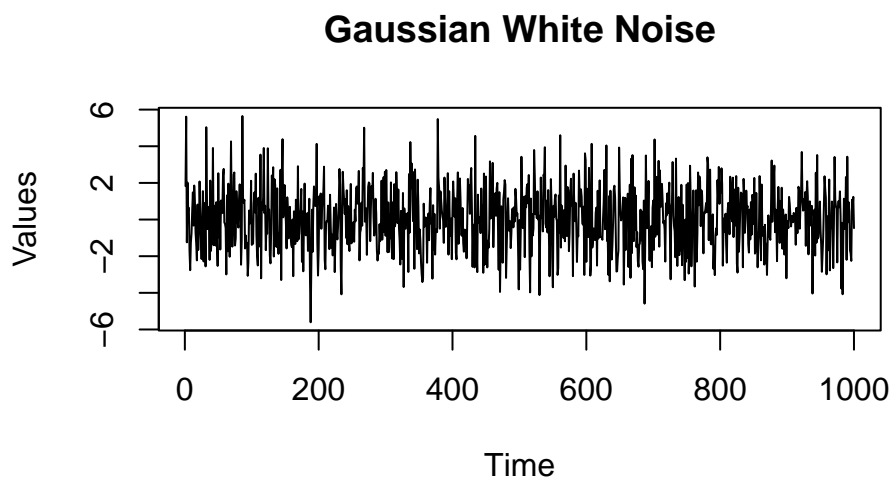
This simulated process is stored in the variable u_t .

```
n <- 1000  
u_t <- rnorm(n, mean = 0, sd = 1.7)
```

2. Using a function plot, plot the generated series.

The plot below displays the Gaussian white noises series.

```
plot.ts(u_t, main = "Gaussian White Noise", ylab = "Values", xlab = "Time")
```



The time series plot shows that the data fluctuates randomly around zero without any trend or seasonal pattern, which is typical for a white noise.

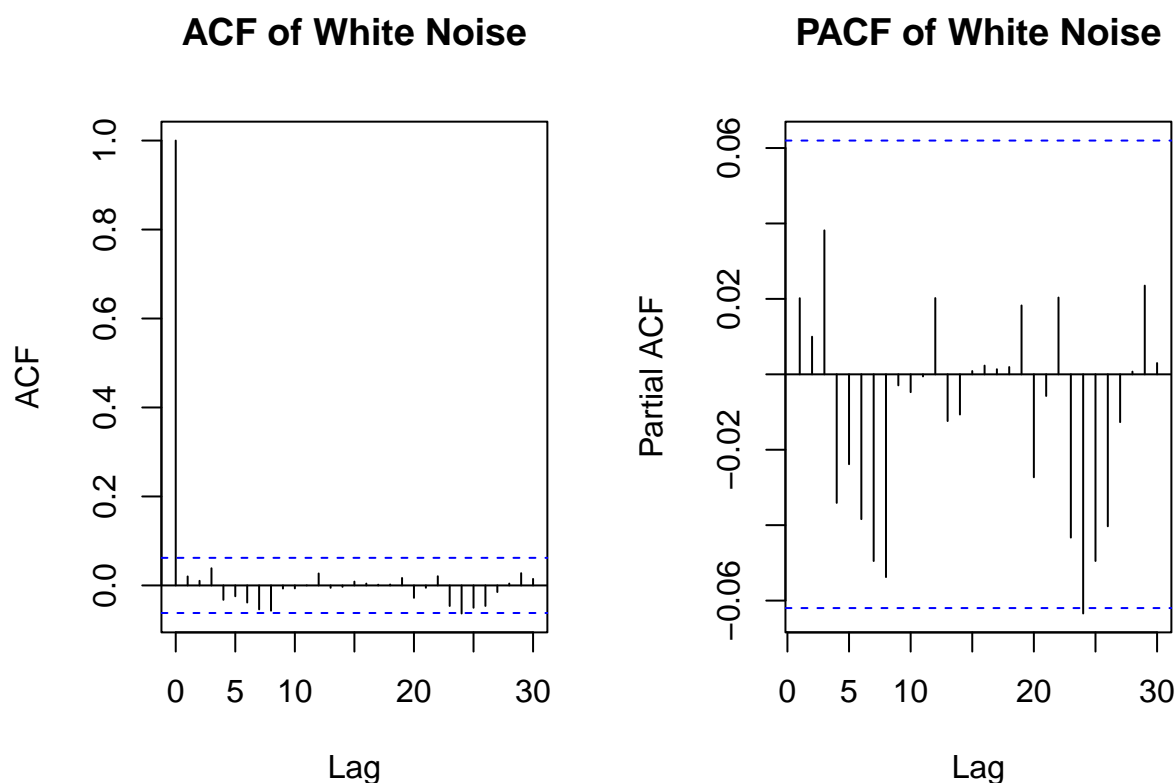
3.Explain what is an autocorrelation function. Plot the ACF of (ut). How do you interpret the ACF results ? Are they aligned with the definition of the white noise.

The autocorrelation function (ACF) measures the correlation between the time series and its lagged versions. For a true white noise process, aside from lag 0 (which is always 1), the correlations at other lags should be close to zero. The partial autocorrelation function (PACF) shows the direct effect of a lag on the series after removing the effects of intermediate lags. For white noise, both the ACF and PACF (beyond lag 0) should not show significant correlations.

The formula for the ACF is:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}$$

```
par(mfrow = c(1, 2));acf(u_t, main = "ACF of White Noise");pacf(u_t, main = "PACF of White Noise")
```



In the ACF plot, we observe a significant spike at lag 0 (which is normal as it's the correlation of the series with itself), while the subsequent lags variate near zero. Similarly, the PACF plot confirms that, once the effect of the intervening lags is removed, there are no significant correlations. This behavior aligns with the definition of white noise.

Exercise 1-b Studying a random walk

We suggest to run the same analysis using a random walk process. Formally speaking, y_t :

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

A random walk can be seen as a cumulative sum of Gaussian white noises increments. We know it is a nonstationnary process: its variance increases over time, and any shock causes the variance to “explode”.

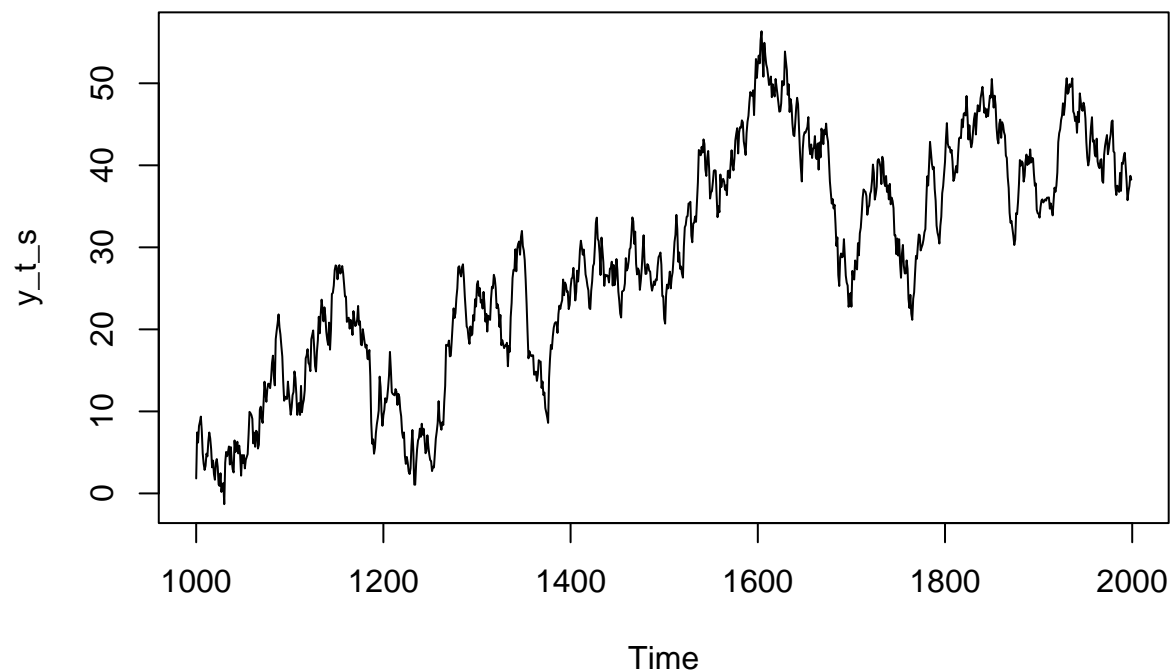
1. Generate a random walk based on the simulated white noise (keep the same seed)

Using the same white noise series u_t , we compute its cumulative sum to create a random walk.

```
y_t <- cumsum(u_t)
```

2. Convert the simulated data into a ts object (choose the the start date and freq). Plot the generated series.

```
y_t_s <- ts(y_t, start = c(1000,1), frequency =1)  
plot(y_t_s)
```



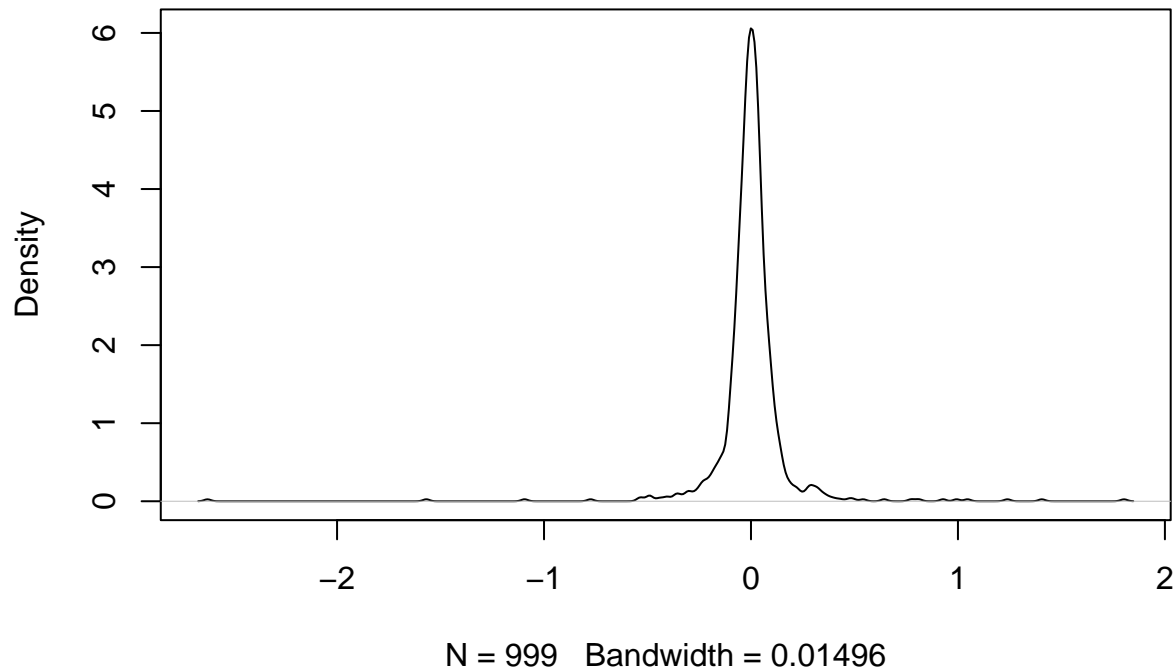
The plotted random walk exhibits a clear trend or drift over time, demonstrating nonstationarity. Unlike white noise, the random walk does not revert to a fixed mean, and its variance increases with time.

Plot its density. What do you observe ?

Since the random walk contains some negative values (we are not dealing with something that is always positive like the price of a stock), we first take the logarithm of the absolute values. Then we compute the differences (log-returns) to analyze the distribution of percentage change

```
log_y <- log(abs(y_t_s)) #  
rend_y <- diff(log_y)  
rend_y <- na.omit(rend_y)  
  
density_plot <- density(rend_y)  
plot(density_plot, main = "Density of Log>Returns")
```

Density of Log>Returns



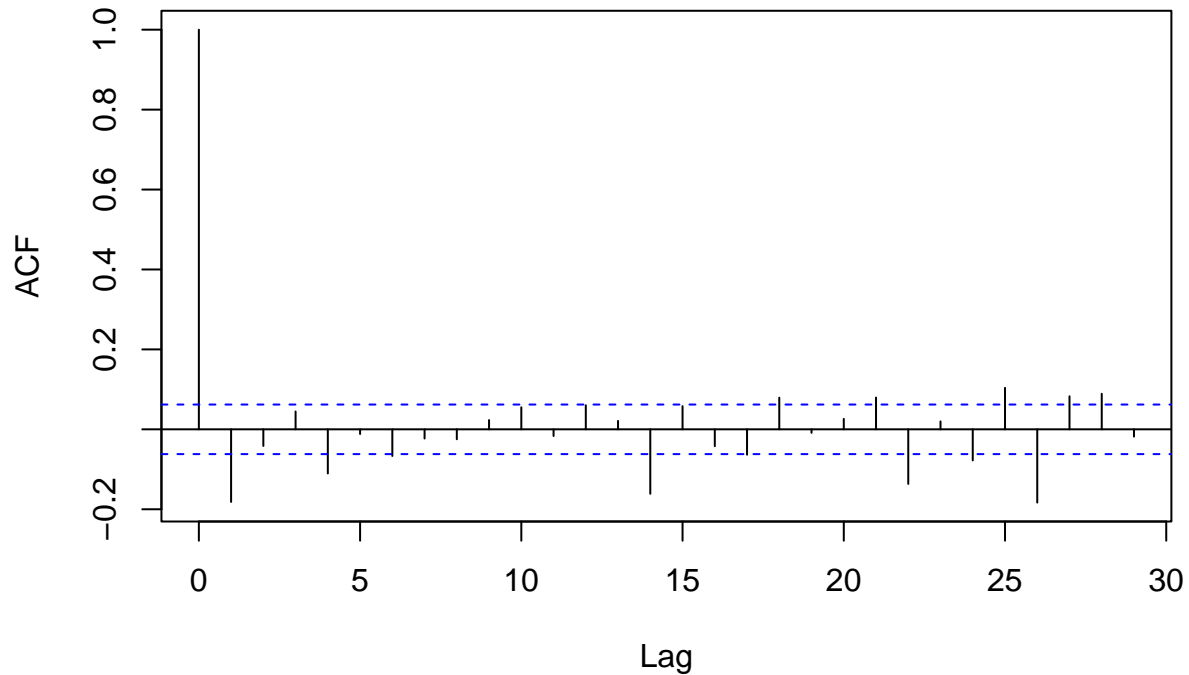
The density plot of the log-returns shows the distribution of the relative changes in the random walk. This distribution is more symmetric and may exhibit heavier tails than a normal distribution.

3. Plot the ACF of (wnt). How do you interpret this results? Can the models seen during the class be used for this type of data? Why?

We now plot the ACF of the log-returns to evaluate the serial correlation in the transformed (and presumably stationary) series.

```
acf(rend_y)
```

Series rend_y



After differencing, the log-returns often appear stationary. The ACF plot shows that most autocorrelations (beyond lag 0) are near zero, which suggests that the differenced series is stationary.

4. Run the Ljung Box test for the processes given in the exercise 1-a and 1-b and for a different lags (a for loop could be usefull). Plot the values of the $Q(m)$ and conclude.

The length of our time series is $N = 1000$. To choose the number of lags m we use $\ln(N) = m$. Since $\ln(1000) \approx 7$, we choose $m = 10$ to be more restrictive The test hypothesis are:

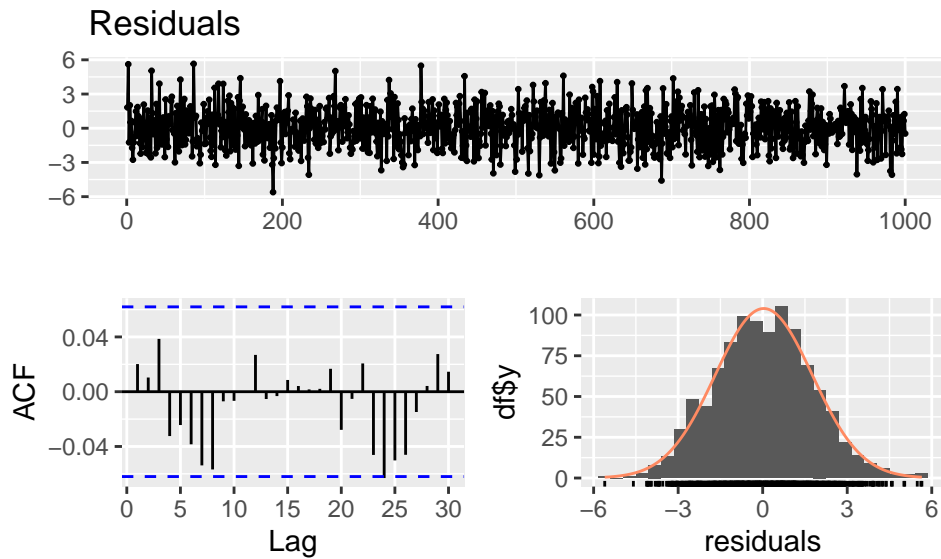
$$\begin{cases} H_0 : \rho(1) = \rho(2) = \dots = \rho(10) = 0 \\ H_1 : \exists i \in \{1, \dots, 10\} \text{ tel que } \rho(i) \neq 0 \end{cases}$$

Here, H_0 represents the absence of autocorrelation, and H_1 the presence of autocorrelation. We reject H_0 if the test statistic $Q^*(k)$ exceeds the critical threshold and the p-value is below our chosen significance level (i.e., the risk of a type I error).

```
library(forecast)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method      from
##   as.zoo.data.frame zoo
```

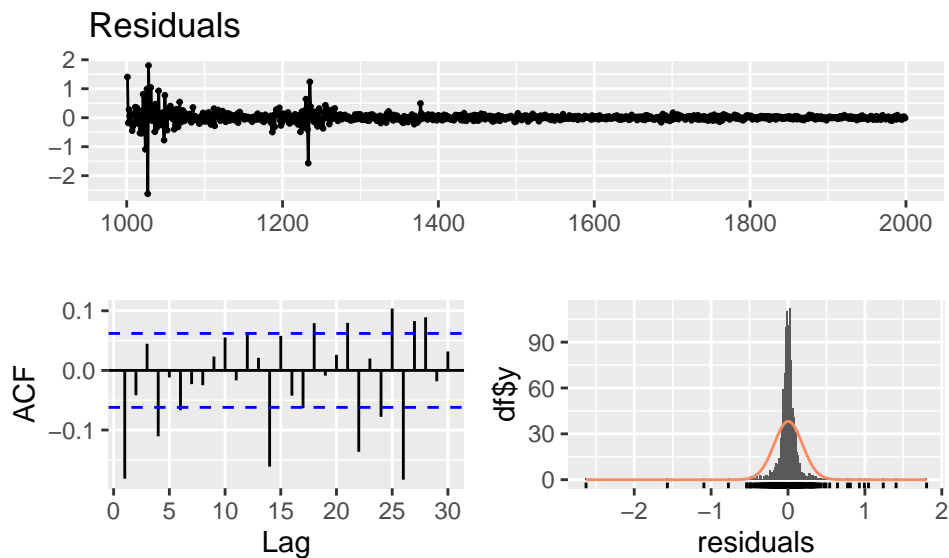
```
test_u_t <- checkresiduals(u_t, lag = 10)
```



```
##
##  Ljung-Box test
##
## data:  Residuals
## Q* = 11.425, df = 10, p-value = 0.3254
##
## Model df: 0.   Total lags used: 10
```

The Ljung-Box test checks for overall randomness based on a specified number of lags. In the white noise process (u_t), the p-value is above 0.05, indicating that we fail to reject the null hypothesis of no autocorrelation.

```
test_rend<-checkresiduals(rend_y,lag = 10)
```



```
##
##  Ljung-Box test
```

```
##
## data: Residuals
## Q* = 58.479, df = 10, p-value = 7.022e-09
##
## Model df: 0.    Total lags used: 10
```

For the log-returns ($rend_y$), if the differencing were fully effective, the series should be stationary, and we would expect the Ljung–Box test to yield a p-value above 0.05 (i.e., we fail to reject H_0). However, our results indicate otherwise—suggesting that there may be some residual autocorrelation (or possible volatility clustering) present.

Exercise 2- Importing, plotting and analyzing data

1. Import the database contained in the cvs file using the read.csv function. This database is a collection of stock prices.

```
data <- read.csv("C:\\Users\\Sebastian\\Downloads\\SP500.csv")
```

3. Plot the density of the selected stock daily returns? Using a quantile plot, check the normality of the empirical distribution of the daily returns. Display the four charts within the same output window

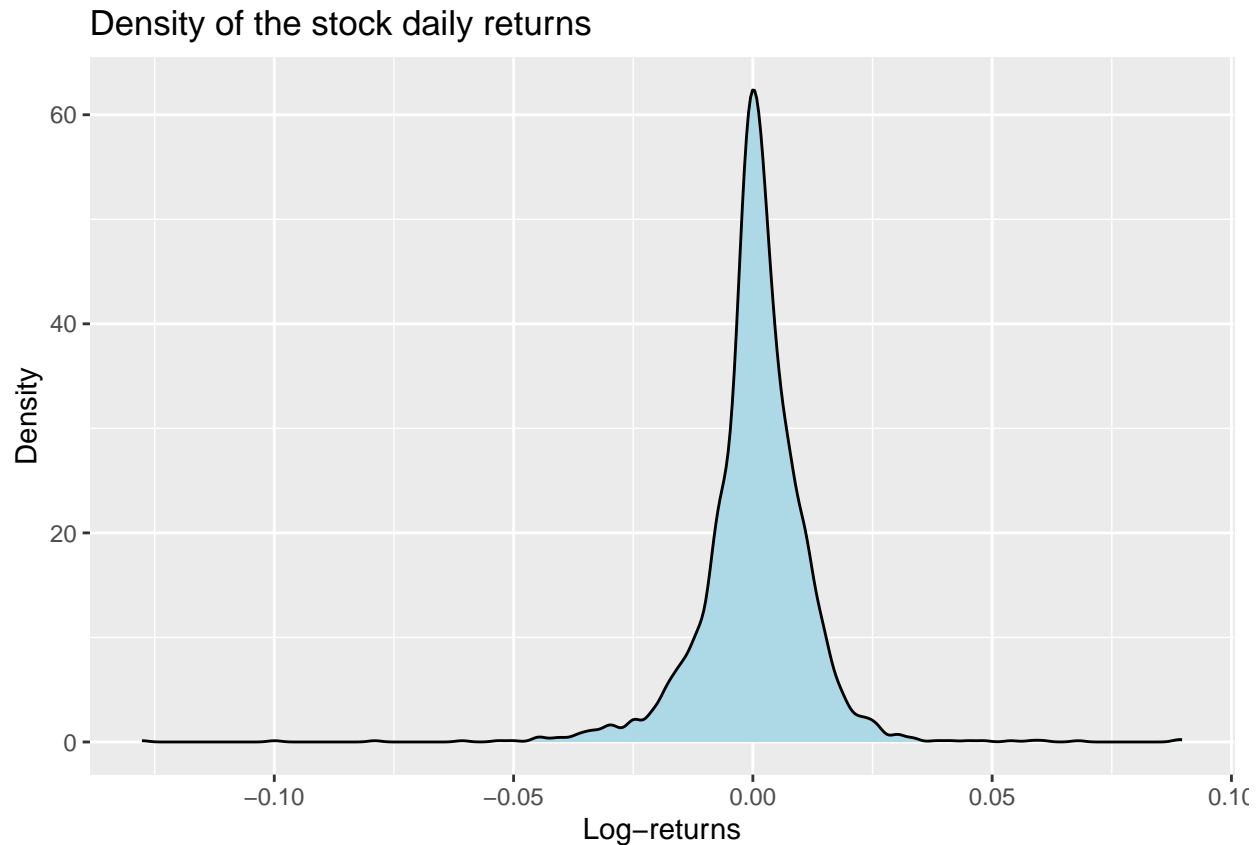
To calculate the selected stock daily returns, we log linearize our daily returns (contained in the column SP500) and then we differentiate them. We store this new variables in the column “log_rendements”

```
data$log_rendements <- c(NA, diff(log(data$SP500)))
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.4.3
```

```
ggplot(data, aes(x = log_rendements)) +
  geom_density(fill = "lightblue", na.rm = TRUE) +
  labs(title = "Density of the stock daily returns",
       x = "Log-returns",
       y = "Density")
```



The log-returns are sharply peaked around zero, with relatively long tails on both sides. In financial data, it is common to see a high peak around zero (many small returns) and occasional large positive or negative returns (fat tails).

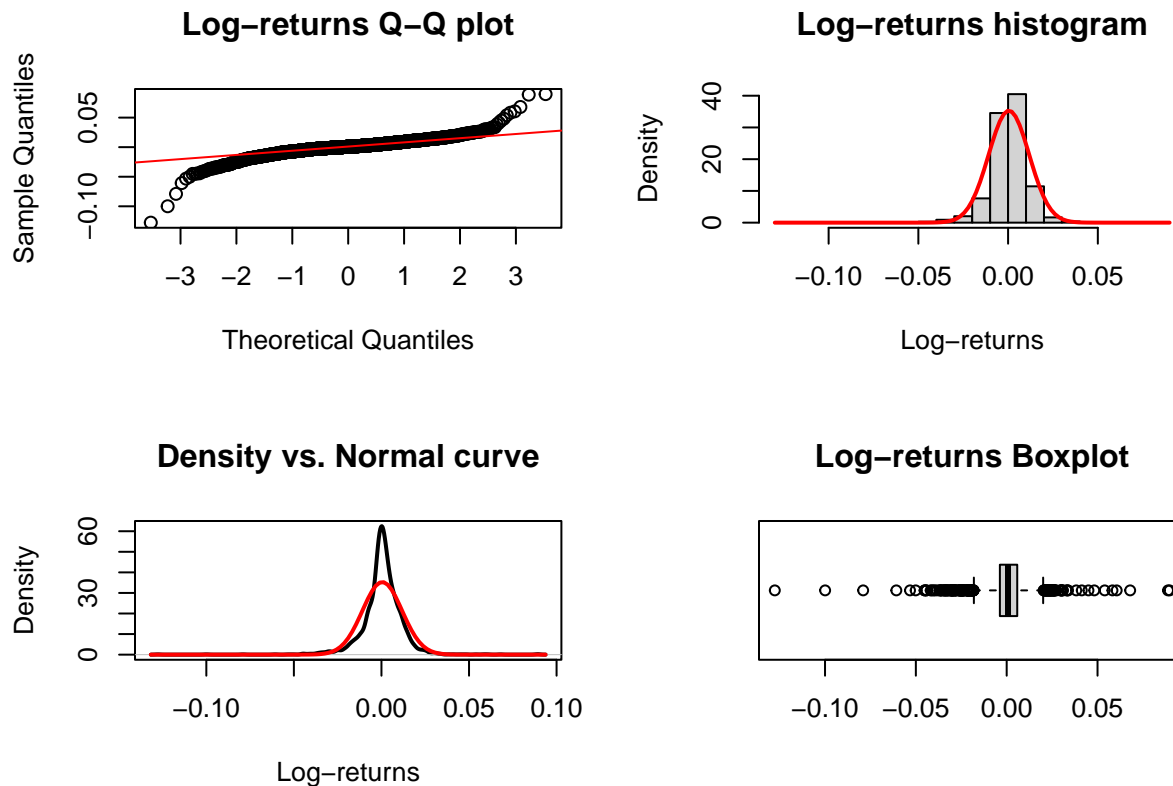
```
returns <- na.omit(data$log_rendements)
par(mfrow = c(2, 2))

qqnorm(returns, main = "Log-returns Q-Q plot")
qqline(returns, col = "red")

hist(returns, breaks = 30, probability = TRUE,
     main = "Log-returns histogram",
     xlab = "Log-returns", col = "lightgray")
curve(dnorm(x, mean = mean(returns), sd = sd(returns)),
      add = TRUE, col = "red", lwd = 2)

plot(density(returns), main = "Density vs. Normal curve",
     xlab = "Log-returns", lwd = 2)
curve(dnorm(x, mean = mean(returns), sd = sd(returns)),
      add = TRUE, col = "red", lwd = 2)

boxplot(returns, horizontal = TRUE, main = "Log-returns Boxplot")
```

Q-Q Plot: The points deviate noticeably from the straight line at both tails, sighting that the distribution is not strictly normal (heavier left tail and right tail).

Log-returns histogram: The histogram is more peaked at the center than a normal curve, again indicating “fat tails” and higher kurtosis.

Density vs. Normal Curve: The black line (empirical density) is more peaked and has thicker tails compared to the red normal curve.

Boxplot: Shows a cluster of data around the median (zero) and many points classified as outliers on both ends. This reaffirms the presence of extreme values.

Conclusion Overall, these four plots show that the distribution of log-returns is not Gaussian and exhibits heavier tails and a higher peak. As stated before, this is a typical feature of financial return data.

4. Calculate the autocovariance and the autocorrelation functions of the daily returns and calculate the Ljung-Box test for the various numbers of lags.

```
library(forecast)

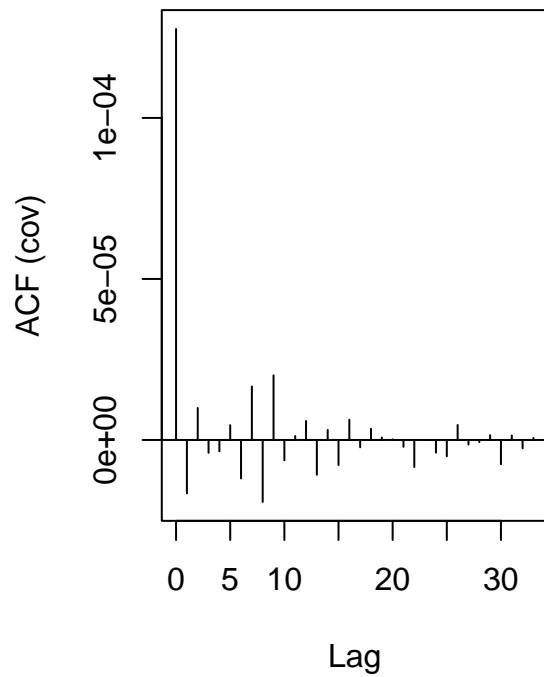
returns <- na.omit(data$log_rendements)

acf_cov <- acf(returns, type = "covariance", plot = FALSE)

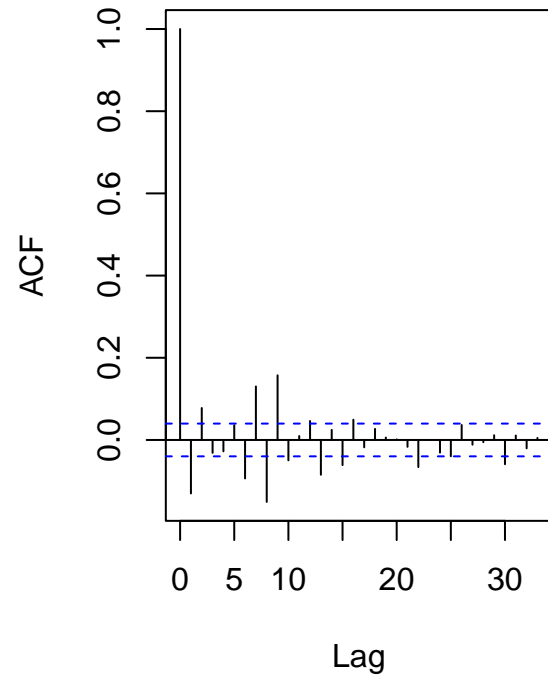
acf_cor <- acf(returns, type = "correlation", plot = FALSE)

par(mfrow = c(1, 2))
plot(acf_cov, main = "Log-returns autocovariance")
plot(acf_cor, main = "Log-returns autocorrelation")
```

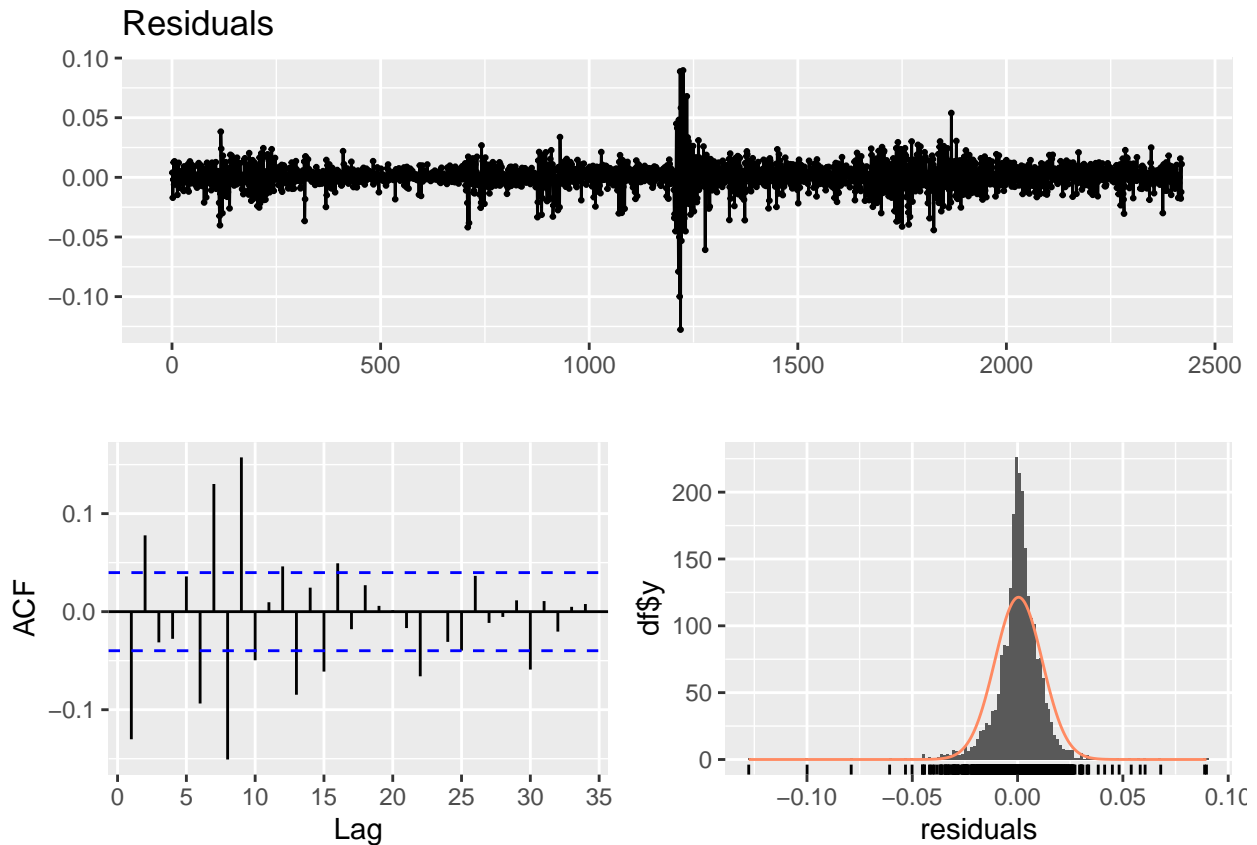
Log-returns autocovariance



Log-returns autocorrelation



```
par(mfrow = c(1, 1))  
test_rend <- checkresiduals(returns)
```



```
##
##  Ljung-Box test
##
## data:  Residuals
## Q* = 247.01, df = 10, p-value < 2.2e-16
##
## Model df: 0.   Total lags used: 10
```

Autocovariance and autocorrelation plots

Both plots show that most autocovariances and autocorrelations are around zero and fall within the confidence bands for the majority of lags. This suggests that the log-returns do not have a strong dependence over time.

Ljung box for daily returns

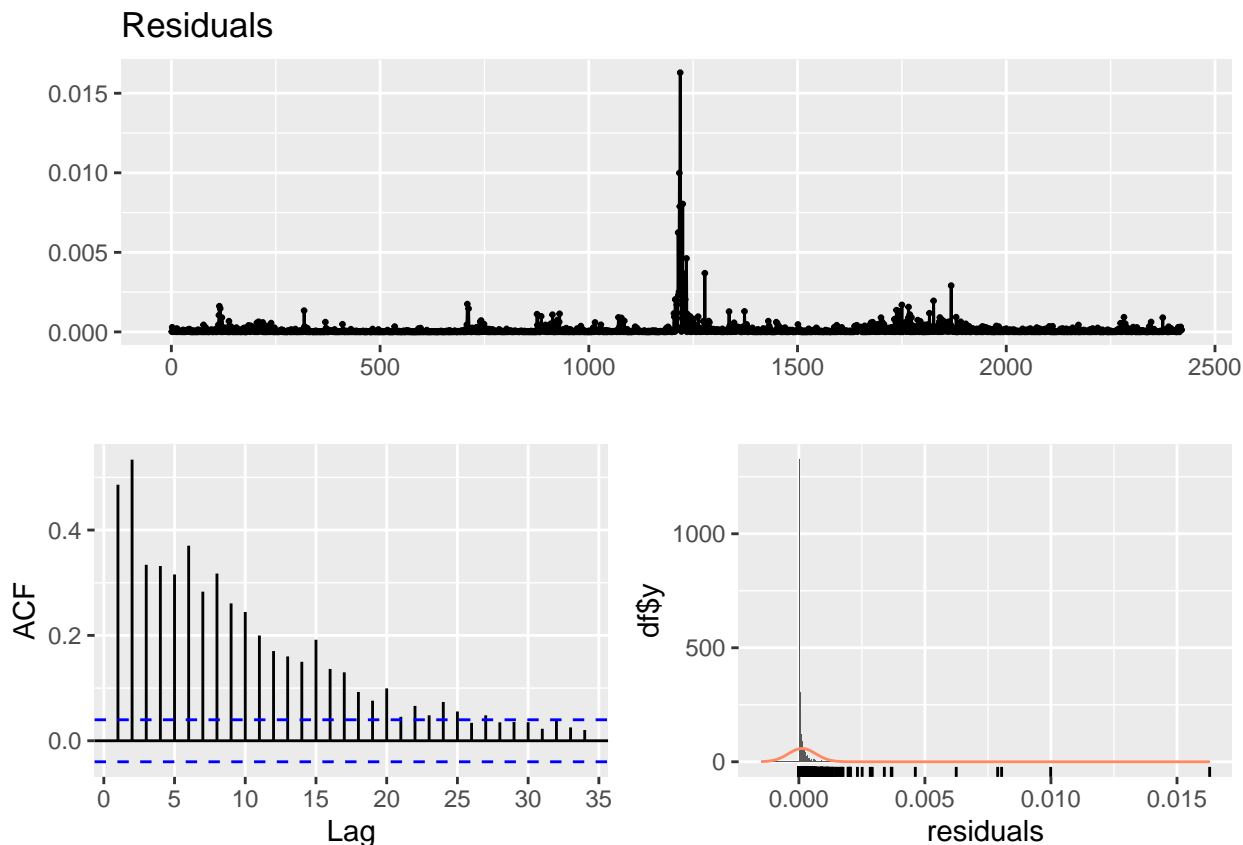
- The residual time series is centered around zero with no obvious pattern, aligning with stationarity in mean.
- The ACF bars do not show significant correlation at most lags.
- The density plot again indicates a tall, narrow peak and fatter tails compared to the superimposed normal curve (orange line).

Overall, the Ljung-Box test confirms that there is no strong evidence of autocorrelation in the raw (log) returns.

5. Plot the same ACF of the daily squared returns and calculate the Ljung-Box test for the various numbers of lags.

Squaring the returns typically amplifies large movements (volatility).

```
returns <- na.omit(data$log_rendements)
squared_returns <- returns^2
test_squared_ret <- checkresiduals(squared_returns)
```



```
##
##  Ljung-Box test
##
## data:  Residuals
## Q* = 3125.2, df = 10, p-value < 2.2e-16
##
## Model df: 0.   Total lags used: 10
```

- The top plot shows occasional spikes in volatility (large values).
- The ACF plot reveals that many lags are above the significance bands, indicating that squared returns are correlated over time. This is a hallmark of volatility clustering commonly seen in financial data.
- The Ljung-Box test on squared returns is usually significant, implying the presence of autocorrelation in volatility (even though the returns themselves may appear uncorrelated).

6. What are your conclusions regarding the stationary of the daily and squared daily returns

Daily Returns (log-returns):

- Appear approximately stationary in mean and variance: the ACF and Ljung-Box tests suggest no strong autocorrelation in the levels of returns.
- Distributions are not normal but rather exhibit heavier tails and a higher peak.

Squared Returns:

- The significant autocorrelation in squared returns indicates volatility clustering, which is typical in financial time series.
- Although the returns themselves are stationary, the variance of the process is not constant over time, reflecting changing volatility over the period.