MSc Business Analytics

Financial Modelling and Analysis

Chapter 2: First impression -exploratory financial data analysis.

Instructor: Roman Matkovskyy

Twitter: @matkovskyy

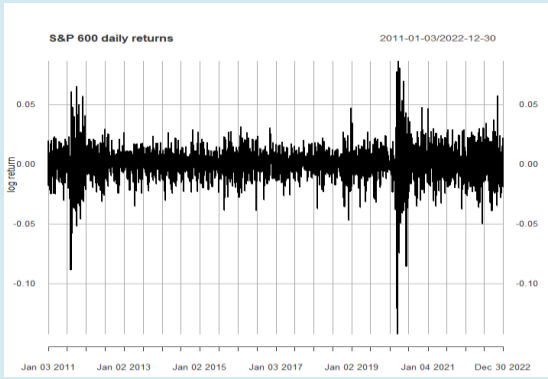
Outline

- The statistical analysis of financial markets data;
- Distributions
- Univariate tests: Shapiro-Wilk, Kolmogorov-Smirnov, Jarque-Bera, D'Agostino test of skewness, Anscombe-Glynn test of kurtosis, Bonett-Seier test of kurtosis, Anderson-Darling goodness of fit test.
- Multivariate tests: Pearson's product moment correlation coefficient t-test, Spearman rank correlation test, Kendall's tau correlation coefficient test
- Two sample t-test for the difference in sample means
- Causality
- Volatility

First impression

A *time series* is a sequence of observations of some quantity or quantities, e.g., equity prices, taken over time, and a *time series plot* is a plot of a time series in chronological order. Figure was produced by the following code:

```
setwd("~/R/Trinity/Session2") # setting a working directory
require (xts)
library(tidyquant) # for yahoo finance

options("getSymbols.warning4.0"=TRUE)
options("getSymbols.yahoo.warning"=TRUE)

getSymbols("^SP600", from = '2010-12-31',
           to = "2023-01-01",warnings = FALSE,
           auto.assign = TRUE) # getting SP600 prices

chart_Series(SP600$SP600.Close) #plotting the series
data.xts = SP600$SP600.Close

return = diff (log(data.xts)) # Log return calculation
return = return [-1] # removing the first empty observation,
received after return calculation
summary (return)

plot(return, main = "S&P 600 daily returns", xlab = "year",
type = "l", ylab = "log return") # plot the graph
```



- *Are the time series stationary? Stationary*, meaning that the nature of its random variation is constant over time.
- We see *volatility* clustering, because there are periods of higher, and of lower, variation within each series.
- Volatility clustering does *not* indicate a lack of stationarity but rather can be viewed as a type of dependence in the conditional variance of each series.

Estimates of Location

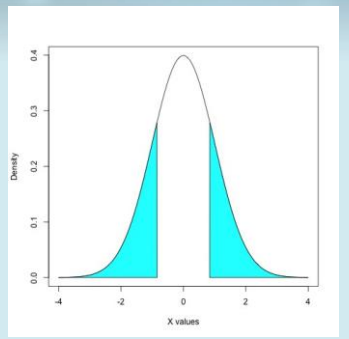
- Variables with measured or count data might have thousands of distinct values.
- A basic step in exploring your data is getting a “typical value” for each feature (variable): an estimate of where most of the data are located (i.e. their central tendency).

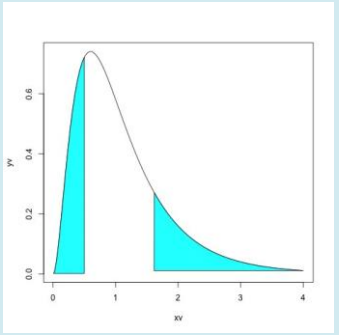
Key Terms for Estimates of Location

- ***Mean***
  - The sum of all values divided by the number of values.
  - *Synonyms:* average
- ***Trimmed Mean***
  - The average of all values after dropping a fixed number of extreme values.
  - *Synonyms :* truncated mean
- ***Median***
  - The value such that one-half of the data lies above and below.
  - *Synonyms :* 50th percentile
- ***Weighted Median***
  - The value such that one-half of the sum of the weights lies above and below the sorted data.
- ***Robust***
  - Not sensitive to extreme values.
  - *Synonyms :* resistant
- ***Outlier***
  - A data value that is very different from most of the data.
  - *Synonyms :* extreme value

Estimates of Location, mean

- ***Mean*** $= \bar{x} = \frac{\sum_{i=1}^{N} x_i}{N}$
  - `mean (return) # calculate mean of the return 0.0003390754`
- ***Trimmed Mean*** $= \bar{x} = \frac{\sum_{i=p+1}^{N-p} x_i}{N-2p}$
  - `mean (return, trim=0.1) # calculate trimmed mean (10 per cent is trimmed) or the return 0.0005515062`
- **How does trimmed mean work**: let's calculate the mean value cutting off 20% of the extreme values.
- Let's consider a sample of 20 observations: 39 92 75 61 45 87 59 51 87 12 8 93 74 16 32 39 87 12 47 50
- First we sort them: 8 12 12 16 32 39 39 45 47 50 51 59 61 74 75 87 87 87 92 93
- The number of observations to remove is **floor**(0.2 * 20) = 4 (return integer number). So we trim 4 observations from each end:
- (8 12 12 16) 32 39 39 45 47 50 51 59 61 74 75 87 (87 87 92 93)
- And we take the mean of the remaining observations, such that our 20% trimmed mean = `mean(c(32,39,39,45,47,50,51,59,61,74,75,87))` = 54.92
- If your goal is to estimate central tendency (a central or typical value for a probability distribution), trimming doesn't discard information, and it actually increases the quality of the information about central tendency





Estimates of Location, mean, cont.

- ***Weighted Mean =*** $\bar{x} = \frac{\sum_{i=1}^{N} w_i x_i}{\sum_{i=1}^{N} w_i}$
  - There are two main motivations for using a weighted mean:
  - **Some values are intrinsically more variable than others, and highly variable observations are given a lower weight**.
    - For example, if we are taking the average from multiple sources and one of the sources is less accurate, then we might downweight the data from that sensor.
  - **The data collected does not equally represent the different groups that we are interested in measure.**
    - For example, because of the way an online experiment was conducted, we may not have a set of data that accurately reflects all groups in the user base. To correct that, we can give a higher weight to the values from the groups that were underrepresented.

```
#weighted mean
x1 <- c(9, 5, 2, 7, 3, 6, 4, 5)  # Create example data: for instance grades per
                                   session test
w1 <- c(0.1, 0.2, 0.1, 0.2, 0.1, 0.1, 0.1, 0.1)# Create example weights
weighted.mean(x1, w1) # calculate weighted mean
[1] 5.3
mean (x1) # compare with the mean value
[1] 5.125
```

Estimates of Location, median, cont.

- The ***median*** **is the middle number on a sorted list of the data**.
- Compared to the mean, which uses all observations, **the median only depends on the values in the center of the sorted data.**

```
median(return) # calculate the median value of the return - 0.0008236158
```

- **The median is referred to as a *robust* estimate of location since it is not influenced by *outliers* (extreme cases) that could skew the results.**
- **An outlier is any value that is very distant from the other values in a dataset**.
- Being an outlier in itself does not make a data value invalid or erroneous
- When outliers are the result of bad data, the mean will result in a poor estimate of location while the median will be still be valid.
- In any case, outliers should be identified and are often worthy of further investigation (*Anomaly Detection*)
- The median is not the only robust estimate of location. In fact, a trimmed mean is widely used to avoid the influence of outliers.

Estimates of Variability

- Location is just one dimension in summarizing a feature.
- **A second dimension, variability, also referred to as dispersion, measures whether the data values are tightly clustered or spread out.**
- At the heart of statistics lies variability: measuring it, reducing it, distinguishing random from real variability, identifying the various sources of real variability and making decisions in the presence of it.

Estimates of Variability: Key Terms for Variability Metrics

- **Deviations** : The difference between the observed values and the estimate of location.
- Synonyms : errors, residuals.

- **Variance** : The sum of squared deviations from the mean divided by N-1 where N is the number of data values.
- Synonyms: mean-squared-error.

- **Standard Deviation** : The square root of the variance.
- Synonyms : l2-norm, Euclidean norm

- **Mean Absolute Deviation** : The mean of the absolute value of the deviations from the mean.
- Synonyms: l1-norm, Manhattan norm

- **Median Absolute Deviation from the Median** : The median of the absolute value of the deviations from the median.

- **Range** : The difference between the largest and the smallest value in a data set.

- **Order Statistics** : Metrics based on the data values sorted from smallest to biggest.
- Synonyms: Ranks

- **Percentile** : The value such that P percent of the values take on this value or less and (100-P) percent take on this value or more.
- Synonyms: quantile

- **Interquartile Range**: The difference between the 75th percentile and the 25th percentile
- Synonyms: IQR

Estimates of Variability : Standard Deviation and Related Estimates

- The most widely used estimates of variation are based on the differences, or **deviations, between the estimate of location and the observed data.**
- For a set of data {1, 4, 4}, the mean is 3 and the median is 4.
- The deviations from the mean are the differences: 1 - 3 = -2, 4 - 3 = 1 , 4 - 3 = 1.
  - *These deviations tell us how dispersed the data is around the central value.*

- One way to measure variability is to estimate a typical value for these deviations.
- **Averaging the deviations themselves would not tell us much - the negative deviations offset the positive ones.**
- **In fact, the sum of the deviations from the mean is precisely zero.**
- Instead, a simple approach is to take **the average of the absolute values of the deviations from the mean.**
- In the above example, the absolute value of the deviations is {2 1 1} and their average is (2+1+1)/3 = 1.33. This is known as **the mean absolute deviation** and is computed using the formula
- $Mean\ Absolut\ deviation = \frac{\sum_{i=1}^{N}|x_i - \bar{x}|}{N}$

Estimates of Variability : Mean absolute deviation

- $Mean\ Absolut\ deviation = \frac{\sum_{i=1}^{N}|x_i - \bar{x}|}{N}$

```
library (DescTools) # Tools for Descriptive Statistics
MeanAD(return)
[1] 0.009804257
```

Estimates of Variability : variance and standard deviation

- **The best known estimates for variability are the variance and the standard deviation which are based on squared deviations.**
- **The variance is an average of the squared deviations and the standard deviation is the square root of the variance.**
- Variance = $s^2 = \frac{\sum(x-\bar{x})^2}{N}$. In R: **`var(x, na.rm = FALSE)`**
- Standard deviation = $\sqrt{variance}$. In R: **`sd(x, na.rm = FALSE)`**
- ***The standard deviation is much easier to interpret than the variance since it is on the same scale as the original data***.
- Still, with its more complicated and less intuitive formula, it might seem peculiar that the standard deviation is preferred in statistics over the mean absolute deviation.
- Its owes its preeminence to statistical theory: mathematically, it turns out that working with squared values is much more convenient than absolute values, especially for statistical models.

```
var(return)
[1,] 0.0002022992
sd(return)
[1] 0.01422319
```

Estimates of Variability : Median Absolution Deviation

- **Neither the variance, the standard deviation nor the mean absolute deviation are robust to outliers and extreme values**.
- The variance and standard deviation are especially sensitive to outliers since they are based on the squared deviations.
- **An robust estimate of variability is the median absolute deviation from the median, sometimes denoted by MAD**:
- **Median Absolution Deviation =** $\boldsymbol{Median}\ (|\boldsymbol{x_1} - \boldsymbol{m}|, |\boldsymbol{x_2} - \boldsymbol{m}|, |\boldsymbol{x_N} - \boldsymbol{m}|)$**,** where *m* is a median of x.
- Like the median, the MAD is not influenced by extreme values.
- In R:
  - `mad(x, center = median(x), constant = 1.4826, na.rm = FALSE, low = FALSE, high = FALSE)`
- Arguments : **x** - a numeric vector; **center** - optionally, the centre: defaults to the median; **constant** - scale factor; **na.rm** - if TRUE then NA values are stripped from x before computation takes place; **low** - if TRUE, compute the 'lo-median', i.e., for even sample size, do not average the two middle values, but take the smaller one; high - if TRUE, compute the 'hi-median', i.e., take the larger of the two middle values for even sample size.

```
mad (return)
[1] 0.0105435
```

But:

- **The variance, the standard deviation, mean absolute deviation and median absolute deviation from the median are not equivalent estimates, even in the case where the data comes from a normal distribution.**
- In fact, **the standard deviation is always greater than the mean absolute deviation which itself is greater than the median absolute deviation**.
- Sometimes, the median absolute deviation is multiplied by a factor of 1.4826: this puts MAD on the same scale as the standard deviation in the case of a normal distribution.

Estimates Based on Percentiles

- A different approach to estimating dispersion is based on looking at the spread of the sorted data.
- Statistics based on sorted (ranked) data are referred to as **order statistics**.
- The most basic measure is **the range**: **the difference between the largest and smallest number**.
- **The minimum and maximum values themselves are useful to know, and helpful in identifying outliers, but the range is extremely sensitive to outliers and not very useful as a general measure of dispersion in the data.**
- To avoid the sensitivity to outliers, we can look at the range of the data after **dropping values from each end**. **Formally, these types of estimates are based on differences between percentiles.**
- In a dataset, **the P-th percentile is a value such that at least P percent of the values take on this value or less and at least (100-P) percent of the values take on this value or more**.
- For example, to find the 80th percentile, sort the data. Then, starting with the smallest value, proceed 80 percent of the way to the largest value. Note that the median is the same thing as the 50th percentile. The percentile is essentially the same as a quantile, with quantiles indexed by fractions (so the .8 quantile is the same as the 80th percentile).
- A common measurement of variability is the difference between the **25th** percentile and the **75th** percentile, called the **interquartile range** (or IQR).
- Here is a simple example: 3,1,5,3,6,7,2,9. We sort these to get 1,2,3,3,5,6,7,9. The 25th percentile is at 2.5, and the 75th percentile is at 6.5, so the interquartile range is 6.5 - 2.5 = 4. Software can have slightly differing approaches that yield different answers.

Estimates Based on Percentiles, R

Calculating the quartiles : Often statisticians report the first and the third quartile together with the median. These quartiles are, respectively, the 25 percent and 75 percent quantiles, which are the numbers for which one-fourth and three-fourths of the data is smaller. You get these numbers using the **`quantile()`** function, like this:

```
quantile (return)
    0%           25%           50%           75%          100%
-0.1421025606 -0.0063928021  0.0008236158  0.0078450263  0.0862388783

quantile (return, probs=c(0.05, 0.95)) # values in the 5th and 95th quantiles
 5%          95%
-0.02138979  0.02114142

tauseq = seq(.1,.95,.1)
quantile (return, tauseq)
 10%          20%          30%          40%          50%          60%          70%          80%          90%
-0.0148570882 -0.0086862698 -0.0047462825 -0.0017447029  0.0008236158  0.0033101828  0.0061042686  0.0095001238  0.0148609908
```

Note: 0% and 100% are respectively the minimum and maximum values

Probability calculations and combinatorics

- For instance, you want to pick five numbers at random from the set 1:40, then you can write `sample(1:40, 5)`.
- The probability of obtaining a given number as the first one of the sample should be 1/40, the next one 1/39, the next one 1/38 and so forth.
- The probability of a given sample should then be 1/(40 × 39 × 38 × 37 × 36).
- In R, use the **`prod`** function, which calculates the product of a vector of numbers `1/prod(40:36)`
- `prod(40:36) = 40*39*38*37*36`
- **However, notice that this is the probability of getting given numbers in a given order.**
- If this were a Lotto-like game, then you would rather be interested in the probability of guessing a given set of five numbers correctly.
- Thus you need also to include the cases that give the *same numbers in a different order*.
- Since obviously the probability of each such case is going to be the same, all we need to do is to figure out how many such cases there are and multiply by that.
- There are five possibilities for the first number, and for each of these there are four possibilities for the second, three possibilities for the third and so forth; that is, the number is 5 × 4 × 3 × 2 × 1. This number is also written as 5! (5 factorial). So the probability of a "winning Lotto coupon" would be

(5 × 4 × 3 × 2 × 1) /(40 × 39 × 38 × 37 × 36) = `prod(5:1)/prod(40:36)`

Probability calculations and combinatorics, cont

- There is another way of arriving at the same result.
  - All sets of five numbers must have the same probability.
  - So all we need to do is to calculate the number of ways to choose 5 numbers out of 40.
  - This is denoted $\binom{40}{5} = \frac{40!}{5!35!} = 658008$ so that the probability is 1/658008.
  - In R, the `choose` function can be used to calculate the probability of this
- `1/choose(40,5)`

The binomial probability function

- For each trial there are only two possible outcomes, success or failure. **Probability of success, *p*, of each trial is fixed. There are *n* trials.**
- Each trial is independent.
- The binomial probability function defines the probability of ***x*** successes from ***n*** trials.

$$p(x) = \frac{n!}{(n-x)!\,x!} p^x (1-p)^{n-x}$$

The binomial probability function, example

$$p(x) = \frac{n!}{(n-x)!\,x!} p^x (1-p)^{n-x}$$

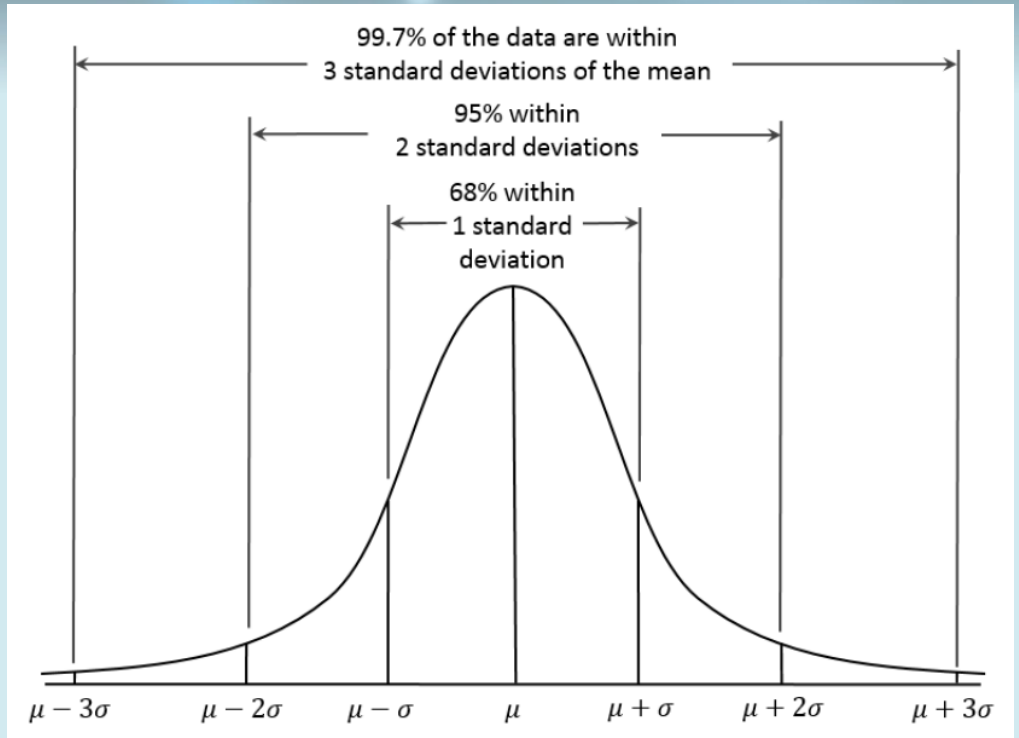
- **You have a pool of stocks having returns either above 5% or below 5%. The probability of selecting a stock with above 5% returns is 0.10. You are going to pick up 5 stocks. Assuming binomial distribution, what is the probability of picking 2 stocks with above 5% returns?**

- Let's define our problem.
- **Success** = Pick stock with above 5% returns
- **p = 0.10** of probaility of being above 5%
- **n = 5** stocks to pick up (trials)
- **x = 2** Success stocks
- $p(X = 2) = p(2) = \frac{5!}{(5-2)!2!} 0.1^2 (1-0.1)^{5-2}$
- In R : **`dbinom(x, size, prob)`** [size - number of trials (zero or more), prob - probability of success on each trial.

```
n=5
x=2
p=0.1

(factorial(n)/(factorial(n-x)*factorial(x)))*p^x*(1-p)^(n-x) # manual calculation as in formula
dbinom(2, size=5, prob=0.1) # R function for binominal distribution
[1] 0.0729
```

Continuous Distributions: normal distribution

- In order to model continuous data, we need to define **random variables that can obtain the value of any real number**.
- Because there are *infinitely many numbers infinitely close*, the probability of any particular value will be zero, so there is no such thing as a point probability as for discrete valued random variables.
- Instead we have the concept of a **density**.
- **The density for a continuous distribution is a measure of the relative probability of "getting a value close to x".**
- **Error**: The difference between a data point and a predicted or average value
- **Standardize**: Subtract the mean and divide by the standard deviation
- **Z score**: The result of standardizing an individual data point
- **Standard normal:** A normal distribution with mean = 0 and standard deviation = 1
- **QQ-Plot:** A plot to visualize how close a sample distribution is to a normal distribution



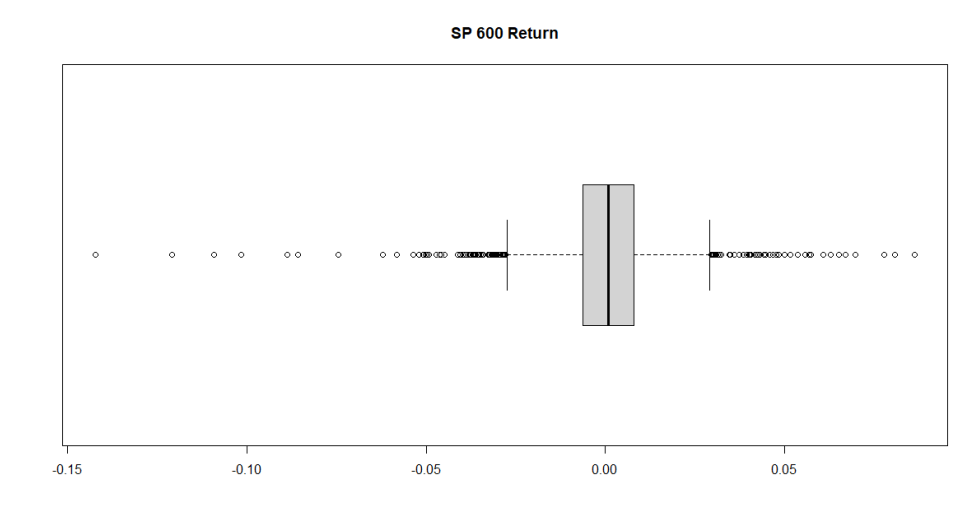
$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ where μ mean and σ standard deviation.

Simple quantiles

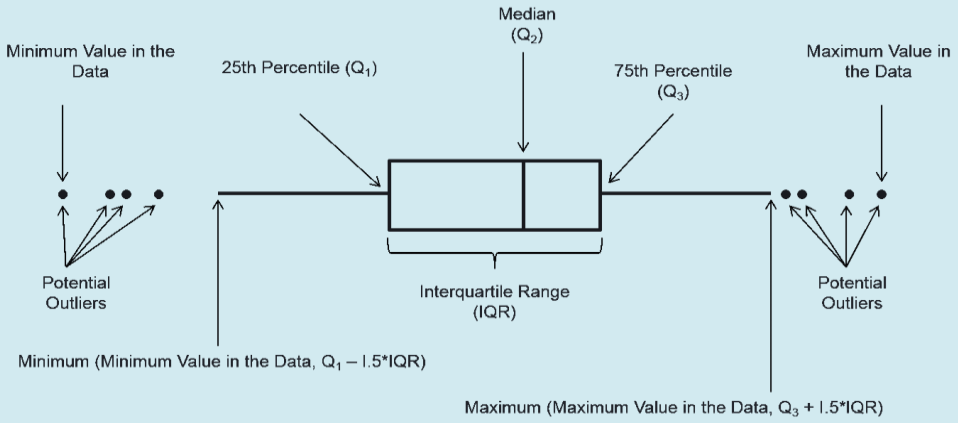
Consider an ordered population of 10 data values {**3**, 6, **7**, 8, 8, **9**, 10, 13, **15**, 16, **20**}. What are the 4 quantiles?

- **Q0 =** (not universally accepted. This is the minimum value of the set, so the zeroth quartile in this example would be **3**.
- **Q1=**The rank of the first quartile is: 11(number of data point) × (**1/4**) = 2.75, which rounds up to 3, which rounds up to 3, meaning that 3 is the rank in the population (from least to greatest values) at which approximately 1/4 of the values are less than the value of the first quartile. The third value in the population is **7**.
- **Q2 =** (**the median**) is 11×(**2/4**) = 5.5, which rounds up to 6 (the rank in the population from least to greatest values) at which approximately 2/4 of the values are less than the value of the second quartile (or median). The sixth value in the population is **9**.
- **Q3 =** 11×(3/4) = 8.25 ~ 9. The ninth value in the population is **15.**
- **Q4 =** (not universally accepted) - the maximum value = **20**.

boxplot




```
boxplot(return2, horizontal=TRUE, main="SP 600 Return")
```



- **The line that divides the box into 2 parts represent the median of the data**. If the median is 10, it means that there are the same number of data points below and above 10.
- **The ends of the box shows the upper (Q3) and lower (Q1) quartiles**. For example, if the third quartile (Q3) is 15, it means that 75% of the observation are lower than 15.
- The difference between Quartiles 1 and 3 is called **the interquartile range (IQR)**
- The extreme line shows **Q1-1.5xIQR** to **Q3+1.5xIQR** (**the lowest and highest value excluding outliers**).
  - low end is min dataset value above Q1-1.5xIQR
  - high end is max dataset value below Q3+1.5xIQR
- Dots (or other markers) beyond the extreme line shows potential outliers.

Histograms and Kernel Density Estimation

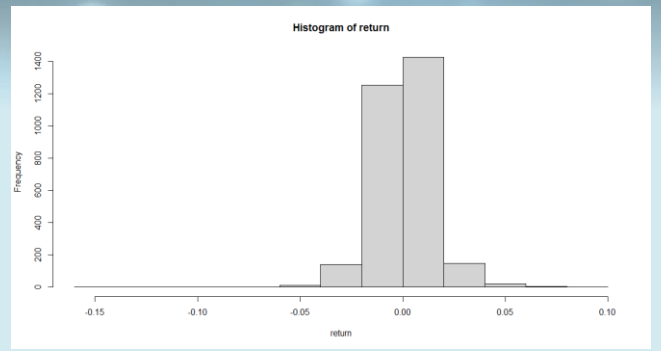
- The histogram is a simple and well-known estimator of probability density functions.
- There are some outliers in this series. We have a long left tail.

- A much better estimator is the *kernel density estimator* (KDE).
- The estimator takes its name from the so-called kernel function, denoted here by *K*, which is a probability density function that is symmetric about 0.
- The standard (means having expectation 0 and variance 1) normal density function is a common choice for *K* and will be used here. The kernel density estimator based on *Y*1, . . . , *Yn* is

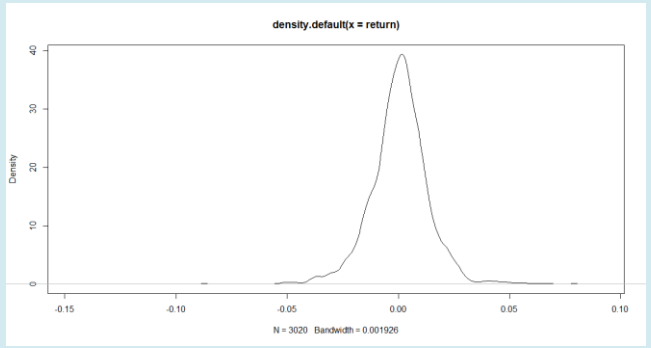
$$\hat{f}(y) = \frac{1}{nb}\sum_{i=1}^{n} K\left(\frac{y - Y_i}{b}\right)$$

where *b*, which is called the bandwidth, determines the resolution of the estimator.

- A small value of *b* allows the density estimator to detect fine features in the true density, but it also permits a high degree of random variation.
- Conversely, a large value of *b* dampens random variation but obscures fine detail in the true density. **Stated differently, a small value of *b* causes the kernel density estimator to have high variance and low bias, and a large value of *b* results in low variance and high bias.**
- Often a kernel density estimate is used to suggest a parametric statistical model. The density estimates in Fig. are not bell-shaped, not suggesting that a normal distribution might be a suitable model.



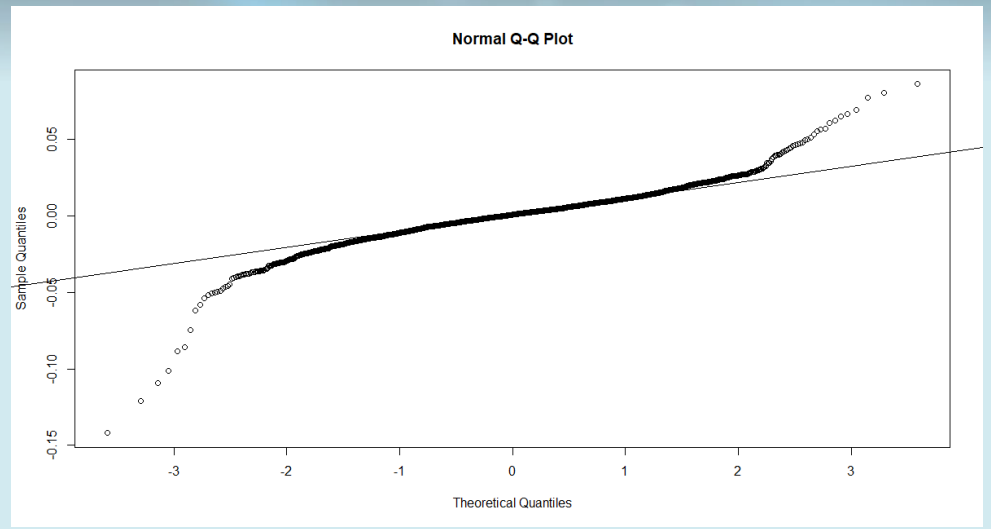

```
hist(return)
```




```
return.density = density(return) # estimate kernel density
plot (return.density) # plot the estimated density
```

Standard Normal and QQ-Plots

- A standard normal distribution is one in which the units on the x-axis are expressed in terms of standard deviations away from the mean.
- **To compare data to a standard normal distribution you subtract the mean then divide by the standard deviation; this is also called normalization or standardization.**
- Note that "standardization" in this sense is unrelated to database record standardization (conversion to a common format).
- The transformed value is termed a **z-score**, and the normal distribution is sometimes called the **z-distribution**.
- A QQ-Plot is used to visually determine how close a sample is to the normal distribution.
- The QQ-Plot orders the z-scores from low to high, and plots each value's z-score on the y-axis; the x-axis is the corresponding quantile of a normal distribution for that value's rank.
- Since the data are normalized, the units correspond to the number of standard deviations away of the data from the mean. **If the points roughly fall on the diagonal line, then the sample distribution can be considered close to normal.**



The plots were produced by the R function qqnorm(). The reference lines pass through the first and third quartiles and were produced by R's qqline() *function:*

```
qqnorm(return)
qqline(return)
```

Converting data to z-scores (i.e. standardizing or normalizing the data) does **not** make the data normally distributed. It just puts the data on the same scale as the standard normal distribution, often for comparison purposes.

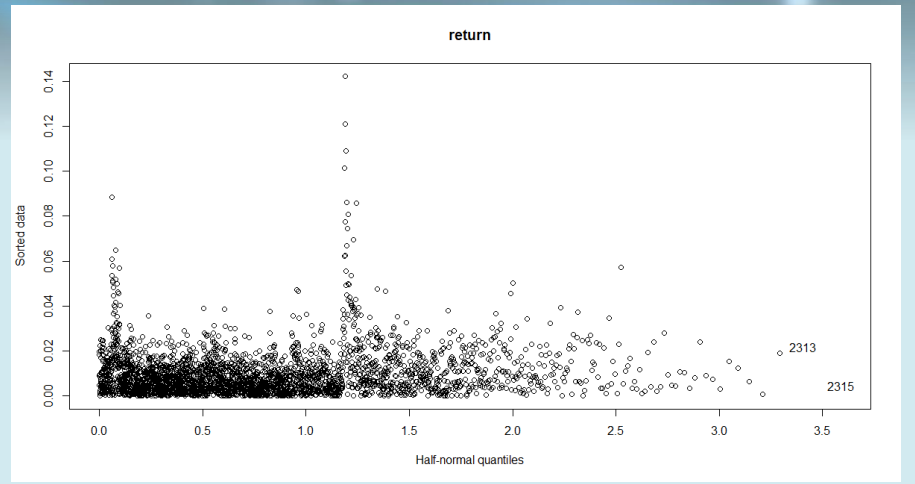
Half-Normal Plots

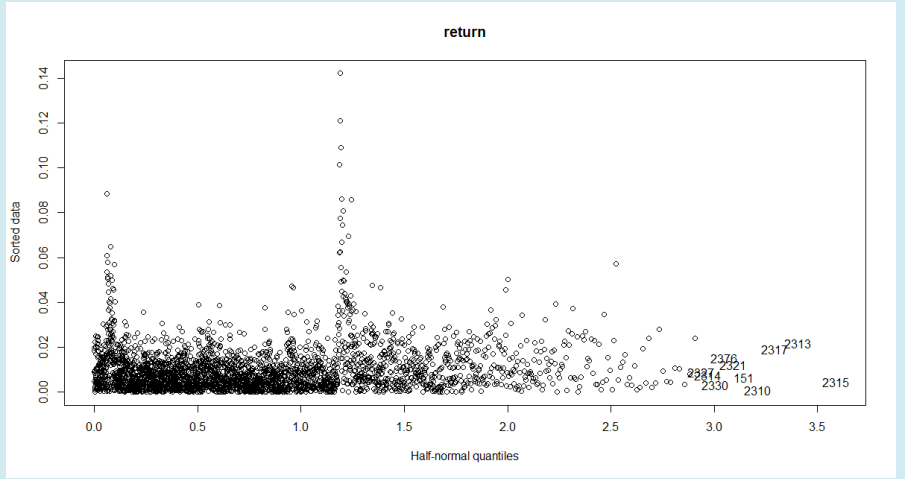
- **The half-normal plot is a variation of the normal plot used for detecting outlying data rather than checking for a normal distribution.**
- **For example, suppose one has data Y1, . . . , Yn and wants to see whether any of the absolute deviations |Y1 −Y |, . . . , |Yn −Y | from the mean are unusual.**
- In a half-normal plot, this deviation is plotted against the quantiles of |Z|, where Z is N(0, 1) distributed.
- More precisely, a half-normal plot is a scatterplot of the order statistics of the absolute values of the data against $\Phi^{-1}\{(n + i)/(2n + 1)\}$, i = 1, . . . , n, where n is the sample size.
- The function **`halfnorm`**( ) in R's **`faraway`** package creates a half-normal plot and labels the nlab most outlying observations, where nlab is an argument of this function with a default value of 2.

Half-Normal Plots, cont.

- Only the two most outlying cases are labeled because the default value of nlab was used.

```
# Half-Normal Plots
library (faraway)
halfnorm(abs(return), main = "return", ylab = "Sorted
data") # label only 2 the most outlying cases

halfnorm(abs(return),nlab = 10, main = "return", ylab =
"Sorted data") # label 10 the most outlying cases
```





Long-Tailed Distributions

- Despite the importance of the normal distribution historically in statistics, and in contrast to what the name would suggest, data are generally **not normally distributed**.

Key Terms for Long-Tail Distribution

- **Tail:** The long narrow portion of a frequency distribution, where relatively extreme values occur at low frequency
- **Skew:** Where one tail of a distribution is longer than the other
- While the normal distribution is often appropriate and useful with respect to the distribution of errors and sample statistics, it typically does not characterize the distribution of raw data.
- Sometimes, the distribution is highly skewed (asymmetric), such as with income data, or the distribution can be discrete, as with binomial data.
- Both symmetric and asymmetric distributions may have long tail(s). The tails of a distribution correspond to the extreme values (small and large).
- Long-tails, and guarding against them, are widely recognized in practical work. Nassim Taleb has proposed the *black swan* theory which predicts that anamolous events, such as a stock market crash, are likely to occur in much greater likelihood than would be predicted by the normal distribution.

Student's t distribution

- **The t-distribution is a normally-shaped distribution, but a bit thicker and longer on the tails.**
- It is used extensively in depicting distributions of sample statistics.
- Distributions of sample means are typically shaped like a t-distribution, and there is a family of t-distributions that differ depending on how large the sample is.
- *The larger the sample, the more normally-shaped the t-distribution becomes*.
- The t-distribution is often called Student's t because it was published in 1908 in Biometrika by W. S. Gossett under the name "Student." Gossett's employer, the Guinness brewery, did not want competitors to know that they were using statistical methods, so insisted that Gossett not use his name on the article.

TESTS for NORMALITY

Tests of Normality

- When viewing a normal probability plot, it is often difficult to judge whether any deviation from linearity is systematic or instead merely due to sampling variation, so a statistical test of normality is useful.
- **The null hypothesis is that the sample comes from a normal distribution and the alternative is that the sample is from a nonnormal distribution.**
- The Shapiro–Wilk test of these hypotheses uses something similar to a normal plot.
- The correlation between this product and the sample order statistics is used as the test statistic. Correlation and covariance matrices will be discussed in greater detail later. For now, only a few facts will be mentioned. The *covariance* between two random variables *X* and *Y*

$$Cov(X,Y) = \sigma_{XY} = E\{X - E(X)\}\{Y - E(Y)\}$$

- and the *Pearson correlation coefficient* between *X* and *Y* is

$$Corr(X,Y) = \rho_{XY} = \sigma_{XY}/\sigma_X\sigma_Y$$

- There are many tests. A recent comparison of eight tests of normality (Yap and Sim 2011) found that the Shapiro-Wilk test was as powerful as its competitors for both short- and long-tailed symmetric alternatives and was the most powerful test for asymmetric alternatives.

Tests of Normality, the Shapiro–Wilk test .

- A correlation equal to 1 indicates a perfect positive linear relationship, where

$$Y = \beta 0 + \beta 1 X \text{ with } \beta 1 > 0.$$

- Under normality, the correlation between sample order statistics and the expected normal order statistics should be close to 1 and the null hypothesis of normality is rejected for small values of the correlation coefficient.
- In R, the Shapiro–Wilk test can be implemented using the shapiro.test() function

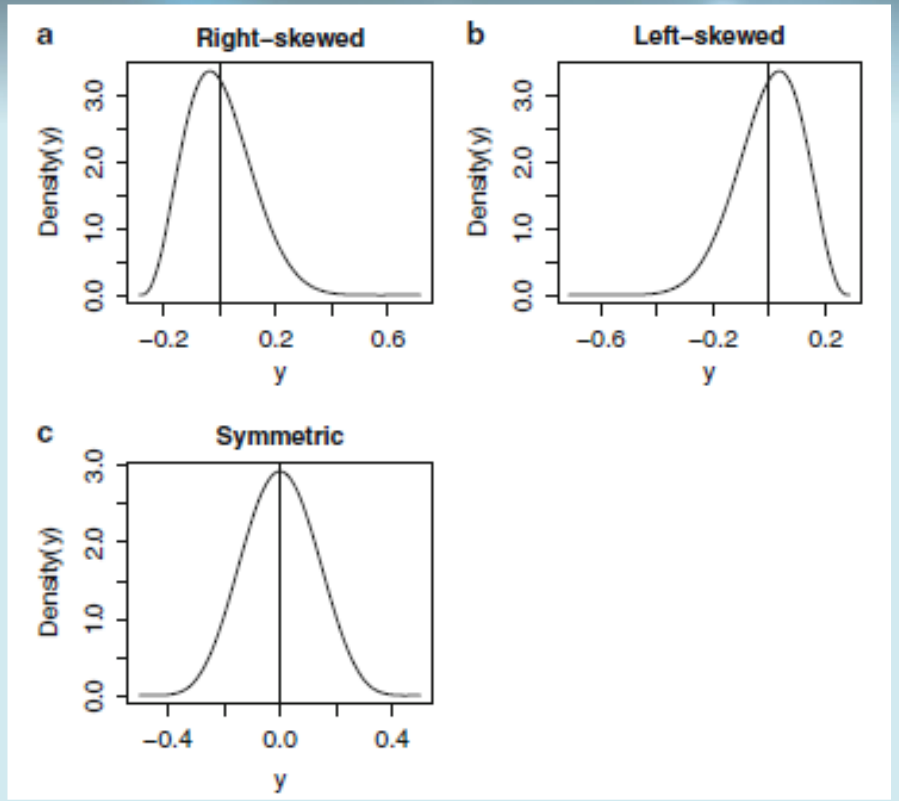
Tests of Normality, the Shapiro–Wilk test cont.

```
# Shapiro-Wilk test
shapiro.test(return2$SP600.Close)
# Shapiro-Wilk test, use a vector as the input, not a dataframe or other object

> shapiro.test(return2)

        Shapiro-Wilk normality test

data:  return2$SP600.Close
W = 0.9143, p-value < 2.2e-16
```

From the output, the p-value < 0.05 implying that the distribution of the data is significantly different from normal distribution. In other words, we can assume the non-normality.

Skewness, Kurtosis, and Moments

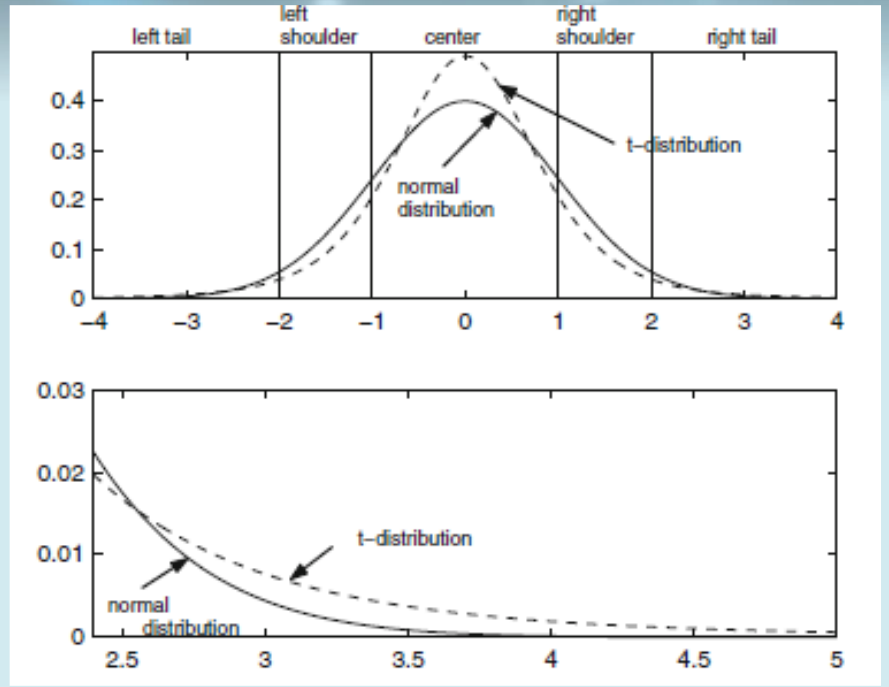
- Skewness and kurtosis help characterize the shape of a probability distribution.
- **Skewness measures the degree of asymmetry, with symmetry implying zero skewness, positive skewness indicating a relatively long right tail compared to the left tail, and negative skewness indicating the opposite.** Figure on the right shows three densities, all with an expectation equal to 0. The densities are right-skewed, left-skewed, and symmetric about 0, respectively, in panels (a)–(c).
- **Kurtosis indicates the extent to which probability is concentrated in the center and especially the tails of the distribution rather than in the "shoulders," which are the regions between the center and the tails.**



*Skewed and symmetric densities. In each case, the mean is zero and is indicated by a vertical line. The distributions in panels (**a**)–(**c**) are beta(4,10), beta(10,4), and beta(7,7), respectively. The* R *function* dbeta() *was used to calculate these densities.*

Center and *shoulder*

- Reasonable definitions of *center* and *shoulder* would be that the center is the region from $\mu - \sigma$ to $\mu + \sigma$, the left shoulder is from $\mu - 2\sigma$ to $\mu - \sigma$, and the right shoulder is from $\mu + \sigma$ to $\mu+2\sigma$. See the upper plot in Fig.
- Because skewness and kurtosis measure shape, they do not depend on the values of location and scale parameters.



Comparison of a normal density and a t-density with 5 degrees of freedom. Both densities have mean 0 and standard deviation 1. The upper plot also indicates the locations of the center, shoulders, and tail regions. The lower plot zooms in on the right tail region.

The skewness of a random variable Y

$$Sk = E\left\{\frac{Y - E(Y)}{\sigma}\right\}^3 = \frac{E\{Y - E(Y)\}^3}{\sigma^3}$$

- **Positive skewness is also called right skewness and negative skewness is called left skewness**.
- A distribution is *symmetric* about a point θ if $P(Y > \theta + y) = P(Y < \theta - y)$ for all $y > 0$. In this case, θ is a location parameter and quals $E(Y)$, provided that $E(Y)$ exists. **The skewness of any symmetric distribution is 0**.
- **According to the theories of skewness preferences** (Mitton and Vorkink, 2007; Barberis and Huang, 2008) **investors prefer positive skewness due to overpricing of positively skewed equities.** Lien and Wang (2015) show that **an increase in skewness reduces the demand for hedging**. Also, markets where going short is not practiced, market returns display significantly less negative skewness.
- In R: **`skewness(data)`**, library e1071.

```
skewness (return2$SP600.Close)
[1] -0.7572442
```

- If negative values - skewed towards the right.

- See more in Jalan, A., Matkovskyy, R. and Urquhart, A., What If Bitcoin Futures Had Never Been Introduced? (November 21, 2019). Available at SSRN: https://ssrn.com/abstract=3491272 or http://dx.doi.org/10.2139/ssrn.3491272

Kurtosis

- The kurtosis of a random variable *Y* is

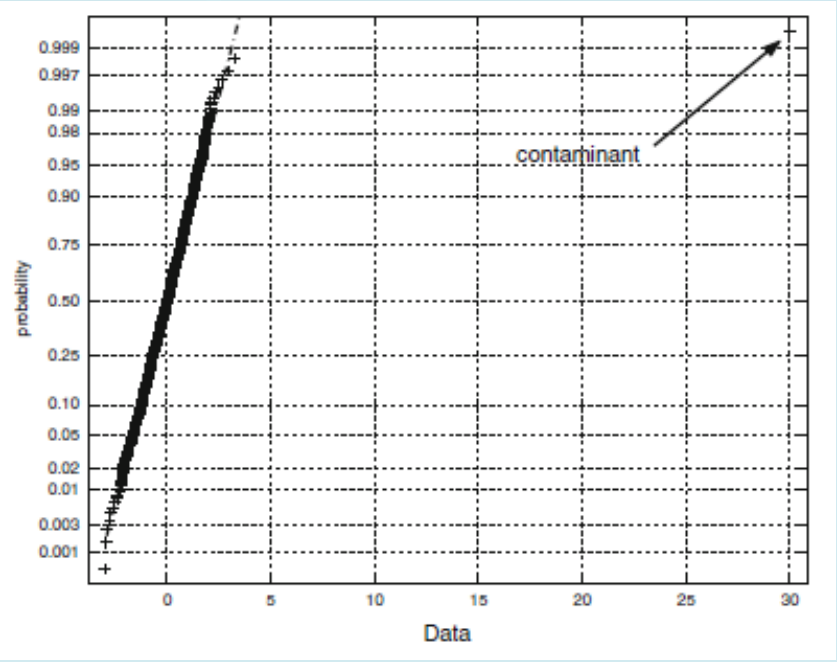
$$Kur = E\left\{\frac{Y - E(Y)}{\sigma}\right\}^4 = \frac{E\{Y - E(Y)\}^4}{\sigma^4}$$

- **The kurtosis of a normal random variable is 3.** The smallest possible value of the kurtosis is 1 and is achieved by any random variable taking exactly two distinct values, each with probability 1/2.
- It is difficult to interpret the kurtosis of an asymmetric distribution because, for such distributions, kurtosis may measure both asymmetry and tail weight.
- In R: `kurtosis(data)` (library e1071)
- Negative excess kurtosis (less than 3) would indicate a *thin-tailed* data distribution, and is said to be **platykurtic**. Positive excess kurtosis (more than 3) would indicate a *fat-tailed* distribution, and is said to be **leptokurtic**.
- **Every normal distribution has a *skewness* coefficient of 0 and a *kurtosis* of 3.**

```
kurtosis(return2$SP600.Close)
[1] 10.38963
```

A word of caution!

- **Skewness and kurtosis are highly sensitive to outliers.**
- Sometimes outliers are due to *contaminants*, that is, bad data not from the population being sampled.
- An example would be a data recording error. A sample from a normal distribution with even a single contaminant that is sufficiently outlying will appear highly nonnormal according to the sample skewness and kurtosis.
- In such a case, a normal plot *will* look linear, except that the single contaminant will stick out. See Fig. which is a normal plot of a sample of 999 $N(0, 1)$ data points plus a contaminant equal to 30.
- This figure shows clearly that the sample is nearly normal but with an outlier. *The sample skewness and kurtosis, however, are 10.85 and 243.04, which might give the false impression that the sample is far from normal.*
- Also, even if there were no contaminants, a distribution could be extremely close to a normal distribution except in the extreme tails and yet have a skewness or excess kurtosis that is very different from 0.



Normal plot of a sample of 999 $N(0, 1)$ *data plus a contaminant.*

D’Agostino test of skewness

- **Question:** Is the sample skewed?
- **When to use the test?** To test for a lack of symmetry (skewness) in a sample. Under the hypothesis of normality, data should be symmetrical (i.e. skewness should be equal to zero). The test is useful for detecting nonnormality caused by asymmetry.
- If a distribution has normal kurtosis but is skewed, the test for skewness may be more powerful than the Shapiro-Wilk test, especially if the skewness is mild.
- Practical Applications:
- Belgrade Stock returns: Djorić and Nikolić-Djorić (2011) investiate the distributions of daily log returns of the Belgrade Stock Exchange index BELEX15. The BELEX15 index is composed of 15 of the most liquid Serbian shares. The sample period covers 1067 trading days from 4 October 2005 to 25 December 2009. Visual inspection of returns indicates the variances may change over time around some level, with large (small) changes tending to be followed by large (small) changes of either sign (volatility tends to cluster). In order to investiate the asymmetry of the data the researchers perform the D’Agostino test of skewness (p-value = 0.1238).

- The null hypothesis of symmetry was not rejected using this test.

D’Agostino test of skewness, R

- The function `agostino.test{moments}` can be used to perform this test. It takes the form

```
agostino.test (sample, alternative = "two.sided" or "less"or "greater").
```

Example:

- Enter the following data:

```
> sample <-c(-1.441,-0.642,0.243,0.154,-0.325,-0.316,0.337,-0.028,1.359,-1.67,-0.42,1.02,-1.15,0.69,-1.18,2.22,1,-1.83,0.01,-0.77,-0.75,-1.55,-1.44,0.58,0.16)
```

- The test can be conducted as follows:

```
> agostino.test(sample, alternative = "two.sided")
D'Agostino skewness test
data: sample
skew = 0.3527, z = 0.5595, p-value = 0.5758
alternative hypothesis: data have a skewness
```

- **The two sided p-value at 0.5758 is greater than 0.05, therefore do not reject the null hypothesis, the data are not skewed.**

References

Djorić, D., & Nikolić-Djorić, E. (2011). Return distributionn and value at risk estimation for BELEX15. The Yugoslav Journal of Operations Research, 21(1).

D'Agostino, R.B. (1970). Transformation to Normality of the Null Distribution of G1. Biometrika, 57, 3, 679-681.

Anscombe-Glynn test of kurtosis

- **Question the test addresses:** *Does the sample exhibit more (or less) kurtosis relative to the normal distribution?*
- **When to use the test?** *The test is useful for detecting nonnormality caused by tail heaviness.*
- **If a distribution is symmetric but heavy-tailed (positive kurtosis), the test for kurtosis may be more powerful than the Shapiro-Wilk test, especially if the heavy-tailedness is not extreme**.
- Some applications: Fagiolo et al (2008) investigate the distribution of US output growth-rate time series. They use quarterly real GDP from 1947Q1 to 2005Q3 (234 observations); monthly industrial production (IP) from January 1921 to October 2005 (1017 observations); and they also look at industrial production (IPS) in the subperiod 1947 to 2005 (702 observations). The Anscombe-Glynn test of kurtosis is applied to each series (GDP p-value = 0.0036, IP p-value <0.001, IPS p-value <0.001). The researchers conclude the growth-rate distributions are markedly nonnormal due to excess kurtosis.

Anscombe-Glynn test of kurtosis, R

- The function anscombe.test{moments} can be used to perform this test. It takes the form anscombe.test (sample, alternative = "two.sided" or "less“ or "greater").

Example:

Enter the following data:

```
> sample <-c(-1.441,-0.642,0.243,0.154,-0.325,-0.316,0.337,-0.028,1.359,-1.67,-
0.42,1.02,-1.15,0.69,-1.18,2.22,1,-1.83,0.01,-0.77,-0.75,-1.55,-1.44,0.58,0.16)
```

The test can be conducted as follows:

```
> anscombe.test (sample, alternative = "two.sided" )
Anscombe-Glynn kurtosis test
data: sample
kurt = 2.5504, z = -0.1187, p-value = 0.9055
alternative hypothesis: kurtosis is not equal to 3
```

- **The two sided p-value at 0.9055 is greater than 0.05, therefore do not reject the null hypothesis, the data do not exhibit excess kurtosis relative to the normal distribution**.

Anscombe, F.J., Glynn, W.J. (1983) Distribution of kurtosis statistic for normal statistics. Biometrika, 70, 1, 227-234

Fagiolo, G., Napoletano, M., & Roventini, A. (2008). Are output growth-rate distributions fat-tailed? Some evidence from OECD countries. Journal of Applied Econometrics, 23(5), 639-669.

Bonett-seier test of kurtosis

- **Question the test addresses:** *Does the sample exhibit more (or less) kurtosis calculated by Geary's measure, relative to the normal distribution?*
- **When to use the test?** *To test for heavy tails (kurtosis) in a sample. This test uses Geary's measure of kurtosis for normally distributed data. Under the null hypothesis of normality the data should have Geary's kurtosis equal to 0.7979.*

Bonett-seier test of kurtosis, R

- The function bonett.test{moments} can be used to perform this test. It takes the form bonett.test (sample, alternative = "two.sided" or "less"or "greater").

Example:

- Enter the following data:

```
> sample <-c(-1.441,-0.642,0.243,0.154,-0.325,-0.316,0.337,-
0.028,1.359,-1.67,-0.42,1.02,-1.15,0.69,-1.18,2.22,1,-1.83,0.01,-
0.77,-0.75,-1.55,-1.44,0.58,0.16)
```

- The test can be conducted as follows:

```
> bonett.test (sample, alternative = "two.sided" )
Bonett-Seier test for Geary kurtosis
data: sample
tau = 0.8400, z = -0.6612, p-value = 0.5085
alternative hypothesis: kurtosis is not equal to sqrt(2/pi)
```

- The two sided p-value at 0.5085 is greater than 0.05, therefore do not reject the null hypothesis, the data do not exhibit excess Geary's measure of kurtosis relative to the normal distribution.

Shapiro-Wilk test

- **Question the test addresses:** *Is the sample from a normal distribution?*
- **When to use the test?** *To investigate whether the observed sample is from a normal distribution*.
- It is used for assessing whether the sample data are randomly obtained from a normally distributed population. It does not require that the mean or variance of the hypothesized normal distribution be specified in advance.

Shapiro-Wilk test, R

- The function shapiro.test{stats} can be used to perform this test. It takes the form shapiro.test (sample) . As an alternative the function shapiroTest{fBasics} can also be used, it takes the form shapiroTest(sample).
- Example: testing against a normal distribution
- Enter the following data:

```
> sample <-c(-1.441,-0.642,0.243,0.154,-0.325,-0.316,0.337,-0.028,1.359,-1.67,-0.42,1.02,-1.15,0.69,-
1.18,2.22,1,-1.83,0.01,-0.77,-0.75,-1.55,-1.44,0.58,0.16)
```

- The test can be conducted as follows:

```
> shapiro.test (sample)
Shapiro-Wilk normality test
data: sample
W = 0.9712, p-value = 0.6767
```

- **Since the p-value is greater than 0.05, do not reject the null hypothesis that the data are from the normal distribution**. Alternatively, we can try:

```
> shapiroTest(sample)
Title:
Shapiro - Wilk Normality Test
Test Results:
STATISTIC:
W: 0.9712
P VALUE:
0.6767
```

- The results are identical, and we cannot reject the null hypothesis.

Kolmogorov-Smirnov test of normality

- **Question the test addresses:** *Is the sample from a normal distribution?*
- **When to use the test?** *To investigate whether the observed sample is from a normal distribution.*
- *It is used for assessing whether the sample data are randomly obtain.*
- Application: Support vector regression: Premanode et al (2013) develop an approach for prediction of foreign exchange time series using support vector regression.
- Daily trading data for the EUR-USD (euro - US dollar) exchange rate was collected over the period January 2, 2001 to June 1, 2012. The Kolmogorov-Smirnov test of normality returned a p-value < 0.01 and the null hypothesis was rejected. The researchers propose an Empirical Mode Decomposition (EMD) de-noising model to model exchange rates. The approach uses to decompose a foreign exchange signal into a new oscillatory signal known as an intrinsic mode function (IMF). For each decomposition a number of IMF's are generated. The researcher report for IMF number 7, a Kolmogorov-Smirnov test of normality p-value of 0.0593, and the null hypothesis of normality cannot be rejected. ed from a normally distributed population.

Kolmogorov-Smirnov test of normality, R

- The function ksnormTest{fBasics} can be used to perform this test. It takes the form ksnormTest(sample). As an alternative the function ks.test{stats} can also be used. It for testing against normality, it takes the form ks.test(sample,"pnorm")
- Example: testing against a normal distribution. Enter the following data:

```
> sample <-c(-1.441,-0.642,0.243,0.154,-0.325,-0.316,0.337,-0.028,1.359,-
1.67,-0.42,1.02,-1.15,0.69,-1.18,2.22,1,-1.83,0.01,-0.77,-0.75,-1.55,-
1.44,0.58,0.16)
```

- The test can be conducted as follows:

```
> ks.test(sample,"pnorm")
One-sample Kolmogorov-Smirnov test
data: sample
D = 0.1549, p-value = 0.5351
alternative hypothesis: two-sided
```

- **The two sided p-value at 0,5351 is greater than 0.05, therefore do not reject the null hypothesis that the data are from the normal distribution**.

Jarque-Bera test

- **Question the test addresses:** *Is the sample from a normal distribution?*
- **When to use the test?** *To test of the null hypothesis that the sample comes from a normal distribution with unknown mean and variance, against the alternative that it does not come from a normal distribution.*
- **Practical Applications:** Robots, humans and the disposition effect: The disposition effect, the fact that investors seem to hold on to their losing stocks to a greater extent than they hold on to their winning stocks, is explored by Da Costa et al (2013). Three groups are exposed to a simulated stock market –experienced investors, inexperienced and robots which make random trade decisions. The Jarque Bera test was used to assess the normality of the disposition effect for each of the three groups. The researchers rejected the null hypothesis of normality for the robots (p-value < 0.05), although it was not rejected for experienced investors and inexperienced investors (p-value > 0.05 for both groups).
- Indian foreign investment inflows: Bhattacharya (2013) studies the relationship between foreign investment inflows and the primary, secondary and tertiary sector of the Indian economy over the period 1996 to 2009. The researcher builds an econometric model (Vector Autoregression model) and uses the Jarque Bera test to assess the normality of the model residuals. The null hypothesis cannot be rejected (p-value=0.51) so the author concludes the model residual series is normally distributed.
- Long memory properties in developed stock markets: The existence of long memory properties in developed stock markets is analyzed by Bhattacharya and Bhattacharya (2013). The daily closing values of the individual indices over the period January 2005 to July 2011 were collected. Daily logarithmic index returns were calculated for ten stock market indices in the Netherlands, Australia, Germany, USA, France, UK, Hong Kong, Japan, New Zealand and Singapore. As part of the analysis the researchers use the Jarque-Bera test. The null hypothesis is rejected to for all ten stock market indices (p-value <0.05). The researchers conclude logarithmic return series of the stock market indices cannot be regarded as normally distributed.

Jarque-Bera test, R

- The function jarqueberaTest{fBasics} or jarque.bera.test{tseries} can be used to perform this test. It takes the form jarqueberaTest(sample) or jarque.bera.test (sample).
- Example: testing against a normal distribution. Enter the following data:

```
> sample <-c(-1.441,-0.642,0.243,0.154,-0.325,-0.316,0.337,-0.028,1.359,-1.67,-0.42,1.02,-
1.15,0.69,-1.18,2.22,1,-1.83,0.01,-0.77,-0.75,-1.55,-1.44,0.58,0.16)
```

- The test can be conducted as follows:

```
> jarqueberaTest(sample)
Title:
Jarque - Bera Normalality Test
Test Results:
STATISTIC:
X-squared: 0.7289
P VALUE:
Asymptotic p Value: 0.6946
```

- Or we can use jarque.bera.test:

```
> jarque.bera.test (sample)
Jarque Bera Test
data: sample
X-squared = 0.7289, df = 2, p-value = 0.6946
```

- **In both cases the two sided p-value at 0.6946 is greater than 0.05, therefore do not reject the null hypothesis that the data are from the normal distribution.**

Anderson-Darling goodness of fit test

- **Question the test addresses:** Is there a significant difference between the observed distribution in a sample and a specified population distribution?
- **When to use the test?** To investigate the null hypothesis that a sample is from a specific distribution. The test compares the fit of an observed cumulative distributioon functioon to a specific cumulative distribution function. It is a modification of the Kolmogorov-Smirnov test giving more weight to the tails of the distribution. Since the test makes use of a specific distribution in calculating critical values it is a more sensitive test than the Kolmogorov-Smirnov test.
- **Some application:** Reducing printer paper waste: Hasan et al (2013) study the effect of teambased feedback on individual printer paper use in an office environment. An email on printer use was sent on a weekly basis to individual participants. The researchers construct a sample based on the difference in printer paper usage before and after the email intervention. In order to check normality of the “difference” sample, the Anderson-Darling test was used (p-value =0.343). The null hypothesis of normality could not be rejected.

Anderson-Darling goodness of fit test, R

- Example: testing against a normal distribution. Enter the following data:

```
> sample <-c(-1.441,-0.642,0.243,0.154,-0.325,-
0.316,0.337,-0.028,1.359,-1.67,-0.42,1.02,-1.15,0.69,-
1.18,2.22,1,-1.83,0.01,-0.77,-0.75,-1.55,-1.44,0.58,0.16)
```

- Let's investigate whether this data are from the lognormal distribution. To do so enter:

```
> ad.test(sample,plnorm)
Anderson-Darling GoF Test
data: sample and plnorm
AD = Inf, p-value = 2.4e-05
alternative hypothesis: NA
```

- **Since the p-value is less than 0.05, reject the null hypothesis that the data are from the lognormal distribution.**

ad.test options, R

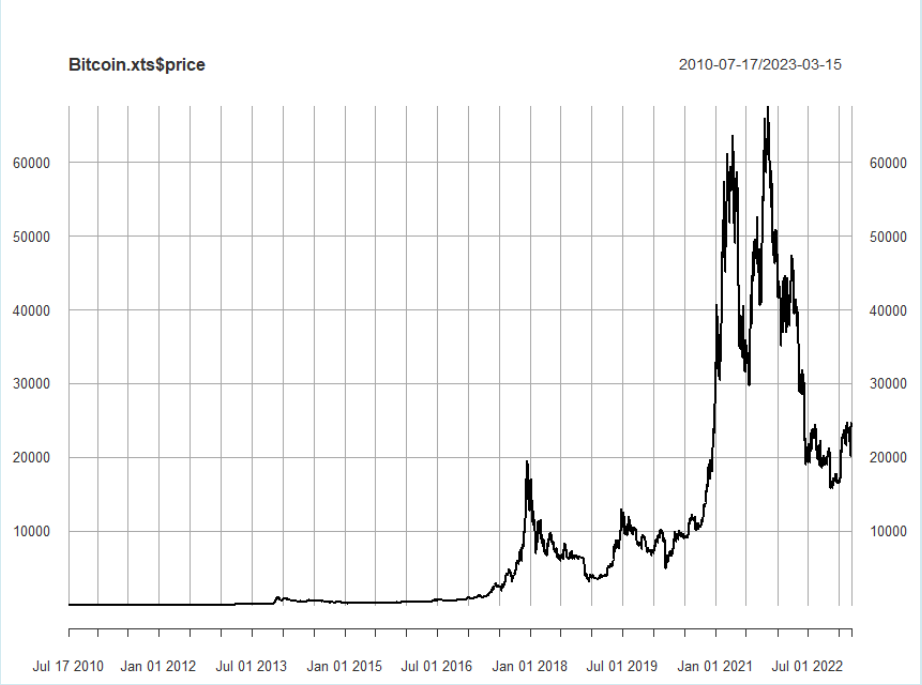
Beta R-code = pbeta	F R-code = pf
Lognormal R-code = plnorm	Uniform R-code = punif
Binomial pbinom R-code =	Gamma R-code = pgamma
Negative Binomial R-code = pnbinom	Tukey R-code = ptukey
Cauchy R-code = pcauchy	Geometric R-code = pgeom
Normal R-code = pnorm	Weibull R-code = pweib
Chisquare R-code = pchisq	Hypergeometric R-code = phyper
Poisson R-code = ppois	Wilcoxon R-code = pwilcox
Exponential R-code = pexp	Logistic R-code = plogis
Student t R-code = pt	

Multivariate tests

General impression

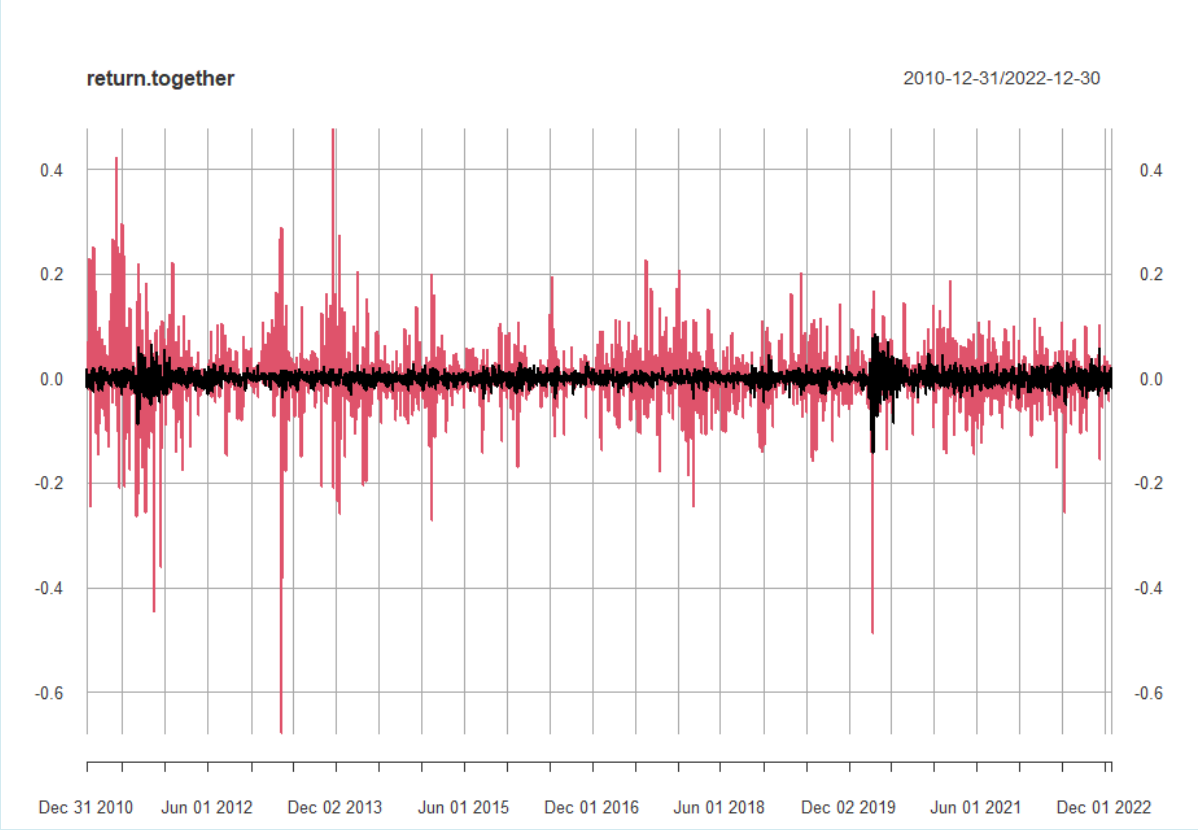
- Daily prices of bitcoin

```
# import a time series of different frequency
Bitcoin = read.csv("BitcoinData.csv")
Bitcoin = Bitcoin[, -1]

Bitcoin$date = as.Date(Bitcoin$date,format="%Y-%m-%d", tz = "")

Bitcoin.xts = xts(Bitcoin[,-1], order.by=Bitcoin[,1])
```



Data preparation

- We clean the data, and calculate returns (5 day weeks), SP600 and BTC



Pearson's product moment correlation coefficient t-test

- **Question the test addresses:** *Is the sample Pearson product moment correlation coefficient between two variables significantly different from zero?*
- **When to use the test?** *To assess the null hypothesis of zero correlation between two variables.*
- Both variables are measured on either an interval or ratio scale.
- However, they do not need to be measured on the same scale, e.g. one variable can be ratio and one can be interval.
- Both variables are assumed to be a paired random sample, approximately normally distributed, their joint distribution is bivariate normal, and the relationship is linear.

Pearson's product moment correlation coefficient t-test, R

- Both cor.test{stats} and correlationTest{fbasics} can be used to perform this test.

cor.test (weekdays.cryptos.xts, data.xts)

```
> cor.test (X, Y,method="pearson",alternative="two.sided",conf.level = 0.95)
        Pearson's product-moment correlation

data:  X and Y
t = 7.7016, df = 3018, p-value = 1.809e-14
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.1036790 0.1736413
sample estimates:
      cor
0.1388334
```

The correlation between x and y is reported as 0.14. Since the p-value is 1.809e-14 and lower than the critical value of 0.05 we reject the null hypothesis of zero correlation. The function also reports the 95% confidence interval as 0.1 to 0.173. It doesn't crosses zero.

As an alternative we could type correlationTest(X, Y)

Spearman rank correlation test

- **Question the test addresses:** *Is the Spearman rank correlation coefficient between two variables significantly different from zero?*
- **When to use the test?** To assess the null hypothesis of zero correlation between two variables.
  - A paired random sample of ordinal or ranked data; or when the data is continuous and it is unreasonable to assume the variables are normally distributed.
  - The relationship between the variables is assumed to be linear.
- **Practical Applications.** Investing: Elton and Gruber (2001) investigate the relationship between marginal tax rates of the marginal stockholder and the firms dividend yield and payout ratio. They examined all stocks listed on the New York Stock Exchange that paid a dividend during April 1, 1966 to March 31, 1967. The Spearman rank correlation between the marginal tax rates of the marginal stock holder and the dividend yield was 0.9152 with a p-value less than 0.01. The null hypothesis of no correlation between the marginal tax rates of the marginal stock holder and the dividend yield was rejected. They also found the Spearman rank correlation between the marginal tax rates of the marginal stock holder and the payout ratio was 0.7939 with a p-value less than 0.01. The null hypothesis of no correlation between the marginal tax rates of the marginal stock holder and the payout ratio was also rejected.
- Elton, J.T. and Gruber, M.J. (2001). Marginal Stockholder Tax Rates and the Clientele Effect. Journal of Economics and Statistics. Volume 52(1), pages 68-74.

Spearman rank correlation test, R

- Both cor.test{stats} and spearman.test{pspearman} can be used to perform this test.

```
cor.test(X,Y, method="spearman",alternative="two.sided")

Spearman's rank correlation rho

data:  X and Y
S = 4173643902, p-value = 5.742e-07
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.0908284
```

The Spearman rank correlation between x and y is .1. Since the p-value is 5.742e-07 and lower than the critical value of 0.05, we do not reject the null hypothesis of zero correlation between two variables and conclude that there is a statistically significant relationship between the variables.

Kendall's tau correlation coefficient test

- **Question the test addresses:** *Is the Kendall tau correlation coefficient between two variables significantly different from zero?*
- **When to use the test?** To assess the null hypothesis of zero tau correlation between two variables.
  - Your sample consists of a paired random sample of ordinal or ranked data; or
  - when the data is continuous and it is unreasonable to assume the variables are normally distributed.
  - The relationship between the variables is assumed to be linear.
- **Practical Applications.** Sports Science: Dayaratna and Miller (2012) test the null hypothesis that goals scored and goals allowed in North American ice hockey are independent. For the Anaheim Ducks and seasons 2008/09, 2009/10 and 2010/11 they report tau correlation of 0.075,-0.105 and 0.008 respectively. The associated p-values were 0.156, 0.078 and 0.450 respectively. The null hypothesis of no correlation between goals scored and goals allowed for the Anaheim Ducks could not be rejected.
- Dayaratna, Kevin D.; Miller, Steven J. (2012). The Pythagorean Won-Loss Formula and Hockey: A Statistical Justification for Using the Classic Baseball Formula as an Evaluative Tool in Hockey. Hockey Research Journal: A Publication of the Society for International Hockey Research. Fall

Kendall's tau correlation coefficient test, R

- The function cor.test{stats} can be used to perform this test.

```
> cor.test(x,y,method="kendal",alternative="two.sided")
     Kendall's rank correlation tau

data:  X and Y
z = 5.0341, p-value = 4.801e-07
alternative hypothesis: true tau is not equal to 0
sample estimates:
       tau
0.06110555
```

- **Since the p-value = 4.801e-07 and is not greater than 0.05, we do not reject the null hypothesis of zero tau correlation between two variables.**

Two sample t-test for the difference in sample means

- **Question the test addresses:** *Is the difference between the mean of two samples significantly different from zero?*
- **When to use the test?** *You want to assess the extent to which* ***the mean of two independent samples are different*** *from each other.*
- The test assumes the sample observations are normally distributed, and the sample variances are equal.

Two sample t-test for the difference in sample means, R

- t.test(x,y, alternative = "two.sided", var.equal=TRUE)

```
t.test(X,Y, alternative = "two.sided", var.equal=TRUE)
Two Sample t-test

data:  X and Y
t = -2.9876, df = 6038, p-value = 0.002823
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.005427761 -0.001126861
sample estimates:
   mean of x    mean of y
0.0003390754 0.0036163866
```

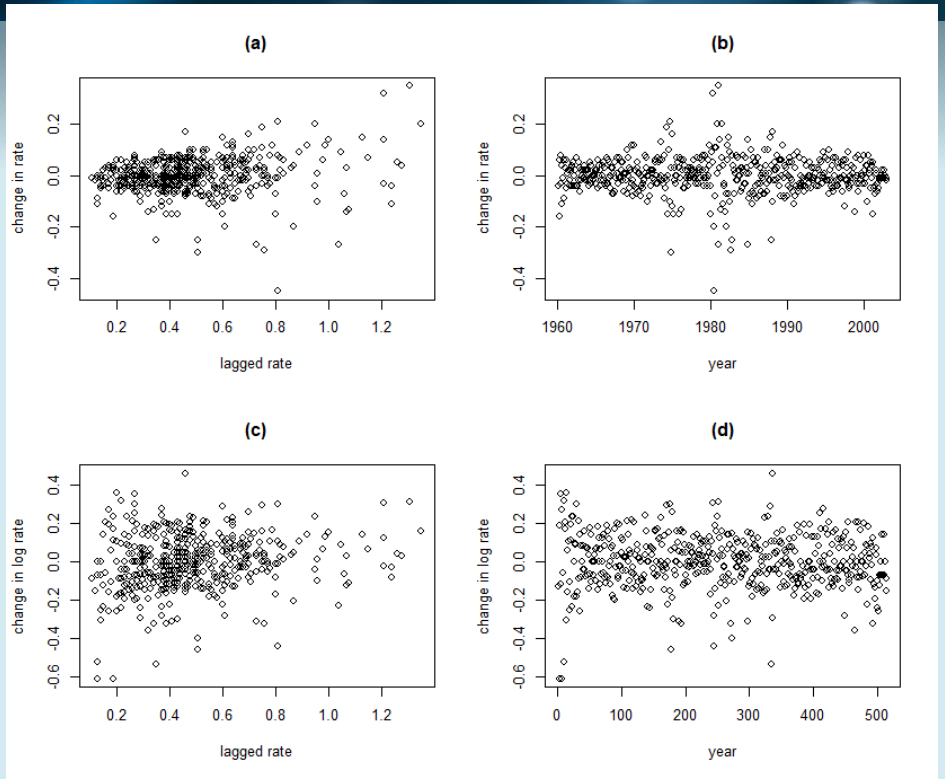
- **Since the p-value is = 0.002823, we reject the null hypothesis (the null hypothesis states that the population means of the two samples are equal), and conclude that there is a statistically significant difference between the means of the two samples.**

Data Transformation

- There are a number of reasons why data analysts often work not with the original variables, but rather with transformations of the variables such as **logs, square roots, or other power transformations**.
- Many statistical methods work best when the data are *normally distributed* or at least *symmetrically distributed* and have a *constant variance*
- *T*he transformed data often exhibits less skewness and a more constant variance compared to the original variables, especially if the transformation is selected to induce these features.
- **A transformation is called variance stabilizing if it removes a dependence between the conditional variance and the conditional mean of a variable**.
- For example, if Y has a distribution with a conditional mean depending on X, then its conditional variance is equal to the conditional mean.
- A transformation *h* would be variance-stabilizing for *Y* if the conditional variance of *h(Y)* did not depend on the conditional mean of *h(Y)*.

Data Transformation, cont.

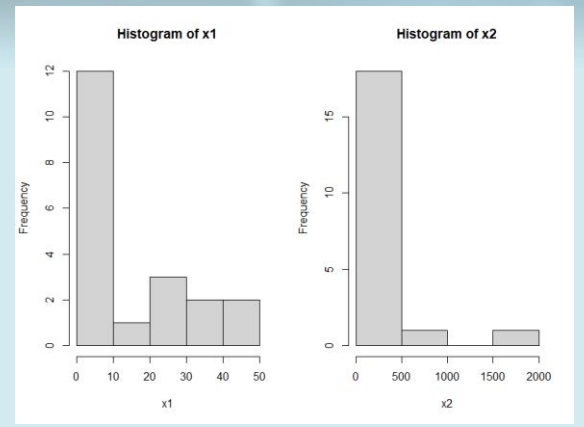
- The ***logarithm transformation*** is probably the most widely used transformation in data analysis, though the square root is a close second.
- The ***log*** **stabilizes the variance of a variable whose conditional standard deviation is proportional to its conditional mean**.
- This is illustrated in Fig., which plots monthly changes in the risk-free return (top row) and changes in the log of the return (bottom row) against the lagged risk-free return (left column) or year (right column).
- Notice that the changes in the return are more variable when the lagged return is higher.
- This behavior is called *nonconstant conditional variance or conditional heteroskedasticity*.
- **We see in the bottom row that the changes in the log return have relatively constant variability, at least compared to changes in the return.**

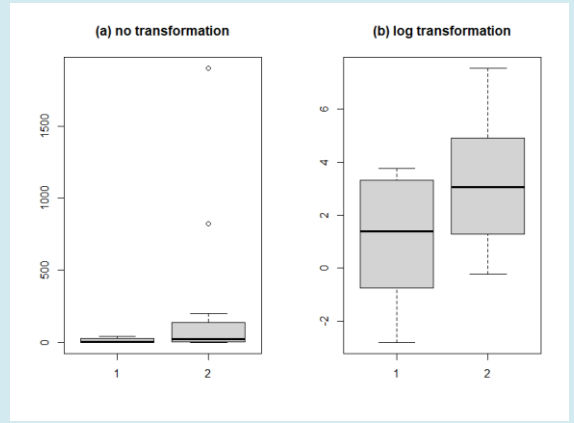


Changes in risk-free rate (top) and changes in the logarithm of the riskfree rate (bottom) plotted against time and against lagged rate.
The risk-free returns are the variable rf of the Capm data set in R's Ecdat package.
(a) Change in riskfree rate versus change in lagged rate.
(b) Change in rate versus year.
(c) Change in log(rate) versus lagged rate.
(d) Change in log(rate) versus year

Example: boxplot and t-Tests with transformations

- This example shows the deleterious effect of skewness and nonconstant variance on hypothesis testing and how a proper data transformation can remedy this problem.
- The boxplots on the panel (a) are of independent samples from lognormal(1,4) (left) and lognormal(3,4) distributions.
- Panel below shows boxplots of the log-transformed data.



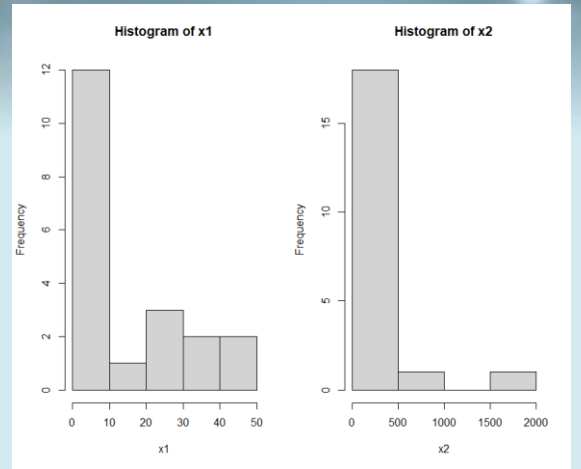


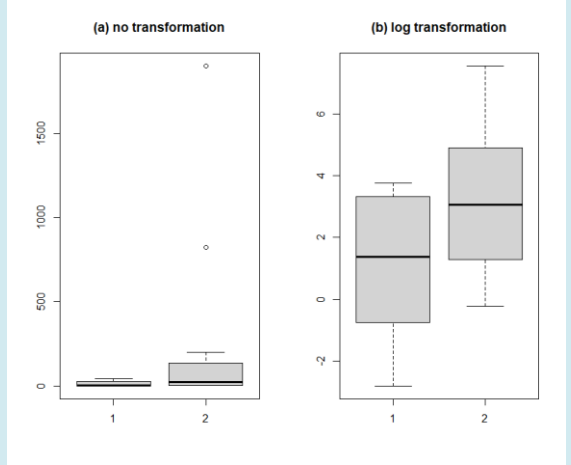

```
set.seed("119933")
par(mfrow=c(1,2))
x1 = rlnorm(20,meanlog=1,sdlog=2)
hist(x1)
x2 = rlnorm(20,meanlog=3,sdlog=2)
hist(x2)
boxplot(list(x1,x2),main="(a) no transformation")
boxplot(list(log(x1),log(x2)),main="(b) log transformation")
t.test(x1,x2,equal.var=F)
t.test(log(x1),log(x2))
```

Example: boxplot and t-Tests with transformations

```
> t.test(x1,x2,equal.var=F)
            Welch Two Sample t-test
data:  x1 and x2
t = -1.7042, df = 19.047, p-value = 0.1046
alternative hypothesis: true difference in means
is not equal to 0
95 percent confidence interval:
 -377.36913   38.61457
sample estimates:
mean of x mean of y
 13.30654 182.68382
```

```
> t.test(log(x1),log(x2))
            Welch Two Sample t-test
data:  log(x1) and log(x2)
t = -3.0057, df = 38, p-value = 0.004675
alternative hypothesis: true difference in means
is not equal to 0
95 percent confidence interval:
 -3.4159709 -0.6664116
sample estimates:
mean of x mean of y
 1.206306  3.247497
```





- Suppose one wants to test the null hypothesis that the two populations have the same means against a two-sided alternative.
- The transformed data satisfy the assumptions of the t-test that the two populations are normally distributed with the same variance, but of course the original data do not meet these assumptions.
- Two-sided independent-samples t-tests have p-values of 0.105 and 0.00467 using the original data and the log-transformed data, respectively.
- **These two p-values lead to rather different conclusions, for the first test that the means are not significantly different at the usual α = 0.05, and not quite significant even at α = 0.1, and for the second test that the difference is highly significant.**
- The first test reaches an incorrect conclusion because its assumptions are not met.

CORRELATIONS BETWEEN CATEGORICAL AND CONTINUOUS VARIABLES

Point-Biserial Correlation (a Population Product-Moment Correlation)

- $\rho_{pb}(X,Y)=E[(X-\mu X)(Y-\mu Y)]/\sigma X\sigma Y$

- Given that $\mu Y=P(Y=1) = p$ and $\sigma Y=\sqrt{p(1-p)}p$, the point-biserial correlation is:

- $\rho_{pb}(X,Y)=E[(X-\mu X)(Y-p)]/\sigma Xp\sqrt{p(1-p)}$
- In R: `biserial.cor {ltm}`
- `biserial.cor(x, y, use = c("all.obs", "complete.obs"), level = 1)`

Arguments

- **x**- a numeric vector representing the continuous variable.
- **y** - a factor or a numeric vector (that will be converted to a factor) representing the dichotomous variable.
- **use** - If use is "all.obs", then the presence of missing observations will produce an error. If use is "complete.obs" then missing values are handled by casewise deletion.
- **level** - which level of y to use.

Point-Biserial Correlation (a Population Product-Moment Correlation), cont

Most people can be trusted	**Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?**	**1 Most people can be trusted** **2 Need to be very careful** **-1 Don´t know** **-2 No answer** **-4 Not asked**	**Q57**

variables	min	Q1	mean	median	Q3	max
Bitcoin Tweets	-0.22051	-0.06917	-0.06737	-0.06917	-0.06917	0.286717
Ethereum Tweets	-0.31626	0.030754	0.030907	0.030754	0.030754	0.470436
Litecoin Tweets	-0.61174	-0.16768	-0.1638	-0.16768	-0.16768	3.474929
Bitcoin Google Trend	-0.26045	-0.0597	-0.05735	-0.0597	-0.0597	0.377595
Ethereum Google Trend	-0.36047	-0.20786	-0.20039	-0.20786	-0.20786	0.836737
Litecoin Google Trend	-0.30028	-0.21775	-0.21096	-0.21775	-0.21775	0.500349

Point-Biserial Correlation with “Need to be very careful”

Variables (growth)	Biserial correlation
Bitcoin Tweets	-0.02025
Ethereum Tweets	-0.01051
Litecoin Tweets	-0.01108
Bitcoin Google Trends	-0.00967
Ethereum Google Trends	-0.01381
Litecoin Google Trends	-0.00815

```
results.biserial = as.data.frame (biserial.cor(ys$BitcoinTweets, controls$Q57,
use = "complete.obs", level = 2))
```

Causality

- Discussions of this topic often starts with Hume (a Scottish philosopher whose works on this subject were published in around 1740).
- He argued that the causal relationship consists of three components;
- Temporal Priority, that is a cause proceeds in effect.
- Contiguity, the cause and effect or spacio-temporally connected.
- Constant Conjunction, that is repeated applications of the cause lead to the same effect.

Causality, cont

- Overall we can distinguish two main approaches that can be associated respectively with the two different types of studies, **observational versus experimental** continued in isolation.
- They also lead to very different ways of thinking: **In observational studies, researchers typically started out with an effective interests**. They then proceeded to ask, what were the causes of this effect?
- This led them to measure as many possible causes that they could think of and to build multivariate models in which the purported joint effects of the causes were estimated, by way of contrast.
- **In the experimental literature**, researchers focused on estimating the effect of the causal variable that was actually manipulated in the experiment, ignoring other causes that were not manipulated.
- **The observational studies focuses on the causes of an effect and the experimental study on the effect of a cause.**

Causality, cont.

- In 1895 the statistician George Yule, invented the partial correlation coefficient to distinguish between spurious correlations and correlations that evidence causation.
- The idea is this, suppose variable x is prior to and correlated with variable y. Possibly x causes y but it is also possible that a third variable z that is prior to x and y causes both x and y. In which case, conditional on z, the correlation between x and y should vanish.
  - This approach to causal inference was subsequently resurrected or reinvented many times in the 20th century. By Lazarsfeld in the social sciences, Reichenbacher and Sups in philosophy and in the time series setting by the econometrician Clive Granger.
- The ones, who use regression analysis and more complicated statistical methods for example structural equation models, use this kind of reasoning when they make inference it's about causation.
- Granges work led to a large literature on Granger causality which still features prominently in various disciplines for example finance and neuroscience.

Granger causality test

- **Is one time series useful in forecasting another?**
- When to use the test?
- Granger causality is a statistical concept of causality. Whenever a "surprise“ in an independent variable leads to a later increase in the dependent variable we call this variable "Granger causal."
- The test is based on prediction, if an independent variable "Granger-causes" a dependent variable, then past values of the independent variable should contain information that helps predict the dependent variable above and beyond the information contained in past values of the dependent variable alone.
- For example, a time series X is said to Granger-cause Y if it can be shown—usually through a series of t-tests and F-tests on lagged values of X (and with lagged values of Y also included)—that those X values provide statistically significant information about future values of Y
- Or in other words: Granger causality exists when using lags of x next to the lags of y for forecasting y delivers better forecast accuracy than using only the lags of y (without the lags of x)
- How to calculate in R
- The function `grangertest {lmtest}`.
- It takes the form `grangertest(x, y, order = 1, na.action = na.omit, ...)`, p is the order of the test (the number of lags to utilize. The default value is 1)
- This test generates an F test statistic along with a p-value. We can reject the null hypothesis and infer that time series X Granger causes time series Y if the p-value is less than a particular significance level (e.g. =.05).

Granger causality test, example

```
library(yfR)
#  options
my_ticker = c("VXX", "AMZN", "ABNB")
first_date = Sys.Date() - 365*2 #First day
last_date = Sys.Date() #last day at the moment of running

# get  data
df_yf = yf_get(tickers = my_ticker,
               first_date = first_date,
               last_date = last_date)

unique.tikers = unique(df_yf$ticker)

ABNB= df_yf[df_yf$ticker==unique.tikers[1], ] #"ABNB"
AMZN= df_yf[df_yf$ticker==unique.tikers[2], ] # "AMZN"
VXX= df_yf[df_yf$ticker==unique.tikers[3], ] # "VXX«

library (lmtest)
grangertest(AMZN$price_open ~ VXX$price_adjusted, order = 1)
grangertest(AMZN$price_open ~ ABNB$price_close, order = 1)
grangertest(AMZN$price_open ~ ABNB$volume, order = 1)
```

```
> grangertest(AMZN$price_open ~ VXX$price_adjusted, order = 1)
Granger causality test

Model 1: AMZN$price_open ~ Lags(AMZN$price_open, 1:1) +
Lags(VXX$price_adjusted, 1:1)
Model 2: AMZN$price_open ~ Lags(AMZN$price_open, 1:1)
  Res.Df Df      F Pr(>F)
1    500
2    501 -1 1.3605  0.244
```

The null hypothesis for the Granger causality test is that X does not Granger-cause Y. We may not reject the null hypothesis of the test because the p-value is > than 0.05, and then we cannot infer that knowing the values of `VXX$price_adjusted` is valuable for forecasting the future values of `AMZN$price_open`.

```
> grangertest(AMZN$price_open ~ ABNB$price_close, order = 1)
Granger causality test

Model 1: AMZN$price_open ~ Lags(AMZN$price_open, 1:1) +
Lags(ABNB$price_close, 1:1)
Model 2: AMZN$price_open ~ Lags(AMZN$price_open, 1:1)
  Res.Df Df      F  Pr(>F)
1    500
2    501 -1 4.1616 0.04188 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Causality based on VAR

- A Vector Autoregression (VAR) model is a multivariate time series model used to capture the linear interdependencies among multiple time series. It extends the univariate autoregressive model to include multiple time series, allowing for the simultaneous modeling of the dynamics of each variable as a function of its own past values and the past values of other variables in the system.

- The general form of a VAR model of order p (VAR(p)) for a set of k variables can be written as:

$$Yt = A1 * Y\{t-1\} + A2 * Y\{t-2\} + ... + Ap * Y\{t-p\} + ut$$

where:

Yt is a k x 1 vector of the time series variables at time t.
Ai (i = 1, 2, ..., p) are k x k matrices of coefficients to be estimated.
Y{t-i} (i = 1, 2, ..., p) are k x 1 vectors of the time series variables at lag i.
ut is a k x 1 vector of error terms, which are assumed to be white noise and follow a multivariate normal distribution with mean 0 and constant covariance matrix Σ.

The main objectives of a VAR model are to:

- Forecast the future values of the time series variables.
- Investigate the dynamic relationships among the variables, such as impulse response functions and variance decompositions, to understand how shocks to one variable propagate through the system and affect other variables.
- The main advantage of VAR models is that they provide a flexible and straightforward framework for analyzing the complex interrelationships among multiple time series without imposing strong theoretical assumptions or causal relationships among the variables.

Causality based on VAR

causality() function from the 'vars' package:

- The causality() function is used with Vector Autoregression (VAR) models.
- It tests for Granger causality in a multivariate time series context, where you have more than one time series variable.
- The test is based on the coefficients of a VAR model fitted to the data.
- The causality() function is applied to a fitted VAR model object and returns the results of the Granger causality test for specified variables.
- You have to interpret the test results considering the interactions of all variables in the VAR model.
  - The presence of more variables may increase the chances of detecting Granger causality, but it can also make the test more susceptible to model misspecification or other issues that may arise from working with multivariate time series data.
- The function provides both the test statistic and the p-value for the Granger causality test, and a low p-value indicates evidence of Granger causality between the specified variables.

grangertest() function from the 'lmtest' package:

- The grangertest() function is used in a bivariate time series context, meaning you have only two time series variables.
  - While it might be easier to detect causality between two variables in a simpler context, it may miss indirect relationships or spillover effects from other variables not included in the bivariate model.
- This function applies a Wald test on the coefficients of a bivariate autoregressive model.
- The grangertest() function takes two time series variables as inputs and returns the results of the Granger causality test.
- a low p-value suggesting Granger causality between the two variables.

- In summary, use causality() from the 'vars' package for multivariate time series data and when you have already fitted a VAR model. Use grangertest() from the 'lmtest' package for bivariate time series data when you want to test Granger causality between just two time series variables.

Causality based on VAR

- Choosing the best criterion for selecting the lag order in a VAR model depends on the specific context and objectives of your analysis. Each criterion has its own strengths and weaknesses:

- **AIC (Akaike Information Criterion)**:
  - AIC is a commonly used criterion that balances the goodness of fit and model complexity.
  - It tends to select models with more parameters, making it more likely to overfit the data.
  - However, it can be advantageous when working with large datasets or when the primary objective is to forecast.

- **HQ (Hannan-Quinn Criterion):**
  - HQ is a more conservative criterion compared to AIC.
  - It also balances the goodness of fit and model complexity but penalizes models with more parameters more heavily.
  - It's less likely to overfit the data compared to AIC but may have a slightly worse forecasting performance.

- **SC (Schwarz Criterion) or BIC (Bayesian Information Criterion):**
  - SC or BIC is an even more conservative criterion than HQ.
  - It has the highest penalty for model complexity among these criteria, making it least likely to overfit the data.
  - SC/BIC is more suitable when the primary objective is to find the true data-generating process or when the sample size is relatively small.

- **FPE (Final Prediction Error)**:
  - FPE is similar to AIC but has a slightly different penalty term.
  - It's also focused on forecasting performance, and its results are often very close to those obtained with AIC.

- Ultimately, the choice of criterion should be based on your specific goals and the characteristics of your data.
- If you are primarily interested in forecasting, AIC or FPE might be more appropriate.
- If you want to identify the true data-generating process and are working with a smaller sample, SC/BIC could be a better choice.
- HQ is a middle ground between AIC and SC/BIC and can be a good compromise.

Causality based on VAR

- Choosing the best criterion for selecting the lag order in a VAR model depends on the specific context and objectives of your analysis. Each criterion has its own strengths and weaknesses:

- **AIC (Akaike Information Criterion)**:
  - AIC is a commonly used criterion that balances the goodness of fit and model complexity.
  - It tends to select models with more parameters, making it more likely to overfit the data.
  - However, it can be advantageous when working with large datasets or when the primary objective is to forecast.

- **HQ (Hannan-Quinn Criterion):**
  - HQ is a more conservative criterion compared to AIC.
  - It also balances the goodness of fit and model complexity but penalizes models with more parameters more heavily.
  - It's less likely to overfit the data compared to AIC but may have a slightly worse forecasting performance.

- **SC (Schwarz Criterion) or BIC (Bayesian Information Criterion):**
  - SC or BIC is an even more conservative criterion than HQ.
  - It has the highest penalty for model complexity among these criteria, making it least likely to overfit the data.
  - SC/BIC is more suitable when the primary objective is to find the true data-generating process or when the sample size is relatively small.

- **FPE (Final Prediction Error)**:
  - FPE is similar to AIC but has a slightly different penalty term.
  - It's also focused on forecasting performance, and its results are often very close to those obtained with AIC.

- Ultimately, the choice of criterion should be based on your specific goals and the characteristics of your data.
- If you are primarily interested in forecasting, AIC or FPE might be more appropriate.
- If you want to identify the true data-generating process and are working with a smaller sample, SC/BIC could be a better choice.
- HQ is a middle ground between AIC and SC/BIC and can be a good compromise.

Causality based on VAR

```
#causality based on VAR
library(vars)
data = cbind(return.together.df$SP600.Close, return.together.df$weekdays.cryptos.xts)
# Set the maximum number of lags to consider
max_lags = 10

# Calculate selection criteria for different lag orders
lag_selection = VARselect(data, lag.max = max_lags, type = "both")

# View the selection criteria
print(lag_selection)

# Extract the optimal lag order based on the AIC or BIC criterion
optimal_lag_aic = lag_selection$selection["AIC(n)"]
optimal_lag_bic = lag_selection$selection["SC(n)"]

# Estimate the VAR model with the optimal lag order
model_aic = VAR(data, p = optimal_lag_aic)
model_bic = VAR(data, p = optimal_lag_bic)

#To test for causality within a VAR model, you can use the causality() function:
causal_test = causality(model_aic, cause = "y1")
```

```
> causal_test
$Granger
                Granger causality H0: y1 do not Granger-cause y2
data:  VAR object model_aic
F-Test = 0.52603, df1 = 8, df2 = 5990, p-value = 0.8378

$Instant
                H0: No instantaneous causality between: y1 and y2

data:  VAR object model_aic
Chi-squared = 57.49, df = 1, p-value = 3.397e-14
```

1.Granger causality test:
H0: y1 does not Granger-cause y2
Interpretation: Since the p-value (0.8378) is significantly higher than the commonly used significance level of 0.05, we fail to reject the null hypothesis. This means that there is insufficient evidence to suggest that y1 Granger-causes y2.

2.Instantaneous causality test:
H0: No instantaneous causality between y1 and y2
Interpretation: Since the p-value (3.397e-14) is extremely low and far below the significance level of 0.05, we reject the null hypothesis. This indicates that there is strong evidence of instantaneous causality between y1 and y2.

In summary, based on the test results, y1 does not Granger-cause y2, but there is evidence of instantaneous causality between y1 and y2.

Instantaneous causality refers to the concept that the current value of one variable has an immediate impact on the current value of another variable, without any lag.
This concept is different from **Granger causality**, which is based on the idea that past values of one variable can help predict future values of another variable.

In the context of time series analysis, instantaneous causality can be seen as **a contemporaneous correlation between two variables**.
When there is **instantaneous causality** between two variables, it **implies that the variables are affected by a common factor or set of factors simultaneously.**

Directed Acyclic Graphs

- A directed acyclic graph (DAG) can be thought of as a kind of flowchart that visualizes a whole causal network, linking causes and effects.
- DAGs are considered to be of use for embedding causality in a formal causal framework (Hernán and Robins 2006; Robins 2001; Hernán et al. 2004).

DAGs

- Directed Acyclic Graphs (DAGs) are graphical representations of causal relationships among a set of variables.
- In a DAG, nodes represent variables, and directed edges (arrows) represent causal relationships between the variables.
- A key property of DAGs is that they do not contain cycles, which means that you cannot follow a path of directed edges that starts and ends at the same node.
- In the context of stock prices, DAGs can be used to analyze the causal relationships between various financial variables, such as stock prices, interest rates, economic indicators, or other financial assets.
- To estimate causality in stock prices using DAGs in R, you can use the `pcalg` package, which offers functions to learn the structure of DAGs from data.
- Here's a simple example to get you started:

DAGs: key concepts

- **Conditional Independence:** In a DAG, two variables X and Y are conditionally independent given a set of variables Z if and only if knowing the values of Z makes X and Y statistically independent.
  - Mathematically, this is expressed as P(X, Y | Z) = P(X | Z) * P(Y | Z), where P represents the joint probability distribution of the variables.

- **d-separation:** A key concept in DAGs is d-separation, which helps determine whether two variables are conditionally independent given a set of other variables.
  - In a DAG, two variables X and Y are d-separated by a set of variables Z if there is no "active" path between X and Y when conditioning on Z.
    - An active path is an undirected path that allows for the propagation of dependence between variables.

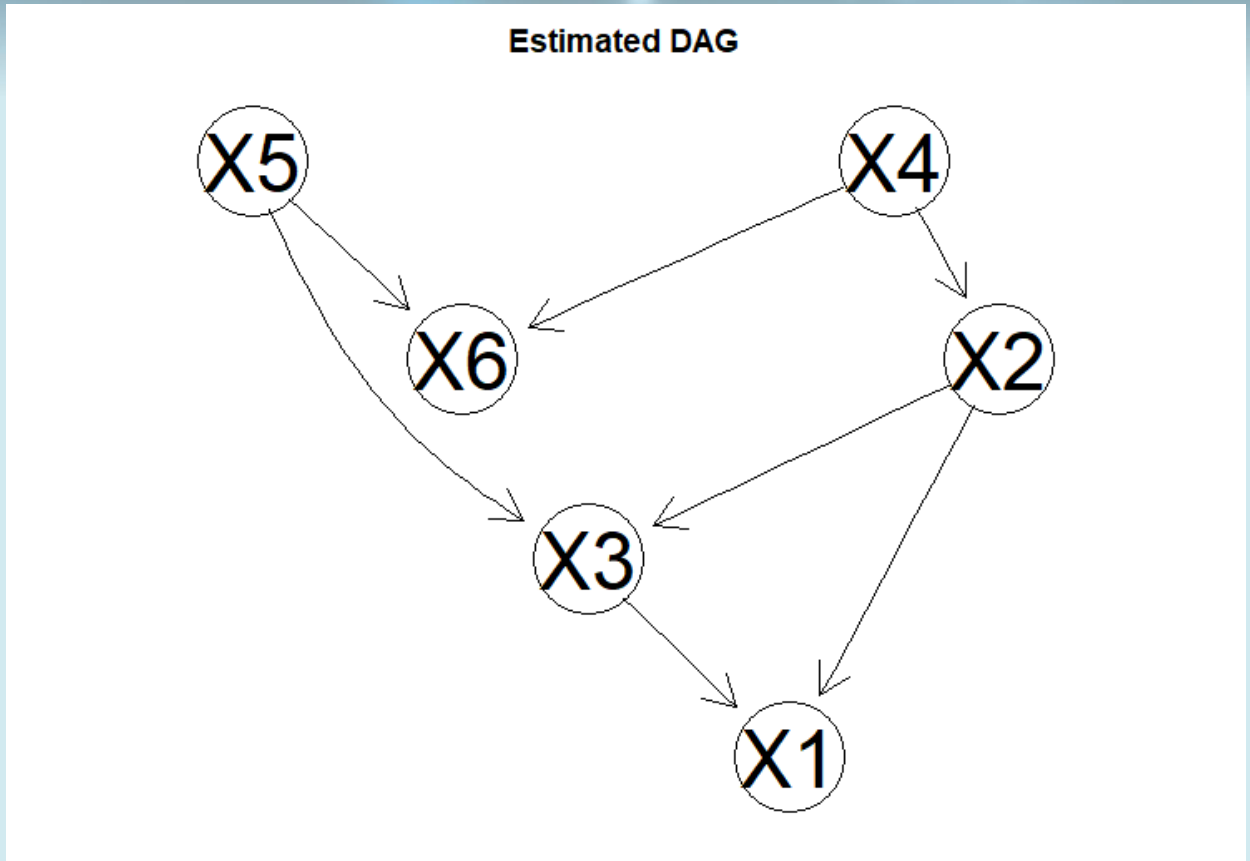
- **Markov Condition:** In a DAG, a variable is conditionally independent of its non-descendants given its parents. This is also known as the Markov condition or the local Markov property.
  - Mathematically, it means that for any variable X, P(X | parents(X), non-descendants(X)) = P(X | parents(X)), where parents(X) represent the set of parent variables of X in the DAG.
  - parent variables are those variables that have a direct causal influence on another variable
  - non-descendants are variables that are not part of the set of variables that are causally influenced, either directly or indirectly, by a given variable
    - For instance: if there is an arrow from variable A to variable B, it indicates that A has a causal effect on B, and B is a descendant of A.
    - If B, in turn, has a causal effect on variable C, then C is also a descendant of A, as it is indirectly influenced by A through B.
    - Descendants include both direct and indirect causal effects of a variable.
    - Non-descendants, on the other hand, are variables that are not part of the set of variables that are causally influenced by a given variable.
    - They are not affected, either directly or indirectly, by the given variable. In the example above, any variable that is not part of the set {B, C} would be a non-descendant of A.

- **Causal Inference:** DAGs enable causal inference by helping researchers identify the appropriate set of variables to condition on when estimating causal effects. Using the concept of d-separation, researchers can identify which variables should be controlled for to obtain unbiased estimates of causal effects.

- **Intervention and Counterfactuals:** DAGs are also useful for reasoning about interventions and counterfactuals.
  - An intervention is an action that sets a variable to a specific value, irrespective of its parents. Counterfactuals are hypothetical scenarios where one or more variables take different values than what was observed.
  - DAGs provide a framework for reasoning about these situations and estimating causal effects under different scenarios.

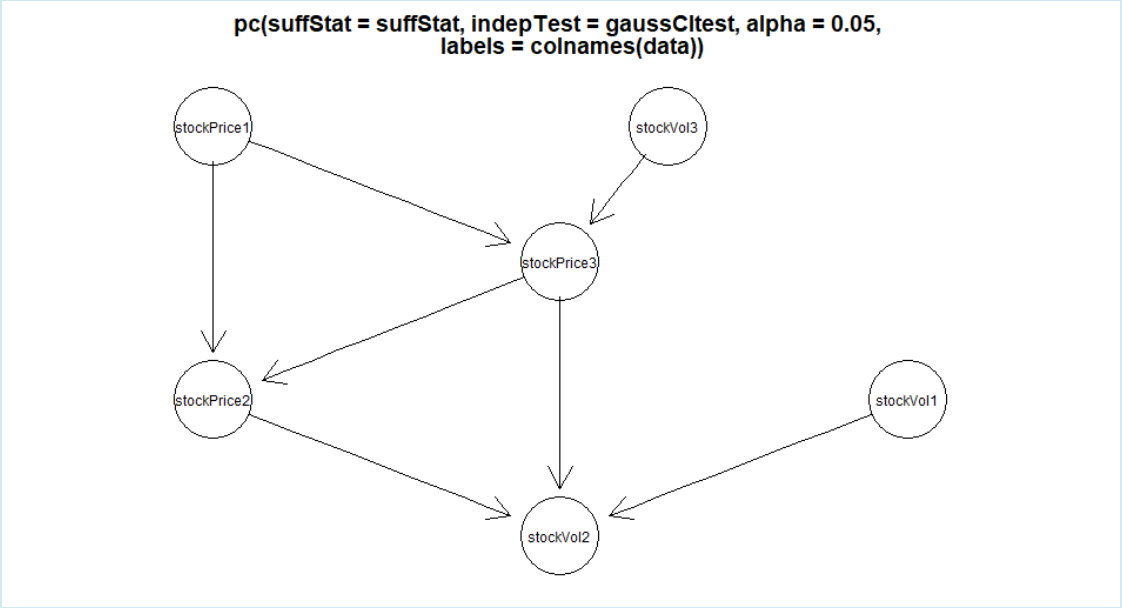
DAG, example

```
## Causality based on DAG
if (!require("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install("graph")
BiocManager::install("RBGL")
BiocManager::install("Rgraphviz")

library(pcalg); library(graph)

# Load the required packages
library(pcalg); library(ggm)

# Set the seed for reproducibility
set.seed(42)

# Number of observations
n = 1000

# Simulate data for six variables
X1 = rnorm(n)
X2 =  0.6 * X1 + rnorm(n, sd=0.8)
X3 = 0.7 * X1 - 0.3 * X2 + rnorm(n, sd=0.9)
X4 = 0.5 * X2 + rnorm(n, sd=1.0)
X5 = -0.8 * X3 + rnorm(n, sd=1.1)
X6 = 0.4 * X4 + 0.4 * X5 + rnorm(n, sd=1.2)

# Combine variables into a data frame
data  = data.frame(X1, X2, X3, X4, X5, X6)

# Compute the covariance matrix
cov_matrix = cor(data)

# Estimate the DAG structure using the PC algorithm
suffStat = list(C = cov_matrix, n = n)
dag = pc(suffStat, indepTest = gaussCItest, alpha = 0.05, labels = colnames(data))

# Plot the DAG
plot(dag, main="Estimated DAG")
```



DAGs, example2

```
if (!require("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install("graph")
BiocManager::install("RBGL")
BiocManager::install("Rgraphviz")

library(pcalg)
library(graph)

library(yfR)
my_ticker = c("VXX", "AMZN", "ABNB")
first_date = Sys.Date() - 365 #First day
last_date = Sys.Date() #last day at the moment of running

df_yf = yf_get(tickers = my_ticker,
               first_date = first_date,
               last_date = last_date)

unique.tikers = unique(df_yf$ticker)

stock1= df_yf[df_yf$ticker==unique.tikers[1], ] # ABNB
stock2= df_yf[df_yf$ticker==unique.tikers[2], ] # AMZN
stock3= df_yf[df_yf$ticker==unique.tikers[3], ] # VXX

data = cbind(stock1$price_close, stock2$price_close, stock3$price_close, stock1$volume,
stock2$volume, stock3$volume)
colnames(data) = c("stockPrice1", "stockPrice2", "stockPrice3", "stockVol1", "stockVol2",
"stockVol3")

# Estimate the DAG:
suffStat = list(C = cor(data), n = nrow(data))
dag = pc(suffStat, indepTest = gaussCItest, alpha = 0.05, labels = colnames(data))

# Plot the estimated DAG:
library(Rgraphviz)
plot(dag)
```



Directed Acyclic Graphs, further direction

- http://www.dagitty.net/
- https://cran.r-project.org/web/packages/ggdag/vignettes/intro-to-dags.html

Spillovers

- Total and directional volatility spillovers Diebold and Yilmaz (2009, 2012) https://cran.r-project.org/web/packages/Spillover/Spillover.pdf
- The Diebold-Yilmaz (DY) spillover index is a measure of the interconnectedness and transmission of shocks across variables in a multivariate time series. It is based on the forecast error variance decomposition (FEVD) obtained from a Vector Autoregressive (VAR) model.
- The DY spillover index is widely used in finance and macroeconomics to analyze the interconnectedness among financial markets, sectors, or countries.

Spillovers

- **Estimate a VAR** model for the multivariate time series of interest. The VAR(p) model is expressed as:

$$Yt = A1*Y\{t-1\} + A2*Y\{t-2\} + ... + Ap * Y\{t-p\} + \varepsilon t$$

- where Yt is a vector of endogenous variables at time t, Ai (i = 1, 2, ..., p) are matrices of coefficients, and εt is a vector of error terms.
- **Compute the impulse response functions (IRFs) of the VAR model**.
  - IRFs capture the responses of each variable in the system to a one-unit shock in each of the other variables, taking into account the dynamic interdependencies among the variables.
- **Compute the forecast error variance decomposition (FEVD) based on the IRFs**.
  - FEVD measures the proportion of the forecast error variance of each variable due to shocks from each of the other variables in the system.
- **Calculate the DY spillover index based on the FEVD.**
  - The total spillover index is computed as the sum of all off-diagonal elements of the FEVD matrix divided by the sum of all elements of the FEVD matrix:
- Total Spillover Index = (sum of off-diagonal elements) / (sum of all elements)
- The pairwise spillover from variable i to variable j is the (i, j)-th element of the FEVD matrix, and the directional spillover from variable i to all other variables is the sum of the i-th row elements excluding the diagonal element.

Diebold and Yilmaz spillover

```
library(quantmod)
library(Spillover)

# Define the stock indices and their respective symbols
indices <- list(
  "S&P 500" = "^GSPC",
  "Nikkei 225" = "^N225",
  "DAX" = "^GDAXI"
)

# Download historical data for the stock indices
data = NULL
for (symbol in indices) {
  prices = getSymbols(symbol, auto.assign = FALSE, from = "2020-01-01")
  if (is.null(data)) {
    data = Ad(prices)
  } else {
    data = cbind(data, Ad(prices))
  }
}
# Calculate the daily returns
returns = na.omit(ROC(data, type = "discrete"))

## Estimate VAR
library(vars)

# Select the optimal lag order for the VAR model
lag_selection =  VARselect(returns, lag.max = 10, type = "both")
optimal_lag = lag_selection$selection[1]

# Estimate a VAR model
var_model <- VAR(returns, p = optimal_lag)
summary (var_model)

# Estimate the spillover index
spillover_index <- G.spillover(var_model, n.ahead = 10)

# Print the spillover index
print(spillover_index)
```

	GSPC.Adjusted	N225.Adjusted	GDAXI.Adjusted	**C.from others**
GSPC.Adjusted	20.263692	3.396560	9.673081	13.06964
N225.Adjusted	7.417803	18.495624	7.419906	14.83771
GDAXI.Adjusted	9.892884	4.153887	19.286562	14.04677
C. to others (spillover)	17.310687	7.550447	17.092987	41.95412
C. to others including own	37.574380	26.046070	36.379550	100.00000

1.Diagonal elements: The diagonal elements (**20.26% for GSPC, 18.5% for N225, and 19.29% for GDAXI**) represent **the proportion of the total forecast error variance in each index due to its own innovations or shocks**. These values indicate **how much of the variability in each index is explained by its own past values**.

2. **Off-diagonal elements**: The off-diagonal elements **show the pairwise spillover effects between the indices**.
For example, the value of **3.40%** in the first row and second column indicates that 3.40% of the total f orecast error variance in the N225 index is due to shocks from the GSPC index.
Similarly, the value of **7.42%** in the second row and first column indicates that 7.42% of the total f orecast error variance in the GSPC index is due to shocks from the N225 index.

3. **C. from others:** The "from others" shows **the proportion of the total forecast error variance in each index that is due to shocks from other indices**.
For example, **13.07%** of the total forecast error variance in the GSPC index is due to shocks from the N225 and GDAXI indices.

3. **C. to others (spillover):** The "to others" row shows **the proportion of the total forecast error variance in each index that is transmitted to other indices.**
For example, the GSPC index sends **17.31%** of its total forecast error variance to the N225 and GDAXI indices. Thus, he results indicate that the GSPC index has the highest spillover effects on other indices (**17.31%**), followed by the GDAXI index (**17.09%**) and the N225 index (**7.55%**).

4. **C. to others including own**: The "to others including own" row shows **the proportion of the total forecast error variance in all indices that is due to spillovers**. In this case, the total spillover index is 100%, indicating the overall level of interdependence among the three indices.

Other methods to check causality (time series)

- Short run and long run causality in time series: inference, https://doi.org/10.1016/j.jeconom.2005.02.003
- Non-linear causality test based on Transfer entropy and Vector Auto-Regressive Neural Network- https://rjournal.github.io/archive/2020/RJ-2020-016/RJ-2020-016.pdf ; https://cran.r-project.org/web/packages/NlinTS/NlinTS.pdf
- To assess the causal relationship between time -series in different market states, the Granger test in quantiles following Hong et al.(2009) and Lee and Yang (2014).
  - Hong et al.(2009) Granger causality in risk and detection of extreme risk spillover between financial markets, Journal of Econometrics, 150, 271-287
  - Lee, T.-H. and W. Yang (2014). Granger-causality in quantiles between financial markets: Using copula approach. International Review of Financial Analysis 33, 70–78.
- Time-varying Granger causality tests
  - Lu, F.-B., Hong, Y.-M., Wang, S.-Y., Lai, K.-K., Liu, J. (2014). Time-varying Granger causality tests for applications in global crude oil markets. Energy Economics, 42, 289–298. https://sites.google.com/site/shupingshi/home/codes
- Experimental package: Variable-Lag Time Series Causality Inference Framework: https://github.com/DarkEyes/VLTimeSeriesCausality

Other methods to check causality (time series)

- **Directional predictability test of Han et al. (2016)** is also applied to investigate if the selected series have directional predictability (cross-quantilogram). [Han, H., O. Linton, T. Oka, and Y.-J. Whang (2016). The cross-quantilogram: measuring quantile dependence and testing directional predictability between time series. Journal of Econometrics 193(1), 251–270.; https://doi.org/10.1016/j.jeconom.2016.03.001]
  - https://sites.google.com/site/whangyjhomepage/research/software
  - https://cran.r-project.org/web/packages/quantilogram/index.htm I
- **Impulse response function** https://cran.r-project.org/web/packages/vars/vignettes/vars.pdf , or bvarsv package for Bayesian estimates
- **Impulse response functions using local projections** https://journal.r-project.org/archive/2019/RJ-2019-052/RJ-2019-052.pdf , https://cran.r-project.org/web/packages/lpirfs/vignettes/lpirfs_vignette.html
- **Causality in Continuous Wavelet Transformation**
  - Olayeni, O.R. Causality in Continuous Wavelet Transform Without Spectral Matrix Factorization: Theory and Application. Comput Econ 47, 321–340 (2016). https://doi.org/10.1007/s10614-015-9489-4

Other methods to check causality (time series)

- Bayesian diffusion-regression (state-space) structural time-series model
- The causal impact of an event is the difference between the observed value of the response variable and the unobserved value that would have been obtained had the intervention not taken place (Brodersen et al. 2015; Claveau 2012; Hoover 2012; Antonakis et al. 2010).
- https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/41854.pdf
- https://google.github.io/CausalImpact/CausalImpact.html

TIME-SERIES AND VOLATILITY

Time series

- **A time series is a sequence of observations in chronological order**, for example, daily log returns on a stock or monthly values of the Consumer Price Index (CPI).
- **A common simplifying assumption is that the data are equally spaced with a discrete-time observation index**; however, this may only hold approximately.
- For example, daily log returns on a stock may only be available for weekdays, with additional gaps on holidays, and monthly values of the CPI are equally spaced by month, but unequally spaced by days.
- A stochastic process is a sequence of random variables and can be viewed as the "theoretical" or "population" analogy of a time series—conversely, a time series can be considered a sample from a stochastic process.
- **"Stochastic" is a synonym for random.**
- One of the most useful methods for obtaining parsimony in a time series model is to assume some form of distributional invariance over time, or stationarity, a property discussed next.

Distinguishing time-series components

- Distinguishing different components of a time-series involves breaking down the time-series into its fundamental components: trend, seasonality, and residuals (or irregular component).

- Trend:
  - The long-term movement in the data, which can be upward, downward, or stationary. It is the general direction in which the time-series is moving over time.

- Seasonality:
  - Regular fluctuations in the data that follow a seasonal pattern, such as daily, weekly, monthly, or yearly patterns. Seasonality can be additive or multiplicative, depending on how it affects the data.

- Residuals (Irregular Component):
  - The remaining variation in the data after removing the trend and seasonality components. This component includes random noise and other unexplained fluctuations in the data.

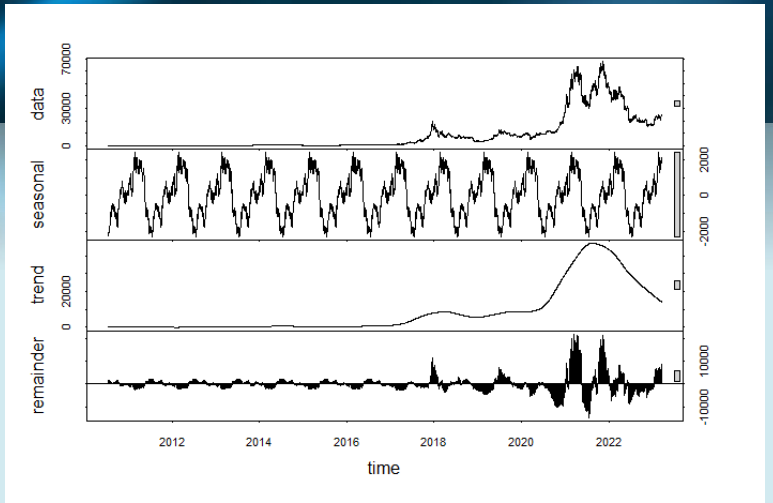
- To distinguish these components, you can use time-series decomposition techniques, such as the classical decomposition, seasonal decomposition of time series (STL), or seasonal-trend decomposition using LOESS (STL).

Decomposition of time series

```
# Load the necessary packages
library(timetk)
library(tidyverse)

Bitcoin = read.csv("BitcoinData.csv")
Bitcoin = Bitcoin[, -1]
Bitcoin$date = as.Date(Bitcoin$date,format="%Y-%m-%d", tz = "")

# Create a time-series object
Bitcoin_ts =  Bitcoin %>%
  tk_ts(start = c(2010, 198), frequency = 365)

# Decompose the time-series using STL
stl_decomposition <- stl(Bitcoin_ts[,1], s.window = "periodic")
# LOESS (locally estimated scatterplot smoothing).
# It is particularly well-suited for data with a seasonal
component and allows # for changing seasonality patterns over
time.

# Plot the decomposed time-series components
plot(stl_decomposition)
```

date
2011-09-11
2011-09-12
2011-09-13
:

- %Y - year
- %m - month
- %d - day
- %H:%M - hours:minutes



1.Trend component: The trend represents the overall direction and long-term movement of the time series data. A trend can be upward, downward, or stationary. It may also change over time, with periods of increasing or decreasing growth rates. By examining the trend component, you can identify the general pattern of the data and determine if there is a consistent increase or decrease over time.

2.Seasonal component: Seasonality refers to the presence of regular fluctuations in the data that repeat over fixed periods. These fluctuations can be due to natural factors (e.g., weather, daylight hours) or human factors (e.g., holidays, business cycles). Seasonal patterns can be additive or multiplicative, depending on how they interact with the trend component. By analyzing the seasonal component, you can identify the periods during which the time series data tends to be higher or lower than average and adjust your forecasts or decisions accordingly.

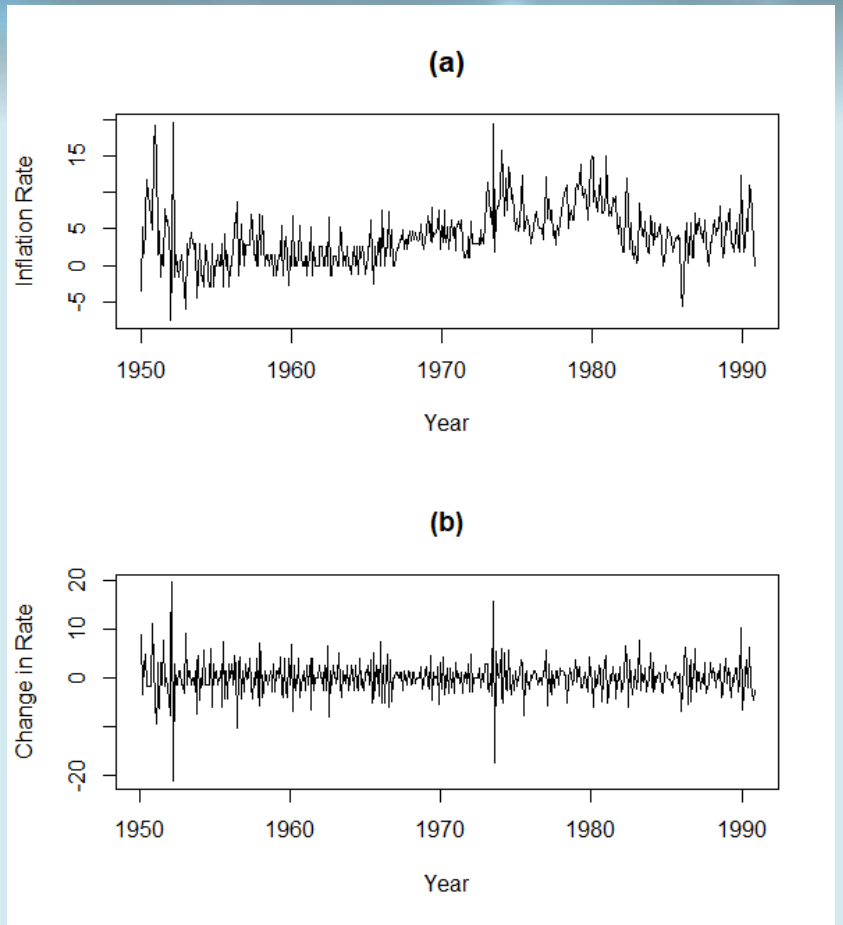
3.Residuals (Random or Irregular component): The residuals represent the unexplained variation in the data after removing the trend and seasonal components. These variations can be due to random noise, unexpected events, or other factors that are not captured by the trend and seasonal components. In a well-modeled time series, the residuals should be random, uncorrelated, and have a mean of zero. Analyzing the residuals can help you identify any remaining patterns or dependencies in the data that were not captured by the decomposition, and possibly improve your model.

Stationary Processes

- When we observe a time series, **the fluctuations appear random, but often with the same type of stochastic behavior from one time period to the next**.
- For example, returns on stocks or changes in interest rates can be very different from the previous year, **but the mean, standard deviation, and other statistical properties often are similar from one year to the next**.
- Similarly, the demand for many consumer products, such as sunscreen, winter coats, and electricity, as random as well as seasonal variation, but each summer is similar to past summers, each winter to past winters, at least over shorter time periods.
- ***Stationary stochastic processes are probability models for time series with time-invariant behavior.***
- ***A process is said to be strictly stationary if all aspects of its behavior are unchanged by shifts in time.***
- Mathematically, stationarity is defined as the requirement that for every m and n, the distributions of (Y1, . . . , Yn) and (Y1+m, . . . , Yn+m) are the same; that is, the probability distribution of a sequence of n observations does not depend on their time origin.
- Strict stationarity is a very strong assumption, because it requires that "all aspects" of stochastic behavior be constant in time. Often, it will suffice to assume less, namely, weak stationarity.
- ***A process is weakly stationary if its mean, variance, and covariance are unchanged by time shifts.***
- **For example, if the process is weakly stationary, then the covariance between Y2 and Y5 is the same as the covariance between Y7 and Y10, since each pair is separated by three units of time. Their mean and variance are also the same**
- The adjective "weakly" in "weakly stationary" refers to the fact that we are only assuming **that means, variance, and covariances, not other distributional characteristics such as quantiles, skewness, and kurtosis, are stationary**. Weakly stationary is also sometimes referred to as covariance stationary. The term stationary will sometimes be used as a shorthand for strictly stationary.

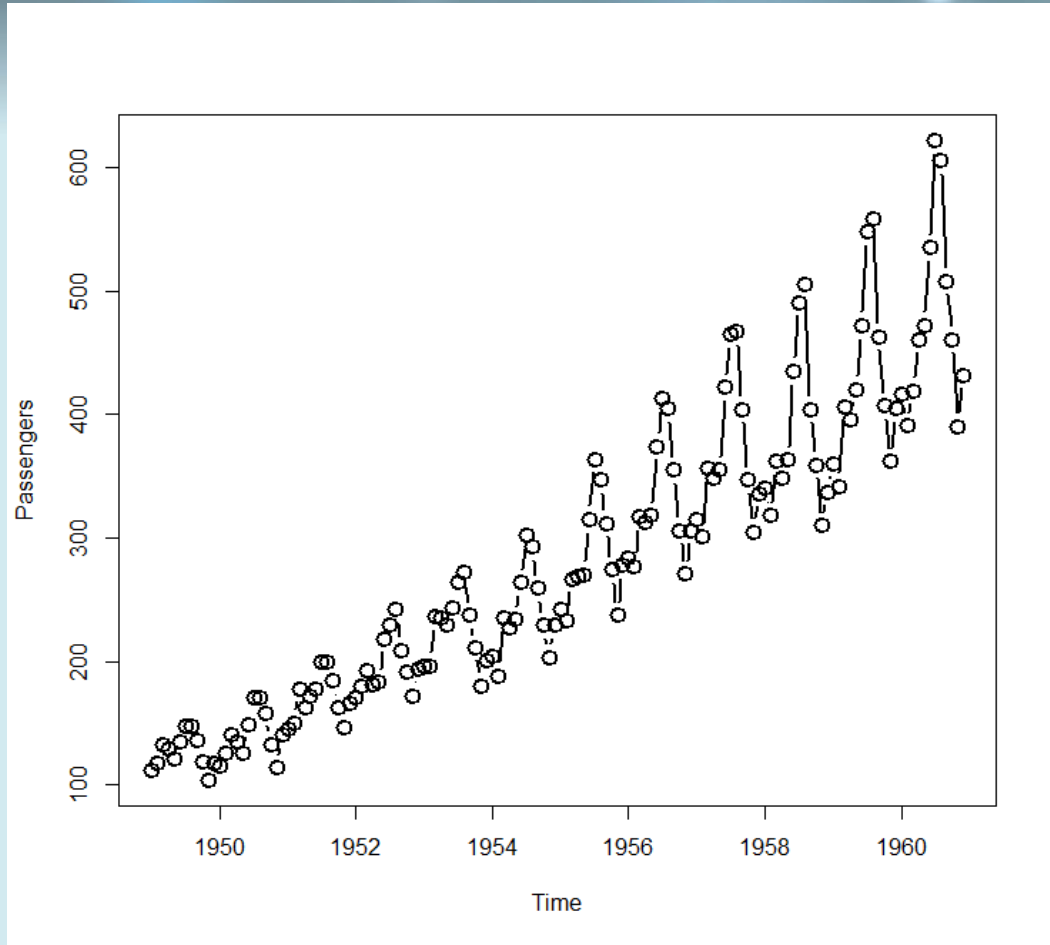
Stationary Processes, cont.

- The beauty of a stationary process is that it can be modelled with relatively few parameters.
- For example, **we do not need a different expectation for each Yt; rather they all have a common expectation, μ. This implies that μ can be estimated accurately by Y .**
- If instead we did not assume stationarity and each Yt had its own unique expectation, μt, then it would not be possible to estimate μt accurately—μt could only be estimated by the single observation Yt itself.
- A time series plot of a stationary series should show random oscillation around some fixed level, a phenomenon called **mean-reversion**. **If the series wanders without returning repeatedly to some fixed level, then the series should not be modeled as a stationary process** (mean reversion is discussed more in this paper: Yarovaya, L., Matkovskyy, R. and Jalan, A., The COVID-19 Black Swan Crisis: Reaction and Recovery of Various Financial Markets (May 27, 2020). Available at SSRN: https://ssrn.com/abstract=3611587 or http://dx.doi.org/10.2139/ssrn.3611587)
- For instance: Inflation rates and changes in inflation rates—time series plots.
- In panel (b), the first differences, that is, the changes from one month to the next, are shown. In contrast to the original series, the differenced series certainly oscillate around a fixed mean that is 0 %, or nearly so.
- The differenced series appears stationary, but whether or not the original series is stationary needs further investigation



Stationary Processes, cont.

- Figure is a plot of monthly total international airline passengers for the years 1949 to 1960.
- The data come from the AirPassengers data set in R's Datasets package.
- There are three types of nonstationarity seen
  - First is the obvious upward trend,
  - Second is the seasonal variation with local peaks in summer and troughs in winter months, and
  - Third is the increase over time in the size of the seasonal oscillations.



Estimating Parameters of a Stationary Process

- **Suppose we observe $Y_1, \ldots, Y_n$ from a weakly stationary process. To estimate the mean μ and variance σ^2 of the process, we can use the sample mean Y and sample variance s^2.**
- The covariance between Yt and its j-th lag, Yt−j, is called the j-th *autocovariance* of the series Yt.
  - **The j-th *autocorrelation coefficient*, also called the *serial correlation coefficient*, measures the correlation between Yt and Yt−j.**
- To estimate the autocovariance function (the sequence of covariances of a stationary process), we use the sample autocovariance function $\hat{\gamma}(h) = n^{-1}\sum_{t=1}^{n-h}(Y_{t+h} - \bar{Y})(Y_t - \bar{Y}) = n^{-1}\sum_{t=h+1}^{n}(Y_t - \bar{Y})(Y_{t-h} - \bar{Y})$.
- This equation is an example of the usefulness of parsimony induced by the stationarity assumption.
- Because the covariance between Y_t and Y_{t+h} does not depend on t, all n − h pairs of data points that are separated by a lag of h time units can be used to estimate $\hat{\gamma}(h)$.
- Some authors define $\hat{\gamma}(h)$ with the factor n−1 in replaced by $(n - h)^{-1}$, but this change has little effect if n is reasonably large and h is small relative to n, as is typically the case.

ACF Plots and the Ljung–Box Test

- To estimate ρ(·), the sample autocorrelation, we use the sample autocorrelation function (sample *ACF*) defined as

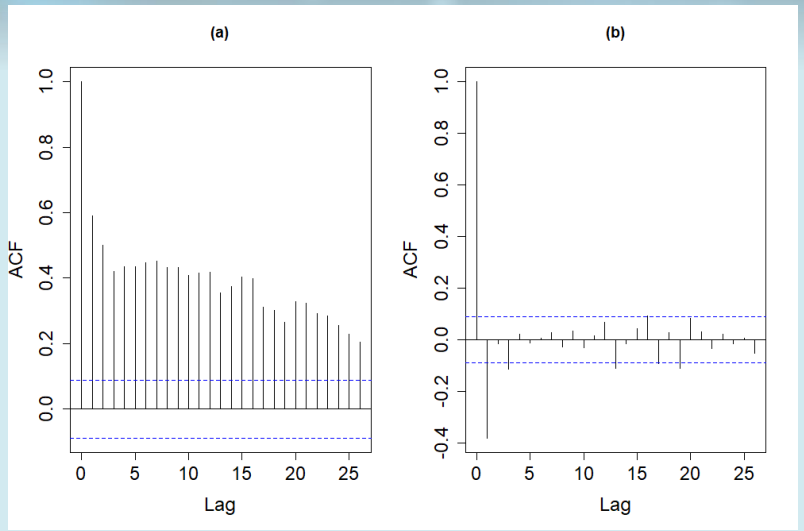
$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} = \frac{\widehat{cov}(Y_t, Y_{t-h})}{\widehat{var}(Y_t)}$$

- Most statistical software will plot a sample ACF with test bounds.
- These bounds are used to test the null hypothesis that an autocorrelation coefficient is 0.
- ***The null hypothesis is rejected if the sample autocorrelation is outside the bounds.***
- The usual level of the test is 0.05, so one can expect to see about 1 out of 20 sample autocorrelations outside the test bounds simply by chance.
- An alternative to using the bounds to test the autocorrelations one at a time is to use a simultaneous test.
- A simultaneous test is one that tests whether a group of null hypotheses are all true versus the alternative that at least one of them is false.
- **The null hypothesis of the Ljung–Box test is H0 : ρ(1) = ρ(2) = · · · = ρ(K) = 0 for some K, say K = 5 or 10. If the Ljung–Box test rejects, then we conclude that one or more of ρ(1), ρ(2), · · · , ρ(K) is nonzero.**
- The Ljung–Box test is sometimes called simply the Box test, though the former name is preferable since the test is based on a joint paper of Ljung and Box.

ACF Plots and the Ljung–Box Test, cont.

- EXAMPLE: Inflation rates and changes in the inflation rate—sample ACF plots and the Ljung–Box test:

```
data(Mishkin, package = "Ecdat")
y = as.vector(Mishkin[,1])
par(mfrow=c(1,2))
acf(y)
acf(diff(y))
```



Sample ACF plots of the one-month inflation rate (a) and changes in the inflation rate (b).

- **In (a) we see that the sample ACF decays to zero slowly. This is a sign of either nonstationarity or possibly of stationarity with long-memory dependence.**
- In contrast, **the sample ACF in (b) decays to zero quickly, indicating clearly that the differenced series is stationary.**
- **Thus, the sample ACF plots agree with the conclusions that the differenced series is stationary and the original series might not be.**
- Several of the autocorrelations of the rate change series fall outside the test bounds, which suggests that the series is not white noise.
- To check, the Ljung–Box test was implemented using R's Box.test() function. K is called lag when Box.test() is called and df in the output, and we specify type="Ljung-Box" for the Ljung–Box test.

```
Box.test(diff(y), lag=10, type="Ljung-Box")
Box-Ljung test

data:  diff(y)
X-squared = 79.915, df = 10, p-value = 5.217e-13
```

- The Ljung–Box test statistic with K = 10 is 79.92, which has an extremely small p-value, **5.217e–13. A significant p-value in this test rejects the null hypothesis that the time series isn't autocorrelated..** Other choices of K give similar results.
- **[In general, what is important here is to keep in mind that p-value < 0.05 lets you reject of the null-hypothesis, but a p-value > 0.05 does not let you confirm the null-hypothesis. In particular, you can not proof the independence of the values of Time Series using the Ljung-Box test. You can only prove the dependence**.]

AR(1) Processes

- Time series models with correlation can be constructed from white noise.
- **The simplest correlated stationary processes are autoregressive processes, where Yt is modeled as a weighted average of past observations plus a white noise "error," which is also called the "noise" or "disturbance." We start with AR(1) processes, the simplest autoregressive processes.**
- Let ϵ_1 , ϵ_2 . . . be weak real valued continuous-time stochastic process $(0,\sigma^2)$. We say that Y_1, Y_2, . . . is an AR(1) process if for some constant parameters μ and ϕ,

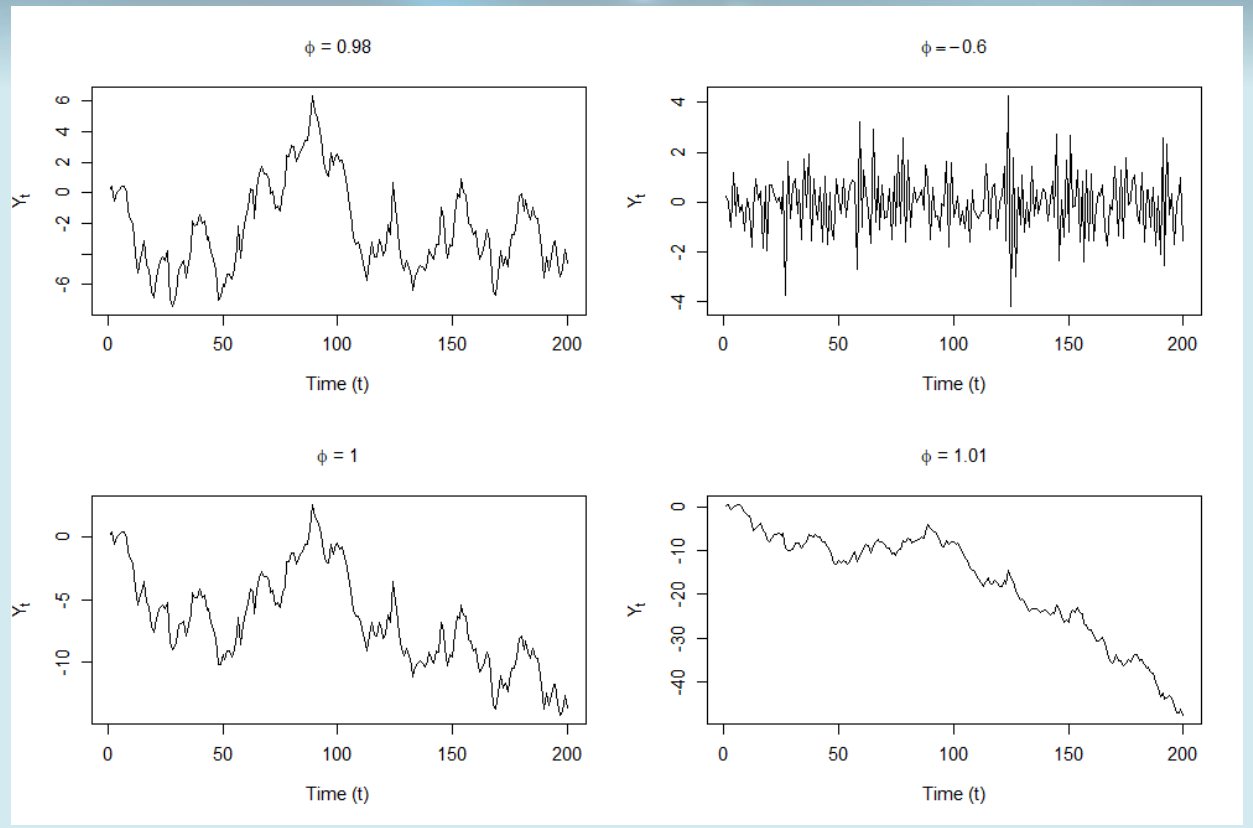
$$Y_t - \mu = \phi(Y_{t-1} - \mu) + \epsilon_t \text{ for all t.}$$

- The parameter μ is the mean of the process, hence $(Y_{t-1} - \mu)$ has mean zero for all t. We may interpret the term $\phi(Y_{t-1} - \mu)$ as representing "memory" or "feedback" of the past into the present value of the process. The process is correlated because the deviation of Y_{t-1} from its mean is fed back into Y_t.
- The parameter ϕ determines the amount of feedback, with a larger absolute value of ϕ resulting in more feedback and $\phi = 0$ implying that $Y_t = \mu + \epsilon_t$.
- **In applications in finance, one can think of ϵ_t as representing the effect of "new information."**
- For example, if Yt is the log return on an asset at time t, then ϵ_t represents the effect on the asset's price of business and economic information that is revealed at time t.
- Information that is truly new cannot be anticipated, so the effects of today's new information should be independent of the effects of yesterday's news. This is why we model new information as white noise.

AR(1) Processes, cont.(optional)

When an AR(1) process is weakly stationary, which implies that $|\varphi| < 1$, then

- $E(Yt) = \mu \ \forall t$,
- $\text{Var}(Yt) = \gamma(0) = \sigma_Y^2$
- $\text{Cov}(Yt, Yt+h) = \gamma(h) \ \forall t$ and $\forall h$, and
- $\text{Corr}(Yt, Yt+h) = \rho(h)$
- If $\varphi = 1$, then $Y_t = Y_{t-1} + \epsilon_t$ for all t. and the process is not stationary (a random walk).
- When $|\phi| > 1$, an AR(1) process has explosive behavior.
- This figure shows simulations of 200 observations from AR(1) processes with various values of φ.
- The explosive case where $\phi = 1.01$ clearly is different from the other cases where $|\varphi| \leq 1$.
- However, the case where $\varphi = 1$ is not that much different from $\varphi = 0.98$ even though the former is nonstationary while the latter is stationary.
- Longer time series would help distinguish between $\varphi = 0.98$ and $\varphi = 1$.



Simulations of 200 *observations from AR*(1) *processes with various values of* φ *and* $\mu = 0$. *The white noise process* $\epsilon_1, \epsilon_2, \ldots, \epsilon_{200}$ *is the same for all four AR*(1) *processes.*

Estimation of AR processes

- R has the function `arima()` for fitting AR and other time series models.
- The function `arima()` and similar functions in other software packages have two primary estimation methods, conditional least-squares and maximum likelihood.
- In this classes, we use the default method in R's `arima()`, which is the MLE with the conditional least-squares estimate used as the starting value for computing the MLE by iterative nonlinear optimization.

Example: Daily returns for BMW stock—ACF plots and AR fit

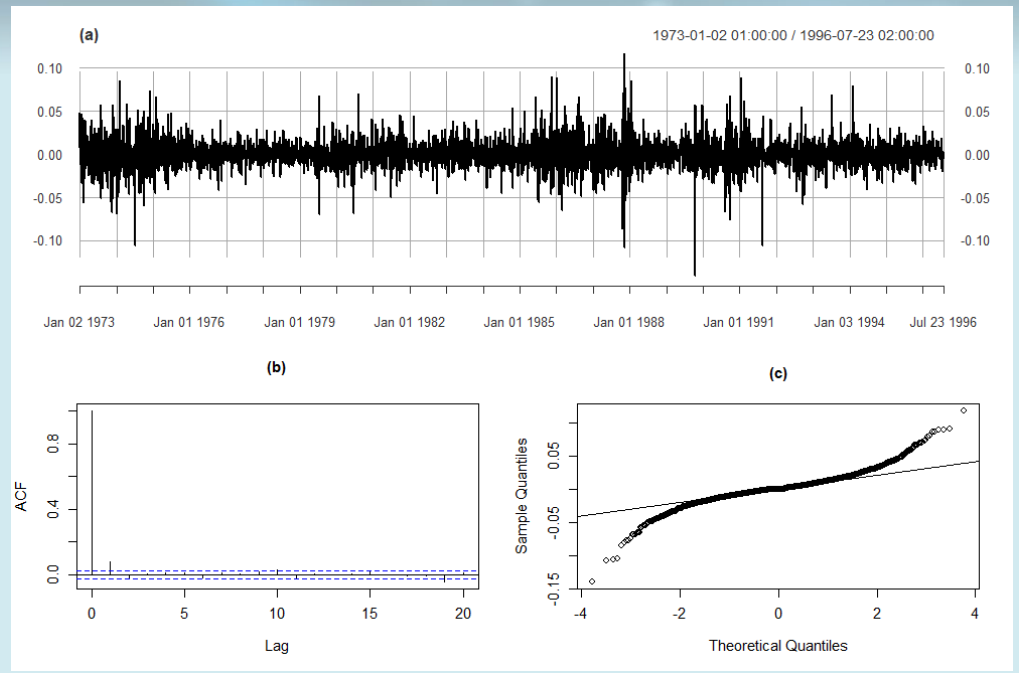
- The daily log returns for BMW stock between January 1973 and July 1996 from the bmw data set in R's evir package are shown in Fig. a. Their sample ACF and quantiles are shown in Fig. b and c, respectively.
- The estimated autocorrelation coefficient at lag 1 is well outside the test bounds, so the series has some dependence. Also, the Ljung–Box test that the first lag autocorrelations are 0 was performed using R's `Box.test()` function.

```
data(bmw, package = "evir")
Box.test(bmw, lag = 5, type = "Ljung-Box")
```

- The parameter lag, which specifies the number of autocorrelation coefficients to test, was set equal to 5, though other choices give similar results. The output was

```
Box-Ljung test
data: bmw
X-squared = 44.987, df = 5, p-value = 1.460e-08
```

- The p-value is very small, indicating that at least one of the first five autocorrelations is nonzero.
- Whether the amount of dependence is of any practical importance is debatable, but an AR(1) model to account for the small amount of autocorrelation might be appropriate.



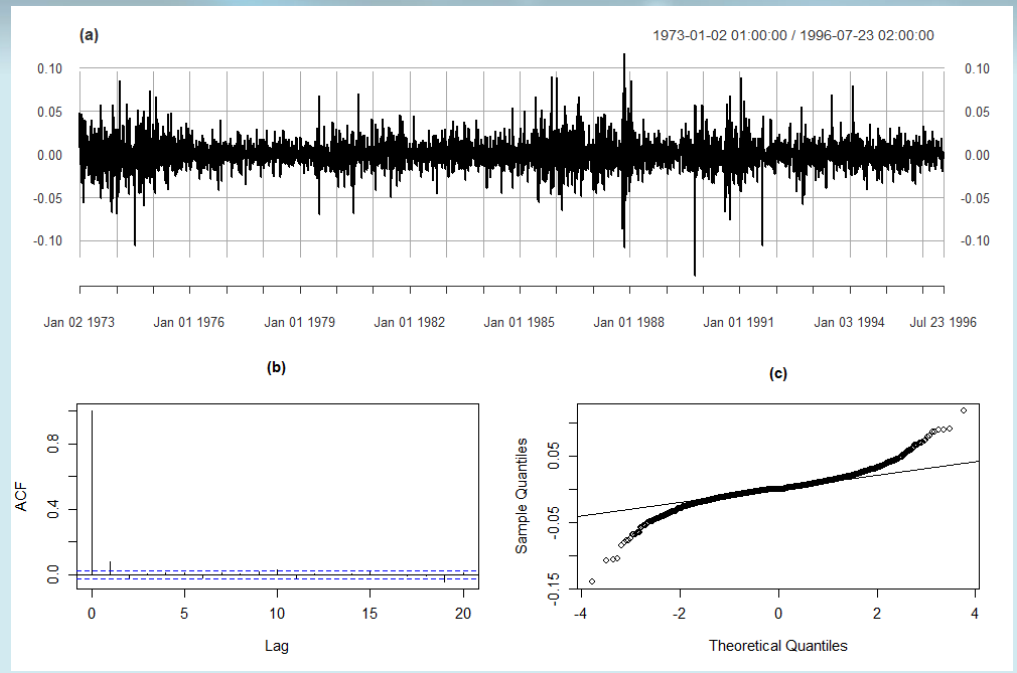
(a) Daily log returns for BMW stock from January 1973 until July 1996, and their (b) sample ACF and (c) sample quantiles relative to the normal distribution.

Example: Daily returns for BMW stock—ACF plots and AR fit

- The order parameter will be explained later, but for an AR(1) process it should be c(1,0,0). A summary of the output is below.

```
fitAR1 = arima(bmw, order = c(1,0,0))
print(fitAR1)
Call:
arima(x = bmw, order = c(1, 0, 0))

Coefficients:
          ar1  intercept
      0.081116   0.000340
s.e.  0.012722   0.000205
sigma^2 estimated as 0.00021626:  log likelihood =
17212.34,  aic = -34418.68
```

- We see that $\hat{\phi} = 0.081$ and $\hat{\sigma}^2 = 0.00022$. Although $\hat{\phi}$ is small, it is statistically significant since it is 6.4 times its standard error 0.013, so its p-value is near zero.
- As just mentioned, whether this small, but nonzero, value of $\hat{\phi}$ is of practical significance is another matter.
- **A non-zero value of ϕ means that there is some information in today's return that could be used for prediction of tomorrow's return, but a small value of ϕ means that the prediction will not be very accurate. The potential for profit might be negated by trading costs**.



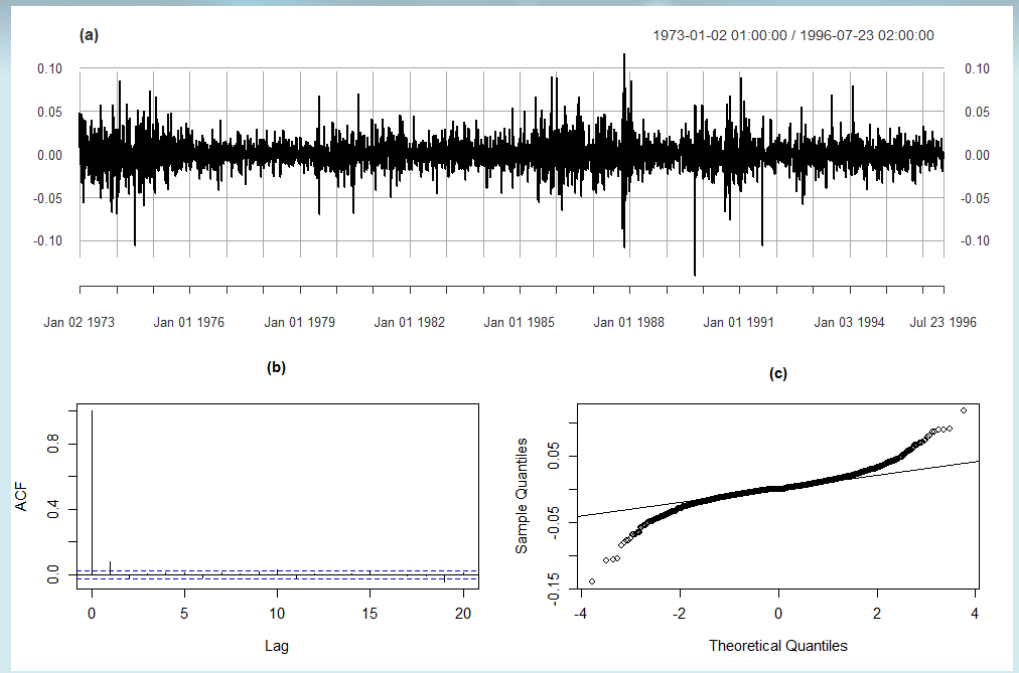
(a) Daily log returns for BMW stock from January 1973 until July 1996, and their (b) sample ACF and (c) sample quantiles relative to the normal distribution.

Example: Daily returns for BMW stock—ACF plots and AR fit

- The sample **ACF of the residuals is plotted in Fig. b**. One of the autocorrelations at low lags are outside the test bounds.
- A few at higher lags are outside the bounds, but this type of behavior is expected to occur by chance or because, with a large sample size, very small but nonzero true correlations can be detected.
- The Ljung–Box test was applied, with lag equal to 5 and fitdf=1. (When the Ljung–Box test is applied to residuals, a correction is needed to account for the use of $\hat{\phi}$ in place of the unknown ϕ. Some software makes this correction automatically. In R the correction is not automatic but is done by setting the fitdf =1 parameter in Box.test())

```
Box.test(residuals(fitAR1), lag = 5, type = "Ljung-
Box", fitdf = 1)
Box-Ljung test
data: residuals(fitAR1)
X-squared = 6.8669, df = 4, p-value = 0.1431
```

- **The large p-value indicates that we should accept the null hypothesis that the residuals are uncorrelated, at least at small lags.**
- This is a sign that the AR(1) model provides an adequate fit.
- However, the Ljung–Box test was repeated with lag equal to 10, 15, and 20 and the p-values were 0.041, 0.045, and 0.004, respectively.
- These values are “statistically significant” using the conventional cutoff of 0.05. The sample size is 6146, so it is not surprising that even a small amount of autocorrelation can be statistically significant. The practical significance of this autocorrelation is very doubtful.



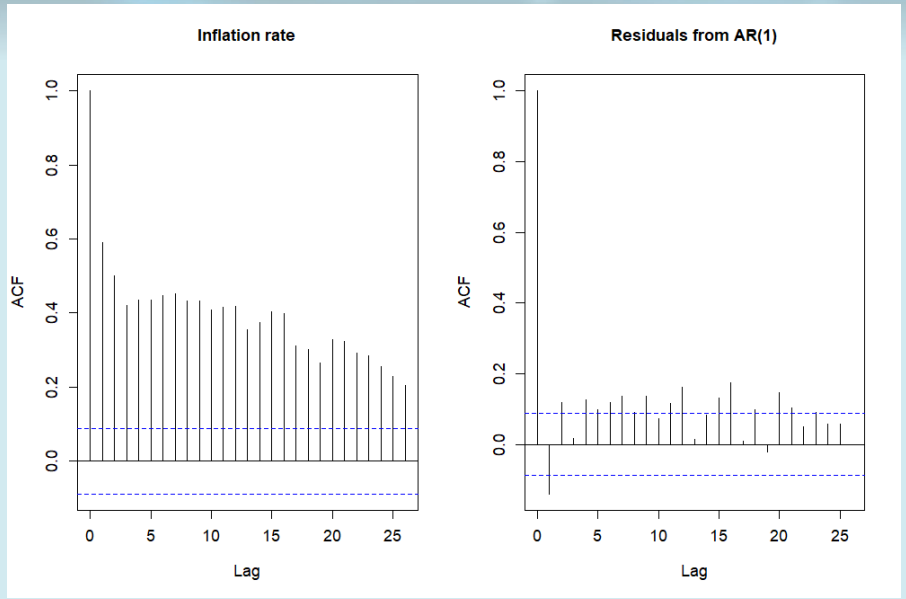
We conclude that the AR(1) model is adequate for the BMW daily returns, but at longer lags some slight amount of autocorrelation appears to remain.
However, the time series plot and normal quantile plot of the AR(1) residuals in Fig. a and c show volatility clustering and heavy tails.

Example: Inflation rate—AR(1) fit and checking residuals

- This example uses the inflation rate time series. Although there is some doubt as to whether this series is stationary, we will fit an AR(1) model.
- The ACF of the residuals are shown in Fig. and there is considerable residual autocorrelation, which indicates that the AR(1) model is not adequate.

```
A Ljung-Box test confirms this result.
data(Mishkin, package = "Ecdat")
y = as.vector(Mishkin[,1])
fit = arima(y, order = c(1,0,0))
Box.test(fit$resid, type = "Ljung", lag = 24,
fitdf = 1)
Box-Ljung test

data:  fit$resid
X-squared = 138.5776, df = 23, p-value < 2.22e-16
```

- One might try fitting an AR(1) to the changes in the inflation rate, since this series is clearly stationary. However, the AR(1) model also does not fit the changes in the inflation rate.




```
par(mfrow=c(1,2))
acf(y,main="Inflation rate")
acf(fit$resid,main="Residuals from AR(1)")
```

The class of ARCH models

- The class of autocorrelated conditional heteroscedastic (ARCH) models was introduced in the seminal paper by Engle (1982).
- ARCH stands for
  - Autoregressive (lags in volatility)
  - Conditional (any new values depend on others)
  - Heteroscedasticity (Greek for varying volatility, here time-varying)
- This type of model has since been modified and extended in several ways.
- The articles by Engle and Bollerslev (1986), Bollerslev et al. (1992), and Bera and Higgins (1993) provide an overview of the model extensions.

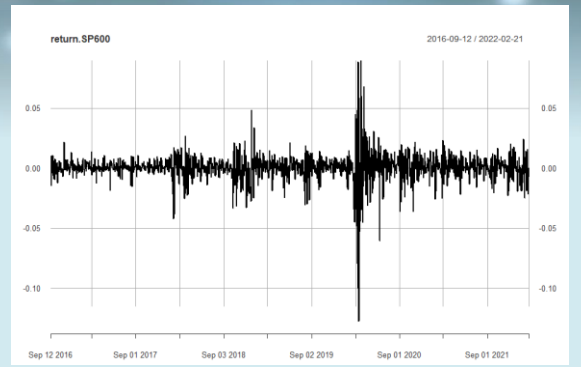
ARCH simplest case – ARCH(1)

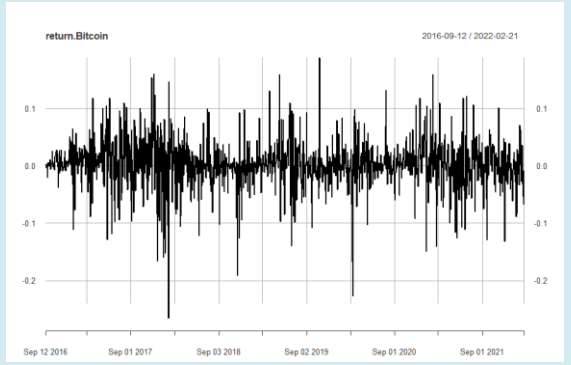
- Typically, **for financial series, the return, does not have a constant conditional variance, and highly volatile periods tend to be clustered together**.
- In other words, there is a strong dependence of sudden bursts of variability in a return on the series own past.

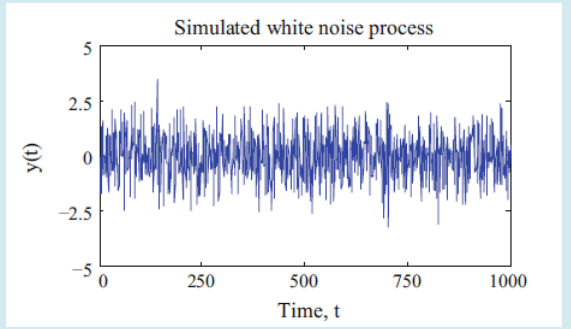
$$r_t = \sigma_t \epsilon_t$$

$$\sigma_t^2 = a_0 + a_1 r_{t-1}^2$$

- Where r_t is return, ϵ_t is standard Gaussian white noise, σ_t^2 is a variance, $a_0 > 0$ and $a_i \geq 0$, $i = 1,\ldots, q$. These parameter restrictions guarantee a positive conditional variance.
- Thus, we can rewrite $r_t | r_{t-1} \sim N(0, a_0 + a_1 r_{t-1}^2)$
- Estimation of the parameters a_0 and a_1 of the ARCH(1) model is typically accomplished by conditional MLE.







GARCH model

- ARCH can be extended to the general ARCH (p) model

$$r_t = \sigma_t \epsilon_t$$
$$\sigma_t^2 = a_0 + a_1 r_{t-1}^2 + \cdots + a_p r_{t-p}^2$$

- Bollerslev developed a GARCH(1, 1) model

$$r_t = \sigma_t \epsilon_t$$
$$\sigma_t^2 = a_0 + a_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

- That can be generalized to GARCH(*p*,*q*)

$$r_t = \sigma_t \epsilon_t$$
$$\sigma_t^2 = a_0 + \sum_{j=1}^{p} a_j r_{t-j}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$$

GARCH (1,1) – example in r

```
library (tseries)
mp <- garch(Y, order =c(1,1))
summary (mp)

> summary (bitcoin.garch.1)

Call:
garch(x = Y, order = c(1, 1))

Model:
GARCH(1,1)

Residuals:
     Min       1Q   Median       3Q      Max
-4.98558 -0.34828  0.06712  0.58893  5.04846

Coefficient(s):
    Estimate  Std. Error  t value Pr(>|t|)
a0 1.084e-04   1.336e-05    8.111 4.44e-16 ***
a1 1.192e-01   1.207e-02    9.869  < 2e-16 ***
b1 8.252e-01   1.667e-02   49.490  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Diagnostic Tests:
             Jarque Bera Test

data:  Residuals
X-squared = 844.3, df = 2, p-value < 2.2e-16


             Box-Ljung test

data:  Squared.Residuals
X-squared = 0.20427, df = 1, p-value = 0.6513
```

- Ideally all p-values are above 0.05.
- the **Jarque Bera Test** for the null hypothesis that the residuals are normally distributed Tests.
- **The Jarque-Bera test statistic is always positive, and if it is not close to zero, it shows that the sample data do not have a normal distribution**
- The **Ljung-Box test** provides a means of testing for auto-correlation within the GARCH model's standardized residuals.
- If the GARCH model has done its job **there should be NO auto-correlation within the residuals.**
  - The null-hypothesis of the Ljung-Box test is that the auto-correlation between the residuals for a set of lags k = 0. If at least one auto-correlation for a set of lags k > 0 then the test statistic indicates that the null-hypothesis may be rejected.
  - **Should the p-value be <= 0.05 (your significance level \alpha) then the null hypothesis should be rejected meaning that the GARCH model has not captured the auto-correlation**.
- In some packages the ARCH LM test is present.
- Similar to the Ljung-Box Test, the ARCH LM test provides a means of testing for serial dependence (auto-correlation) due to a conditional variance process by testing for auto-correlation within the squared residuals. The null hypothesis is that the auto-correlation between the residuals for a set of lags k = 0.
  - Autocorrelation and autoregressive conditional heteroskedasticity are not the same. You can have one, the other or both in a time series. Hence, you should not be surprised if some tests find presence of one but not the other.

GARCH (1,1) – example in r

```
library (rugarch)
bitcoin.garch.spec =
ugarchspec(mean.model=list(armaOrder=c(0,0)),
distribution="norm")
bitcoin.garch.2 = ugarchfit(bitcoin.garch.spec, Y)
summary(bitcoin.garch.2)
```

$$(r_t - \text{mu}) = \text{ar1}(r_t - \text{mu}) + r_t + \text{ma1} r_{t-1}$$
$$r_t = \sigma_t \epsilon_t$$
$$\sigma_t^2 = \text{omega} + \text{alpha1} r_{t-1}^2 + \text{beta1} \sigma_{t-1}^2$$
$$\epsilon_t \sim i.i.d.(0, 1, shape)$$

```
Optimal Parameters
------------------------------------
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.002673    0.000990   2.7001 0.006932
omega   0.000105    0.000027   3.8276 0.000129
alpha1  0.115958    0.022723   5.1030 0.000000
beta1   0.829969    0.032217  25.7618 0.000000

Robust Standard Errors:
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.002673    0.001212   2.2064 0.027356
omega   0.000105    0.000060   1.7400 0.081856
alpha1  0.115958    0.043930   2.6396 0.008301
beta1   0.829969    0.065940  12.5868 0.000000
```

The robust standard errors are based on the method of White (1982) which produces asymptotically valid condence intervals.

GARCH (1,1) – example in r

```
Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic p-value
Lag[1]                     0.1692  0.6808
Lag[2*(p+q)+(p+q)-1][5]    2.4494  0.5167
Lag[4*(p+q)+(p+q)-1][9]    4.7008  0.4746
d.o.f=2

Weighted ARCH LM Tests
------------------------------------
            Statistic Shape Scale P-Value
ARCH Lag[3]     1.528 0.500 2.000  0.2164
ARCH Lag[5]     2.949 1.440 1.667  0.2973
ARCH Lag[7]     4.211 2.315 1.543  0.3171

Nyblom stability test
------------------------------------
Joint Statistic:  1.5699
Individual Statistics:
mu     0.4303
omega  0.2265
alpha1 0.1190
beta1  0.1266

Asymptotic Critical Values (10% 5% 1%)
Joint Statistic:               1.07 1.24 1.6
Individual Statistic:          0.35 0.47 0.75

Sign Bias Test
------------------------------------
                   t-value   prob sig
Sign Bias           0.6863 0.4926
Negative Sign Bias  0.9960 0.3194
Positive Sign Bias  0.2924 0.7700
Joint Effect        1.3808 0.7100


Adjusted Pearson Goodness-of-Fit Test:
------------------------------------
  group statistic p-value(g-1)
1    20     197.1    1.301e-31
2    30     220.3    2.614e-31
3    40     233.1    1.738e-29
4    50     255.4    1.076e-29
```

- The ARCH-LM test is a weighted portmanteau test for testing the null hypothesis of adequately fitted ARCH process, whilst the Ljung-Box is another portmanteau test with null the adequacy of the ARMA fit.
- The `signbias` calculates the Sign Bias Test of Engle & Ng (1993), and is also displayed in the summary.
- This tests the presence of leverage effects in the standardized residuals (to capture possible misspecification of the GARCH model), by regressing the squared standardized residuals on lagged negative and positive shocks
- The Nyblom stability test (Nyblom 1989) provides a means of testing for structural change within a time series. A structural change implies that the relationship between variables changes overtime e.g. for the regression $y_t = x_t'\beta$ beta changes over time. The null hypothesis is that the parameter values are constant i.e. zero variance, the alternative hypothesis is that their variance > 0.
- The gof calculates the chi-squared goodness of fit test, which compares the empirical distribution of the standardized residuals with the theoretical ones from the chosen density. The implementation is based on the test of Palm (1996) which adjusts the tests in the presence on non-i.i.d. observations by reclassifying the standardized residuals not according to their value (as in the standard test), but instead on their magnitude, calculating the probability of observing a value smaller than the standardized residual, which should be identically standard uniform distributed.

ARIMA(p_M, d, q_M)+GARCH(p_V , q_V)

- GARCH models permit a wider range of behavior, in particular, more persistent volatile ty. The GARCH(p, q) model is

$$r_t = \sigma_t \epsilon_t$$

$$\sigma_t = \sqrt{a_0 + \sum_{j=1}^{p} a_j r_{t-j}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2}$$

$$\epsilon_t \sim i.i.d.(0, 1, shape)$$

- **Because past values of the σ_t process are fed back into the present value (with nonnegative coefficients βj), the conditional standard deviation can exhibit more persistent periods of high or low volatility than seen in an ARCH process.**
- In the stationary case, the process at is uncorrelated with a constant unconditional mean and variance and r_t^2 has an ACF like an ARMA process
- A very general time series model lets at be GARCH(pV , qV) and uses at as the noise term in an ARIMA(pM, d, qM) model.
- The subscripts on p and q distinguish between the conditional variance (V) or GARCH parameters and the conditional mean (M) or ARIMA parameters.
- We will call such a process an **ARIMA(pM, d, qM)+GARCH(pV , qV)** model.
- When one fits an ARIMA(pM, d, qM)+GARCH(pV , qV) model to a time series Yt, **there are two types of residuals**.
- The ordinary residual, denoted $\hat{r}_t$, is the **difference between Yt and its conditional expectation**.
- As the notation implies, $\hat{r}_t$ estimates r_t.
- A standardized residual, denoted $\hat{\epsilon}_t$, is an ordinary residual $\hat{r}_t$ divided by its estimated conditional standard deviation $\hat{\sigma}_t$.
- A standardized residual estimates ϵ_t. The standardized residuals should be used for model checking.
- If the model fits well, then neither $\hat{\epsilon}_t$ nor $\hat{\epsilon}_t^2$ should exhibit serial correlation.
- Moreover, if ϵ_t has been assumed to have a normal distribution, then this assumption can be checked by a normal plot of the standardized residuals $\hat{\epsilon}_t$.

Fitting ARMA+GARCH Models

- This example uses the daily BMW stock log returns.
- The ugarchfit() function from R's rugarch package is used to fit an AR(1)+GARCH(1,1) model to this series.
- Although ugarchfit() allows the white noise to have a nonGaussian distribution, we begin this example using Gaussian white noise (the default).
- First the model is specified using the ugarchspec() function;
- for an AR(1)+GARCH(1,1) model we specify armaOrder=c(1,0) and
- garchOrder=c(1,1).
- The commands and abbreviated output are the following:

```
library(rugarch)
data(bmw, package="evir")
arma.garch.norm = ugarchspec(mean.model=list(armaOrder=c(1,0)),
variance.model=list(garchOrder=c(1,1)))
bmw.garch.norm = ugarchfit(data=bmw, spec=arma.garch.norm)

show(bmw.garch.norm)
GARCH Model : sGARCH(1,1)
Mean Model : ARFIMA(1,0,0)
Distribution : norm
Optimal Parameters
------------------------------------
          Estimate Std. Error t value Pr(>|t|)
mu         0.000453 0.000175 2.5938 0.009493
ar1        0.098135 0.014261 6.8813 0.000000
omega      0.000009 0.000000 23.0613 0.000000
alpha1     0.099399 0.005593 17.7730 0.000000
beta1      0.863672 0.006283 137.4591 0.000000
LogLikelihood : 17752

Information Criteria
------------------------------------
Akaike -5.7751
Bayes -5.7696
Shibata -5.7751
Hannan-Quinn -5.7732
```

Fitting ARMA+GARCH Models

- In the output, $\varphi 1$ is denoted by ar1, the estimated mean μ is mean, and α_0 is called omega (in some formulas ω *is used instead)*
- Note that **$\varphi 1 = 0.0981$** and is statistically significant, **implying that there is a small amount of positive autocorrelation.**
- Both **$\alpha 1$** and **$\beta 1$** are highly significant and **$\beta 1 = 0.8636$**, **which implies rather persistent volatility clustering.**
- There are two additional information criteria reported, Shibata's information criterion and Hannan–Quinn information criterion (HQIC).
  - These are less widely used than AIC and BIC and will not be discussed here.
- In the output from ugarchfit(), the AIC and BIC values have been normalized by dividing by *n*, so these values should be multiplied by $n = 6146$ to have their usual values.
- In particular, AIC and BIC will not be so close to each other after multiplication by 6146 (number of observations).
- The daily BMW stock log return series *Yt*, with two estimated conditional standard deviations superimposed, and the estimated conditional standard deviation series σt

```
library(rugarch)
data(bmw, package="evir")
arma.garch.norm = ugarchspec(mean.model=list(armaOrder=c(1,0)),
variance.model=list(garchOrder=c(1,1)))
bmw.garch.norm = ugarchfit(data=bmw, spec=arma.garch.norm)

show(bmw.garch.norm)
GARCH Model  : sGARCH(1,1)
Mean Model   : ARFIMA(1,0,0)
Distribution : norm
Optimal Parameters
------------------------------------

        Estimate  Std. Error  t value    Pr(>|t|)
mu      0.000453  0.000175     2.5938    0.009493
ar1     0.098135  0.014261     6.8813    0.000000
omega   0.000009  0.000000    23.0613    0.000000
alpha1  0.099399  0.005593    17.7730    0.000000
beta1   0.863672  0.006283   137.4591    0.000000
LogLikelihood : 17752

Information Criteria
------------------------------------
Akaike -5.7751
Bayes -5.7696
Shibata -5.7751
Hannan-Quinn -5.7732
```

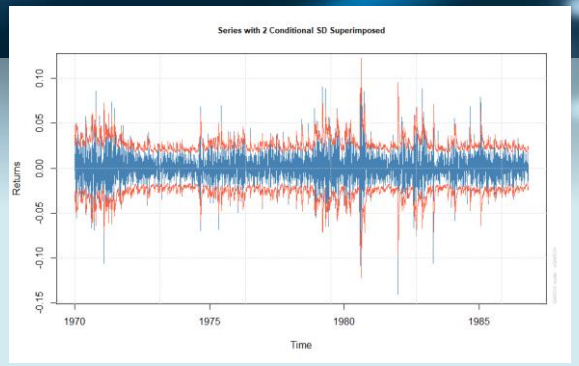
Fitting ARMA+GARCH Models

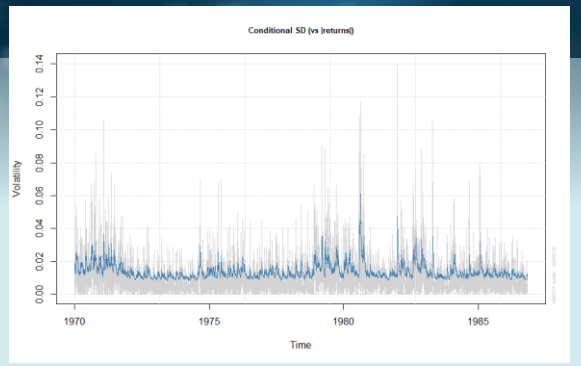
- The output also includes the following tests applied to the standardized and squared standardized residuals.
- **Weighted versions of the Ljung-Box (and ARCH-LM) test statistics and their approximate p-values all indicate that the estimated model for the conditional mean and variance are adequate for removing serial correlation from the series and squared series, respectively.**
- The Goodness-of-Fit tests compare the empirical distribution of the standardized residuals with the theoretical ones from the specified density, which is Gaussian by default. The small p-values strongly reject the null hypothesis that the white noise standardized innovation process $\hat{\epsilon}_t$ is Gaussian.

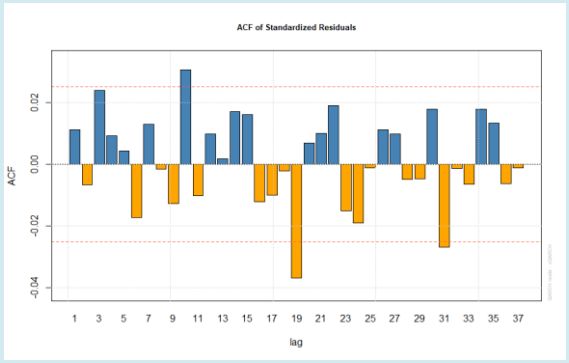
```
Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic p-value
Lag[1]                     0.7790  0.3775
Lag[2*(p+q)+(p+q)-1][2]    0.9161  0.7890
Lag[4*(p+q)+(p+q)-1][5]    3.3274  0.3536
d.o.f=1
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic p-value
Lag[1]                     0.2769  0.5988
Lag[2*(p+q)+(p+q)-1][5]    1.0262  0.8537
Lag[4*(p+q)+(p+q)-1][9]    1.7209  0.9356
d.o.f=2

Weighted ARCH LM Tests
------------------------------------
            Statistic Shape Scale P-Value
ARCH Lag[3]    0.1922 0.500 2.000  0.6611
ARCH Lag[5]    1.1094 1.440 1.667  0.7008
ARCH Lag[7]    1.2290 2.315 1.543  0.8737

Adjusted Pearson Goodness-of-Fit Test:
------------------------------------
  group statistic p-value(g-1)
1    20     493.2    1.477e-92
2    30     513.4    4.975e-90
3    40     559.3    2.484e-93
4    50     585.5    5.611e-93
```

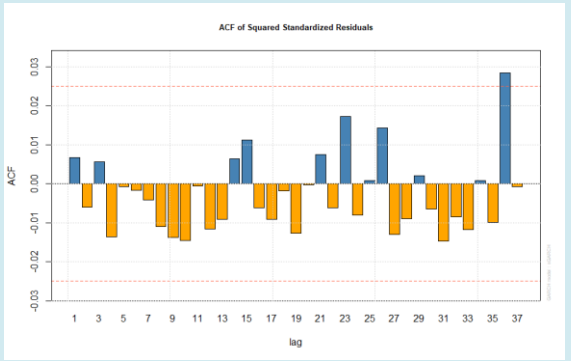
Fitting ARMA+GARCH Models

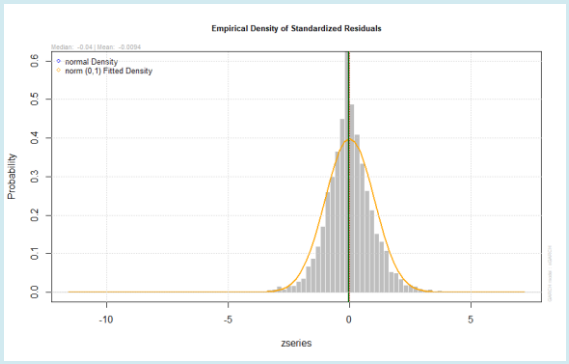
- The daily BMW stock log return series Yt, with two estimated conditional standard deviations superimposed;
- the estimated conditional standard deviation σt series (vs. the absolute value of the log return series |Yt|);
- the sample ACF of the standardized residuals $\hat{\epsilon}_t$ and the squared standardized residuals $\hat{\epsilon}_t^2$;
- empirical density estimates of the standardized residuals $\hat{\epsilon}_t$;
- and a normal quantile plot of the standardized residuals $\hat{\epsilon}_t$.
- The figures indicate that potentially t-distribution can be fit
- Let's estimate with the student distribution

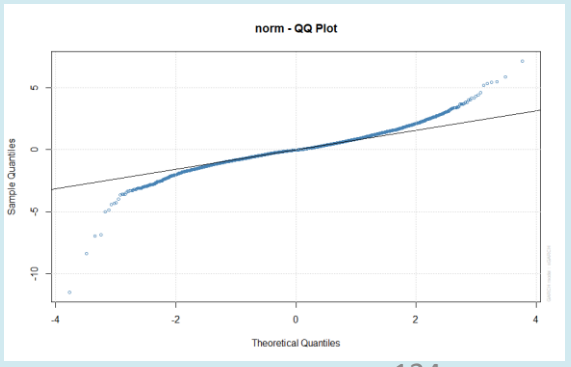












Fitting ARMA+GARCH Models

```
arma.garch.t = ugarchspec(mean.model=list(armaOrder=c(1,0)),
variance.model=list(garchOrder=c(1,1)), distribution.model =
"std")
bmw.garch.t = ugarchfit(data=bmw,spec=arma.garch.t)
show(bmw.garch.t)
*---------------------------------*
*          GARCH Model Fit        *
*---------------------------------*

Conditional Variance Dynamics
-----------------------------------
GARCH Model     : sGARCH(1,1)
Mean Model      : ARFIMA(1,0,0)
Distribution    : std

Optimal Parameters
------------------------------------
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.000135    0.000144   0.9398 0.347322
ar1     0.063910    0.012521   5.1043 0.000000
omega   0.000006    0.000003   1.6996 0.089211
alpha1  0.090592    0.012477   7.2607 0.000000
beta1   0.889886    0.014634  60.8108 0.000000
shape   4.070208    0.301286  13.5094 0.000000

Robust Standard Errors:
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.000135    0.000144  0.93980 0.347318
ar1     0.063910    0.014543  4.39441 0.000011
omega   0.000006    0.000017  0.34927 0.726885
alpha1  0.090592    0.044379  2.04133 0.041218
beta1   0.889886    0.056238 15.82362 0.000000
shape   4.070208    1.040138  3.91314 0.000091
LogLikelihood : 18151.53
```

The weighted Ljung–Box tests for the residuals have small *p*-values. These are due to small autocorrelations that should not be of practical importance (the sample size here is 6146 so, not surprisingly, small autocorrelations are statistically significant).

```
Information Criteria
------------------------------------
```

both AIC and BIC increased a bit;

```
Akaike       -5.9048
Bayes        -5.8983
Shibata      -5.9048
Hannan-Quinn -5.9026

Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic   p-value
Lag[1]                      9.641 1.903e-03
Lag[2*(p+q)+(p+q)-1][2]     9.654 3.362e-09
Lag[4*(p+q)+(p+q)-1][5]    11.984 1.454e-04
d.o.f=1
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared
Residuals
------------------------------------
                        statistic p-value
Lag[1]                      0.564  0.4527
Lag[2*(p+q)+(p+q)-1][5]     1.296  0.7898
Lag[4*(p+q)+(p+q)-1][9]     2.015  0.9032
d.o.f=2

Weighted ARCH LM Tests
------------------------------------
            Statistic Shape Scale P-Value
ARCH Lag[3]    0.3563 0.500 2.000  0.5506
ARCH Lag[5]    1.1291 1.440 1.667  0.6950
ARCH Lag[7]    1.2436 2.315 1.543  0.8710
```

```
Nyblom stability test
------------------------------------
Joint Statistic:  2.6328
Individual Statistics:
mu      0.07989
ar1     0.07605
omega   0.29147
alpha1  0.23024
beta1   0.23009
shape   0.23577

Asymptotic Critical Values (10% 5% 1%)
Joint Statistic:                 1.49 1.68 2.12
Individual Statistic:            0.35 0.47 0.75

Sign Bias Test
------------------------------------
                   t-value   prob sig
Sign Bias          0.07924 0.9368
Negative Sign Bias 1.58367 0.1133
Positive Sign Bias 0.58050 0.5616
Joint Effect       5.06157 0.1673


Adjusted Pearson Goodness-of-Fit Test:
------------------------------------
  group statistic p-value(g-1)
1    20     229.0    5.460e-38
2    30     279.6    8.428e-43
3    40     313.8    1.230e-44
4    50     374.6    1.037e-51

Elapsed time : 0.9029431
```

The goodness-of-fit test statistics are much smaller but still significant; the large sample size again makes rejection likely even when the discrepancies are negligible from a practical standpoint.

The refit model with a *t* conditional distribution does not offers a significant improvement over the original fit with a Gaussian conditional distribution.

APARCH Models

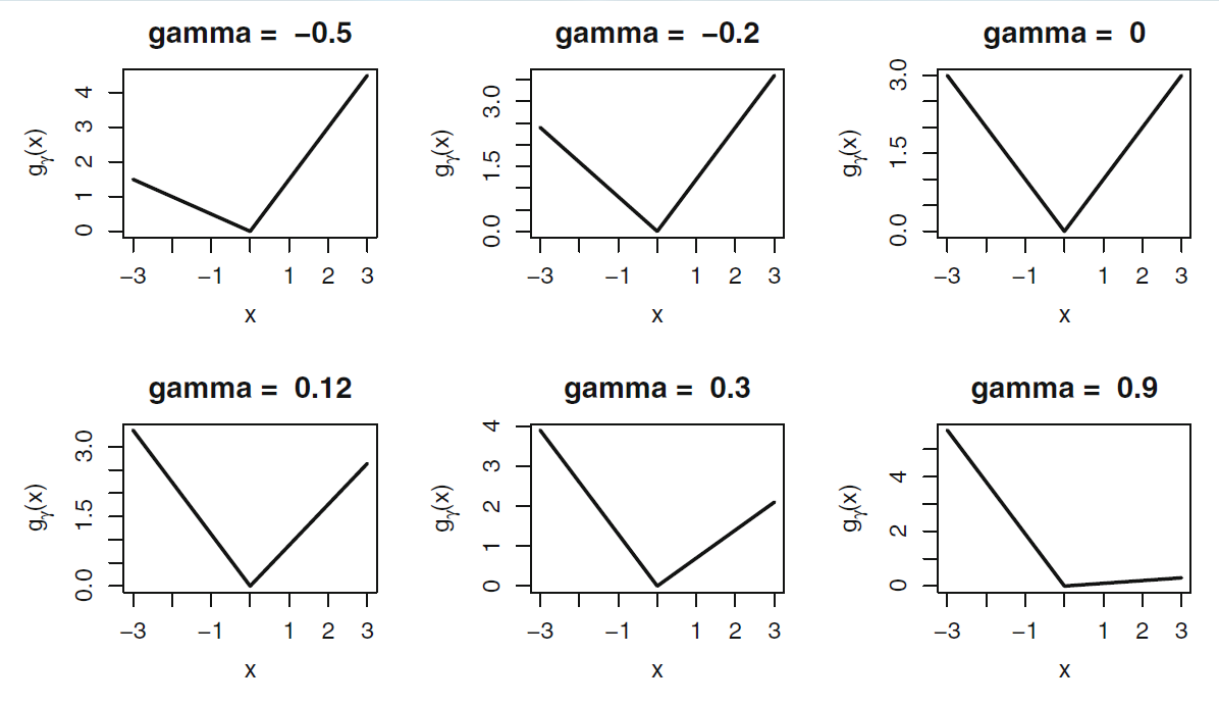
- **In some financial time series, large negative returns appear to increase volatility more than do positive returns of the same magnitude.**
- This is called the leverage effect.
- Standard GARCH models cannot model the leverage effect because they model σ_t as a function of past values of r_t^2—whether the past values of r_t are positive or negative is not taken into account.
  - The problem here is that the square function is symmetric.
- The solution is to replace the square function with a flexible class of nonnegative functions that include asymmetric functions.
- The APARCH (asymmetric power ARCH) models do this.
- They also offer more flexibility than GARCH models by modeling σ_t^δ where $\delta > 0$ is another parameter.
- The APARCH(p, q) model for the conditional standard deviation is

$$\sigma_t^\delta = \omega + \sum_{i=1}^{p} r_i(|r_{i-i}| - \gamma_i r_{i-i})^\delta + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^\delta$$

- where $\delta > 0$ and $-1 < \gamma_i < 1$, i = 1, . . . , p. Note that $\delta = 2$ and $\gamma = \cdots = \gamma_i = 0$ give a standard GARCH model.

APARCH Models, cont

- The effect of r_{i-i} upon σ_t is through the $|r_{i-i}| - \gamma_i r_{i-i}$
- AR(1)+APARCH(1,1) fit to daily BMW stock log returns



Plots of $|r_{i-i}| - \gamma_i r_{i-i}$ for various values of γ.

APARCH Model - BMW

```
arma.aparch.t = ugarchspec(mean.model=list(armaOrder=c(1,0)),
variance.model=list(model="apARCH",
garchOrder=c(1,1)))

bmw.aparch.t = ugarchfit(data=bmw, spec=arma.aparch.t)
show(bmw.aparch.t)

Conditional Variance Dynamics
-----------------------------------
GARCH Model      : apARCH(1,1)
Mean Model       : ARFIMA(1,0,0)
Distribution     : norm

Optimal Parameters
------------------------------------
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.000244    0.000179   1.3671 0.171602
ar1     0.095180    0.014000   6.7987 0.000000
omega   0.000037    0.000019   1.8816 0.059885
alpha1  0.085311    0.010162   8.3951 0.000000
beta1   0.901068    0.012806  70.3646 0.000000
gamma1  0.197643    0.038233   5.1695 0.000000
delta   1.569984    0.118734  13.2227 0.000000

Robust Standard Errors:
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.000244    0.000189  1.29474 0.195411
ar1     0.095180    0.014520  6.55502 0.000000
omega   0.000037    0.000049  0.74194 0.458126
alpha1  0.085311    0.031354  2.72092 0.006510
beta1   0.901068    0.037824 23.82267 0.000000
gamma1  0.197643    0.079769  2.47769 0.013223
delta   1.569984    0.239676  6.55045 0.000000

LogLikelihood : 17774.24

Information Criteria
------------------------------------

Akaike       -5.7817
Bayes        -5.7741
Shibata      -5.7817
Hannan-Quinn -5.7791
```

The estimate of δ is 1.57 with a standard error of 0.11, so there is strong evidence that δ is not 2, the value under a standard GARCH model.

Also, γ_1 is 0.2 with a standard error of 0.038, so there is a statistically significant leverage effect, since we reject the null hypothesis that $\gamma_1 = 0$.

However, the leverage effect is not too strong, as can be seen in the plot (plot (bmw.aparch.t))