1 INTRODUCTION TO FINANCIAL ENGINEERING

aferreiro@mat.uab.cat
ferreiroalb@bancsabadell.cat

MODULE HANDBOOK

Module covers an Introduction to Finance and the statistical methods used in computational finance. Is a mixture of:

i) Computation (mainly in R) & Numerical Algorithms.
ii) Finance
iii) Maths & Stats

Lectures → 2h / week (more like seminars)
Computer Lab → 1h / week

Assigments (course work) aprox. 1 per week.

Evaluation

$$\max\left\{ \underbrace{(0.7x+0.3y)\mathbb{1}_{\{x\geq 4\}}}_{:=w},\ (0.7x+0.3z)\mathbb{1}_{\{w<5\}} \right.$$

x = Assignments
y = Final Exam
z = Recovery Exam.

KEY TOPICS COVERED

① Introduction to math Finance
↳ Interest Rates, Capitalization, Discount Factors, ...
↳ Financial Markets & Products.

② Time Series : Macroeconomic series.
↳ Review on ARMA, ARCH and GARCH models.

③ Stochastic Calculus : Valuation of financial derivatives.
↳ Introduction to Financial Derivatives and Fair Value
↳ Discrete Models
↳ The Continuous Models : Brownian Motion
↳ Monte Carlo Methods.

④ Mathematical Optimization : Portfolio Management
↳ Modern Portfolio Management
↳ Portfolio Optimization : Lagrange Multipliers.
↳ CAPM

⑤ Probability : Risk Estimation
↳ Typology of Risks
↳ VaR Calculations

⑥ Financial Disasters : Lessons.

The main goal is to show the student the different applications of mathematical and statistical concepts in financial engineering, focusing on the proper use and result interpretation (.... unfortunately we will not prove theorems, on the other hand we will see the many places a professional mathematician can work in the financial industry ... !!!)

Time Series.
Macroeconomic forecast: GDP, Employment Rate, ... House prices.

Stochastic Calculus:
Pricing derivatives in a Front Office, Trading Activities, Market Brokers, ...

Mathematical Optimization
Portfolio Optimization and management, Fund mangement

Probability
Risk estimation, 3rd line of defence activities, Regulatory Activity

Financial Disasters
Sometimes bad practices, miss knowledge and greed lead to bad ends....

MODULE WORKFLOW

Reading books references will be given during the lecture ~~nots.~~ sessions. Therefore a set of not comprehensive lecture nots will be available. It is expected that you will need to consult these and other reference to complete coursework

ASSIGNMENTS : RULES & SUBMISSION

* You may work in cooperation, but the solutions you submit must be entirely your own work.
* You must give citations for all sources that you use.
* All programs should be written in R (if not otherwise stated).
* It is not acceptable to copy, modify or adapt someone else's code and submit it as your own.
* Submissions must be in electronic form (Moodle)
* The submission must be the source code (.r) and the answer sheet (.pdf)

* It is not acceptable to submit files with other extension or in zip format.

* You should keep a copy of everything you submit

* It is not acceptable to submit assignments after the deadline.

* Fail in any of the above rules might lower you assignment score.

FINAL REMARKS

Your code should compile without errors and should be properly commented.

The accompanying write-up should be written as a report and should demonstrate that you understand the underlying theory.

DEFINITION OF FINANCE

Finance is the science underlying the answer to the following question:

$\boxed{\text{How much is } 1\,€ \text{ worth in one year time?}}$

Finance := Time Value of Money

Money changes value because:

- Oportunity cost
- Counterparty risk
- Associated costs
- ......

FAIR VALUE

FINANCE $\simeq$ GAME OF CHANCE

Finance has a lot of connections with game theory and thus probability and stochastic calculus have become one of the main tools for professionals in financial markets.

Imagine you are faced with an investment oportunity. An investment of EUR 1.000 today can go great and get EUR 1.500 in one year time or go bad and recover EUR 500. The two outcomes are equally probable.
Would you invest?

- Is the investment fair?
- What are the interest rate in 1 year?
- Is there an oportunity cost?

PRODUCTS & MARKETS

EQUITIES :

The most basic financial instrument also known as stock or share.

This is the ownership of a small piece of a company. Companies issue share to rise capital in exchange of ownership and hence rights to benefits or dividends.

COMMODITIES :

Usually raw products such as precious metals, oil, food, ...

Prices of products often show seasonal effects and most of the trading activity is done by means of future contracts

CURRENCIES :

Exchange rates, foreign exchange or FX markets are the markets that exchange one currency for another.

This market is an active market for arbitrageurs since the market "should" show consistency :

$$\frac{EUR}{GBP} \times \frac{GBP}{USD} \simeq \frac{EUR}{USD}$$

INDICES:

Are means of measuring how the stock market / economy is doing as a whole. A typical index is made of a basket of representative stocks or products (IBEX 35, S&P 500, NASDAQ, ...)

You can trade on indices aswell.

FIXED - INCOME SECURITIES:

Securities that pay a fixed income over the life of the financial instrument. Also known as bonds, buying a bond means lending money to the counterpart who has the obligation to pay back the principal and interest stipulated in the bond's statement.

FORWARDS & FUTURES:

A forward contract is an agreement where one party promises to buy an asset from another party at some specified time in the future. No money changes hands until the delivery date or maturity.

Future contracts are similar to forward contracts but they trade through an exchange market in a standarized form.

DERIVATIVES :

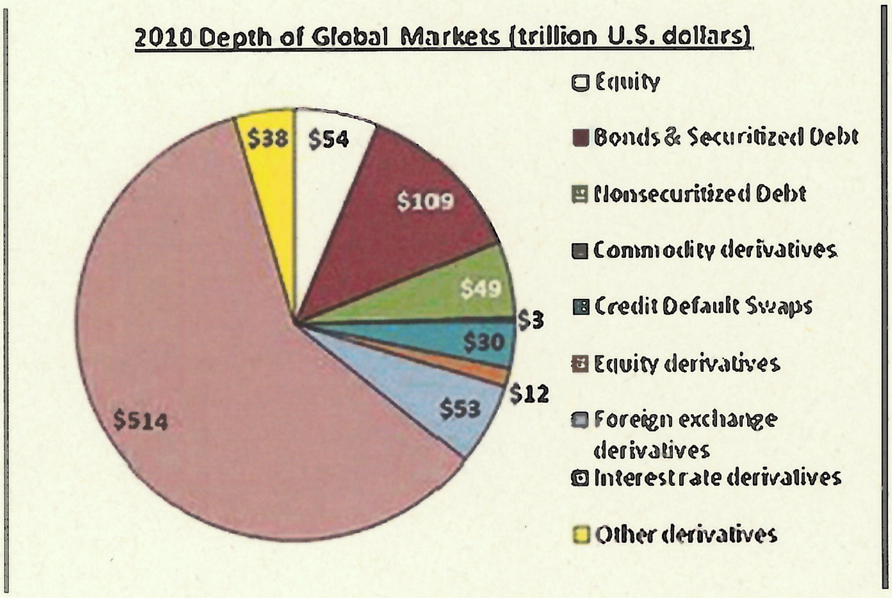
Instruments whose value depend upon the value evolution of other products. Virtually anything that has not been defined before.

This ~~course~~ will focus mainly on equities and simple derivatives of ~~equity~~ equities.

MODULE DISCLAIMER

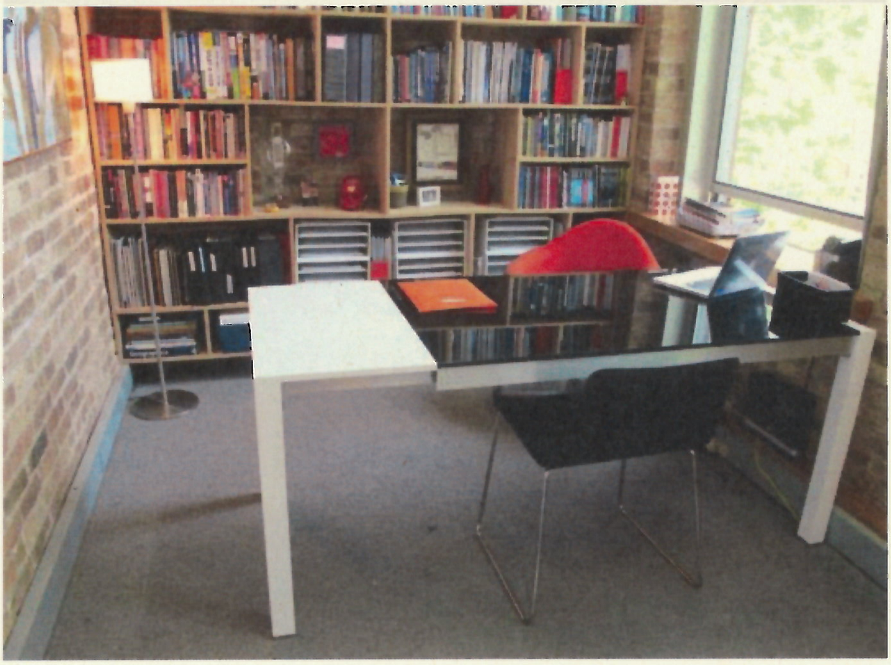
This course aims to be a gentle introduction to finance, showing the building blocks of modern financial engineering techniques. There is a HUGE gap between theory and practition!!

See Fig 1.1.



2010 Depth of Global Markets	(trillions USD)	(%)	
Equity	54	6%	Introduction to Financial Engineering
Bonds & Securitized Debt	109	13%	Advance Financial Engineering
Nonsecuritized Debt	49	6%	
Commodity Derivatives	3	0%	
Credit Default Swaps	30	3%	
Equity Derivatives	12	1%	Introduction to Financial Engineering
Foreign Exchange Derivatives	53	6%	
Interest Rate Derivatives	514	60%	Advance Financial Engineering
Other Derivatives	38	4%	
	862	100%	


NEED HELP?



FINANCIAL MATHEMATICS : The basics

The aim of this chapter is to introduce the basic mathematical operations in finance .... at the end of the chapter you will be able to compute the monthly payment of a morgatge ..... for 90% of world's population that investment will be the most important in their lifes.

We won't use these concepts until chapter 3 of the course, nevertheless these are basic concepts not to be forgoten

Contents :

i) Basic Definitions
ii) Types of financial laws
   a) Simple interest
   b) Compound interest
iii) Compounding frequencies
iv) Loans & Mortgages.

BASIC DEFINITIONS

We name financial capital to the value of an asset at the moment it becomes available.

Interest rate is the return ~~demanding~~ demanded for renouncing to consume now, in exchange for a promise of future consumption.

→ Any interest rate comprises two components.

- Risk free interest rate: reflecting the time value of money
- Risk premium: Reflects the extra return demanded for assuming the risk that the promise of future payment will not be honored.

FINANCIAL LAWS

Are models for (valuing) moving money over time.
If 1€ today is more/less than 1€ in 1 year time, how can we compare cash flows in different time periods?

Example

i) 1€ today
ii) 1.1€ in 1 year

{ what is best?
depend on interest rates.

A financial equivalence is the correspondence between two financial capitals under a given financial law.

i) Capitalization : Moving money forward in time, we should determine the final value that is equivalent to a certain initial value.

ii) Discounting : Moving money backwards, we have to determine the initial value that is equivalent to a certain initial value

Let's formulate the capitalization laws $\begin{cases} \text{simple interest} \\ \text{compound interest} \end{cases}$

Simple interest Capitalization

$$C_T = C_0 (1 + i * t)$$

C_T = Final Capital (\$)
C_0 = Initial Capital (\$)
i = interest rate (%) annualized
t = maturity ("years")

Simple interest Discount

$$C_0 = C_T / (1 + i * t)$$

Example

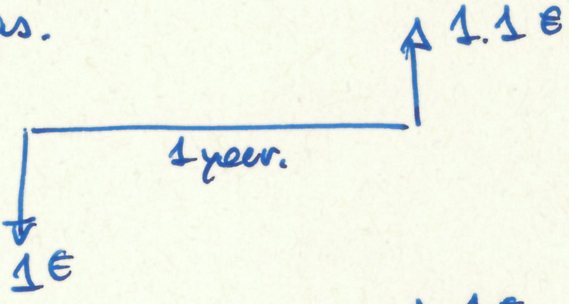
Assume an interest rate of 10%. Then if I give 1€ to a solvent counterparty what can I expect to get back in 1 year?

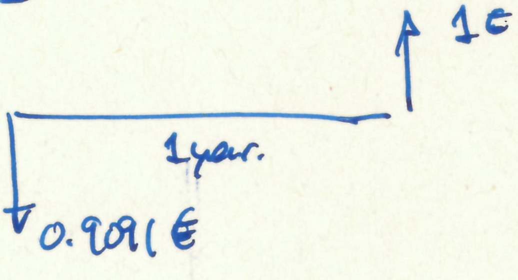
$$\left.\begin{array}{l} C_0 = 1€ \\ i = 10\% \\ t = 1 \end{array}\right\} \Rightarrow C_T = C_0(1 + i \cdot t) = 1.1\ €$$

Assume I want to get 1€ in 1 year, how much should I invest?

$$\left.\begin{array}{l} C_T = 1€ \\ i = 10\% \\ t = 1y \end{array}\right\} \Rightarrow C_0 = C_T/(1+it) = 0.9091\ €$$

Graphically we will depict the above financial equivalences as.





In simple interest formulas interest rate is only applied to initial capital, but is common use to apply interest on interest in long investments.

Compound interest Capitalization

$$C_t = C_0 \left(1 + i/n\right)^{t*n}$$

n = payment frequency of interest over a year
Same definition as before for other variables

Compound interest Discount

$$C_0 = C_t \Big/ \left(1 + i/n\right)^{t*n}$$

Example

Assume an investment pays interest semiannually of 5%.
If I invest 1€, how much should I expect in 3 years?

$C_0 = 1€$
$i = 5\%$
$n = 2$
$t = 3.$

$$C_t = C_0 \left(1 + i/n\right)^{t*n} = 1.1597€$$

There is still another type of compounding interest used in pricing derivatives (Chapter 3) which is the Continuously Compounded interest.
Lets assume that our investment produces instantaneous interest continuously, then we would have to

take $n \to \infty$ in the compounded interest formula

$$C_T = C_0 \left(1 + i/n\right)^{t \cdot n} \xrightarrow[n \to \infty]{} C_0 e^{rt}$$

Continuously Compounded interest formula

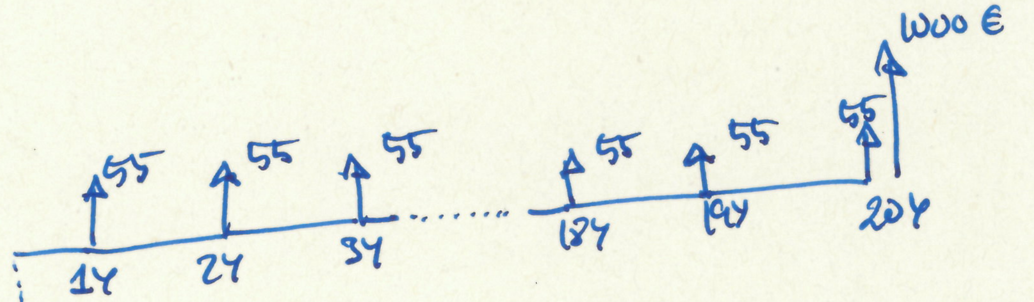
$$\boxed{C_T = C_0 e^{rt}}$$

Example

Consider a bond which makes regular payments plus principal at maturity with the following characteristics

Principal 1000 €
Risk-free rate 6%
Maturity 20 years
Annual coupon 5.5%

what would be the price of the bond today?



?

← this is what we are asked.

We just need to discount each cash flow to today

$$\text{Price} = \sum_{i=1}^{20} \frac{55}{(1+0.06)^i} + \frac{1000}{(1+0.06)^{20}} = \underline{942.65\,€}$$

LOANS

A loan is a financial operation in which the lender gives a sum termed principal to the borrower who undertakes to amortize the principal and pay interest until the whole debt has been repaid.

Does not the bond definition fall in the above definition?
Yes, only a ~~bond~~ bond has a very special way to amortize principal which is to repay in full amount at the maturity. Typical loans for retail pay a constant payment over the life of the loan which is split between principal and amortization.

This way of amortizing principal is named French Amortization.

LOANS & MORTGAGE REPAYMENT FORMULA

Lets assume you take a loan of \$D repayable over T years at R% interest rate on a monthly repayment.
Every monthly instalment ~~has~~ to pay off interests accumulated by the loan and an amount to reduce the outstanding debt.
Let us summarize the instalments as follow.

Month	Debt at start	Interest	Capital Repayment
1	D	$D \times R/12$	X
2	$D - X$	$(D-X)\,R/12$	Y
⋮			

If we want all instalments to be equal, then

$$DR\frac{1}{12} + X = (D-X)\frac{R}{12} + Y$$

and hence $\quad Y = (1 + R/12)\,X$

Therefore the previous table can be rewritten as.

Month	Debt at start	Interest	Capital Repayment
1	D	$D \times R/12$	X
2	$D-X$	$(D-X)\,R/12$	$X(1+R/12)$
⋮	⋮	⋮	
$12*T$			$(1+R/12)^{12T}\,X$

Since Capital Repayment should sum up to initial debt we have

$$\sum_{i=0}^{12*T} x\,(1+R/12)^i = D$$

and hence $$x = \frac{D\,R/12}{\left[(1+R/12)^{12T}-1\right]}$$

Finally, to know the monthly quote you just need to add the interest component, for instance the installment for the first period is.

$$\text{Monthly instalment} = D \times R/12 + \frac{D\,R/12}{(1+R/12)^{12T}-1}$$

$$= \frac{D\frac{R}{12}(1+R/12)^{12T}}{\left[1+R/12\right]^{12T}-1}$$

The above is far from beeing the most important formula in this course, but believe me it is <u>the</u> most important formula for a large part of our society!

Simulador Hipotecas

Informe de resultados de la simulación (29/08/2019)

Simulación Hipoteca Fija

Datos de tu simulación

Valor compraventa:	**300.000 €**
Valor estimado de tasación vivienda:	**300.000 €**
Importe de la hipoteca:	**240.000 €**
Tipo de inmueble:	**Vivienda nueva**
Ubicación del inmueble:	**Barcelona**
Plazo total de la hipoteca:	**30 años**
Comisión de apertura:	**0 € (0,00)**
Finalidad:	**Vivienda habitual**

Cuota simulada para préstamo hipotecario

	Primer año	Resto de años cumpliendo condiciones
Tipo de Interés Nominal	1,99	1,99[1]
Cuota Mensual	**885,89 €**	**885,89 €**

TAE(1): 2,58

2 TIME SERIES : MACROECONOMIC STUDIES

The objective of this chapter is to introduce the aplications of ARMA and GARCH models in finance. The most common aplication of such models ~~is~~ is to produce predictions of macroeconomic time series such, are GDP (Gross Domestic Product), inflation, unemployment rates, house prices, ... As you may imagine such data is very important for long term plans in finance industry.

The aim of the chapter is to present a quick overview of the mathematical models and use built-in packages from R to calibrate real data.

Further Reading

Ch 9, D. Ruppert. Statistics and Data Analysis for Financial Engineering

STOCHASTIC PROCESS & STATIONARY

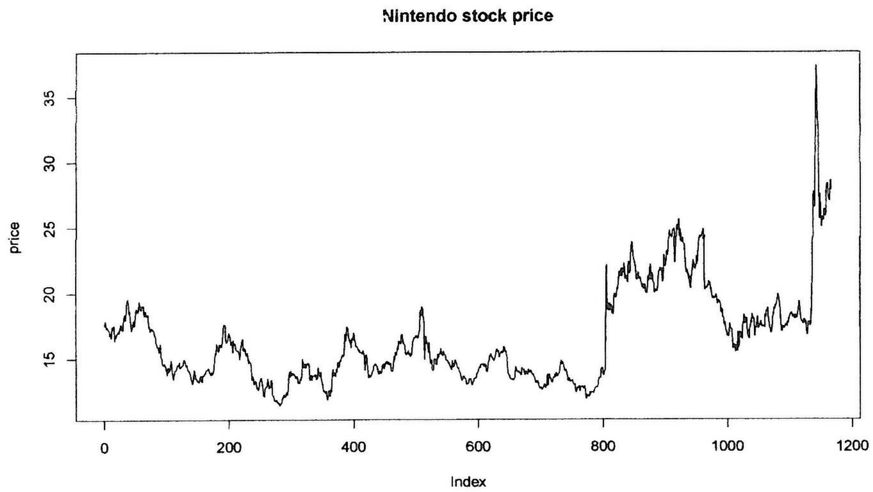
A time series is a sequence of observations in chronological order (for example, daily stock returns).

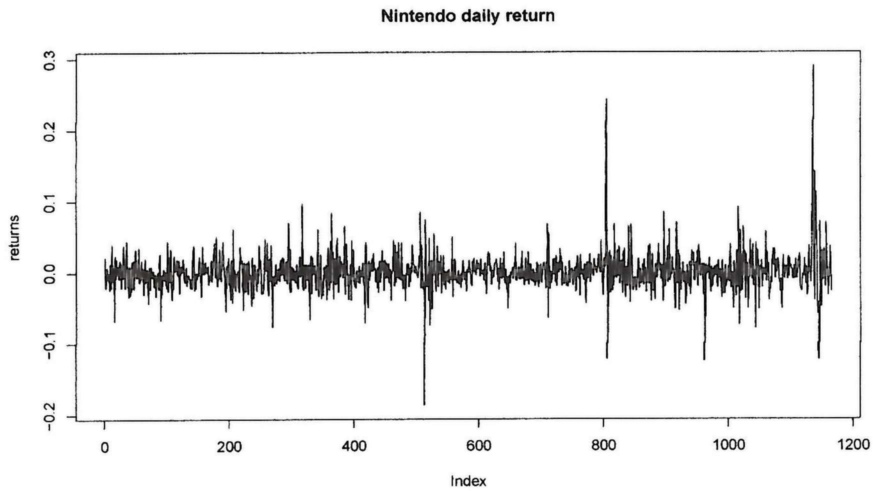
Stochastic Process: A sequence of random variables, hence a time series can be viewed as a realization (sample) from a stochastic process.

One of the most useful methods for obtaining parsimony (describing data with as few as possible parameters) in a time series is to assume stationarity. This property is observed in a time series when the fluctuations appear random but often with the same type of stochastic properties. For example, stock returns appear random from one period to the next, but often statistics like mean and variance seem to behave less randomly.

A stationary process is a ~~[illegible]~~ probability model with time invariant behaviour, in other words all aspects of its behaviour are invariant under a time-shift.

```r
# Stock evolution of Nintendo
# Albert Ferreiro Castilla
# 21/08/2016

rm(list=ls())                                    # Removes all variables from workspace
library("quantmod")                              # Library to get prices from Yahoo
options("getSymbols.warning4.0"=FALSE)           # set off warnings
cat("\f")                                        # Clear console


getSymbols('NTDOY',scr='yahoo',from="2012-01-01",to="2016-08-20")
nint_close=NTDOY$NTDOY.Close

price=as.numeric(nint_close)
returns=log(price[2:length(price)]/price[1:length(price)-1])


pdf("nintendo01.pdf",width=10,height=6)
plot(price, type="l",main="Nintendo stock price")
dev.off()

pdf("nintendo02.pdf",width=10,height=6)
plot(returns,type="l", main="Nintendo daily return")
dev.off()
```





Strong Stationary Process: Let $\{Y_n\}_n$ be a stochastic process and $F(Y_{n_1}, Y_{n_2}, \ldots, Y_{n_k})$ be the cumulative distribution function of $\{Y_n\}_n$ at times $n_1, n_2, \ldots n_k$. then $\{Y_n\}_n$ is said to be strongly stationary if for any n and m

$$F(Y_{n_1}, Y_{n_2}, \ldots, Y_{n_k}) = F(Y_{n_1+m}, Y_{n_2+m}, \ldots, Y_{n_k+m})$$

This property is called time-invariance

Weak Stationary Process: Let $\{Y_n\}_n$ be a stochastic process, then $\{Y_n\}_n$ is said to be weak stationary if

$$\mathbb{E}[Y_i] = \mu \qquad \text{"similar first moments"}$$
$$Var(Y_i) = \sigma^2$$
$$Corr(Y_i, Y_j) = \rho(|i-j|)$$

for all i, j and some function $\rho(h)$.

The function $\rho(h)$ is called the autocorrelation function. Another relevant function is the autocovariance function:

$$Cov(Y_i, Y_j) = \gamma(|i-j|)$$

Note that $\rho(h) = \sigma^2 \gamma(h)$.

A stationary time series should show oscillations around the same fixed level, called mean-reversion. If the series wonders around without repeatedly returning to the same fixed level, then the series should not be model as a stationary time series.

As seen before, many financial time series are not stationary but often, transformations on them are. For example transforming prices to log-returns.

White noise: is the simplest example of stationary process. The sequence $\{Y_n\}_n$ is a weak $WN(\mu, \sigma^2)$ if
$\mathbb{E}[Y_i] = \mu$; $Var(Y_i) = \sigma^2$; $Corr(Y_i, Y_j) = 0 \quad \forall i \neq j$

By definition a WN is weakly stationary with autocorrelation function $\begin{cases} \rho(0) = 1 \\ \rho(h) = 0 & ; h \neq 0 \end{cases}$

Note that a i.i.d WN (independent and identically distributed) white noise is strongly stationary

ESTIMATING PARAMETERS

Suppose we observe $Y_1, \dots, Y_n$ from stationary process. To estimate the mean and variance we can use the sample mean and variance

$$\hat{\mu} = \bar{Y} \quad ; \quad \hat{\sigma}^2 = s^2$$

To estimate the autocovariance function, we use the sample autocovariance function

$$\hat{\gamma}(h) = \frac{\sum_{j=1}^{n-h} (Y_{j+h} - \bar{Y})(Y_j - \bar{Y})}{n}$$

and from here we can estimate the autocorrelation function

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

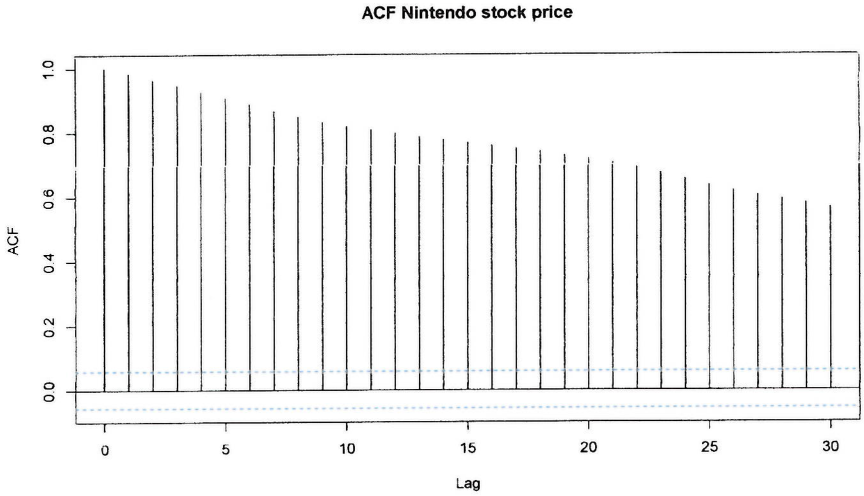
ACF plots and the Ljung-Box ~~test~~ are the R built-in tools to plot the autocorrelation function. The ACF plot is pictured with test bounds, which are used to test the null hypothesis that the autocorrelation ~~coef~~ coefficient is 0. The usual level of the test is 0.05 and hence we can expect 1 out of 20 sample autocorrelations outside the bounds simply by chance.

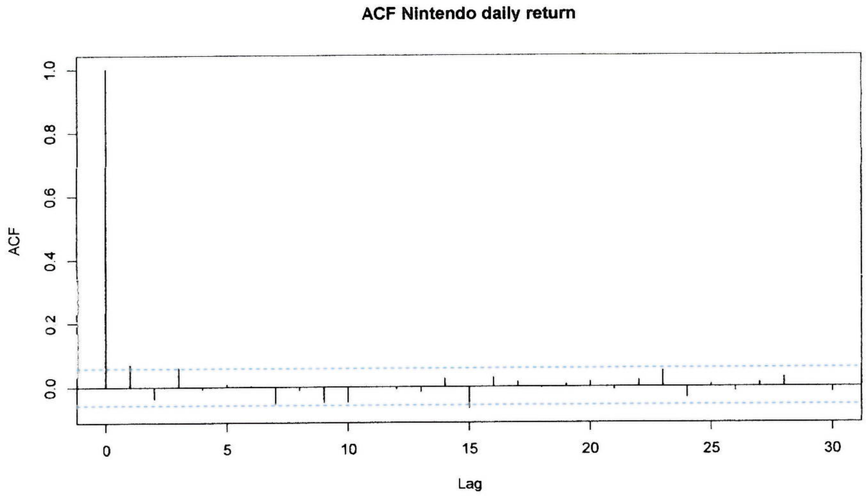
The alternative is to use a simultaneous test, to test the null hypothesis $H_0 := \rho(1) = \rho(2) = \dots = \rho(k) = 0$

for some k

```
...

pdf("nintendo01_ACF.pdf",width=10,height=6)
acf(price, main="ACF Nintendo stock price")
dev.off()

pdf("nintendo02_ACF.pdf",width=10,height=6)
acf(returns, main="ACF Nintendo daily return")
dev.off()

Box.test (returns, lag = 20, type = "Ljung")
```

```
     Box-Ljung test

data:  returns
X-squared = 29.1, df = 20, p-value = 0.08581
```





AR(p): Autoregressive Processes

Although stationary processes are somehow parsimonious with parameters, they have still an infinite number of parameters $\{\rho(i)\}_i$. What we need is a class of stationary processes with a finite number of parameters but a wide range of behaviours, those are ARIMA models. To construct ARIMA models we first look at the smaller subclass of AR processes, which essentially are ~~models~~ autoregressive models

AR(1)

Let $\{\varepsilon_i\}_i$ be a WN$(0,\sigma^2)$. We say that $\{Y_n\}_n$ is an AR(1) process if for some constant parameters μ and ϕ

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + \varepsilon_t$$

Parameter μ is the mean of the process and $\phi(Y_{t-1}-\mu)$ represents the memory of the process. The parameter ϕ determines the velocity of the mean reversion.

In finance $\{\varepsilon_i\}_i$ can be interpreted as information shocks. Information that is truly new cannot be anticipated so effects of today's information should be independent from yesterday's news.

observe that if Y is weakly stationary process, then $|\phi|<1$.

> To see this note that the variance of $Y_t-\mu$ and $Y_{t-1}-\mu$ should be equal, say σ_Y^2, hence
> $$\sigma_Y^2 = \phi^2 \sigma_Y^2 + \sigma_\varepsilon^2$$

We can also rewrite the definition of AR(1) as

$$Y_t = (1-\phi)\mu + \phi Y_{t-1} + \varepsilon_t$$

which resembles a regression. We can iterate the process to state

$$Y_t = \mu + \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \dots = \sum_{h=0} \phi^h \varepsilon_{t-h}$$

which is a way of looking at an AR(1) as the weighted average of all past white noise shocks.

~~For AR(1) with $|\phi|<1$ (ie stationary)~~

For an stationary AR(1) process (ie $|\phi|<1$), then

$$\mathbb{E}[Y_t]=\mu$$

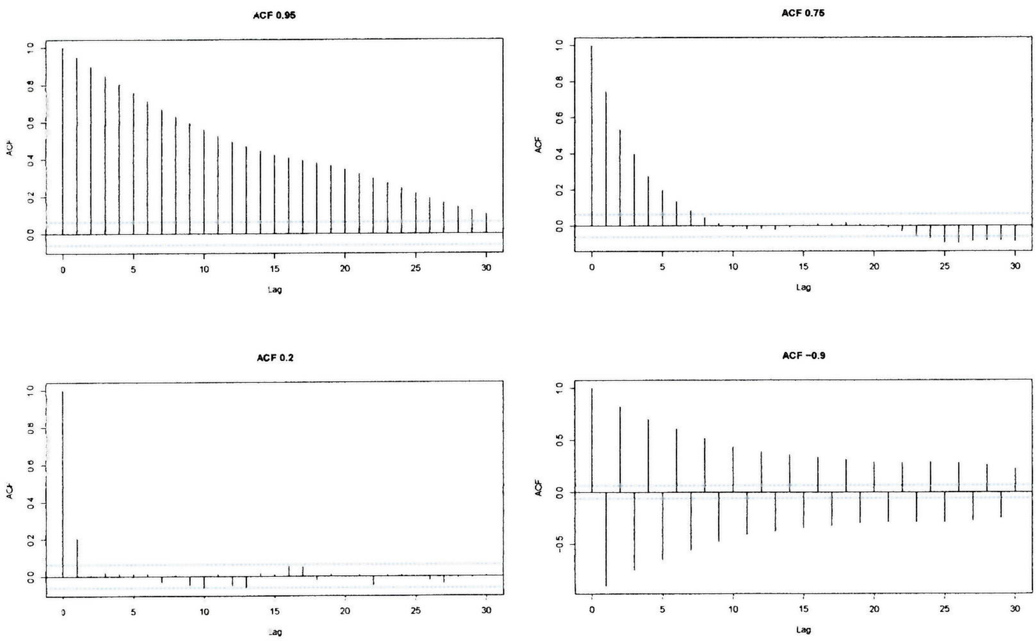
$$\gamma(0) = Var(Y_t) = \sigma_\varepsilon^2 / 1-\phi^2$$

$$\gamma(h) = \frac{\sigma_\varepsilon^2}{1-\phi^2}\phi^{|h|}$$

$$\rho(h) = \phi^{|h|}$$

The ACF plot for an AR(1) depends upon only one parameter ϕ. This is a remarkable level of parsimony but with high price. Only a limited amount of ACFs shapes are achievable with an AR(1) process.

```r
rm(list=ls())                         # Removes all variables from workspace
cat("\f")                             # Clear console

y.95 <- arima.sim(model=list(ar=.95), n=1000)

pdf("AR_ACF_1.pdf",width=10,height=6)
acf(y.95, main="ACF 0.95")
dev.off()

y.75 <- arima.sim(model=list(ar=.75), n=1000)

pdf("AR_ACF_2.pdf",width=10,height=6)
acf(y.75, main="ACF 0.75")
dev.off()

y.2 <- arima.sim(model=list(ar=.2), n=1000)

pdf("AR_ACF_3.pdf",width=10,height=6)
acf(y.2, main="ACF 0.2")
dev.off()

y.99 <- arima.sim(model=list(ar=-.9), n=1000)

pdf("AR_ACF_4.pdf",width=10,height=6)
acf(y.99, main="ACF -0.9")
dev.off()
```



WEEK 3

If $|\phi|>1$ the AR(1) process is not stationary, and the mean, variance and correlation are not constant.

Random Walk

When $\phi = 1$, the process is called a random walk

$$Y_t = Y_{t-1} + \varepsilon_t$$

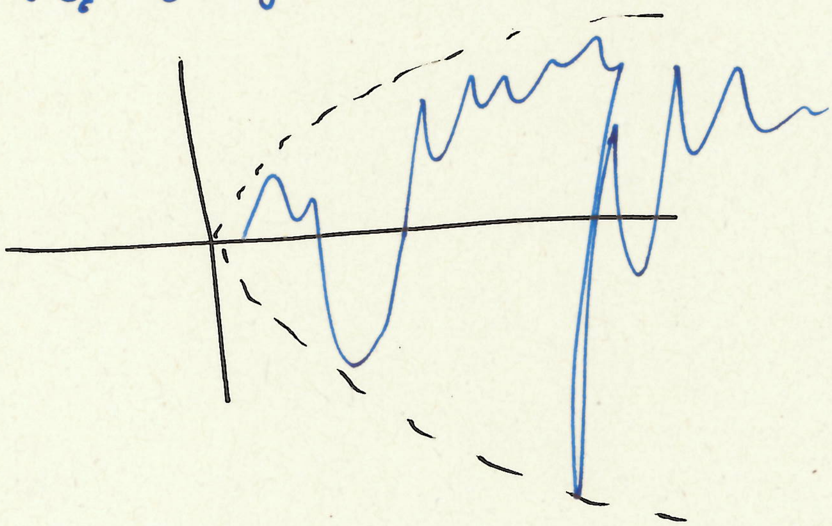
It is easy to write the following equation

$$Y_t = Y_0 + \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t$$

from which we derive

$\mathbb{E}[Y_t \mid Y_0] = Y_0$ (mean-reversion)

$Var(Y_t) = t\sigma_\varepsilon^2$ (long excursions from mean



When $|\phi|>1$ the process has explosive behaviour.

R has a built-in package to estimate AR process and other time series. The method we will use is the maximum likelihood estimation with least-square estimate as the starting point

One will usually analyse the autocorrelation function of the residuals, as any positive autocorrelation in the residuals is an evidence against the AR(1) model

The following example fits a time series of BMW prices. In the example R fits the following parameters

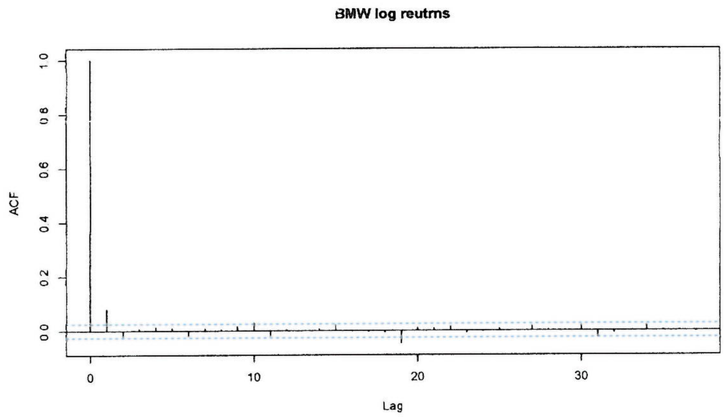
$$\hat{\phi} = 0.0811$$
$$\hat{\mu} = 0.00034$$

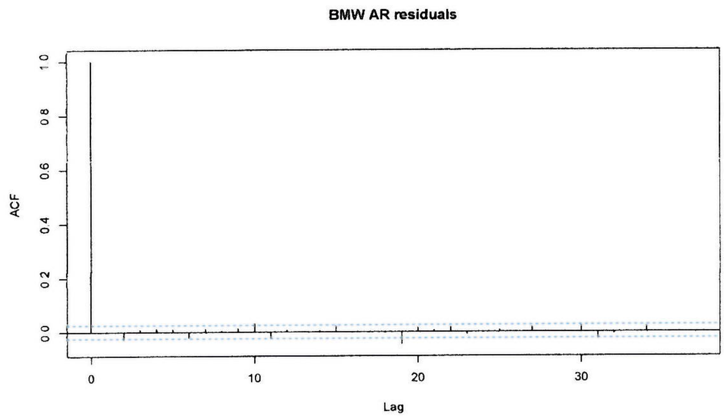
```r
rm(list=ls())                                  # Removes all variables from workspace
cat("\f")                                      # Clear console
library("evir")
data(bmw,package="evir")                       # Load BMW log prices

pdf("BMW_1.pdf",width=10,height=6)             # ACF for log prices
acf(bmw, main="BMW log reutrns")
dev.off()
Box.test (bmw, lag = 5, type = "Ljung")        # Evidence of not being a white noise
fit<-arima(x=bmw,order=c(1,0,0))               # Fit of AR(1)
fit                                            # Showing the parameters
                                               # Final analysis of the residuals
pdf("BMW_2.pdf",width=10,height=6)
acf(fit$resid, main="BMW AR residuals")
Box.test (fit$resid, lag = 5, type = "Ljung",fitdf = 1)
dev.off()
```

```
    Box-Ljung test
data:  bmw
X-squared = 44.987, df = 5, p-value = 1.46e-08
```

```
Call:
arima(x = bmw, order = c(1, 0, 0))
Coefficients:
         ar1   intercept
      0.0811       3e-04
s.e.  0.0127       2e-04

sigma^2 estimated as 0.0002163:  log likelihood = 17212.34,  aic = -34418.68
```

```
    Box-Ljung test
data:  fit$resid
X-squared = 6.8669, df = 4, p-value = 0.1431
```





WEEK 3

We have seen that the ACF for an AR(1) process decays geometrically to zero and alternate signs if $\phi < 0$, but this is a limited range of behaviour for the ACF function. To get an extended class of autocorrelation functions we extend AR(1) to regress to older values of the process.

AR(p)

Let $\{\varepsilon_i\}_i$ be a $WN(0, \sigma^2)$. We say that $\{Y_n\}_n$ is an AR(p) process if for some constant parameters μ and $\{\phi_i\}_{i=1}^{p}$,

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \dots + \phi_p(Y_{t-p} - \mu) + \varepsilon_t$$

The general formula can be rewritten as

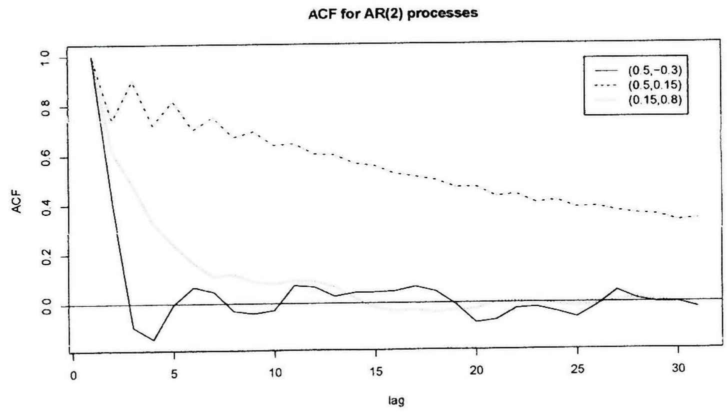
$$Y_t = \beta_0 + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t$$

where $\beta_0 = [1 - (\phi_1 + \dots + \phi_p)]\mu$. It can be showed that if Y is a stationary process, then

$$\beta_0 / \mu > 0$$

Most of the concepts we have discussed for AR(1) generalize for AR(p). Let us plot some ACF functions for AR(2):

```r
y <- arima.sim(model=list(order=c(2,0,0), ar=c(0.5,-0.3)), n=1000)
z <- arima.sim(model=list(order=c(2,0,0), ar=c(0.5,0.15)), n=1000)
v <- arima.sim(model=list(order=c(2,0,0), ar=c(0.15,0.8)), n=1000)

yy<-acf(y)
zz<-acf(z)
vv<-acf(v)

pdf("AR2_ACF.pdf",width=10,height=6)
plot(yy$acf, type="l",main="ACF for AR(2) processes",xlab="lag",ylab="ACF")
lines(zz$acf,lty=3)
lines(vv$acf,lty=4)
legend("topright", inset=.05, cex = 1, c("(0.5,-0.3)","(0.5,0.15)","(0.15,0.8)"), horiz=FALSE, lty=c(1,2,3))
abline(h=0,lty=1,lwd=1)
dev.off()
```



MA Processes : Moving Average

There is a potential need for large values of p when fitting an AR(p) model, the remedy is to introduce a moving average component to construct an ARMA process.

Before we introduce ARMA process, let's look at MA processes in more detail.

MA

Let $\{\varepsilon_t\}_t$ be a $WN(0, \sigma_\varepsilon^2)$. We say that $\{Y_t\}_t$ is a MA(q) process if for some constant parameters μ and $\{\theta_i\}_{i=1}^{q}$,

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

For the MA(1) process defined as

$$Y_t = \mu + \varepsilon_t + \theta \varepsilon_{t-1}$$

we can easily derive

$$\mathbb{E}[Y_t] = \mu$$

$$Var(Y_t) = \sigma_\varepsilon^2 (1 + \theta^2)$$

$$\gamma(1) = \theta \sigma_\varepsilon^2$$

$$\gamma(h) = 0 \quad \text{for } |h| > 1$$

It can be showed that for a general MA(q) process

$$\gamma(h) = 0 \quad \text{for } |h| > q$$

ARMA Processes

Stationary time series with complex autocorrelation behaviour often are more parsimoniously modeled by mixed autoregressive and moving average (ARMA) process. For example, some times an ARMA(1,1) - a process with one parameter in the AR part and one parameter in the MA part -, fits better than a pure AR or MA process.

Before we introduce ARMA process, let us define two simple operators to ease the notation.

Backward Operator

It is defined for $k \geq 0$ by

$$B^k Y_t = Y_{t-k}$$

Differencing Operator

It is defined as $\Delta = 1 - B$, hence

$$\Delta Y_t = (1-B) Y_t = Y_t - Y_{t-1}$$

$$\Delta^k Y_t = (1-B)^k Y_t = \sum_{i=1}^{k} \binom{k}{i} Y_{t-i}$$

ARMA

Let $\{\varepsilon_i\}_i$ be a $WN(0, \sigma_\varepsilon^2)$. We say that $\{Y_n\}_n$ is an ARMA(p, q) process if for some constant parameters μ, $\{\phi_i\}_{i=1}^{p}$ and $\{\theta\}_{j=1}^{q}$,

$$Y_t - \mu = \phi_1 (Y_{t-1} - \mu) + \dots + \phi_p (Y_{t-p} - \mu) + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

A useful notation using the background operator is

$$(1 - \phi_1 B - \dots - \phi_p B^p)(Y_t - \mu) = (1 + \theta_1 B + \dots + \theta_q B^q)\varepsilon_t$$

Note that the WN is an ARMA(0,0). The ARMA(1,1) is used in practice and yet simple enough to study theoretically, among other expressions one can derive:

$$\rho(1) = \frac{(1 + \theta\phi)(\theta + \phi)}{1 + \theta^2 + 2\theta\phi}$$

$$\rho(k) = \phi\,\rho(k-1) \quad \text{for } k \geq 2$$

In particular, one can observe that after one lag, the autocorrelation function decays as an AR(1).

The following code fits several ARMA processes to the same time series and extracts its AIC and BIC.

When fitting ARMA models, one way to decide to initially choose between different models are the AIC and BIC criteria, which prevents overfitting.

- AIC : Akaike information criteria
- BIC : Bayesian information criteria

Nevertheless, after choose a few models we still need to evaluate ACF plots and residuals to finally decide between models.

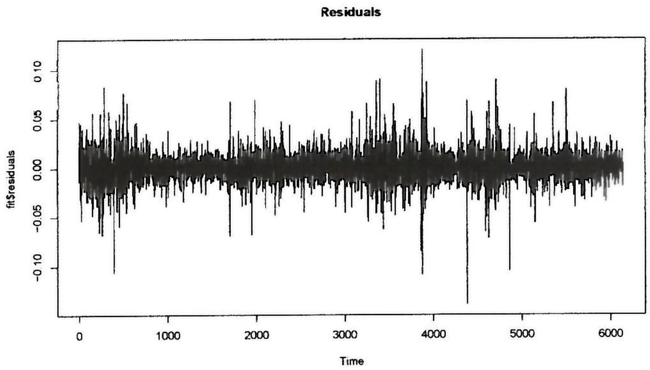
In our example the ARMA(1,1) seems a satisfactory model, the following are the ACF plot, and the QQ plot ~~with~~ which shows heavy tails although the ACF plot shows no sign of short-term autocorrelation. The residual plot also shows a bit of clustering.

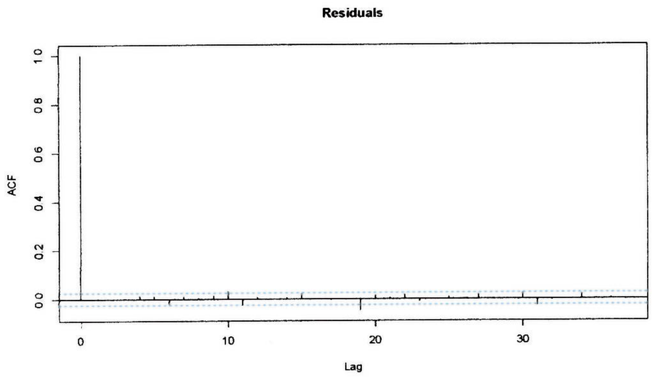
```
rm(list=ls())                                   # Removes all variables from workspace
cat("\f")                                       # Clear console

library("evir")
data(bmw, package="evir")                       # Load BMW log prices

aic_bic<-array(dim=c(9,4))
k=1
for (i in 0:2) {
  for (j in 0:2){
    fit<-arima(x=bmw, order=c(i,0,j))
    aic_bic[k,1]=i
    aic_bic[k,2]=j
    aic_bic[k,3]=AIC(fit)+34500
    aic_bic[k,4]=AIC(fit,k=log(length(bmw)))+34500
    k=k+1
  }
}
colnames(aic_bic) <- c("p","q","AIC","BIC")
aic_bic
```

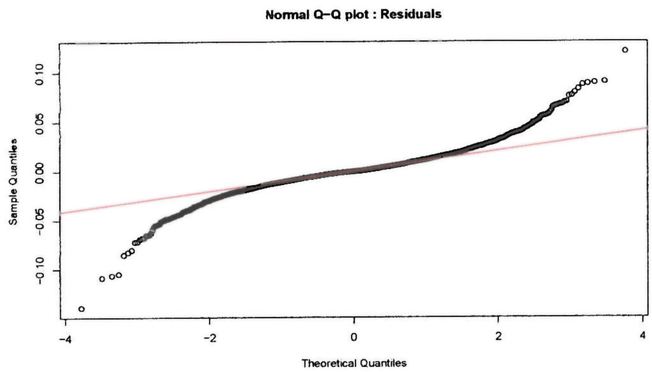
```
 aic_bic
      p q       AIC       BIC
 [1,] 0 0 119.82867 133.27578
 [2,] 0 1  79.21393  99.38461
 [3,] 0 2  78.36773 105.26195
 [4,] 1 0  81.31575 101.48642
 [5,] 1 1  78.25941 105.15364
 [6,] 1 2  80.44557 114.06336
 [7,] 2 0  78.73415 105.62838
 [8,] 2 1  80.58248 114.20026
 [9,] 2 2  82.22671 122.56805
```

WEEK 3

```r
rm(list=ls())                              # Removes all variables from workspace
cat("\f")                                  # Clear console

library("evir")
data(bmw, package="evir")                  # Load BMW log prices

fit<-arima(x=bmw, order=c(1,0,1))

pdf("BMW_res_plot.pdf",width=10,height=6)
plot(fit$residuals , type="l",main="Residuals")
dev.off()

pdf("BMW_res_acf.pdf",width=10,height=6)
acf(fit$residuals , main="Residuals")
dev.off()

pdf("BMW_res_qqplot.pdf",width=10,height=6)
qqnorm(fit$residuals ,main="Normal Q-Q plot : Residuals")
qqline(fit$residuals ,col=2)
dev.off()
```







ARIMA Processes

Often the first or second difference of nonstationary time series are stationary. Autoregressive Integrated Moving Average (ARIMA) process are introduced by the following simple definition

ARIMA(p,d,q)

A time series $\{Y_t\}_t$ is said to be an ARIMA(p,d,q) if the process $\Delta^d Y_t$ is an ARMA(p,q).

By definition we set the ARIMA$(p,0,q)$ as ARMA(p,q)

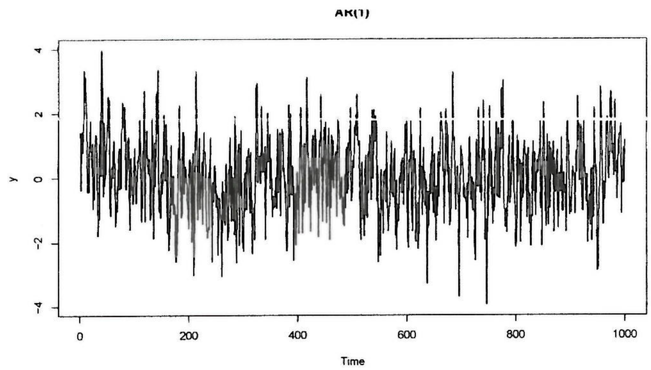
For example if log-returns are ARMA(p,q), then the log prices are ARIMA$(p,1,q)$. Note that if ARIMA(p,d,q) is stationary, then $d=0$.

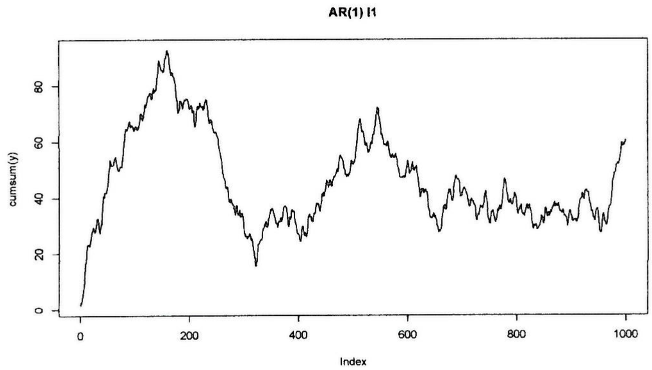
The inverse of differentiating is integration. The integral of a process Y_t is the process X_t, where

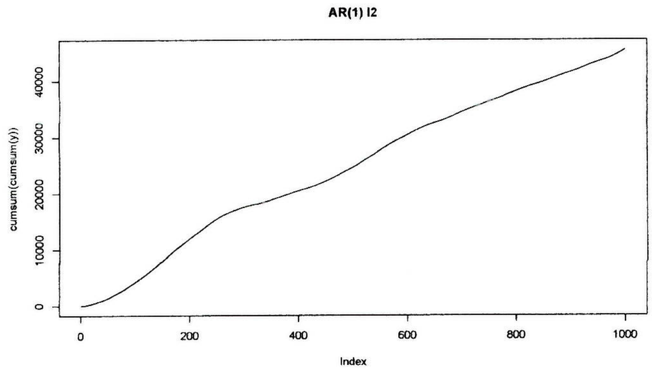
$$X_t = X_{t_0} + Y_{t_0} + Y_{t_0+1} + \cdots + Y_t$$

Indeed $\Delta X_t = Y_t$

```r
rm(list=ls())                    # Removes all variables from workspace
cat("\f")                        # Clear console

y <- arima.sim(model=list(order=c(1,0,0), ar=c(0.5)), n=1000)

pdf("AR1.pdf",width=10,height=6)
plot(y, type="l",main="AR(1)")
dev.off()

pdf("AR1_I1.pdf",width=10,height=6)
plot(cumsum(y), type="l",main="AR(1) I1")
dev.off()

pdf("AR1_I2.pdf",width=10,height=6)
plot(cumsum(cumsum(y)), type="l",main="AR(1) I2")
dev.off()
```







- The original time series shows mean-reversion behaviour.
- The integrated series behaves much more like a random walk
- The twice-integrated series shows what is known as momentum, tendency to move in a particular direction onces the movement is started – this is a sign of second order integration

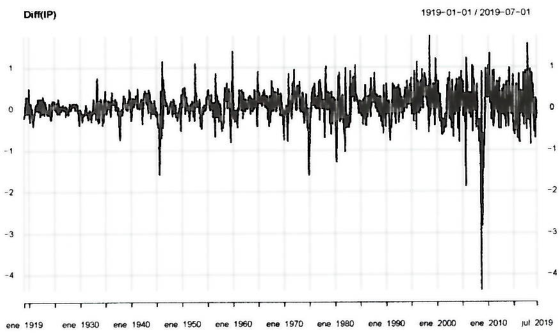
Fitting an ARIMA model is easy as we have done with ARMA. The following data shows Industrial Production Index fitted to ARIMA (1, 1, 1)

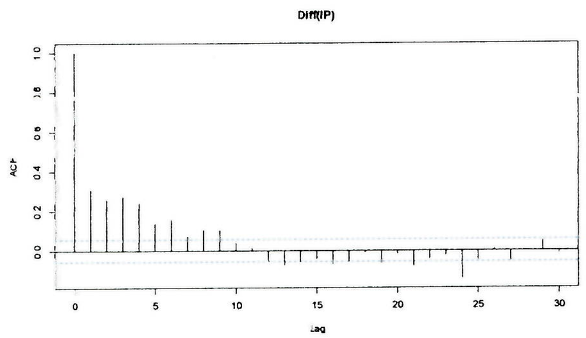
- The first two plots shows the original series and the differenticting one.
- Plotting the ACF for ΔY_t process seems to fit an ARMA to it.
- ARMA (1,1) is chosen based on BIC criteria
- We finally check ACF for residuals, concluding a satisfactory fit.

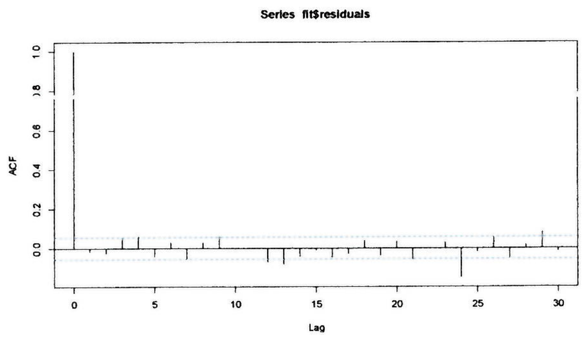
If a nonstationary process has a constant mean, then the first difference of its process have mean zero. The "arima" function in R make this assumption. If instead of a constant mean, the process has a deterministic linear trend the differentiated process has mean different to zero. In order to address this use "auto-arima" function.

```r
rm(list=ls())                                    # Removes all variables from workspace
cat("\f")                                        # Clear console

library("quantmod")                              # Library to get prices from Yahoo
options("getSymbols.warning4.0"=FALSE)           # set off warnings
cat("\f")                                        # Clear console

getSymbols.FRED('INDPRO',env=globalenv())

pdf("IP.pdf",width=10,height=6)
plot(INDPRO,main="Industrial Production (IP)")
dev.off()

pdf("IP_diff.pdf",width=10,height=6)
plot(diff(INDPRO),main="Diff(IP)")
dev.off()

pdf("IP_diff_ACF.pdf",width=10,height=6)
acf(diff(INDPRO),na.action = na.pass,main="Diff(IP)")
dev.off()

fit<-arima(x=diff(INDPRO),order=c(1,0,1))
fit

pdf("IP_diff_ACF_res.pdf",width=10,height=6)
acf(fit$residuals,na.action = na.pass)
dev.off()
```

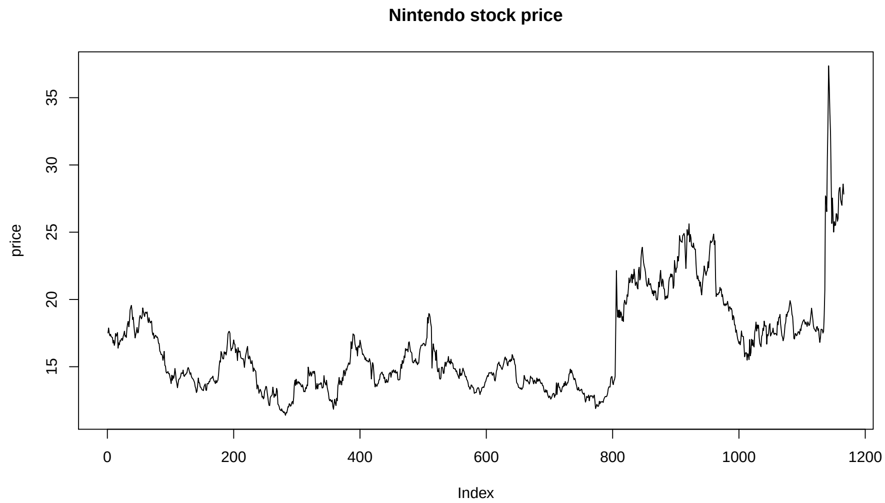
```
Call:
arima(x = diff(INDPRO), order = c(1, 0, 1))

Coefficients:
         ar1      ma1  intercept
      0.8492  -0.6189     0.0857
s.e.  0.0279   0.0398     0.0275

sigma^2 estimated as 0.1399:  log likelihood = -509.76,  aic = 1027.52
```

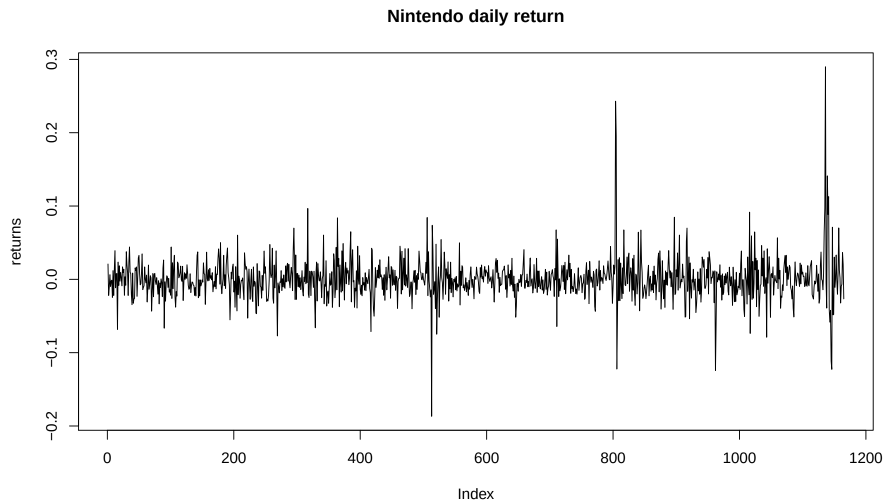






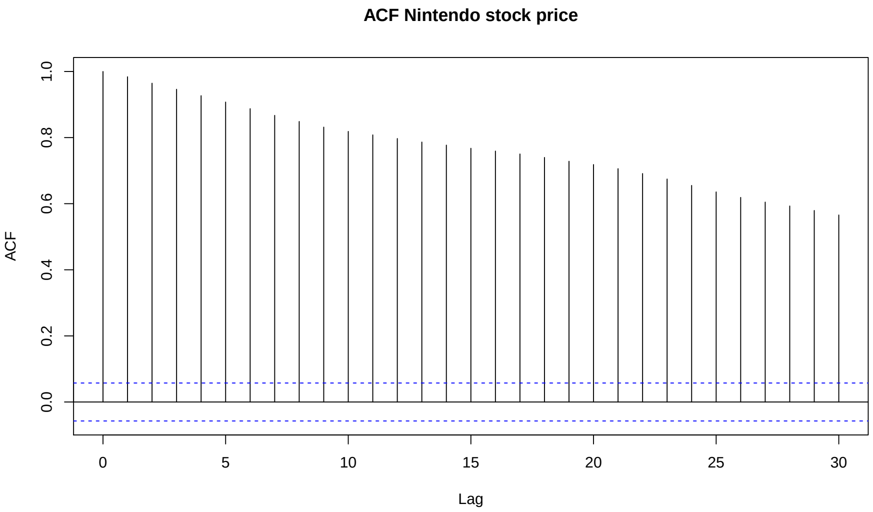
```r
# Stock evolution of Nintendo
# Albert Ferreiro Castilla
# 21/08/2016

rm(list=ls())                              # Removes all variables from workspace
library("quantmod")                        # Library to get prices from Yahoo
options("getSymbols.warning4.0"=FALSE)     # set off warnings
cat("\f")                                  # Clear console


getSymbols('NTDOY',scr='yahoo',from="2012-01-01",to="2016-08-20")
nint_close=NTDOY$NTDOY.Close

price=as.numeric(nint_close)
returns=log(price[2:length(price)]/price[1:length(price)-1])


pdf("nintendo01.pdf",width=10,height=6)
plot(price, type="l",main="Nintendo stock price")
dev.off()

pdf("nintendo02.pdf",width=10,height=6)
plot(returns,type="l", main="Nintendo daily return")
dev.off()
```

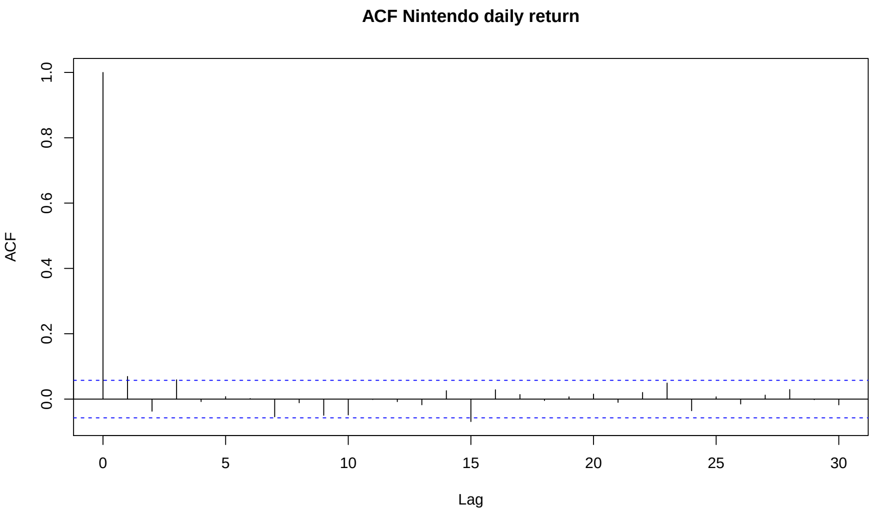




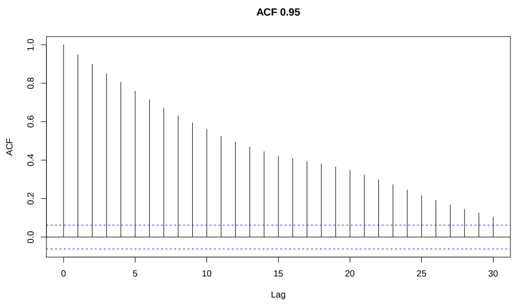
```
...

pdf("nintendo01_ACF.pdf",width=10,height=6)
acf(price, main="ACF Nintendo stock price")
dev.off()

pdf("nintendo02_ACF.pdf",width=10,height=6)
acf(returns, main="ACF Nintendo daily return")
dev.off()

Box.test (returns, lag = 20, type = "Ljung")
```

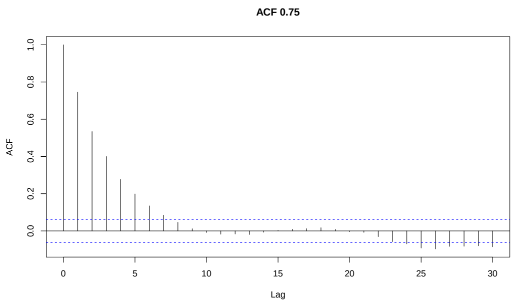
```
      Box-Ljung test

data:  returns
X-squared = 29.1, df = 20, p-value = 0.08581
```

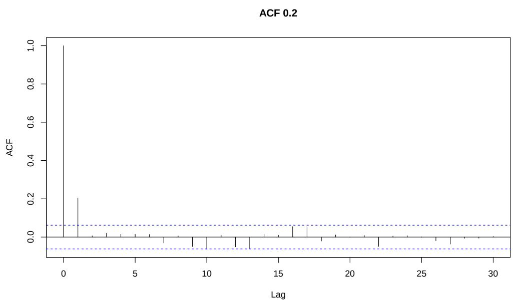


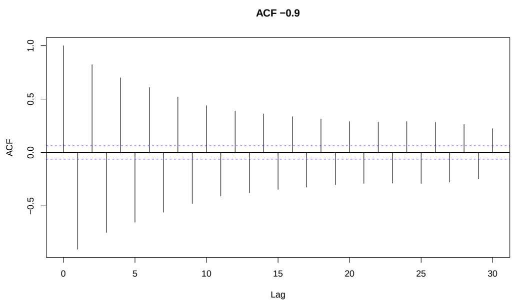


```
rm(list=ls())                                    # Removes all variables from workspace
cat("\f")                                        # Clear console

y.95 <- arima.sim(model=list(ar=.95), n=1000)

pdf("AR_ACF_1.pdf",width=10,height=6)
acf(y.95, main="ACF 0.95")
dev.off()

y.75 <- arima.sim(model=list(ar=.75), n=1000)

pdf("AR_ACF_2.pdf",width=10,height=6)
acf(y.75, main="ACF 0.75")
dev.off()

y.2 <- arima.sim(model=list(ar=.2), n=1000)

pdf("AR_ACF_3.pdf",width=10,height=6)
acf(y.2, main="ACF 0.2")
dev.off()

y.99 <- arima.sim(model=list(ar=-.9), n=1000)

pdf("AR_ACF_4.pdf",width=10,height=6)
acf(y.99, main="ACF -0.9")
dev.off()
```







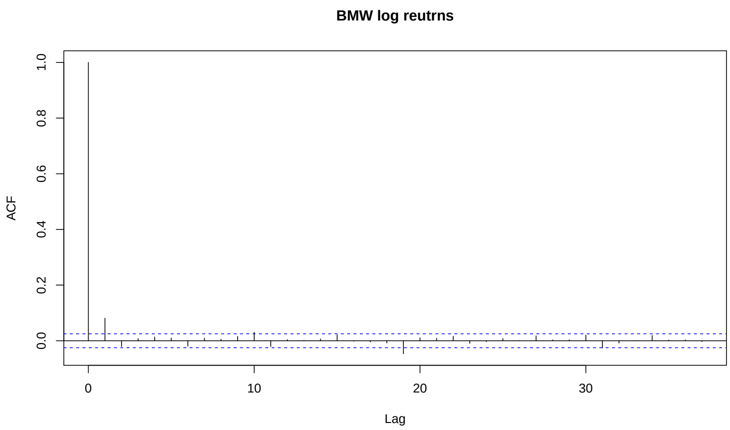


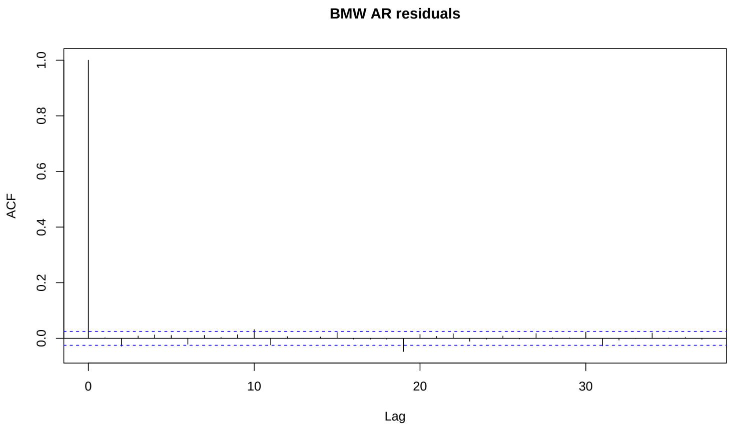
```
rm(list=ls())                        # Removes all variables from workspace
cat("\f")                            # Clear console
library("evir")
data(bmw,package="evir")             # Load BMW log prices

pdf("BMW_1.pdf",width=10,height=6)   # ACF for log prices
acf(bmw, main="BMW log reutrns")
dev.off()
Box.test (bmw, lag = 5, type = "Ljung")     # Evidence of not being a white noise
fit<-arima(x=bmw,order=c(1,0,0))     # Fit of AR(1)
fit                                  # Showing the parameters
                                     # Final analysis of the residuals
pdf("BMW_2.pdf",width=10,height=6)
acf(fit$resid , main="BMW AR residuals")
Box.test (fit$resid , lag = 5, type = "Ljung",fitdf = 1)
dev.off()
```

```
      Box-Ljung test
data:  bmw
X-squared = 44.987, df = 5, p-value = 1.46e-08
```

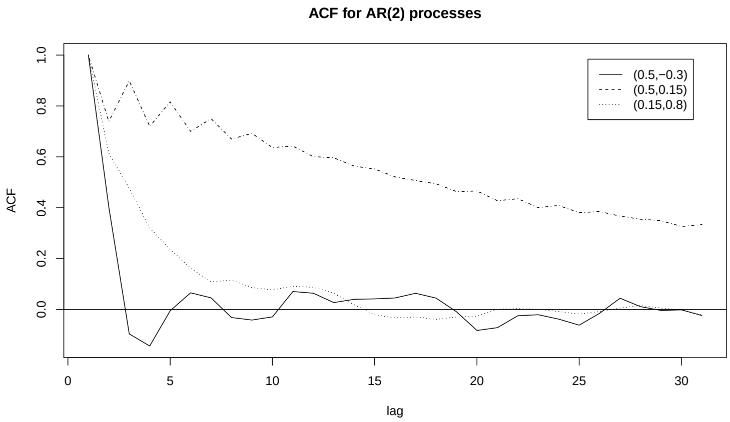
```
Call:
arima(x = bmw, order = c(1, 0, 0))
Coefficients:
         ar1   intercept
      0.0811       3e-04
s.e.  0.0127       2e-04

sigma^2 estimated as 0.0002163:  log likelihood = 17212.34,  aic = -34418.68
```

```
      Box-Ljung test
data:  fit$resid
X-squared = 6.8669, df = 4, p-value = 0.1431
```





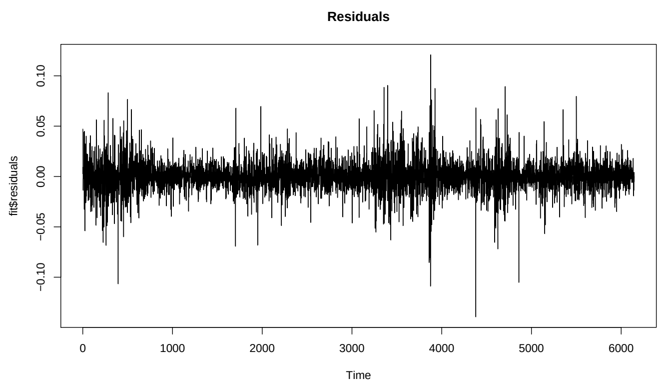
```r
y <- arima.sim(model=list(order=c(2,0,0), ar=c(0.5,-0.3)), n=1000)
z <- arima.sim(model=list(order=c(2,0,0), ar=c(0.5,0.15)), n=1000)
v <- arima.sim(model=list(order=c(2,0,0), ar=c(0.15,0.8)), n=1000)

yy<-acf(y)
zz<-acf(z)
vv<-acf(v)

pdf("AR2_ACF.pdf",width=10,height=6)
plot(yy$acf, type="l",main="ACF for AR(2) processes",xlab="lag",ylab="ACF")
lines(zz$acf,lty=3)
lines(vv$acf,lty=4)
legend("topright", inset=.05, cex = 1, c("(0.5,-0.3)","(0.5,0.15)","(0.15,0.8)"), horiz=FALSE, lty=c(1,2,3))
abline(h=0,lty=1,lwd=1)
dev.off()
```

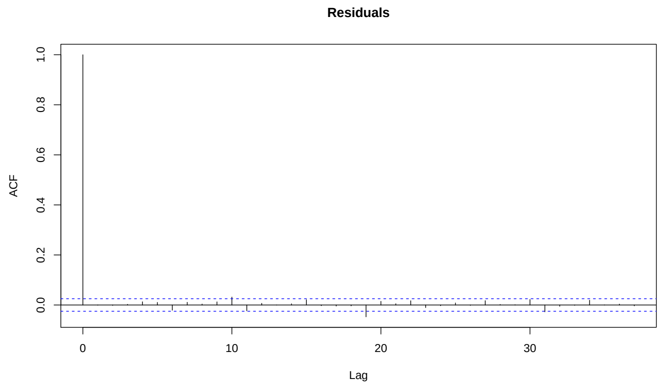


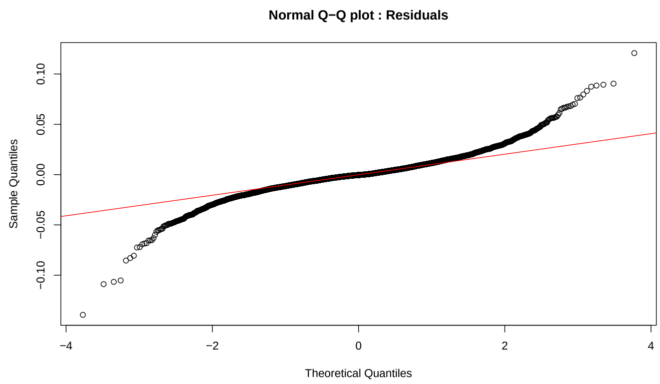
```
rm(list=ls())                                    # Removes all variables from workspace
cat("\f")                                        # Clear console

library("evir")
data(bmw, package="evir")                        # Load BMW log prices

aic_bic<-array(dim=c(9,4))
k=1
for (i in 0:2) {
  for (j in 0:2){
    fit<-arima(x=bmw, order=c(i,0,j))
    aic_bic[k,1]=i
    aic_bic[k,2]=j
    aic_bic[k,3]=AIC(fit)+34500
    aic_bic[k,4]=AIC(fit,k=log(length(bmw)))+34500
    k=k+1
  }
}
colnames(aic_bic) <- c("p","q","AIC","BIC")
aic_bic
```

```
 aic_bic
      p q       AIC       BIC
 [1,] 0 0 119.82867 133.27578
 [2,] 0 1  79.21393  99.38461
 [3,] 0 2  78.36773 105.26195
 [4,] 1 0  81.31575 101.48642
 [5,] 1 1  78.25941 105.15364
 [6,] 1 2  80.44557 114.06336
 [7,] 2 0  78.73415 105.62838
 [8,] 2 1  80.58248 114.20026
 [9,] 2 2  82.22671 122.56805
```

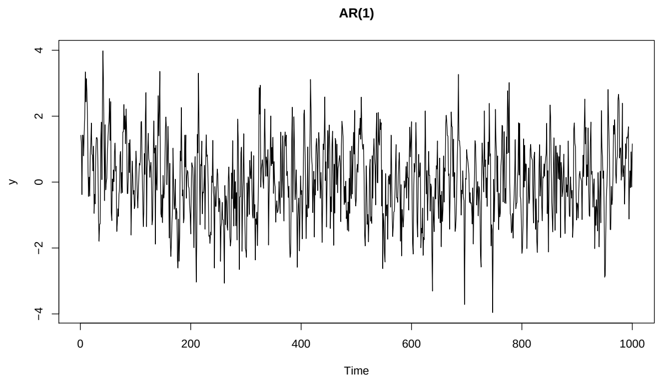
```r
rm(list=ls())                              # Removes all variables from workspace
cat("\f")                                  # Clear console

library("evir")
data(bmw, package="evir")                  # Load BMW log prices

fit<-arima(x=bmw, order=c(1,0,1))

pdf("BMW_res_plot.pdf",width=10,height=6)
plot(fit$residuals , type="l",main="Residuals")
dev.off()

pdf("BMW_res_acf.pdf",width=10,height=6)
acf(fit$residuals , main="Residuals")
dev.off()

pdf("BMW_res_qqplot.pdf",width=10,height=6)
qqnorm(fit$residuals ,main="Normal Q-Q plot : Residuals")
qqline(fit$residuals ,col=2)
dev.off()
```

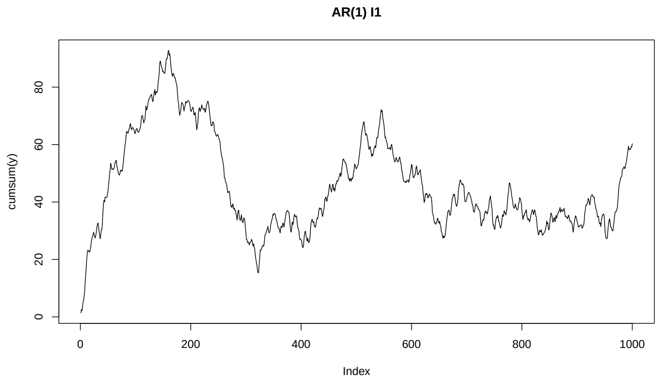


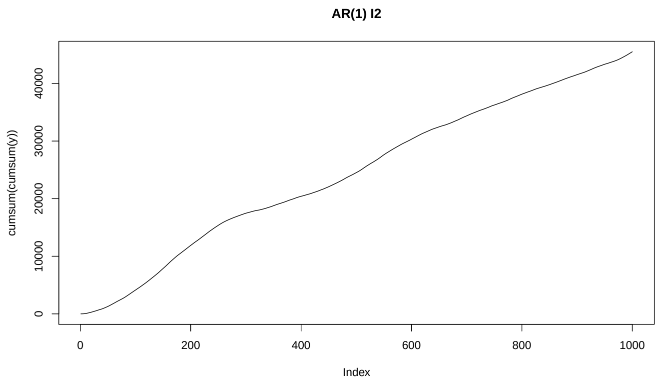




```
rm(list=ls())                                    # Removes all variables from workspace
cat("\f")                                        # Clear console

y <- arima.sim(model=list(order=c(1,0,0), ar=c(0.5)), n=1000)

pdf("AR1.pdf",width=10,height=6)
plot(y, type="l",main="AR(1)")
dev.off()

pdf("AR1_I1.pdf",width=10,height=6)
plot(cumsum(y), type="l",main="AR(1) I1")
dev.off()

pdf("AR1_I2.pdf",width=10,height=6)
plot(cumsum(cumsum(y)), type="l",main="AR(1) I2")
dev.off()
```

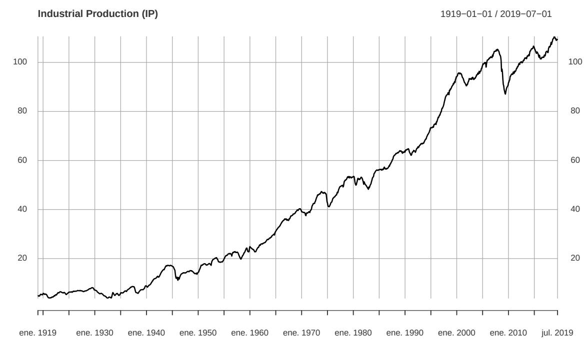


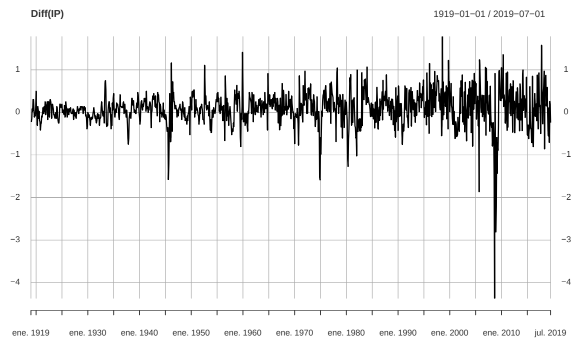


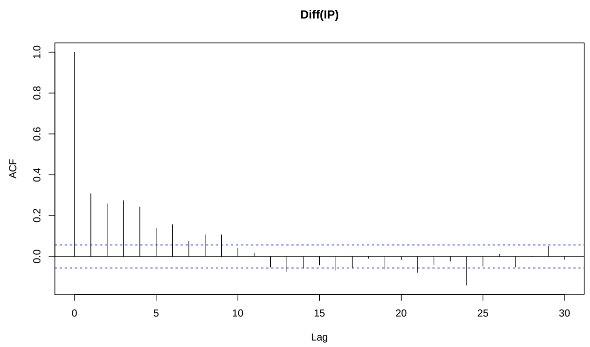


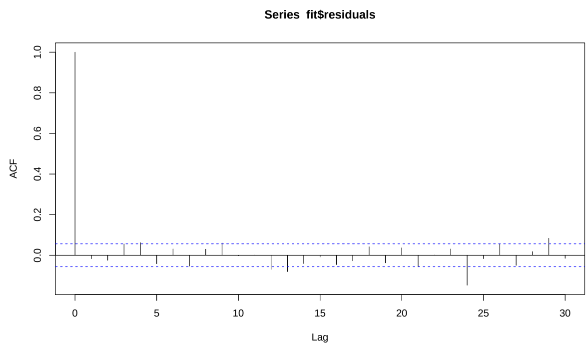
```r
rm(list=ls())                          # Removes all variables from workspace
cat("\f")                              # Clear console

library("quantmod")                    # Library to get prices from Yahoo
options("getSymbols.warning4.0"=FALSE) # set off warnings
cat("\f")                              # Clear console

getSymbols.FRED('INDPRO',env=globalenv())

pdf("IP.pdf",width=10,height=6)
plot(INDPRO,main="Industrial Production (IP)")
dev.off()

pdf("IP_diff.pdf",width=10,height=6)
plot(diff(INDPRO),main="Diff(IP)")
dev.off()

pdf("IP_diff_ACF.pdf",width=10,height=6)
acf(diff(INDPRO),na.action = na.pass,main="Diff(IP)")
dev.off()

fit<-arima(x=diff(INDPRO),order=c(1,0,1))
fit

pdf("IP_diff_ACF_res.pdf",width=10,height=6)
acf(fit$residuals,na.action = na.pass)
dev.off()
```

```
 Call:
arima(x = diff(INDPRO), order = c(1, 0, 1))

Coefficients:
         ar1      ma1  intercept
      0.8492  -0.6189     0.0857
s.e.  0.0279   0.0398     0.0275

sigma^2 estimated as 0.1399:  log likelihood = -509.76,  aic = 1027.52
```









UNIT ROOT TEST

We have seen the difficulty to tell whether a time series is best modeled as stationary or nonstationary. ~~To~~ To help decide between both we use a unit root test.

- Recall that an ARMA (p,q) process can be written as

$$Y_t - \mu = \phi_1 (Y_{t-1} - \mu) + \cdots + \phi_p (Y_{t-p} - \mu) + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

It turns out that the condition for Y_t to be stationary is that all roots of the polynomial

$$1 - \phi_1 x - \cdots - \phi_p x^p$$

have ~~all roots~~ absolute value greater than one.

- There are 3 main unit root tests

Dickey-Fuller: The null hypothesis is that there is a unit root, and the alternative that the process is stationary

Phillips-Perron: Similar to Dickey-Fuller with minor changes.

KPPS: The null hypothesis is that the process is stationary and the alternative that there is a unit root

The function "auto.arima" in R executes the KPPS test recursively until a differentiated series cannot reject the null hypothesis and fits an ARMA to it.

```
rm(list=ls())                              # Removes all variables from workspace
cat("\f")                                  # Clear console

library("quantmod")                        # Library to get prices from Yahoo
library("forecast")
options("getSymbols.warning4.0"=FALSE)     # set off warnings
cat("\f")                                  # Clear console

getSymbols.FRED('INDPRO',env=globalenv())

fit<-auto.arima(x=INDPRO)
fit
```

```
Series: INDPRO
ARIMA(1,1,1) with drift

Coefficients:
         ar1      ma1   drift
      0.8491  -0.6185  0.0860
s.e.  0.0280   0.0399  0.0275

sigma^2 estimated as 0.1403:  log likelihood=-509.76
AIC=1027.52   AICc=1027.56   BIC=1047.78
```

FORECASTING

Forecasting means predicting the future values of a time series using the current information set, which is the set of present and past values of the time series.

→ Consider forecasting an AR(1). Assume known $Y_1, \dots, Y_n$ and estimates $\hat{\mu}$ and $\hat{\phi}$. We know that

$$Y_{n+1} = \mu + \phi(Y_n - \mu) + \varepsilon_{n+1},$$

~~since~~ since ε_{n+1} is independent of the past and the present Y_n, the best prediction of ε_{n+1} is its mean 0. Therefore our best guess is

$$\hat{Y}_{n+1} = \hat{\mu} + \hat{\phi}(Y_n - \hat{\mu})$$

And the reasoning can be iterated to obtain the k-th estimator

$$\hat{Y}_{n+k} = \hat{\mu} + \hat{\phi}(Y_n - \hat{\mu})$$

Note that if $|\hat{\phi}|<1$, as for stationary processes, then as k increases, the forecasts values converge exponentially to the mean.

The same sort of arguments are valid for forecasting more complex processes, for example an ARIMA(1,1,0). One first fits ΔY_t for an AR(1) and forecast ΔY_{t+k}, then we simply use

$$\hat{Y}_{n+1} = Y_n + \widehat{\Delta Y}_{n+1}$$

or iterate the recursion to obtain

$$\hat{Y}_{n+2} = Y_n + \hat{\Delta} Y_{n+1} + \Delta \hat{Y}_{n+2}$$

Finally consider the forecasting for an ARMA(1,1), since the model follows the equation

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + \varepsilon_t + \theta \varepsilon_{t-1}$$

then we use

$$\hat{Y}_{n+1} = \hat{\mu} + \hat{\phi}(Y_n - \hat{\mu}) - \hat{\theta}\hat{\varepsilon}_n$$

and for $k \geq 2$

$$\hat{Y}_{n+k} = \hat{\mu} + \hat{\phi}(Y_{n+k-1} - \hat{\mu})$$

To obtain the uncertainty interval of the forcast we simply compute the forcast standard deviation times $z_{\alpha/2}$, assuming ε_i is a Gaussian white noise. This is the typicall output by most statistical software and R.

The R function "prediet" receeves a model (as of given by the function "arima") and returns the prediction

Partial Autocorrelation Function
the PACF can be useful to identify the order of the AR process. The k-th partial autocorrelation function $\phi_{k,k}$ for stationery process Y_t is the correlation of Y_t and Y_{t+k} condition on given $Y_{t+1}, \dots Y_{t+k-1}$
There is a way to compute PACF reccursivelly, but this was more usefull when resources were scarce, now a days BIC and AIC criteria are more common.

For example an AR(p) process has $\phi_{kk}=0$ for $k>p$.

GARCH Models

As seen previously, financial market data exhibit volatility clustering, meaning that periods of high volatility tend to give pass to periods of low volatility. In other words, it seems convenient to model a process with time_varying volatility. As we have seen the ARMA process are used to model conditional expectation given the past but the conditional variance given the past is constant. GARCH models adress this issue.

First we will introduce ARCH models, AutoRegressive Conditional Heteroscedasticity.

ARCH(1)
Let ~~the~~ $\{\varepsilon_t\}_t$ be a Gaussian $WN(0,1)$. We say $\{Y_t\}_t$ is an ARCH(1) process if for some constant parameter $\alpha \geq 0$ and $\omega > 0$,

$$Y_t = \sqrt{\omega + \alpha Y_{t-1}^2}\, \varepsilon_t$$

- For Y to be stationary we require $\alpha < 1$.
- The equation driving an ARCH(1) model is very similar to an AR model, only that we have now a multiplicative noise.

① Since ε_t is independent of Y_{t-1}, we have

$$\mathbb{E}[Y_t / Y_{t-1}, \ldots] = 0$$

and

$$\sigma_t : Var(Y_t / Y_{t-1}, \ldots) = \omega + \alpha Y_{t-1}^2 \quad \left(\approx \text{ similar to AR(1)}\right)$$

From the last equation we can deduce the formation of volatility clusters. This is an straight forward example of an uncorrelated process with dependence.

Although the process Y_t is uncorrelated and hence has a trivial autocorrelation function, the process Y_t^2 has a more interesting ACF function:

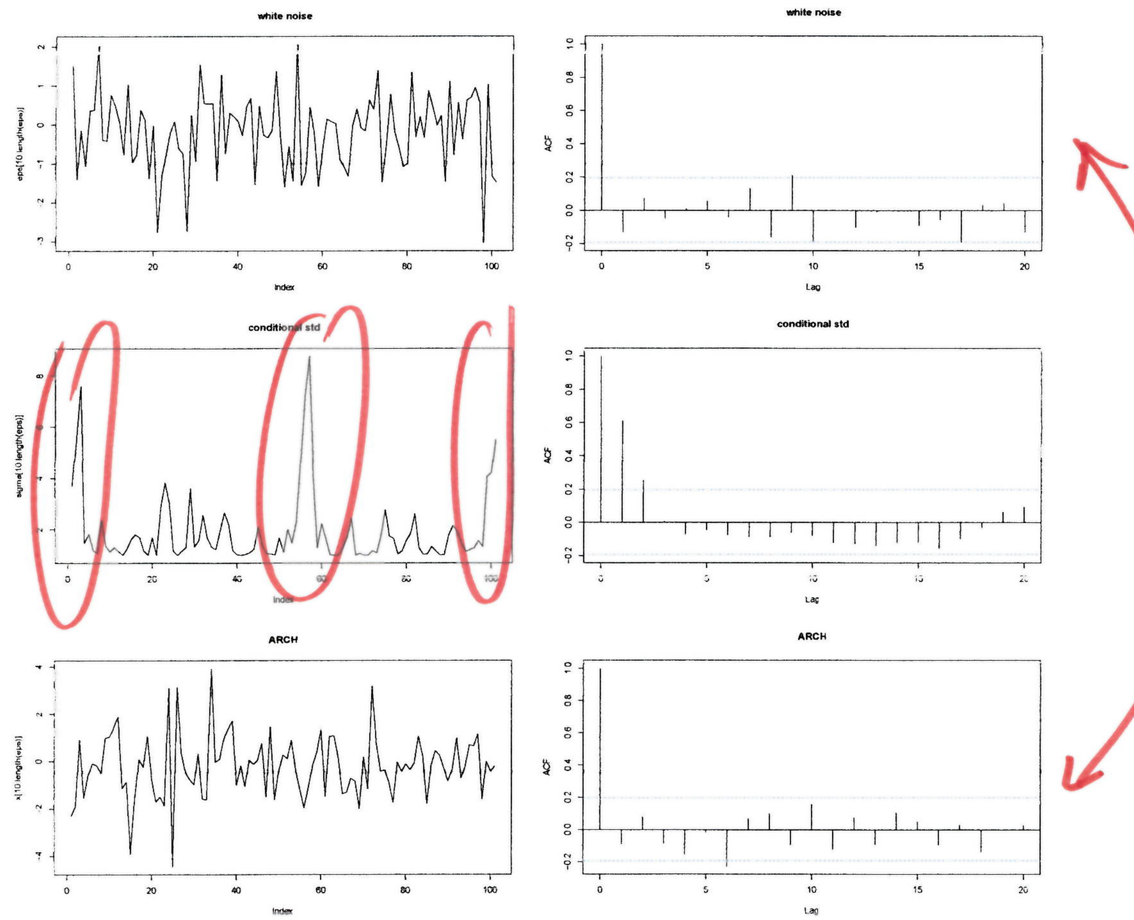
$$\rho(h) = \alpha_1^{|h|}$$

for $\alpha < 1$, otherwise the process is nonstationary or has infinite variance which in either case means not having an ACF:

In the next code we show:

① ~~IID~~ White noise
② The conditional standard deviation process
③ The ARCH(1) process.

```r
eps=rnorm(110)
x <- vector(mode="numeric", length=110)
sigma <- vector(mode="numeric", length=110)
x[1]<-0
sigma[1]<-0

for (i in 2:110){
  x[i]<-sqrt(1+0.95*(x[i-1])^2)*eps[i]
  sigma[i]<-sqrt(1+0.95*(x[i-1])^2)
}

plot(eps[10:length(eps)],main="white noise",type="l")
acf(eps[10:length(eps)],main="white noise")
plot(sigma[10:length(eps)],main="conditional std",type="l")
acf(sigma[10:length(eps)],main="conditional std")
plot(x[10:length(eps)],main="ARCH",type="l")
acf(x[10:length(eps)],main="ARCH")
```



Quite similar but completely different processes!

As we have seen the AR(1) has a nonconstant conditional mean but a constant conditional variance.
ARCH(1) is just the opposite.

If both the conditional mean and variance depend on the past, we can combine both process.

AR(1) / ARCH(1)

Let a_t be an ARCH(1) process, then an AR(1) / ARCH(1) process $\{Y_t\}_t$ is defined by the equation:

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + a_t$$

The noise process is a weak white noise, and because it is uncorrelated, the ACF of an AR(1) is the same as an AR(1) / ARCH(1)

The above ideas can be generalised to higher orders:

ARCH(p)

Let $\{\varepsilon_i\}_i$ be a Gaussian WN(0,1). We say that $\{Y_n\}_n$ is an ARCH(p) process if for some constant parameters $\alpha_i \geq 0$ and $\omega \geq 0$,

$$Y_t = \sqrt{\omega + \sum_{i=1}^{p} \alpha_i Y_{t-i}^2}\, \varepsilon_t$$

Like an ARCH(1) process, the above is uncorrelated and has a constant mean (both conditional and unconditional). For the process to be weakly stationary we require

$$\sum_{i=1}^{p} \alpha_i < 1.$$

It has constant unconditional variance, but the conditional variance is not constant. In fact the ACF for the squared process Y_t^2 is the same as for an AR(q)

The drawback on ARCH(p) process is that the conditional volatility behaves with shocks. To keep the volatility smoother but still exhibiting clusters we introduce the GARCH(p,q) model, which feeds the past volatility terms in an ARCH process.

GARCH(p,q)

Let $\{\varepsilon_t\}_t$ be a Gaussian WN(0,1). We say that $\{Y_n\}_n$ is a GARCH(p,q) process if for some constant parameters $\alpha_i \geq 0$, $\beta_i \geq 0$ and $\omega > 0$

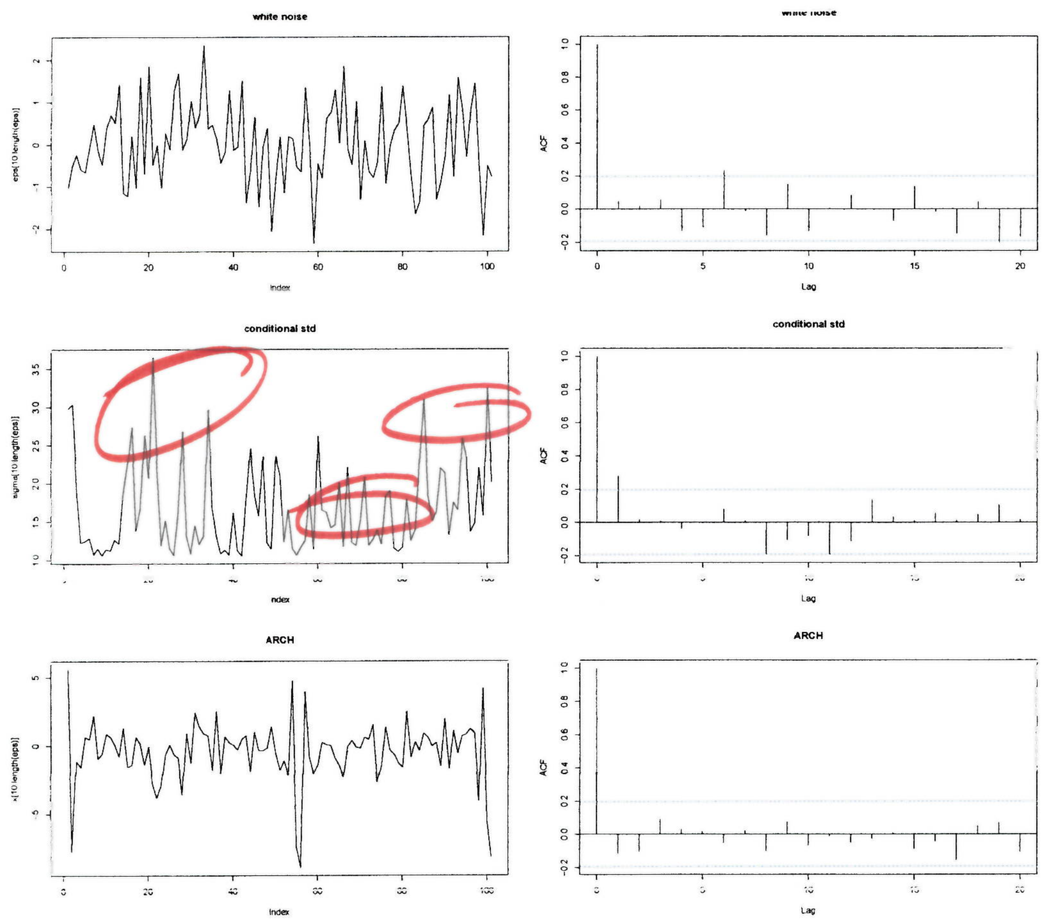
$$Y_t = \sigma_t \varepsilon_t$$

where $$\sigma_t = \sqrt{\omega + \sum_{i=1}^{p} \alpha_i Y_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2}$$

For the process to be stationary weakly we require

$$\sum_{i=1}^{p} \alpha_i + \sum_{i=1}^{q} \beta_i < 1$$

```r
rm(list=ls())                          # Removes all variables from workspace
library(tseries)
cat("\f")

eps=rnorm(110)

x <- vector(mode="numeric", length=110)
sigma <- vector(mode="numeric", length=110)
x[1]<-0
sigma[1]<-0
for (i in 2:110){
  sigma[i]<-sqrt(1+0.8*(x[i-1])^2+0.09*(sigma[i-1])^2)
  x[i]<-sigma[i]*eps[i]
}

plot(eps[10:length(eps)],main="white noise",type="l")
acf(eps[10:length(eps)],main="white noise")
plot(sigma[10:length(eps)],main="conditional std",type="l")
acf(sigma[10:length(eps)],main="conditional std")
plot(x[10:length(eps)],main="GARCH",type="l")
acf(x[10:length(eps)],main="GARCH")
```



The process y^2 has similar ACF to an ARMA process and exhibits more persistent periods of volatility clustering.

Now, we just need to combine the all the processes we know

$\boxed{\text{ARIMA}(p,d,q) / \text{GARCH}(P,Q)}$

Let a_t be an GARCH(P,Q) process, then an ARIMA(p,d,q) / GARCH(P,Q) $\{Y_n\}_n$ is an ARIMA(p,d,q) process where the noise term is given by a GARCH(P,Q) process.

When one fits an ARIMA(p,d,q) / GARCH(P,Q) process we find two types of residuals.

- → The ordinary residual $\hat{Y}_t$ is the difference between Y_t and its conditional expectation.
- → The standarized residual error, $\hat{\varepsilon}_t$, is an ordinary residual divided by its standard deviation $\hat{\sigma}_t$

One of the key features of GARCH models are conditional heteroscedasticity and heavy tails, two features we ~~alr~~ already have seen in financial returns time series

The following code, fits an ARMA/GARCH to the already seen BMW time series.

```
library(fGarch)
data(bmw, package="evir")
fit<-garchFit(formula=~arma(1,0)+garch(1,1), data=bmw, cond.dist="norm")
```

```
Call:
 garchFit(formula = ~arma(1, 0) + garch(1, 1), data = bmw, cond.dist = "norm")

Mean and Variance Equation:
 data ~ arma(1, 0) + garch(1, 1)
<environment: 0x10fd753b8>
 [data = bmw]

Conditional Distribution:
 norm

Coefficient(s):
        mu          ar1        omega       alpha1       beta1
4.0092e-04   9.8596e-02   8.9043e-06   1.0210e-01   8.5944e-01

Std. Errors:
 based on Hessian

Error Analysis:
         Estimate  Std. Error  t value Pr(>|t|)
mu      4.009e-04   1.579e-04    2.539   0.0111 *
ar1     9.860e-02   1.431e-02    6.888 5.65e-12 ***
omega   8.904e-06   1.449e-06    6.145 7.97e-10 ***
alpha1  1.021e-01   1.135e-02    8.994  < 2e-16 ***
beta1   8.594e-01   1.581e-02   54.348  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Log Likelihood:
 17757.16    normalized:  2.889222
```

$\hat{\phi}$ is statistical significant also α_1 and β_1 from the GARCH component.

$\hat{\beta}_1$ is large $\Rightarrow$ shows persisting volatility clusters

```
Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  11377.99  0
 Shapiro-Wilk Test  R    W      NA        NA
 Ljung-Box Test     R    Q(10)  15.15693  0.1264445
 Ljung-Box Test     R    Q(15)  20.09345  0.168377
 Ljung-Box Test     R    Q(20)  30.54788  0.06144917
 Ljung-Box Test     R^2  Q(10)  5.032717  0.8889818
 Ljung-Box Test     R^2  Q(15)  7.539272  0.9409245
 Ljung-Box Test     R^2  Q(20)  9.277229  0.9794681
 LM Arch Test       R    TR^2   6.03254   0.9144329
```

```
library(fGarch)
data(bmw, package="evir")
fit<-garchFit(formula=~arma(1,1)+garch(1,1), data=bmw, cond.dist="std")
```

```
Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  13355.05  0
 Shapiro-Wilk Test  R    W      NA        NA
 Ljung-Box Test     R    Q(10)  21.93247  0.01545233
 Ljung-Box Test     R    Q(15)  26.50076  0.03307727
 Ljung-Box Test     R    Q(20)  36.7898   0.01239977
 Ljung-Box Test     R^2  Q(10)  5.828522  0.8294584
 Ljung-Box Test     R^2  Q(15)  8.09067   0.9200962
 Ljung-Box Test     R^2  Q(20)  10.73302  0.9528543
 LM Arch Test       R    TR^2   7.009039  0.8570157
```

The output also includes the following test applied to the residuals and squared residuals.

$\rightarrow$ The Jarque-Bera test for normality strongly rejects the null hypothesis, ie the white noise ε_t is not Gaussian.

$\rightarrow$ one way to adress the above issue is to fit the model to a t-distribution

dist = "std"

FITTING ARCH MODELS

```r
#Fitting and Diagnostic Checking for ARCH Models

library("forecast")
library("rugarch")
library("tseries")
require(fBasics)

rm(list=ls())
cat("\f")

n <- 600
arch2.spec = ugarchspec(variance.model = list(garchOrder=c(2,0)),
                        mean.model = list(armaOrder=c(0,0)),
                        fixed.pars=list(mu = 0, omega=0.25, alpha1=0.6, alpha2=0.35))
arch2.sim <- ugarchpath(arch2.spec, n.sim=n)
x <- drop(arch2.sim@path$seriesSim)[101:600]
time <- 1:length(x)
```

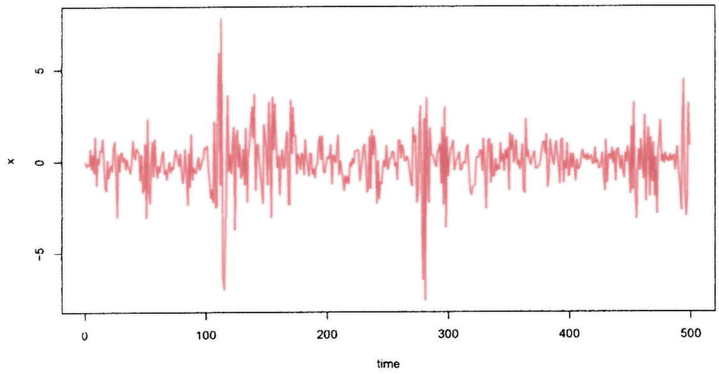
In the following slides we will follow a ~~step~~-by-step process to fit an check ARCH models for a given time series.

The procedure do not restrict to ARCH, ~~but~~ and can be used generally.

Assumed that you are given the above generated time series, obviously you DO NOT know the model that generate the data.

Let's try to identify the model generator!!

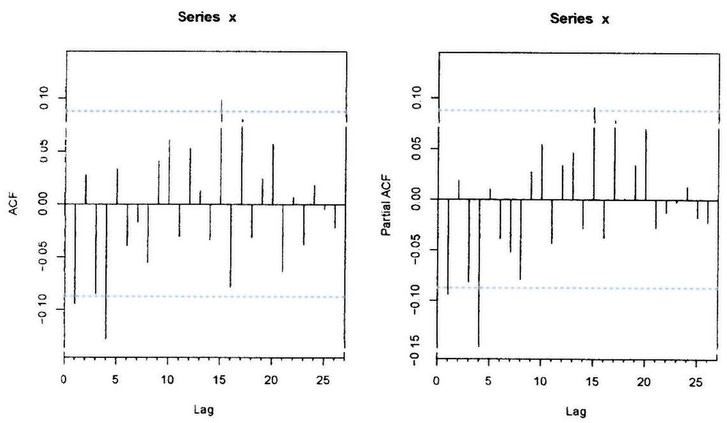
```r
plot(time, x, type="l", col=2)
```



We first plot the time series. Which stylized features can you see from the plot?

- Mean reversion ✓
- Stationary
- Volatility clustering ✓
- Integration ✗
- ....

```
par(mfrow=c(1,2))
Acf(x)
Pacf(x)
```

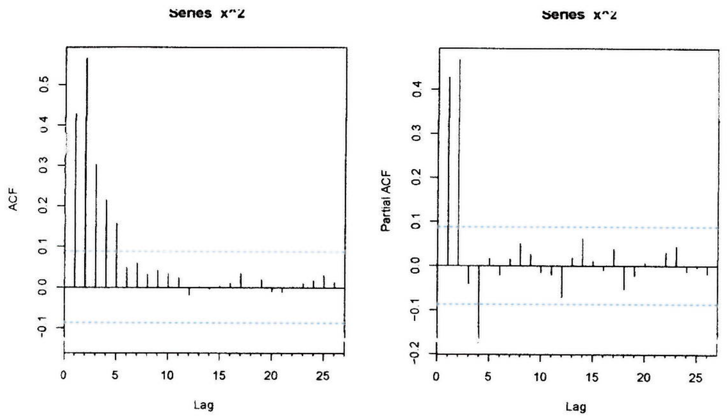


Lets plot the ACF and PACF function for the original time series.

- The ACF shows no need of differentiation
- The PACF is compatible with WN

Hence no need for an ARMA term.

```
par(mfrow=c(1,2))
Acf(x^2)
Pacf(x^2)
```



Lets plot the ACF & PACF for the squared process.

- Sharp decline for PACF, meaning AR term with lower order.
- ACF no so much compatible with MA, but not completely.

⇒ There is no trend part, we just need to fit the volatility part of the process.

Fit ARCH(q) model with "tseries"

```
fit1 <- garch(x, c(0,1), trace=FALSE) # ARCH(1)
fit2 <- garch(x, c(0,2), trace=FALSE) # ARCH(2)
fit3 <- garch(x, c(0,3), trace=FALSE) # ARCH(3)
fit4 <- garch(x, c(0,4), trace=FALSE) # ARCH(4)
fit5 <- garch(x, c(0,5), trace=FALSE) # ARCH(5)

N <- length(x)
loglik1 <- logLik(fit1)
loglik2 <- logLik(fit2)
loglik3 <- logLik(fit3)
loglik4 <- logLik(fit4)
loglik5 <- logLik(fit5)

loglik <- c(loglik1, loglik2, loglik3, loglik4, loglik5)
q <- c(1, 2, 3, 4, 5)
k <- q + 1
aicc <- -2 * loglik  + 2 * k * N / (N - k - 1)

print(data.frame(q, loglik, aicc))
```

```
  q    loglik     aicc
1 1 -705.7442 1415.513
2 2 -666.3854 1338.819
3 3 -729.4100 1466.901
4 4 -734.6751 1479.472
5 5 -729.9601 1472.091
```

Choices are between ARCH(1)
ARCH(2)

Fit ARCH(q) model with "rugarch"

```
spec1 <- ugarchspec(variance.model = list(garchOrder=c(1,0)),
                    mean.model = list(armaOrder=c(0,0), include.mean=FALSE))
fit1 <- ugarchfit(spec1, x)
loglik1 <- fit1@fit$LLH

q <- c(1, 2, 3, 4, 5)
loglik <- rep(NA, length(q))

for (i in 1:length(q)) {
    spec <- ugarchspec(variance.model = list(garchOrder=c(q[i],0)),
                       mean.model = list(armaOrder=c(0,0), include.mean=FALSE))
    fit <- ugarchfit(spec, x)
    loglik[i] <- likelihood(fit)
}

k <- q + 1
aicc <- -2 * loglik  + 2 * k * N / (N - k - 1)

print(data.frame(q, loglik, aicc))
```

```
  q    loglik     aicc
1 1 -707.5218 1419.068
2 2 -669.4327 1344.914
3 3 -669.2573 1346.595
4 4 -669.1381 1348.398
5 5 -668.1957 1348.562
```

Final choice could be ARCH(2).

Model selection with "tseries"


```r
fit.tseries <- garch(x, c(0, 2), trace=FALSE)
summary(fit.tseries)
```

```
Call:
garch(x = x, order = c(0, 2), trace = FALSE)

Model:
GARCH(0,2)

Residuals:
     Min       1Q   Median      3Q     Max
-2.96390 -0.69896 -0.06408 0.66584 2.75127

Coefficient(s):
   Estimate  Std. Error  t value Pr(>|t|)
a0  0.23304     0.04186    5.567 2.59e-08 ***
a1  0.57147     0.10637    5.372 7.77e-08 ***
a2  0.37627     0.08302    4.532 5.84e-06 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Diagnostic Tests:
	Jarque Bera Test

data:  Residuals
X-squared = 2.6731, df = 2, p-value = 0.2627


	Box-Ljung test

data:  Squared.Residuals
X-squared = 0.21612, df = 1, p-value = 0.642
```

ε ✓

Model selection with "rugarch"

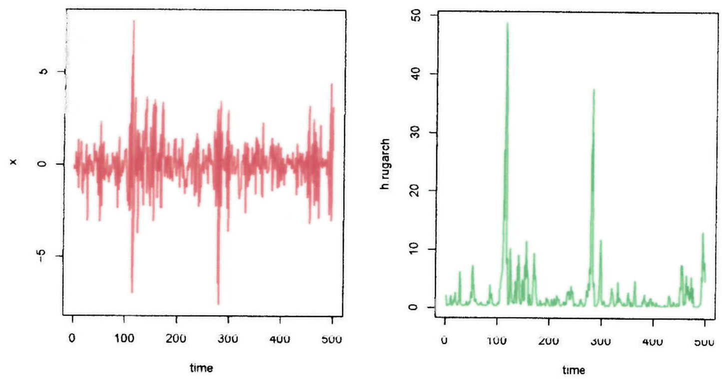
WEEK 4

```
spec <- ugarchspec(variance.model = list(garchOrder=c(2,0)),
                   mean.model = list(armaOrder=c(0,0), include.mean=FALSE))
fit.rugarch <- ugarchfit(spec, x)
print(fit.rugarch)
```

```
*---------------------------------*
*          GARCH Model Fit        *
*---------------------------------*

Conditional Variance Dynamics
-----------------------------------
GARCH Model    : sGARCH(2,0)
Mean Model     : ARFIMA(0,0,0)
Distribution   : norm

Optimal Parameters
------------------------------------
        Estimate  Std. Error  t value Pr(>|t|)
omega    0.23305    0.036103   6.4550    0e+00
alpha1   0.57147    0.102465   5.5772    0e+00
alpha2   0.37627    0.078854   4.7718    2e-06

Robust Standard Errors:
        Estimate  Std. Error  t value Pr(>|t|)
omega    0.23305    0.033225   7.0142    0e+00
alpha1   0.57147    0.107341   5.3238    0e+00
alpha2   0.37627    0.078140   4.8154    1e-06

LogLikelihood : -669.4327

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

Akaike       2.6897
Bayes        2.7150
Shibata      2.6897
Hannan-Quinn 2.6997

Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic p-value
Lag[1]                   0.003676  0.9517
Lag[2*(p+q)+(p+q)-1][2]  0.029782  0.9724
Lag[4*(p+q)+(p+q)-1][5]  0.972187  0.8659
d.o.f=0
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic p-value
Lag[1]                     0.2444  0.6210
Lag[2*(p+q)+(p+q)-1][5]    1.4494  0.7524
Lag[4*(p+q)+(p+q)-1][9]    4.8919  0.4455
d.o.f=2

Weighted ARCH LM Tests
------------------------------------
            Statistic Shape Scale P-Value
ARCH Lag[3]     1.401 0.500 2.000  0.2365
ARCH Lag[5]     1.635 1.440 1.667  0.5576
ARCH Lag[7]     5.517 2.315 1.543  0.1772

Nyblom stability test
------------------------------------
Joint Statistic:  0.8739
Individual Statistics:
omega  0.3208
alpha1 0.1603
alpha2 0.1501

Asymptotic Critical Values (10% 5% 1%)
Joint Statistic:          0.846 1.01 1.35
Individual Statistic:     0.35 0.47 0.75

Sign Bias Test
------------------------------------
                   t-value   prob sig
Sign Bias          0.04125 0.9671
Negative Sign Bias 0.07285 0.9420
Positive Sign Bias 0.52097 0.6026
Joint Effect       0.41880 0.9363


Adjusted Pearson Goodness-of-Fit Test:
------------------------------------
  group statistic p-value(g-1)
1    20     21.04       0.3346
2    30     32.20       0.3112
3    40     36.00       0.6075
4    50     50.40       0.4179


Elapsed time : 0.0921278
```

Plotting the conditional variance

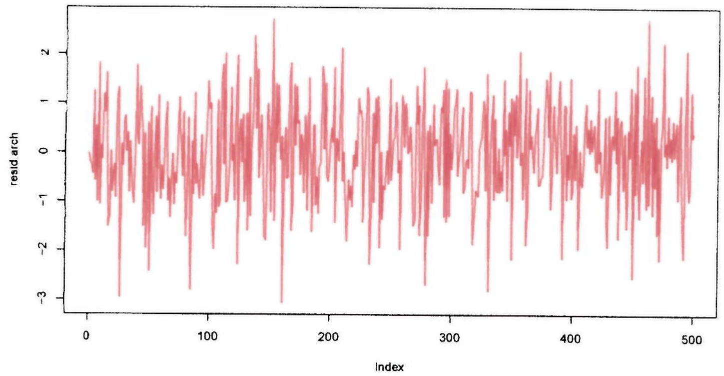

```r
# Using tseries package
h.tseries <- fit.tseries$fit[,1]^2

# Using rugarch package
h.rugarch <- as.numeric(sigma(fit.rugarch)^2)

time <- 1:length(x)
par(mfrow=c(1,2))

plot(time, x, type="l", col=2)
plot(time, h.rugarch, type="l", col=3)
```



- We have chose the model
- We have compute the parameters for the model
- We now need to check the residuals to see that the model is fully explaining the data.

Getting the Residuals

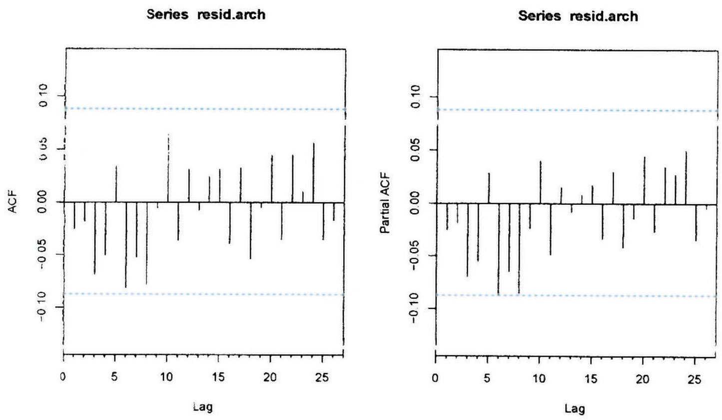
```
# Using tseries package:
resid.tseries <- residuals(fit.tseries)

# Using rugarch package:
resid.rugarch <- as.numeric(residuals(fit.rugarch, standardize=TRUE))

# Note: resid.tseries is equal to (x / sqrt(h.tseries))
#       resid.rugarch is equal to (x / sqrt(h.rugarch))

par(mfrow=c(1,1))
resid.arch <- resid.rugarch
plot(resid.arch, type="l", col=2)
```



Is it compatible with WN?

Check the ACF & PACF

```
par(mfrow=c(1,2))
Acf(resid.arch)
Pacf(resid.arch)
```



It seems compatible with WN.

Let us check it with some tests as well.

Ljung-Box tests.

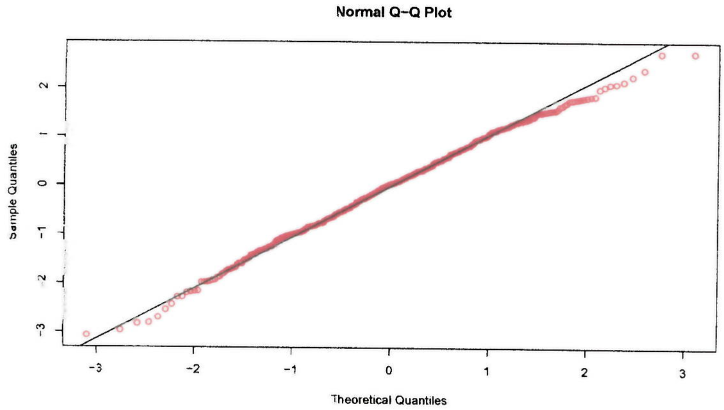
```
Box.test(resid.arch, lag=1, type="Ljung-Box", fitdf=0)
Box.test(resid.arch, lag=12, type="Ljung-Box", fitdf=0)
Box.test(resid.arch, lag=24, type="Ljung-Box", fitdf=0)
Box.test(resid.arch, lag=36, type="Ljung-Box", fitdf=0)
```

```
> Box.test(resid.arch, lag=1, type="Ljung-Box", fitdf=0)

      Box-Ljung test

data:  resid.arch
X-squared = 0.0036762, df = 1, p-value = 0.9517

> Box.test(resid.arch, lag=12, type="Ljung-Box", fitdf=0)

      Box-Ljung test

data:  resid.arch
X-squared = 4.7582, df = 12, p-value = 0.9656

> Box.test(resid.arch, lag=24, type="Ljung-Box", fitdf=0)

      Box-Ljung test

data:  resid.arch
X-squared = 13.96, df = 24, p-value = 0.9476

> Box.test(resid.arch, lag=36, type="Ljung-Box", fitdf=0)

      Box-Ljung test

data:  resid.arch
X-squared = 23.563, df = 36, p-value = 0.945
```

Accept H_0

QQ plot

```
qqnorm(resid.arch, col=2)
qqline(resid.arch, col=1)
```



Normality compatible?

Lets check with a test.

```r
normalTest(resid.arch,method="jb")
```

```
Title:
 Jarque - Bera Normalality Test

Test Results:
  STATISTIC:
    X-squared: 2.5738
  P VALUE:
    Asymptotic p Value: 0.2761

Description:
 Wed Sep 28 22:33:13 2016 by user:
```

WEEK 4

Summary of rules for identifying ARIMA models

Identifying the order of differencing and the constant:

- Rule 1: If the series has positive autocorrelations out to a high number of lags (say, 10 or more), then it probably needs a higher order of differencing.
- Rule 2: If the lag-1 autocorrelation is zero or negative, or the autocorrelations are all small and patternless, then the series does *not* need a higher order of differencing. If the lag-1 autocorrelation is -0.5 or more negative, the series may be overdifferenced. **BEWARE OF OVERDIFFERENCING.**
- Rule 3: The optimal order of differencing is often the order of differencing at which the standard deviation is lowest. (Not always, though. Slightly too much or slightly too little differencing can also be corrected with AR or MA terms. See rules 6 and 7.)
- Rule 4: A model with no orders of differencing assumes that the original series is stationary (among other things, mean-reverting). A model with one order of differencing assumes that the original series has a constant average trend (e.g. a random walk or SES-type model, with or without growth). A model with two orders of total differencing assumes that the original series has a time-varying trend (e.g. a random trend or LES-type model).
- Rule 5: A model with no orders of differencing normally includes a constant term (which allows for a non-zero mean value). A model with two orders of total differencing normally does not include a constant term. In a model with one order of total differencing, a constant term should be included if the series has a non-zero average trend.

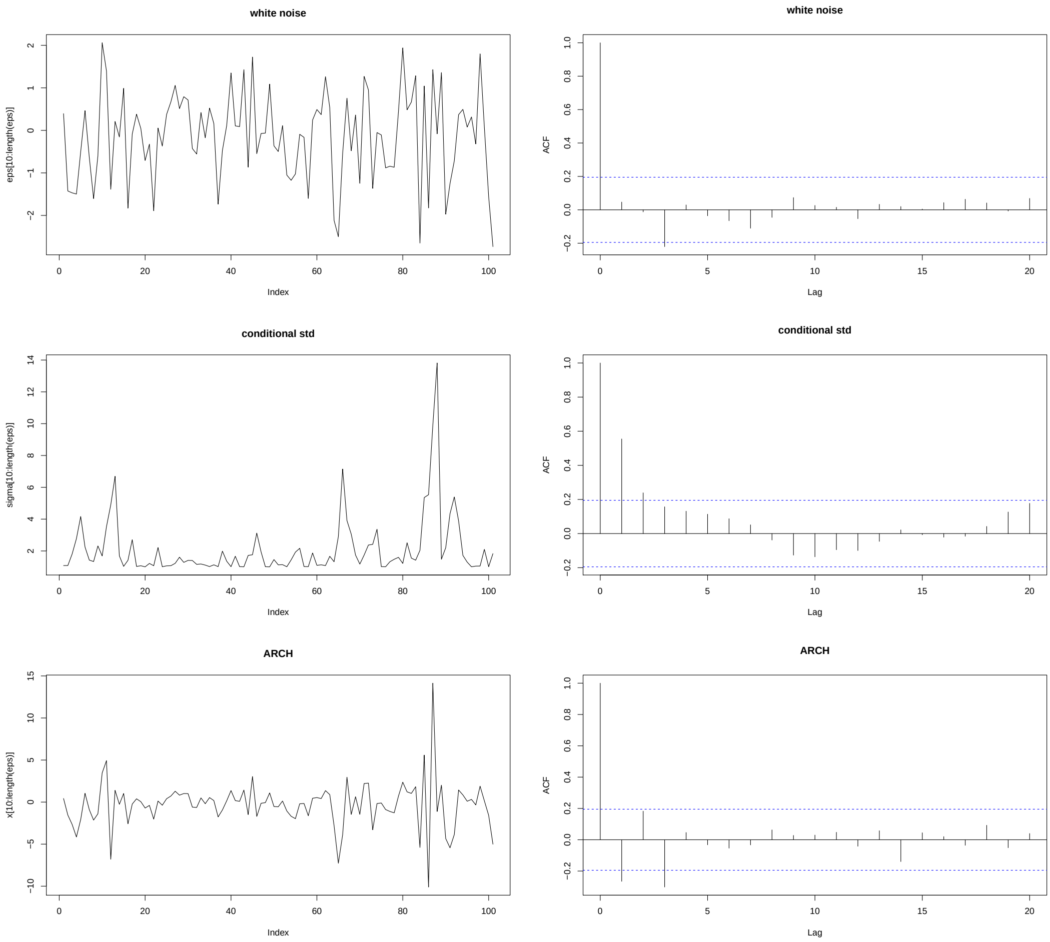
Identifying the numbers of AR and MA terms:

- Rule 6: If the partial autocorrelation function (PACF) of the differenced series displays a sharp cutoff and/or the lag-1 autocorrelation is positive--i.e., if the series appears slightly "underdifferenced"--then consider adding one or more AR terms to the model. The lag beyond which the PACF cuts off is the indicated number of AR terms.
- Rule 7: If the autocorrelation function (ACF) of the differenced series displays a sharp cutoff and/or the lag-1 autocorrelation is negative--i.e., if the series appears slightly "overdifferenced"--then consider adding an MA term to the model. The lag beyond which the ACF cuts off is the indicated number of MA terms.
- Rule 8: It is possible for an AR term and an MA term to cancel each other's effects, so if a mixed AR-MA model seems to fit the data, also try a model with one fewer AR term and one fewer MA term--particularly if the parameter estimates in the original model require more than 10 iterations to converge. **BEWARE OF USING MULTIPLE AR TERMS AND MULTIPLE MA TERMS IN THE SAME MODEL.**
- Rule 9: If there is a unit root in the AR part of the model--i.e., if the sum of the AR coefficients is almost exactly 1--you should reduce the number of AR terms by one and increase the order of differencing by one.
- Rule 10: If there is a unit root in the MA part of the model--i.e., if the sum of the MA coefficients is almost exactly 1--you should reduce the number of MA terms by one and reduce the order of differencing by one.

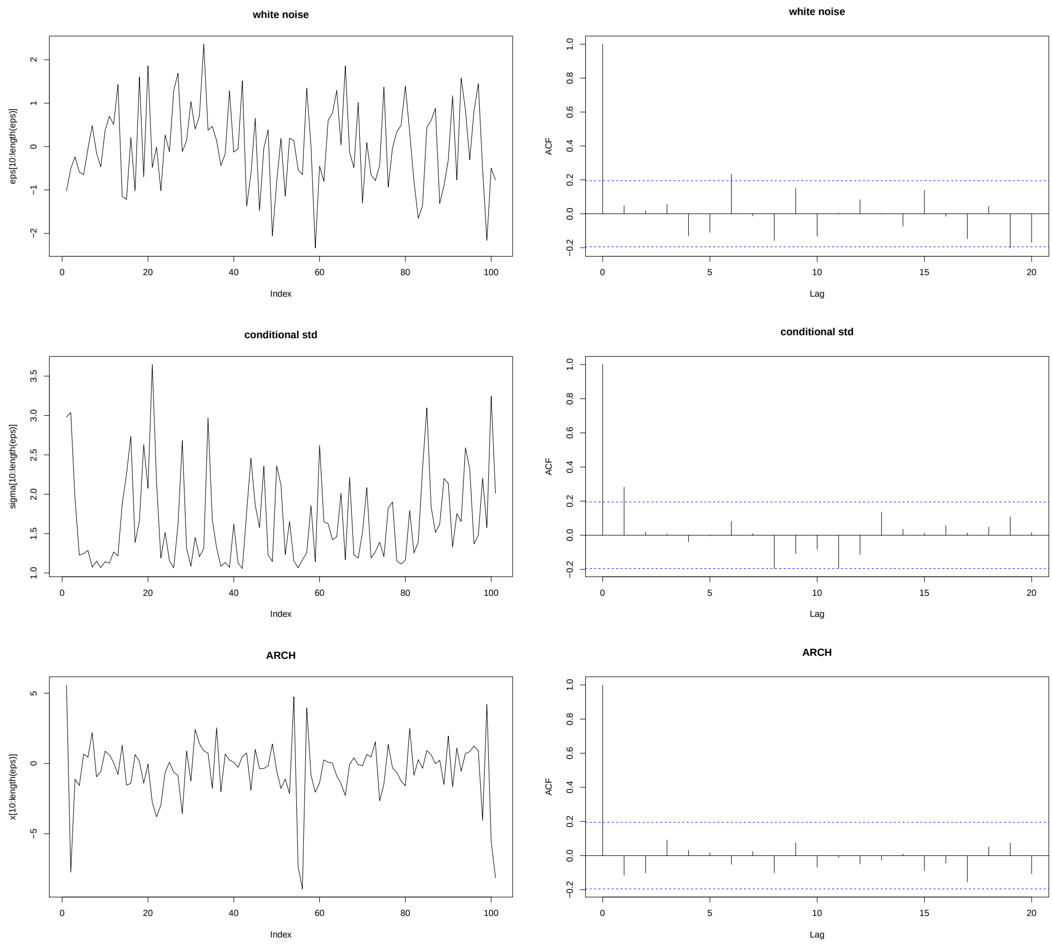
```r
rm(list=ls())                                # Removes all variables from workspace
cat("\f")                                    # Clear console

library("quantmod")                          # Library to get prices from Yahoo
library("forecast")
options("getSymbols.warning4.0"=FALSE)       # set off warnings
cat("\f")                                    # Clear console

getSymbols.FRED('INDPRO',env=globalenv())

fit<-auto.arima(x=INDPRO)
fit
```

```
Series: INDPRO
ARIMA(1,1,1) with drift

Coefficients:
         ar1      ma1    drift
      0.8491  -0.6185   0.0860
s.e.  0.0280   0.0399   0.0275

sigma^2 estimated as 0.1403:  log likelihood=-509.76
AIC=1027.52   AICc=1027.56   BIC=1047.78
```

```r
eps=rnorm(110)
x <- vector(mode="numeric", length=110)
sigma <- vector(mode="numeric", length=110)
x[1]<-0
sigma[1]<-0

for (i in 2:110){
  x[i]<-sqrt(1+0.95*(x[i-1])^2)*eps[i]
  sigma[i]<-sqrt(1+0.95*(x[i-1])^2)
}

plot(eps[10:length(eps)],main="white noise",type="l")
acf(eps[10:length(eps)],main="white noise")
plot(sigma[10:length(eps)],main="conditional std",type="l")
acf(sigma[10:length(eps)],main="conditional std")
plot(x[10:length(eps)],main="ARCH",type="l")
acf(x[10:length(eps)],main="ARCH")
```



```r
rm(list=ls())                                   # Removes all variables from workspace
library(tseries)
cat("\f")

eps=rnorm(110)

x <- vector(mode="numeric", length=110)
sigma <- vector(mode="numeric", length=110)
x[1]<-0
sigma[1]<-0
for (i in 2:110){
  sigma[i]<-sqrt(1+0.8*(x[i-1])^2+0.09*(sigma[i-1])^2)
  x[i]<-sigma[i]*eps[i]
}

plot(eps[10:length(eps)],main="white noise",type="l")
acf(eps[10:length(eps)],main="white noise")
plot(sigma[10:length(eps)],main="conditional std",type="l")
acf(sigma[10:length(eps)],main="conditional std")
plot(x[10:length(eps)],main="GARCH",type="l")
acf(x[10:length(eps)],main="GARCH")
```



```
library(fGarch)
data(bmw, package="evir")
fit<-garchFit(formula=~arma(1,0)+garch(1,1),data=bmw,cond.dist="norm")
```

```
Call:
 garchFit(formula = ~arma(1, 0) + garch(1, 1), data = bmw, cond.dist = "norm")

Mean and Variance Equation:
 data ~ arma(1, 0) + garch(1, 1)
<environment: 0x10fd753b8>
 [data = bmw]

Conditional Distribution:
 norm

Coefficient(s):
        mu          ar1        omega       alpha1        beta1
4.0092e-04  9.8596e-02  8.9043e-06  1.0210e-01  8.5944e-01

Std. Errors:
 based on Hessian

Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
mu     4.009e-04   1.579e-04    2.539   0.0111 *
ar1    9.860e-02   1.431e-02    6.888 5.65e-12 ***
omega  8.904e-06   1.449e-06    6.145 7.97e-10 ***
alpha1 1.021e-01   1.135e-02    8.994  < 2e-16 ***
beta1  8.594e-01   1.581e-02   54.348  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Log Likelihood:
 17757.16    normalized:  2.889222
```

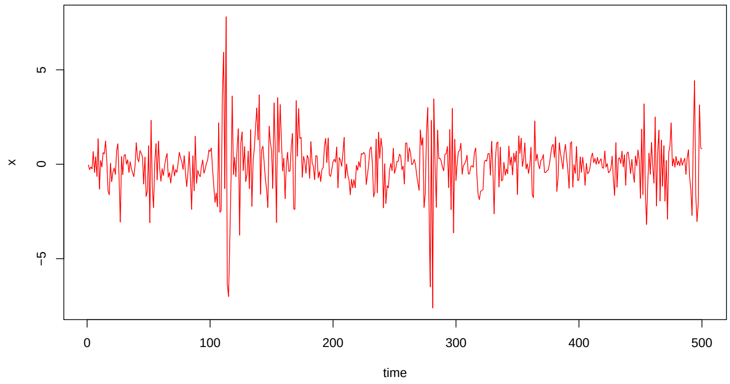
```
Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  11377.99  0
 Shapiro-Wilk Test  R    W      NA        NA
 Ljung-Box Test     R    Q(10)  15.15693  0.1264445
 Ljung-Box Test     R    Q(15)  20.09345  0.168377
 Ljung-Box Test     R    Q(20)  30.54788  0.06144917
 Ljung-Box Test     R^2  Q(10)  5.032717  0.8889818
 Ljung-Box Test     R^2  Q(15)  7.539272  0.9409245
 Ljung-Box Test     R^2  Q(20)  9.277229  0.9794681
 LM Arch Test       R    TR^2   6.03254   0.9144329
```

```
library(fGarch)
data(bmw, package="evir")
fit<-garchFit(formula=~arma(1,1)+garch(1,1),data=bmw,cond.dist="std")
```

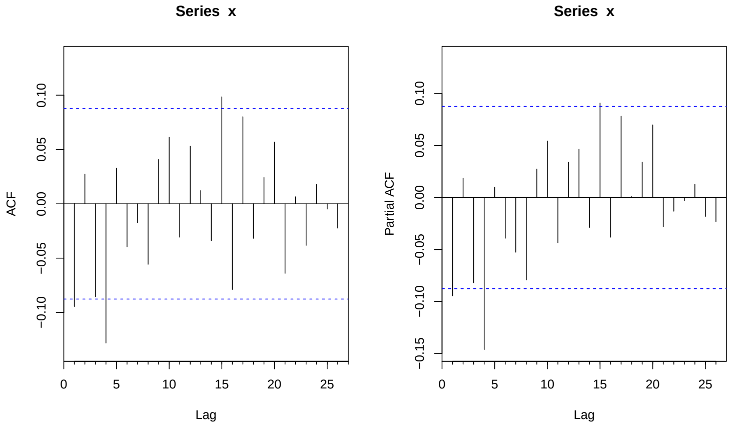
```
Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  13355.05  0
 Shapiro-Wilk Test  R    W      NA        NA
 Ljung-Box Test     R    Q(10)  21.93247  0.01545233
 Ljung-Box Test     R    Q(15)  26.50076  0.03307727
 Ljung-Box Test     R    Q(20)  36.7898   0.01239977
 Ljung-Box Test     R^2  Q(10)  5.828522  0.8294584
 Ljung-Box Test     R^2  Q(15)  8.09067   0.9200962
 Ljung-Box Test     R^2  Q(20)  10.73302  0.9528543
 LM Arch Test       R    TR^2   7.009039  0.8570157
```

```
#Fitting and Diagnostic Checking for ARCH Models

library("forecast")
library("rugarch")
library("tseries")
require(fBasics)

rm(list=ls())
cat("\f")

n <- 600
arch2.spec = ugarchspec(variance.model = list(garchOrder=c(2,0)),
                        mean.model = list(armaOrder=c(0,0)),
                        fixed.pars=list(mu = 0, omega=0.25, alpha1=0.6, alpha2=0.35))
arch2.sim <- ugarchpath(arch2.spec, n.sim=n)
x <- drop(arch2.sim@path$seriesSim)[101:600]
time <- 1:length(x)
```

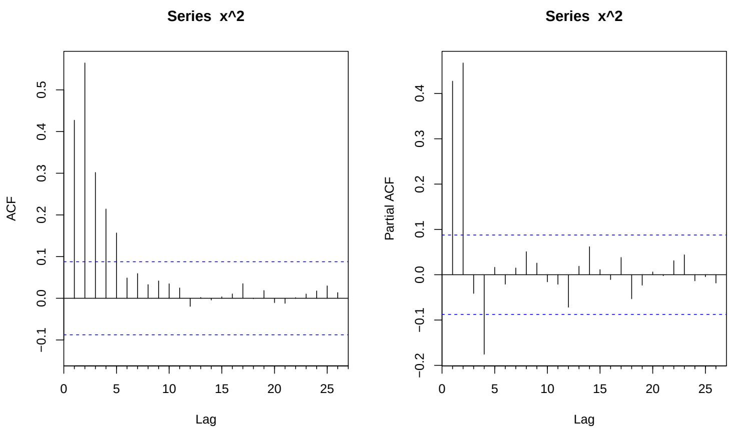
```r
plot(time, x, type="l", col=2)
```



```
par(mfrow=c(1,2))
Acf(x)
Pacf(x)
```



```
par(mfrow=c(1,2))
Acf(x^2)
Pacf(x^2)
```



```r
fit1 <- garch(x, c(0,1), trace=FALSE) # ARCH(1)
fit2 <- garch(x, c(0,2), trace=FALSE) # ARCH(2)
fit3 <- garch(x, c(0,3), trace=FALSE) # ARCH(3)
fit4 <- garch(x, c(0,4), trace=FALSE) # ARCH(4)
fit5 <- garch(x, c(0,5), trace=FALSE) # ARCH(5)

N <- length(x)
loglik1 <- logLik(fit1)
loglik2 <- logLik(fit2)
loglik3 <- logLik(fit3)
loglik4 <- logLik(fit4)
loglik5 <- logLik(fit5)

loglik <- c(loglik1, loglik2, loglik3, loglik4, loglik5)
q <- c(1, 2, 3, 4, 5)
k <- q + 1
aicc <- -2 * loglik  + 2 * k * N / (N - k - 1)

print(data.frame(q, loglik, aicc))
```

```
  q    loglik     aicc
1 1 -705.7442 1415.513
2 2 -666.3854 1338.819
3 3 -729.4100 1466.901
4 4 -734.6751 1479.472
5 5 -729.9601 1472.091
```

```r
spec1 <- ugarchspec(variance.model = list(garchOrder=c(1,0)),
                    mean.model = list(armaOrder=c(0,0), include.mean=FALSE))
fit1 <- ugarchfit(spec1, x)
loglik1 <- fit1@fit$LLH

q <- c(1, 2, 3, 4, 5)
loglik <- rep(NA, length(q))

for (i in 1:length(q)) {
    spec <- ugarchspec(variance.model = list(garchOrder=c(q[i],0)),
                       mean.model = list(armaOrder=c(0,0), include.mean=FALSE))
    fit <- ugarchfit(spec, x)
    loglik[i] <- likelihood(fit)
}

k <- q + 1
aicc <- -2 * loglik  + 2 * k * N / (N - k - 1)

print(data.frame(q, loglik, aicc))
```

```
  q    loglik     aicc
1 1 -707.5218 1419.068
2 2 -669.4327 1344.914
3 3 -669.2573 1346.595
4 4 -669.1381 1348.398
5 5 -668.1957 1348.562
```

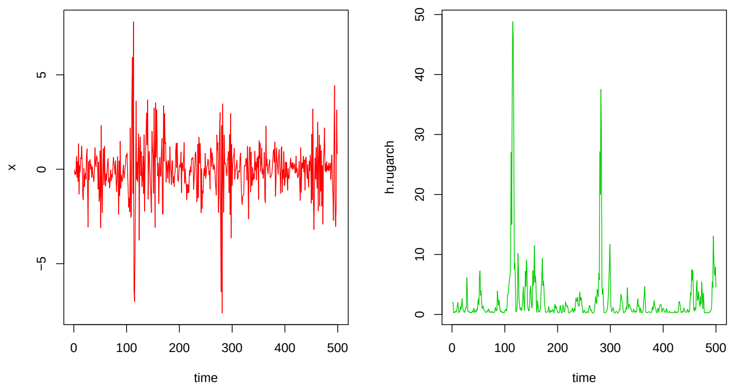
```r
fit.tseries <- garch(x, c(0, 2), trace=FALSE)
summary(fit.tseries)
```

```
Call:
garch(x = x, order = c(0, 2), trace = FALSE)

Model:
GARCH(0,2)

Residuals:
      Min        1Q    Median        3Q       Max
-2.96390  -0.69896  -0.06408   0.66584   2.75127

Coefficient(s):
    Estimate  Std. Error  t value Pr(>|t|)
a0   0.23304     0.04186    5.567 2.59e-08 ***
a1   0.57147     0.10637    5.372 7.77e-08 ***
a2   0.37627     0.08302    4.532 5.84e-06 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Diagnostic Tests:
     Jarque Bera Test

data:  Residuals
X-squared = 2.6731, df = 2, p-value = 0.2627


     Box-Ljung test

data:  Squared.Residuals
X-squared = 0.21612, df = 1, p-value = 0.642
```

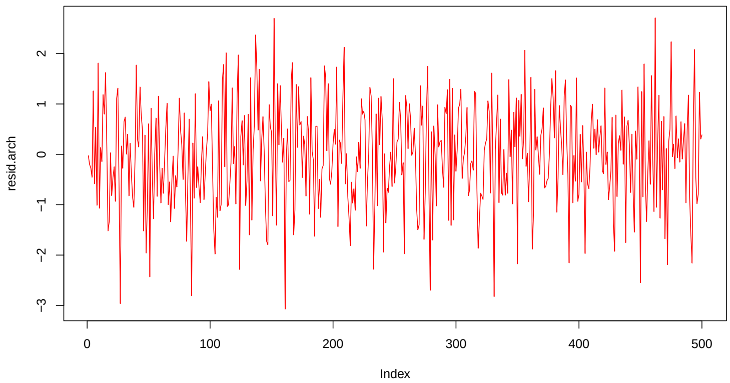
```
spec <- ugarchspec(variance.model = list(garchOrder=c(2,0)),
                   mean.model = list(armaOrder=c(0,0), include.mean=FALSE))
fit.rugarch <- ugarchfit(spec, x)
print(fit.rugarch)
```

```
*---------------------------------*
*          GARCH Model Fit        *
*---------------------------------*

Conditional Variance Dynamics
-----------------------------------
GARCH Model     : sGARCH(2,0)
Mean Model      : ARFIMA(0,0,0)
Distribution    : norm

Optimal Parameters
------------------------------------
        Estimate  Std. Error  t value Pr(>|t|)
omega    0.23305    0.036103   6.4550    0e+00
alpha1   0.57147    0.102465   5.5772    0e+00
alpha2   0.37627    0.078854   4.7718    2e-06

Robust Standard Errors:
        Estimate  Std. Error  t value Pr(>|t|)
omega    0.23305    0.033225   7.0142    0e+00
alpha1   0.57147    0.107341   5.3238    0e+00
alpha2   0.37627    0.078140   4.8154    1e-06

LogLikelihood : -669.4327

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

Akaike       2.6897
Bayes        2.7150
Shibata      2.6897
Hannan-Quinn 2.6997

Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic p-value
Lag[1]                   0.003676  0.9517
Lag[2*(p+q)+(p+q)-1][2]  0.029782  0.9724
Lag[4*(p+q)+(p+q)-1][5]  0.972187  0.8659
d.o.f=0
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic p-value
Lag[1]                     0.2444  0.6210
Lag[2*(p+q)+(p+q)-1][5]    1.4494  0.7524
Lag[4*(p+q)+(p+q)-1][9]    4.8919  0.4455
d.o.f=2

Weighted ARCH LM Tests
------------------------------------
            Statistic Shape Scale P-Value
ARCH Lag[3]     1.401 0.500 2.000  0.2365
ARCH Lag[5]     1.635 1.440 1.667  0.5576
ARCH Lag[7]     5.517 2.315 1.543  0.1772

Nyblom stability test
------------------------------------
Joint Statistic:  0.8739
Individual Statistics:
omega  0.3208
alpha1 0.1603
alpha2 0.1501

Asymptotic Critical Values (10% 5% 1%)
Joint Statistic:          0.846 1.01 1.35
Individual Statistic:     0.35 0.47 0.75

Sign Bias Test
------------------------------------
                   t-value   prob sig
Sign Bias          0.04125 0.9671
Negative Sign Bias 0.07285 0.9420
Positive Sign Bias 0.52097 0.6026
Joint Effect       0.41880 0.9363


Adjusted Pearson Goodness-of-Fit Test:
------------------------------------
  group statistic p-value(g-1)
1    20     21.04       0.3346
2    30     32.20       0.3112
3    40     36.00       0.6075
4    50     50.40       0.4179


Elapsed time : 0.0921278
```

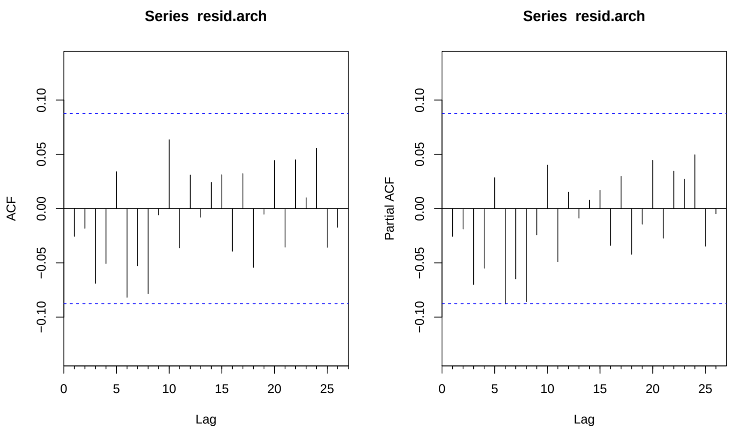
```
# Using tseries package
h.tseries <- fit.tseries$fit[,1]^2

# Using rugarch package
h.rugarch <- as.numeric(sigma(fit.rugarch)^2)

time <- 1:length(x)
par(mfrow=c(1,2))

plot(time, x, type="l", col=2)
plot(time, h.rugarch, type="l", col=3)
```



```r
# Using tseries package:
resid.tseries <- residuals(fit.tseries)

# Using rugarch package:
resid.rugarch <- as.numeric(residuals(fit.rugarch, standardize=TRUE))

# Note: resid.tseries is equal to (x / sqrt(h.tseries))
#       resid.rugarch is equal to (x / sqrt(h.rugarch))

par(mfrow=c(1,1))
resid.arch <- resid.rugarch
plot(resid.arch, type="l", col=2)
```



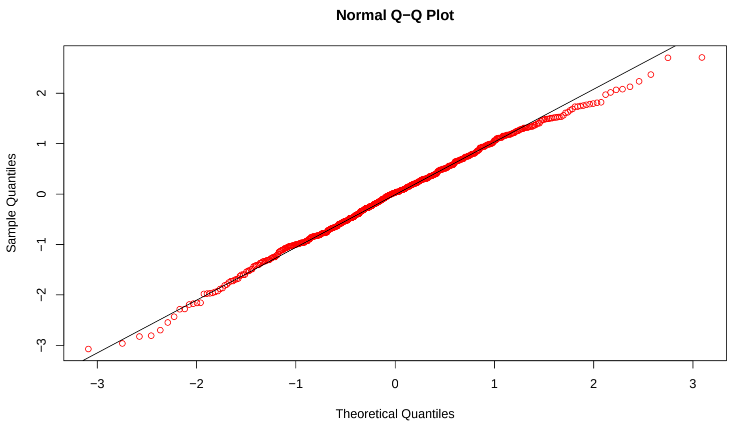
```
par(mfrow=c(1,2))
Acf(resid.arch)
Pacf(resid.arch)
```



```
Box.test(resid.arch, lag=1, type="Ljung-Box", fitdf=0)
Box.test(resid.arch, lag=12, type="Ljung-Box", fitdf=0)
Box.test(resid.arch, lag=24, type="Ljung-Box", fitdf=0)
Box.test(resid.arch, lag=36, type="Ljung-Box", fitdf=0)
```

```
> Box.test(resid.arch, lag=1, type="Ljung-Box", fitdf=0)

	Box-Ljung test

data:  resid.arch
X-squared = 0.0036762, df = 1, p-value = 0.9517

> Box.test(resid.arch, lag=12, type="Ljung-Box", fitdf=0)

	Box-Ljung test

data:  resid.arch
X-squared = 4.7582, df = 12, p-value = 0.9656

> Box.test(resid.arch, lag=24, type="Ljung-Box", fitdf=0)

	Box-Ljung test

data:  resid.arch
X-squared = 13.96, df = 24, p-value = 0.9476

> Box.test(resid.arch, lag=36, type="Ljung-Box", fitdf=0)

	Box-Ljung test

data:  resid.arch
X-squared = 23.563, df = 36, p-value = 0.945
```

```
qqnorm(resid.arch, col=2)
qqline(resid.arch, col=1)
```



```
normalTest(resid.arch,method="jb")
```

```
Title:
 Jarque - Bera Normalality Test

Test Results:
  STATISTIC:
    X-squared: 2.5738
  P VALUE:
    Asymptotic p Value: 0.2761

Description:
 Wed Sep 28 22:33:13 2016 by user:
```

3) STOCHASTIC CALCULUS.

VALUATION OF FINANCIAL DERIVATIVES

The aim of this chapter is to introduce you to the very basic tools used to price financial derivatives. The techniques used here are common practice on front-office departments, market risk divisions or in general which ever activity involved directly to trading activities.

Although we already know that equity markets is a small portion of financial market it is convenient to introduce these new concepts through a simple product as the equity is. Nevertheless the ideas explained here can be applied to a ~~broader~~ wider variety of products.

We will also simplify our exposition by restricting the scope of derivatives to what is known in the market as "plain vanilla" derivatives.

The contents of this chapter can be split into:

- Review on product, market & Derivatives.
- Discrete Models
- Continuous Models.

Basic Reference

Paul Willmott, Ch. 1-6, "Introduction to Quantitative Finance"

PRODUCTS, MARKETS & DERIVATIVES

EQUITIES

Equities will be our basic financial instrument in this chapter. They are also known as stocks, shares, or when refered to options on equities, underlying assets.

> **Share**
> Owning a share means partial ownership of the company

The following timeline represents the most common process for a company to go into the market:

- A company is born: When a company is founded is own by the founders, who own the company and run it.
- A company grows up: Founders typically make further investments or ask for loans to banks to expand the company.
- IPO (Initially Public Offering) When a particular investment is too much for the owners to cope with, the company goes public on the stock market and rises capital against offering stocks
  - → Shareholders have a ~~right~~ say in running the company
  - → The recieve part of the benefits of the company in form of dividends.

From that point onwards the company has a CEO which runs the daily business and major decisions are taken by vote with all shareholders.

- **The company could become private again:**

If an investors is able to buy all outstanding ~~shares~~ shares in the market, the company can become private again. This is a very rare event and usually a company ends its life by merging with other companies or bought by another company.

EQUITY PRICES

There are many theories about pricing formation in stock markets, among the most important ones are:

- Dividend Cash Flow: Market reflect market prevision of all future cash flow due to dividends
- Residual value of company liquidation: The residual of all liabilities are substracted from the assets.

One method tends to view the company as a living entity while the later views it in a static picture.

In practice market players use both valuations and hence market price is form by a mixture of those approaches

Even though we could explain the particularities of each model, ~~A~~ where does the randomness of market prices come from?

In other words, if we all agree in a company future cash flows of dividends, why would we disagree in the price of the stock? (Notice that randomness in prices is a manifestation of investors different views on the price)

- Predictability
- Litigation
- Market Evolution
- Consumer response
- Geopolitical events
- ......

On the folling example you will see that essentially the impact of on the above list of variables comes in the form of "news".

A more general paradigm for market equilibrium arises from the assumption that markets are informally efficient. This means that the information available at the time of making an investment is already reflected in the price of the security. This hypothesis is known as Efficient Market Hypothesis (EMH).

There are three types of efficiency:

- Weak: only the price history of the security is the available information.
- Semi-strong: All public information known to the present time is available.
- Strong: All public and private information known up to the present is available.

What do you think is the reality for equity markets?

PAGARÁN EL TRIBUTO SOBRE ACTOS JURÍDICOS

Los bancos se hunden tras la 'bofetada' del Supremo con los impuestos de la hipoteca

El alto tribunal establece que es el banco y no el cliente quien debe abonar el tributo sobre Actos Jurídicos Documentados en las escrituras públicas con garantía hipotecaria

18/10/2018
12:43

Bolsa de Madrid. (EFE)

JTOR

18/10/2018 17:56

11:00
12:00

ARBITAGE MARKETS

In practice markets might be a bit inefficient, but it is extremely difficult to find arbitrage opportunities. Therefore we consider the market as an arbitrage free structure.

Arbitrage

Is the process of trading a single product in two different markets with different prices. The trader would buy the security in the market that shows the cheapest quote and sell it immediately in the market that shows the higher quote and make a profit.

The above is obviously not possible and hence we say:

There is not such a thing as a free lunch

Proposition

- Assuming no arbitrage, to portfolios with equal value at time T must have equal value at all time $t<T$.
- If A and B are two portfolios with $V(A,T) \geq V(B,T)$ then for all $t \leq T$, $V(A,t) \geq V(B,t)$

FORWARDS & FUTURES

Among all types of derivatives, these are the most simple ones. The objective of this section is to give the formal definition of both products and give the first pricing formula of the course!!

Forward An agreement where one party promises to buy an asset from another party at some specific time in the future and at some specific prize. No money changes hands until the end of the contract, and the terms of it make it an obligation.

Future
Similar to a forward contract, but traded through an exchange, which standarizes the terms (maturity, notional, ...)
↳ marke to market: daily deliver of benefits / loses (ie less risky than a forward)

The main uses of both contracts are:
↳ Speculation (Taking a market view, ie very risky)
↳ Hedging (Lock in price to avoid surprises)

COMPUTATION OF FORWARD PRICE

How on earth can we agree on a forward price with all that randomness in prices? !!

This will be the first example of a non-arbitrage theory ¿WTF is this?

SETTINGS

- Consider a formal contract that obliges you to deliver $\$F$ at time T, and to receive the underlying asset $S(T)$

  This is the FWD contract.

- Today's date is t and today's underlying asset is $S(t)$ – also known as spot price –.
  - At maturity the price of the underlying will be $S(T)$

- The profit (or loss) at maturity is $S(T) - F$, which is unknown and depends on the evolution (random) of the underlying.

$$\boxed{\text{The magic in non-arbitrage theory is that we can "eliminate" randomness.}}$$

TRADE (which try to eliminate randomness)

- Enter into the forward contract (which cost us nothing) because the forward price should be fair)

- Go short the underlying (means selling the stock), by the way you can sell something you do not own by borrowing it
- We have the amount $S(t)$ in cash from selling the stock.
- Put the cash in a bank account to recieve interest at a rate $r\%$.

- Our net position is zero.

MATURITY

- We hand over the amount F
- Recieve the asset, which cancels out our short position, we return it to the "friend" from which I borrowed it

- Go to the bank and cash out the account with $S(t)e^{r(T-t)}$

- Our net position is
$$S(t)e^{r(T-t)} - F$$

There is no randomness here!!!

By an argument of no-arbitrage, if I have something worth zero at the begining, then I have zero at the end (there is not such a thing as a free lunch)

$$\Rightarrow \quad S(t)e^{r(T-t)} - F = 0$$

$$\boxed{F = S(t)e^{r(T-t)}}$$

This is the "fair" price

Position	Worth today (t)	Worth at Maturity (T)
Forward	0	$\cancel{S(T)} - F$
- Stock	$-S(t)$	$-\cancel{S(T)}$
Cash (Bank)	$S(t)$	$S(t)e^{r(T-t)}$
Σ	0	$S(t)e^{r(T-t)} - F$

- There is not randomness in the price of a forward, all variables are known at the initial moment $(S(t), r, T)$
- It is independent of the future evolution of the asset $(S(T))$!!
- These type of arguments are called arbitrage-free and are linked to what we known as "fair game"

DERIVATIVES

Derivatives are financial assets whose price depends upon another financial asset. This seems a very ~~vage~~ wide definition, but indeed this is because the derivative market is soo much larger than any other market in the world, you can virtually trade any structure of derivative you could imagine.

Before we start with the definitions, let us introduce some basic concepts:

- Premium: The amount paid for the contract initially. How to find this value is the subject of this chapter.

- Underlying: The financial instrument on which the option value depends (often denoted by S)

- Strike: The amount for which the underlying can be bought or sold (often denoted by K)

- Expiration: Date on which the derivative matures. (often denoted by T)

- Payoff: The liquidation of the derivative if the underlying is at its current level or at a given level. ~~(This is a function o~~

Recall that forwards are written in a way that the trade becomes an obligation. The next derivatives we will introduce are OPTIONS.

Options give the holder the right to trade, but takes away the obligation.

CALL OPTION

A <u>call</u> option is the right to <u>buy</u> a particular asset for an agreed price at a specified time in the future.

PUT OPTION

A <u>put</u> option is the right to <u>sell</u> a particular asset for an agreed price at a specified time in the future

The main objective of this chapter is to price options.

Why would anyone trade options?

PAYOFF DIAGRAMS.

The understanding of options is helped by the visual interpretation of an option's value at expiry.

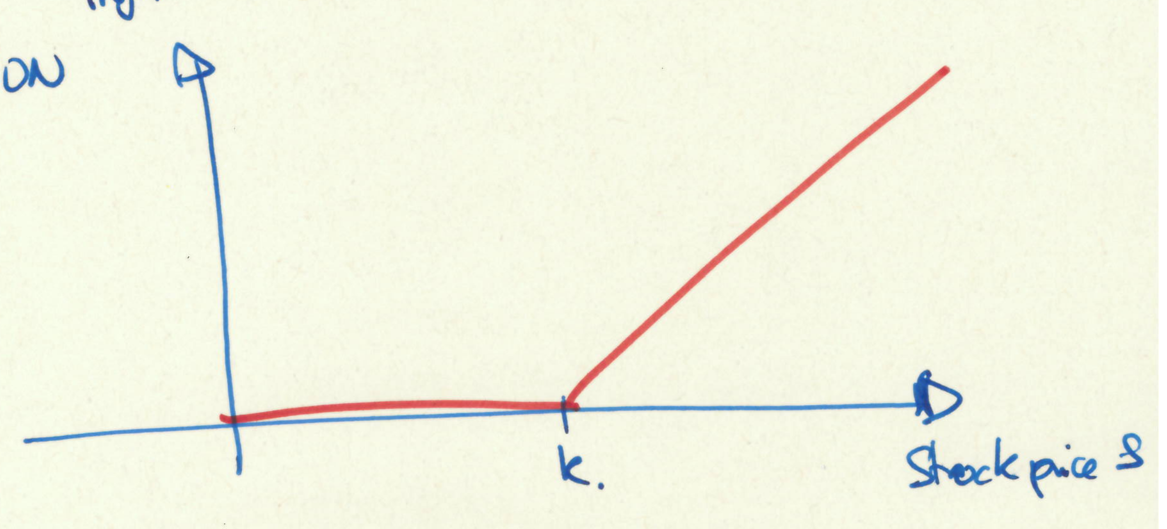
Lets start with a call option. Assume the underlying asset S and the exercice price K (strike). Assume you hold an ~~op~~ call option, then at time $t=T$

- if $S>K$, you excercise the right to buy the asset at price K while market price is S, hence your benefit is $(S-K)$.

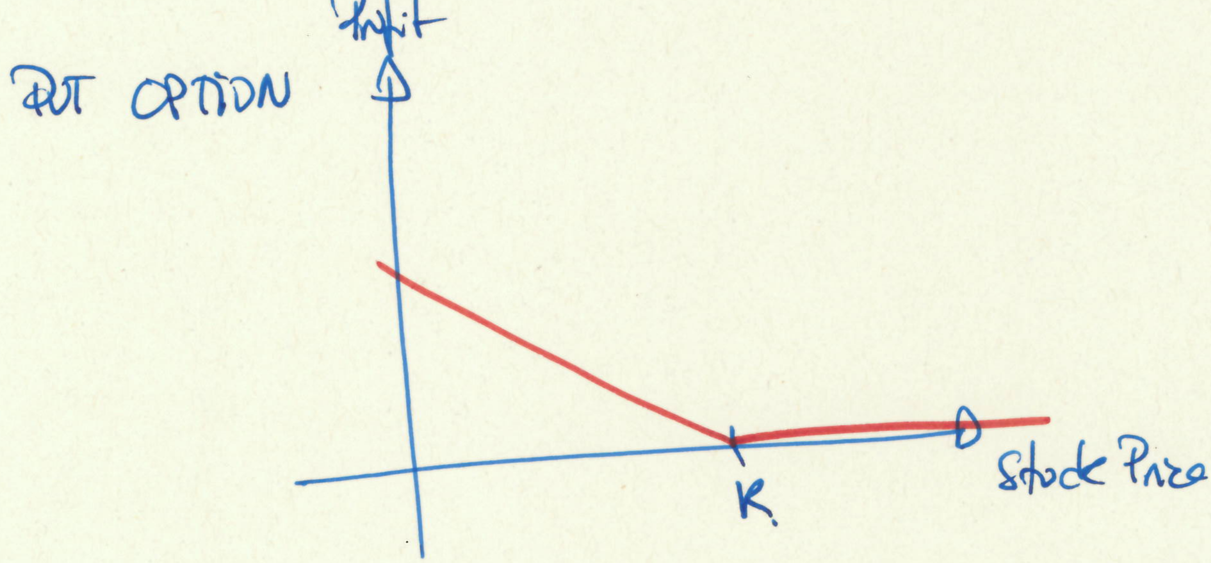
- if $S<K$, you won't exercice the right to buy the asset at price K because you can buy it in the market at S (cheaper), hence your benefit is 0 (you do not buy the asset).

Summarize it on the pay off function $\max(S-K, 0)$

CALL OPTION

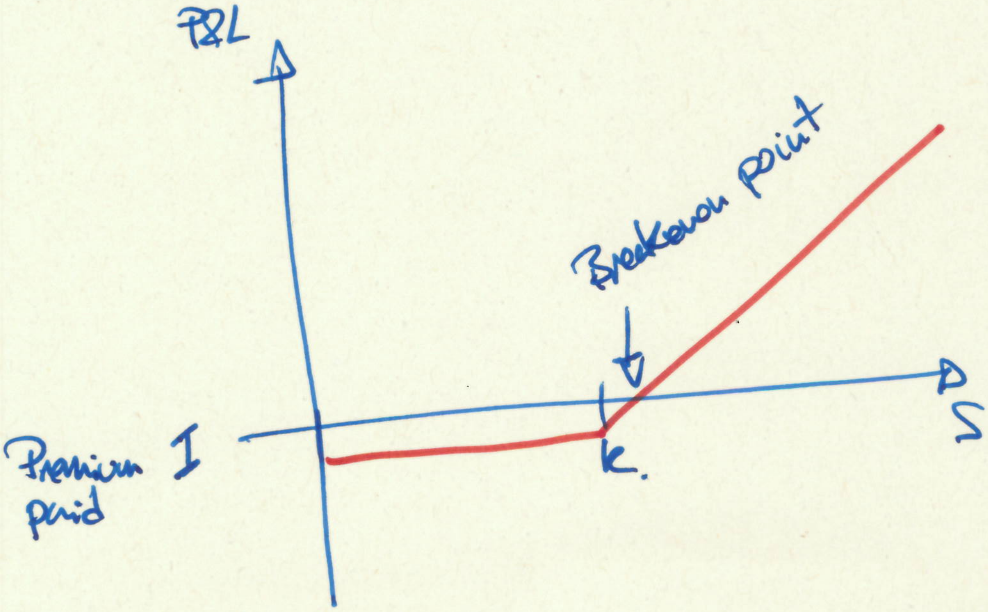


In a similar way the payoff function of a put option would be



Obviously options (either calls or puts) give the holder the right but not the obligation and hence the above diagrams show potencial gains with no possible lost. This is because we have not incorporated the premium, options / rights cost money !!

The following payoff diagram incorporates premium effect



THE VALUE OF AN OPTION BEFORE EXPIRY

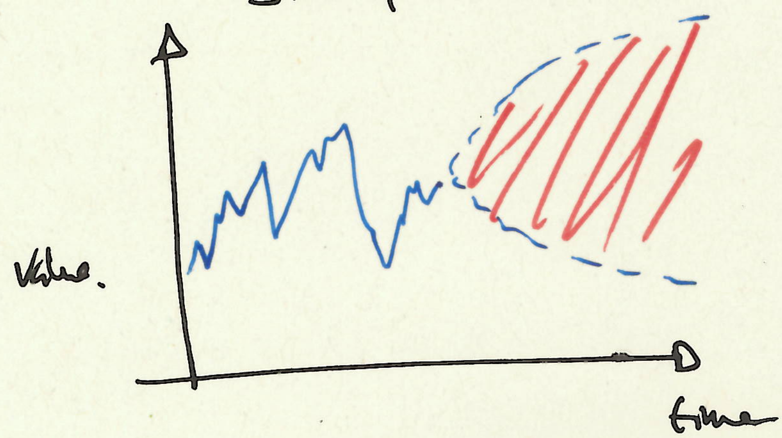
We have seen how much calls and puts worth at expiry, but the central question is how much is the contract worth before expiry.

- At the very least you know that there is no downside to own an option, it gives you rights but no obligations

Two variables seems to determine the value of an option today:

- The value of the asset today, the higher the value is today, the higher we might expect to be at expiry.
- The dependance on time is more subtle, the longer the time to maturity, the longer the time has the asset to rise or fall

Stock price evolution



FACTORS AFFECTING DERIVATIVE PRICES

The two most important factors affecting prices are:

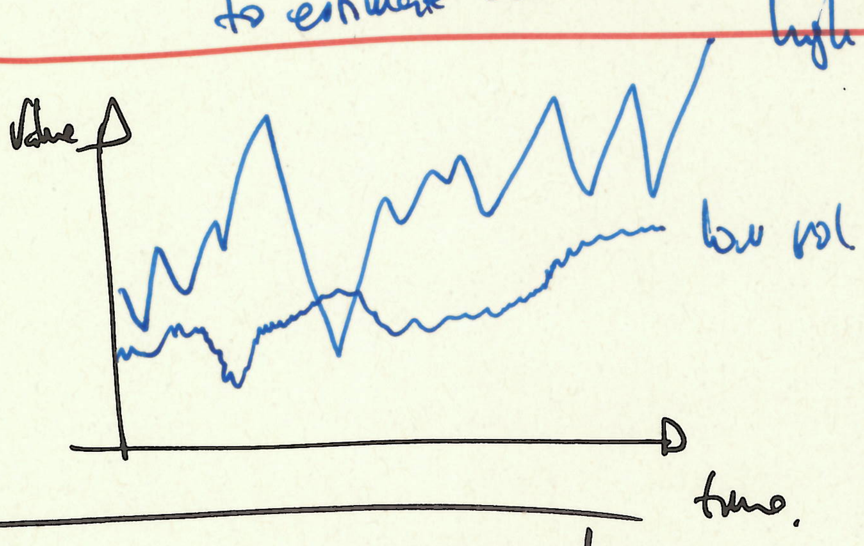
- → Value of underlying asset S
- → Time to expiry $(T-t)$

Other parameters affect prices:

Interest rates → Affecting the time value of the option
Strike price → Affecting the probability of reaching that point.

Volatility → One of the major parameters affecting option prices.

Volatility: measures the amount of fluctuation of an asset as the annualized standard deviation of an asset return. It is very difficult to estimate and changes over time



⇑ Volatility ⇒ ⇑ Option

SPECULATION & GEARING

Options are examples of leveraged instruments (you can make a lot of profit by investing very little)

Example: Today is 14/04 and ABC stock is worth \$666. The cost of a 680 call expiring on 22/08 is \$39.

Buy stock: Buy the stock at \$666 and by mid August the stock has risen to \$730. Then my return is

$$\frac{730-666}{666} \times 100 = 9.6\%$$

Buy a CALL: I buy a call option for \$39 and exercise the call paying \$680 for something worth \$730, hence

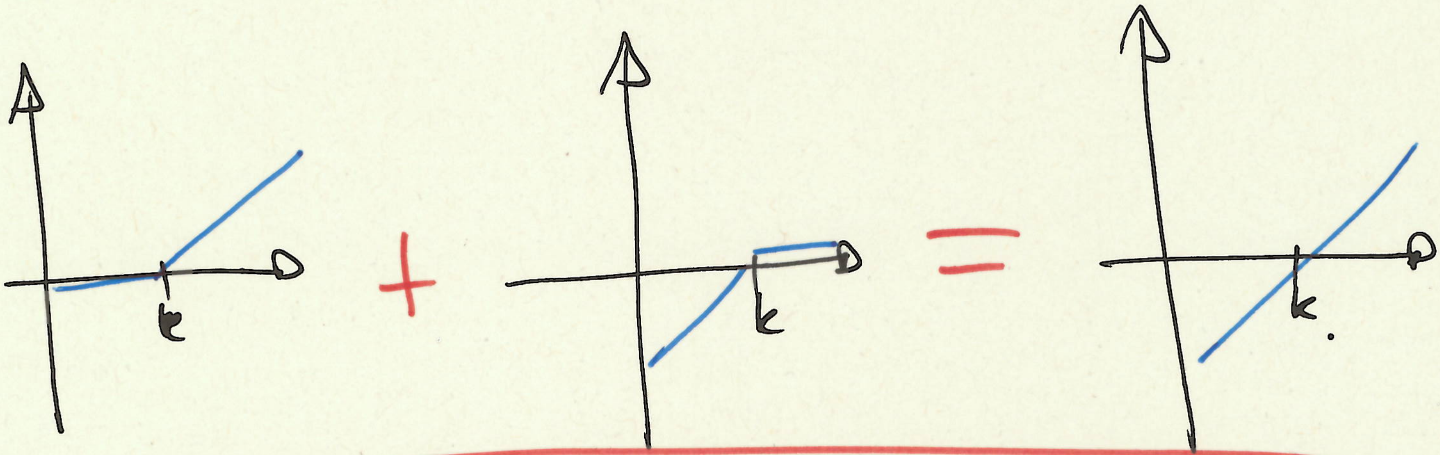
$$\frac{730-680-39}{39} \times 100 = 28\%$$

→ If the stock remains flat at \$666, the first strategy has 0% profit, and the second has 100% loss!!

PUT-CALL PARITY

Another example of no arbitrage theory

Imagine you go long (buy) a call with given strike K and maturity T. You also go short (sell) a put with the same strike and expiry. The agregated payoff at expiry looks like:



$$\boxed{\max(S_T - K, 0) - \max(K - S_T, 0) = S_T - K.}$$

$$C(T) - P(T) = S(T) - K.$$

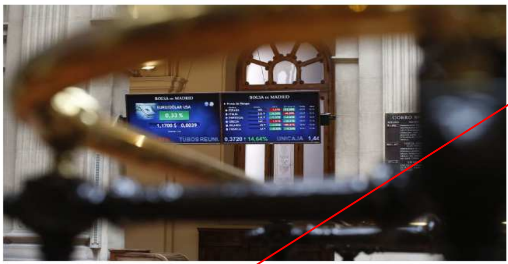
$\Downarrow$ Discounting ⊕ No arbitrage theory

$$\boxed{C(t) - P(t) = S(t) - Ke^{-r(T-t)}}$$

PAGARÁN EL TRIBUTO SOBRE ACTOS JURÍDICOS

Los bancos se hunden tras la 'bofetada' del Supremo con los impuestos de la hipoteca

El alto tribunal establece que es el banco y no el cliente quien debe abonar el tributo sobre Actos Jurídicos Documentados en las escrituras públicas con garantía hipotecaria



Bolsa de Madrid. (EFE)

32

UTOR
ARÍA IGARTUA

18/10/2018 17:55 - ACTUALIZADO: 23/10/2018 22:48

18/10/2018
12:43

11:00
12:00

BINOMIAL MODELS & RANDOM BEHAVIOUR OF ASSETS

Recall that we will start the study of continuous models to the case of modelling equity assets. To this end we will first introduce a discrete model called Binomial Model, and will obtain a continuous model by passing to the limit the discrete case.

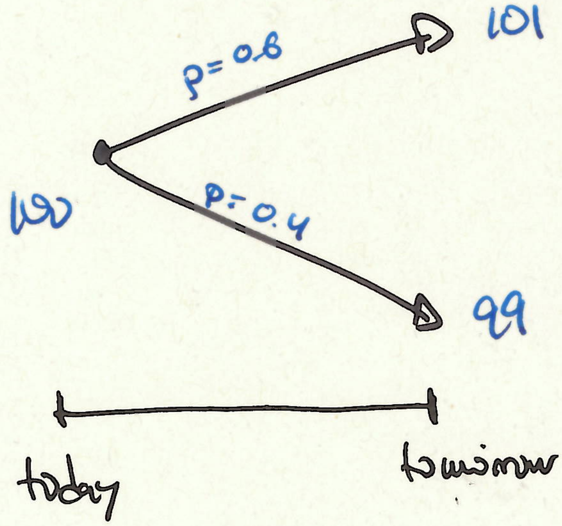
Binomial Model

It is the simplest model for asset prices, yet used in pricing some exotic derivatives.

Let us consider the following situation:

- We have ~~got~~ a stock and a CALL option expiring tomorrow
- The stock can either go up or down by a known amount between today and tomorrow
- Interest rates are zero.

Graphically the situation is as follows:



Assume the stock can rise to 101 or fall to 99, with probability of doing the later 0.4 and probability to rise of 0.6.

Furthermore, assume the call has strike 100.

$\boxed{\text{WHAT IS THE OPTION VALUE?}}$

In order to answer the question we have to eliminate the ~~rest~~ randomness of the experiment. As we did in the derivation of the forward price we have to apply a non-arbitrage argument (this is essencially building up a portfolio without randomness in the outcomes)

Let us construct a MAGIC portfolio consisting of:

- Long one option (Call with strike 100)
- Short 1/2 of the underlying (Let me call this number Δ, ie $\Delta = 1/2$ in this case)

Lets denote the portfolio Π, then our MAGIC portfolio is the following

$$\Pi(t) = C_{100}(t) - \tfrac{1}{2} S_t$$

For $t=T$ we have the folling outcomes

$$\Pi(T) \xrightarrow{S(T)=101} C_{100}(T) - \tfrac{1}{2} S_T = \max(101-100, 0) - \tfrac{1}{2} 101 = -99/2$$

$$\Pi(T) \xrightarrow{S(T)=99} C_{100}(T) - \tfrac{1}{2} S_T = \max(99-100, 0) - \tfrac{1}{2} 99 = -99/2$$

Either way the portfolio is worth $-99/2$ ~~tomorrow~~ no matter how the stock moves .... and NO matter which are the probabilities of rising and falling !!!

Therefore we have a portfolio, $\Pi(t)$, such that $\Pi(T) = -99/2$ for sure, since interest rates are zero.

then
$$\Pi(t) = -99/2 \quad \forall t$$

$\Rightarrow$ it is constant!! For $t < T$ $\Rightarrow$

$$\Pi(t) = C_{100}(t) - 1/2\, S_t = -99/2$$
$$C_{100}(t) - 1/2\, 100 = -99/2$$
$$C_{100}(t) = 1/2$$
$\Rightarrow$ We have found the price of the call option by a no arbitrage theory.

Three questions follow from the above simple argument.

1 - Why this "theoretical" price is right?

2 - How did I choose $\Delta = 1/2$?

3 - What happens if interest rates are zero?

QUESTION 1

The theoretical price is right because "there is not such thing as a free lunch". Since the strategy Π has a known outcome, then

- $C_{100}(t) < 1/2 \Rightarrow \Pi(t) < \Pi(T)$
  We buy the call and sell $1/2$ stock and make a sure profit $(\Pi(T) - \Pi(t))$

- $C_{100}(t) > 1/2 \Rightarrow \Pi(t) > \Pi(T)$
  We do the opposite version of the above strategy and make the same profit.

Therefore there is only one reasonable price for the option $\boxed{C_{100}(t) = 1/2}$.

QUESTION 2

Let us denote by Δ the quantity of the stock that has to be sold to hedge the option.

Hedging means that we end up in the same situations regardless of the path of the stock price

In other words solving :

$$1 - \Delta \times 101 = 0 - \Delta \times 99$$

This is called delta hedge

$$\Delta = 1/2$$

Example

Stock price is 100, can rise 103 or fall 98. Interest rates are 0%. Price a CALL option with strike 100.

i) Consider the portfolio $\Pi(t) = C_{100}(t) - \Delta S_t$

ii) Find Δ-hedge by solving $\Pi_{103}(T) = \Pi_{98}(T)$

$$3 - \Delta \times 103 = 0 - \Delta \times 98 \Rightarrow \Delta = 0.6$$

iv) Solve $\Pi(T) = 3 - 0.6 \times 103 = \dots$

v) Apply no arbitrage theory to solve.

$$\Pi(T) = \Pi(t) = C_{100}(t) - 0.6 \times 100$$

The call option is worth 1.2

QUESTION 3

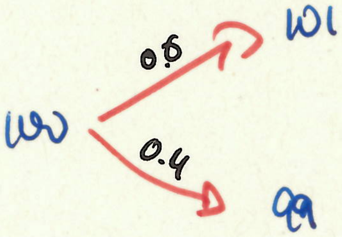
If interest rates are non-zero, then the portfolio $\Pi(t)$ must be discounted before applying a non-arbitrage argument.

In other words, step v) in the previous example must be substituted by

v) $$\Pi(t) = \Pi(T)\, e^{-r(T-t)}$$

We typically use continuous compounding interests to value options.

The Real and Risk-Neutral Worlds



This is one of the most difficult concepts in math finance

In the real world we have used our statistical skills to estimate the probability of possible stock prices.

But we haven't used them to price the option, and what about "fair value" ??

REAL WORLD

- We know how to delta hedge and eliminate risk
- We are sensitive to risk, we expect greater return for taking risk
- Only stock prices matter, not the probability

RISK-NEUTRAL WORLD

- We do not care about risk that can be hedged
- We do not estimate the probabilities of an event
- We believe that everything is priced using simple expectations

In our previous example, the price of the portfolio $\Pi(t)$ should be computed using expectations as.

$$\Pi(t) = \mathbb{E}_{\mathbb{R}}[\Pi(T)]$$

This should be also true for any portfolio, in particular for the portfolio consisting in owning one share, ie:

$$S_t = \mathbb{E}_{\mathbb{R}}[S_T]$$

$$\begin{tikzcd} & 101 \\ 100 \arrow[ur, "p"] \arrow[dr, "1-p"'] & \\ & 99 \end{tikzcd}$$

Therefore

$$100 = p \times 101 + (1-p)\, 99$$

$$\Rightarrow p = 0.5.$$

The above probability is NOT real, is just an invention ~~to~~ since we believe that everything is priced using expectations. Using the above probability what would be the price of the CALL option?

$$\mathbb{E}[C_{100}(T)] = \max(101-100, 0) * 0.5 + \max(99-100, 0) * 0.5$$

$$= 1/2$$

The right price!!!

$\boxed{\text{We use the wrong probabilities to get the right price using the expected payoff}}$

This technique allways works under mild assumptions. Unfourtunately is out of the scope of this course to prove it.

Heuristically speaking, computing prices in a risk-neutral world works because the market is not going to reward you for a risk that is hedgeable.

Therefore

$$\boxed{\text{Risk Neutral} \cong \text{No-arbitrage.}}$$

To complete this section, lets assume that interest rates aren't zero, then the price of an option is

$$C_{(0)}(t) = \mathbb{E}[C_{(0)}(T)]e^{-r(T-t)}$$

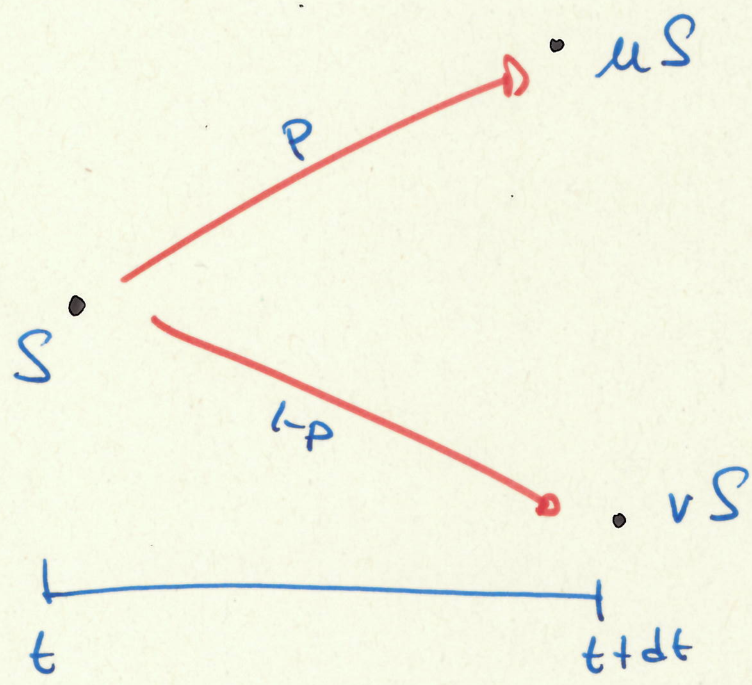
GENERALIZATION OF BINOMIAL MODEL

A binomial model consists of an asset with initial value S, which in a time step dt can:

i) rise to value $u \times S$

ii) fall to value $v \times S$

where $0 < v < 1 < u$. Lets assume that the probability of rise is p. the model is graphed as:

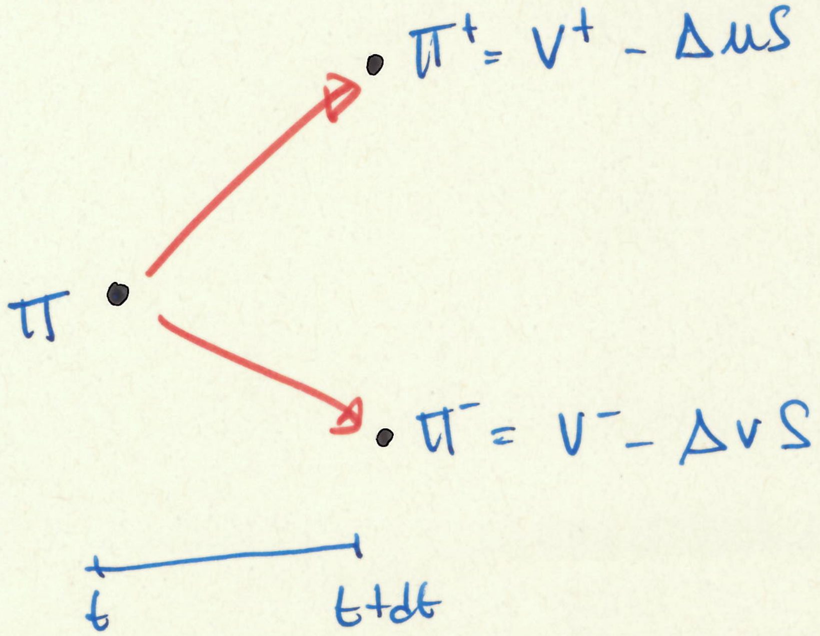


Lets now try to find an equation for the value of an option. Suppose we know the value of an option at $t+dt$ (say, for instance at expiration). Lets now derive the value at t:

We construct a portfolio consisting of a long position in the option an a short position of Δ stocks:

$$\Pi \equiv V - \Delta S$$

At $t+dt$ the option takes values V^+ or V^- depending on the evolution of the underlying (ie, the payoff function). therefore the portfolio is worth



Hedging: We chose Δ so to make the portfolio value equal in either case of stock direction, ie:

$$\Pi^+ = \Pi^-$$

$$V^+ - \Delta u S = V^- - \Delta v S$$

$$\Delta = \frac{V^+ - V^-}{(u-v)\, S}$$

The portfolio value at expiry is then worth

$$\Pi_{t+dt} = \Pi^+ \quad (= \Pi^-)$$

$$= V^+ - \Delta u S \quad (= V^- - \Delta v S)$$

$$= V^+ - \frac{u(V^+ - V^-)}{u-v} \quad \left(= V^- - \frac{v(V^+ - V^-)}{u-v}\right)$$

No-arbitrage: Since there is no randomness in the portfolio value at expiry, the portfolio value at inception is just discounted the deterministic flow at expiry.

A

In other words:

$$\Pi_t = \Pi_{t+dt}\, e^{-r(dt)}$$

→ Once we get to $t=t_0$ we know the value of the stock at t_0 and hence
$\Pi_{t_0} = V_{t_0} - \Delta S_{t_0}$
can be solved for V_t

DO IT !! (with $r=0$)

Counter-intuitive !!!

That option prices don't depend on the direction that the stock is going or its real probabilities can be difficult to accept initially. The key to understand this concept is that hedgeable risk is worthless

Disclaimer: Although binomial models are used in practice for specific purposes, the above construction and derivations do not hold for trinomial trees.

A

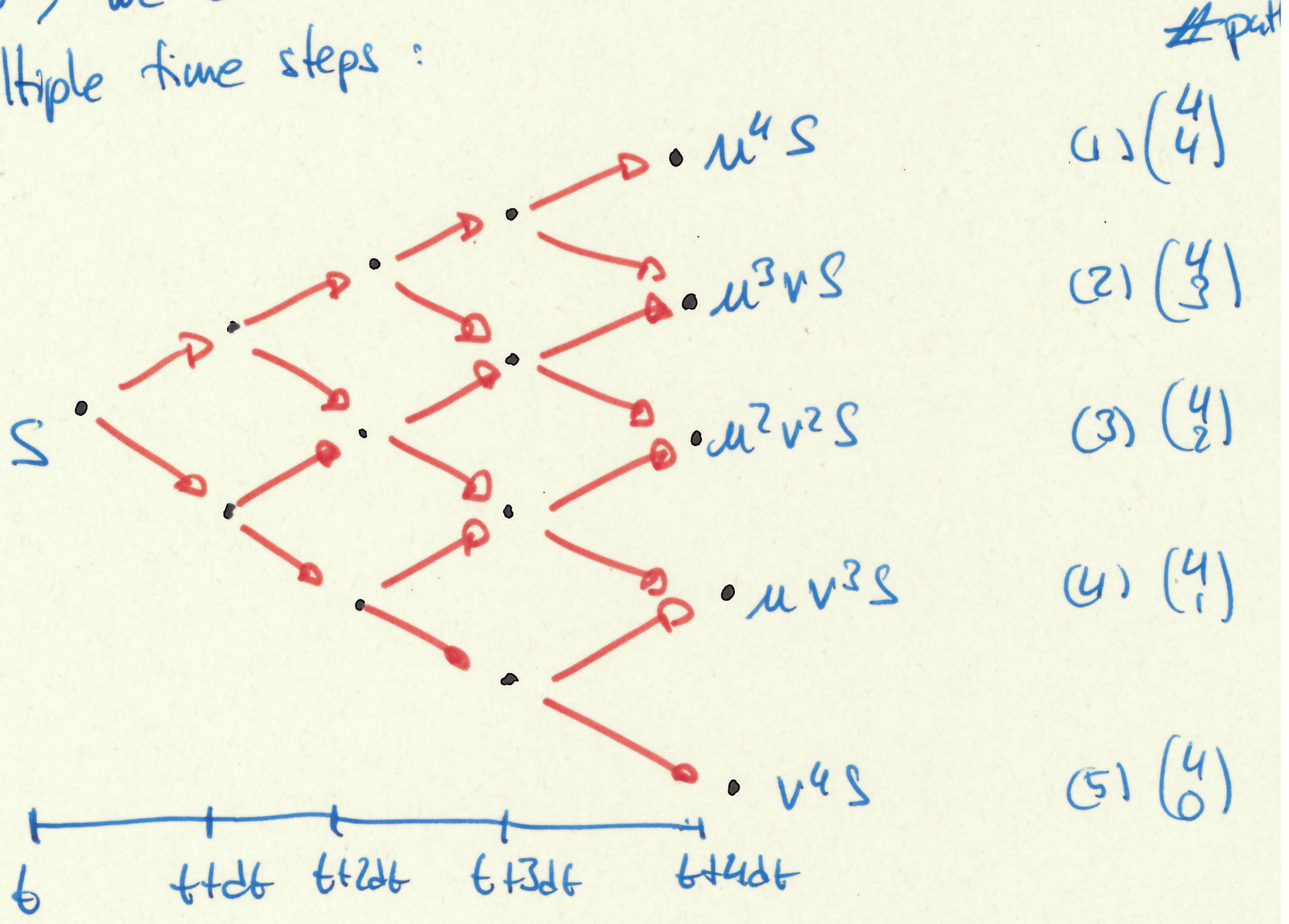
$$\Pi_t = V_t - \Delta S_t = V^+ - \frac{u(V^+ - V^-)}{u-v}$$

$$V_t - \frac{V^+ - V^-}{(u-v)\cancel{S_t}}\cancel{S_t} = V^+ - \frac{u(V^+ - V^-)}{u-v}$$

$$V_t = V^+ - \left(\frac{V^+ - V^-}{u-v}\right)(u-1)$$

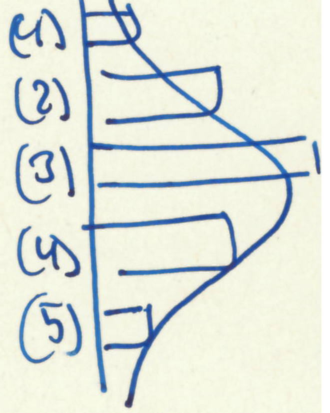
The complete binomial tree

The binomial model just described, the stock moves up and down a prescribed amount in one step; we can extend this model further by allowing multiple time steps:

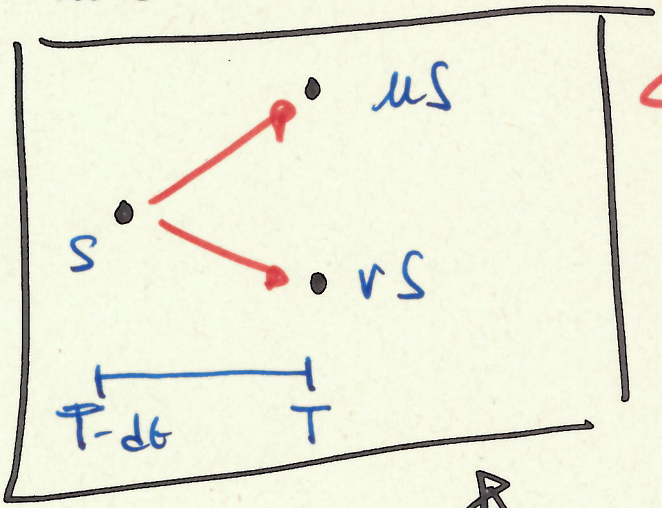


Note that intermediate values are more reacheable than the extreme ones (there are more possible paths to them)

!! The probability of reaching a final node is approximately bell shaped

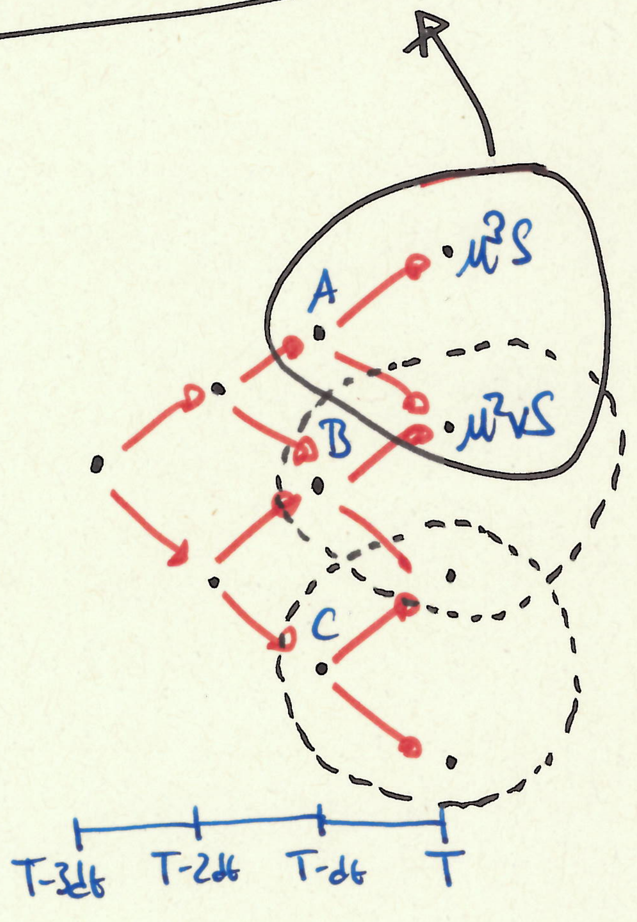


Now we need to know how to price options on large binomial trees. Clearly we know how to price options on "small" trees, so we need to iterate recursively the process:



We know how to price an option if the terminal value is known (payoff) & for one step time

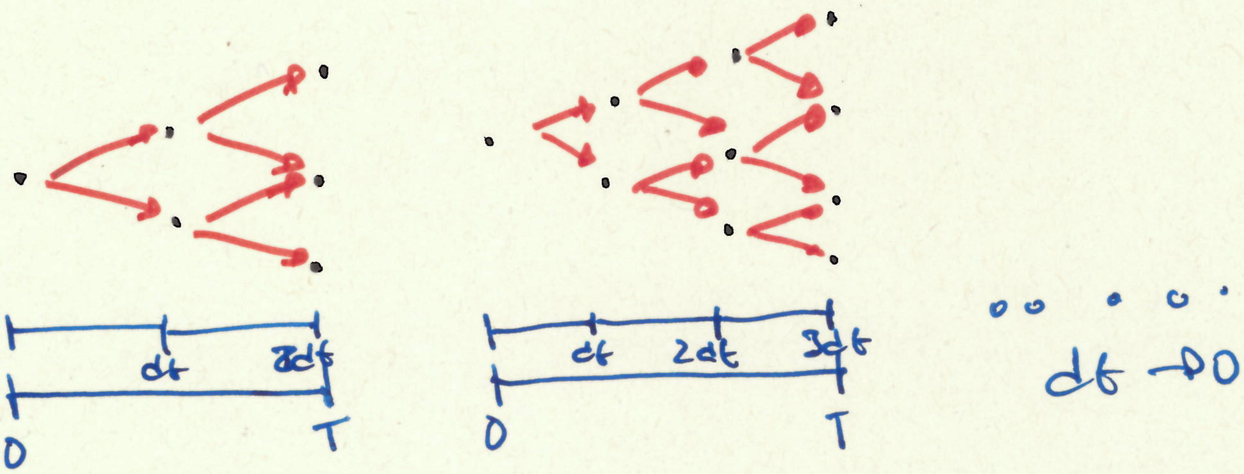
Therefore we know the value of the option at A, B & C



Returns Computation: Revisited

The choice to denote the time step as dt is not arbitrary, in order to construct a continuous model we are going to pass to the limit the binomial model (i.e. $dt \to 0$).

For a given time period $[0, T]$ we are going to increasingly construct more dense binomial trees until we get close to a continuous model.



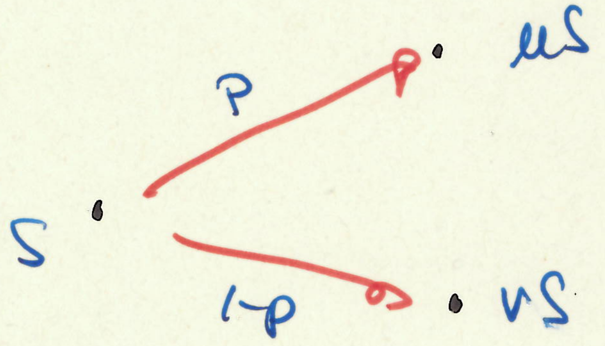
Where does Δ hedge come from?

Recall the derived function for Δ as

$$\Delta = \frac{V^+ - V^-}{(u-v)S} = \frac{V^+ - V^-}{uS - vS} \approx \frac{\partial V}{\partial S} \quad !!!$$

For $v \to 1$, $u \to 1$

Where did the probability "p" went?



And p ¿?
!!!

The probability p is irrelevant as we derived the price by constructing a non random portfolio but we can derive p.

Recall that p is the risk-neutral probability, hence we do not care "about risk", in other words

$$\mathbb{E}[S_{t+dt}] = S_t$$

$$puS + (1-p)vS = S$$

$$p = \frac{1-v}{u-v}$$

Now the following is a very interesting exercise:

① We know that for a one step binomial model

$$\Pi_t \equiv V_t - \Delta S_t = V^+ - \frac{u(V^+ - V^-)}{u - v}$$

where $\Delta = \dfrac{V^+ - V^-}{(u-v)S}$

All quantities in RHS are known!!

② Check that we can compute the value of V, as

$$V_t = \mathbb{E}[V_{t+dt}] = V^+ p + V^-(1-p)$$

where $p = \dfrac{1-v}{u-v}$

same here!!

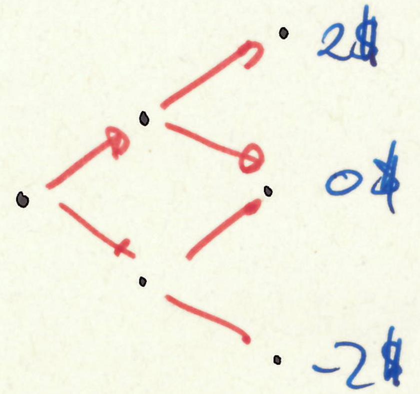
⇒ hence by using the risk neutral probability we can compute the option prices as simple expectation and returning back to the notion of fair value.

BROWNIAN MOTION & BLACK-SCHOLES

In this chapter we will introduce some stochastic calculus to be able to present the Black-Scholes equation. This was a milestone in mathematical finance worth a Nobel Prize. Stochastic Calculus is a very broad topic which need more than one whole course to master, hence we will only be giving the idea behind the main manipulations.

EXAMPLE
Consider the experiment of tossing a coin. If you get a head you earn 1\$ and if you get a toss you get loss 1\$. Each tossing is independent.

This experiment is very similar to a binomial model with equal amount of upward moving and downward moving, and equiprobable probability.



Consider R_i the random amount earned at toss i-th., then

- $\mathbb{E}[R_i]=0$
- $\mathbb{E}[R_i]=1$ & $\mathbb{E}[R_i R_j]=0 \quad \forall\, i\neq j$

and define $S_i=\sum_{j=1}^{i} R_j$, then

- $\mathbb{E}[S_i]=0$
- $\mathbb{E}[S_i^2]=i$

The following properties are easy to check.

Markov Property

The expected value of the random variable S_i condition upon all of the past event only depends on the previous state. (We say it has no memory)

$$\mathbb{E}[S_i / S_0, S_1, \dots, S_{i-1}] = \mathbb{E}[S_i / S_{i-1}]$$

Martingale Property

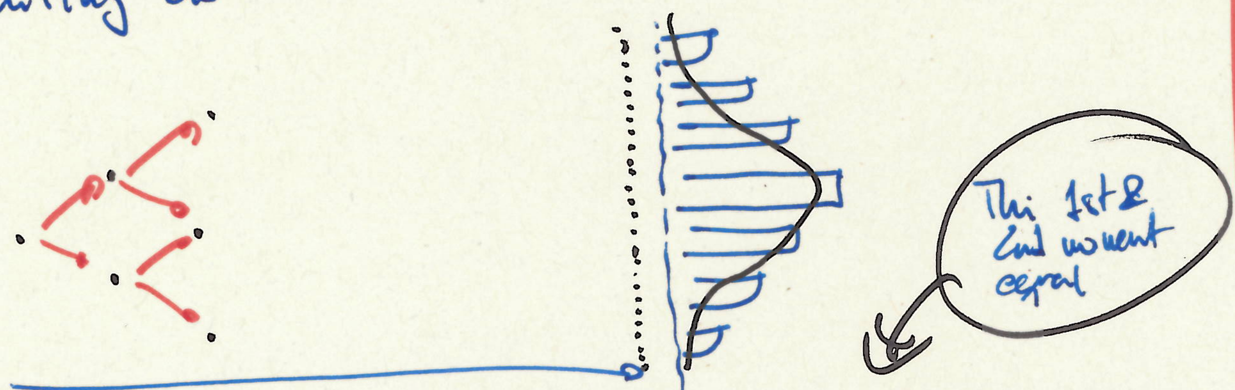
Your expected winnings after any number of tosses is the amount you already hold

$$\mathbb{E}[S_i / S_{i-1}] = S_{i-1}$$

Quadratic Variation
Means the expectation of the squared process increments is finite.

$$\mathbb{E}\left[\sum_{i=1}^{j}(S_i - S_{i-1})^2\right] = j$$

Normality
For large n the binomial ~~approximation~~ $B(n,p)$ is well approximated by $N(np, np(1-p))$, and the limiting distribution as $n \to \infty$ is N



$n \uparrow \infty$

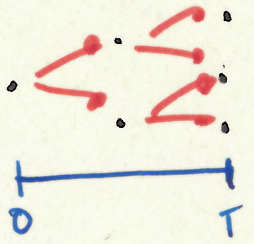
$\mathbb{E}[S] + S_n \sim B(n, 1/2) \sim N(n/2, n/4)$

Let us now change the game a bit. For a fixed time T we now play n games (tossing) every T/n units of time. ~~And in~~ We also modify the winning / lose quantity by $\sqrt{T/n}$. It is easy to see that

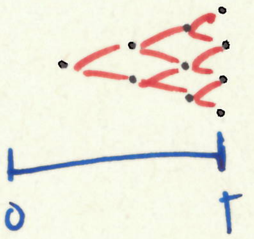
$$\mathbb{E}[S_n] = \mathbb{E}[S_T] = 0$$

$$\mathbb{E}\left[\sum (S_i - S_{i-1})^2\right] = n \times \left(\sqrt{T/n}\right)^2 = T$$

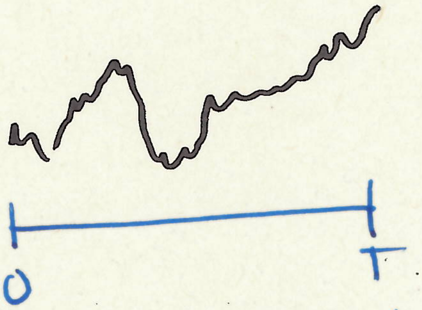
The new game sets up an increasing mesh of binomial models on the interval $[0,T]$.



$n=3$



$n=4$



"One path realization of the binomial dense mesh"

For large n ($n \uparrow \infty$) we will have a process such that:

i) Markovian
ii) Martingale
iii) Finite Quadratic Variation
iv) Normality
v) "Continuity" (it make sense as the mesh is more dense)

$\Downarrow$ By definition $\square$

BROWNIAN MOTION

BROWNIAN MOTION & STOCK RETURNS

We have built a continuous model called Brownian Motion, but is that a good model for stock returns? In other words we are proposing the following model for stock returns

$$\frac{S_{i+1}-S_i}{S_i} = R_i = \text{mean} + \text{standard deviation} * \text{noise}$$

$$= \mu\, dt + \sigma \underbrace{\phi \sqrt{t}}_{\varepsilon_i}$$

- Note that the mean scale by dt
- Note that the variance scale by $\sqrt{dt}$

→ Remember the coin tossing example

In the example of the coin tossing ($\mu = 0$.). But ~~~~ it plausible to see that

- $\mathbb{E}[\varepsilon_i] = 0$; $\mathbb{E}[\varepsilon_i \varepsilon_j] = \delta_{ij}\, dt$
- Martingale, Markov, Normality.

→ ε_t is a Brownian Motion.

For now on we will denote ε_t by a Brownian Motion (W_t). Hence our model is.

$$\frac{S_{i+1}-S_i}{S_i} = \mu dt + \sigma\, dW_t.$$

Yet the question is still to be answer. Is Brownian Motion a good model for stock returns?

$\boxed{\text{The answer is NOT}}$

Real world is not
- Markovian
- Normal
- Not symetric for losses and winings
- ⋮

but it makes our live soo muuuch easy ...!!!
and it is not drametically different.

In practice the approach is to consider the brownian motion a good model and then tune its properties bit by bit to enhance the model and make it explain better the reality .....

For example keep all properties of the Brownian Motion and remove Normality ⟹ you get a completely new familily of processes called Levy Processes!!

Lets recap our stock model equation as

$$dS_t = \mu S_t dt + \sigma S_t dW_t$$

meaning:

$$\boxed{\frac{S_t - S_{t-1}}{S_{t-1}} = R_t = \mu (t_i - t_{i-1}) + \sigma \sqrt{t_i - t_{i-1}}\, \underset{\substack{\uparrow \\ \text{White noise}}}{\varepsilon_t}.}$$

An lets assume a given function $V(t, S_t)$ (for example the price of an option depending on S_t and t)
We can write the Stochastic Differential Equation governing $V(t, S_t)$ using the Chain Rule (only in the stochastic world is known as Itô Lemma)

$$dS_t = \mu S_t dt + \sigma S dW_t$$

$$dV = \frac{\partial V}{\partial t} dt + \frac{\partial V}{\partial S} dS + \frac{1}{2} \sigma S^2 \frac{\partial^2 V}{\partial S^2} dt$$

BLACK - SCHOLES PRICING FORMULA

As done before in the binomial model the formula to obtain the price under a given model is to set a non random portfolio by hedging the payoff function and apply a non-arbitrage theory. The idea here is the same. Let us consider

$$\Pi = V(S,t) - \Delta S$$

To construct the Δ-hedge, we assume the underlying process

$$dS_t = \mu S_t dt + \sigma S_t dW_t$$

Using the above equation and the "special chain rule" we get that

$$d\Pi = \frac{\partial V}{\partial t} dt + \frac{\partial V}{\partial S} dS + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} dt - \Delta dS$$

These are the only sources of randomness hence we need to set

$$\Delta = \frac{\partial V}{\partial S}$$

This means that the evolution of our portfolio Π no longer random and it follows the equation

$$d\Pi = \frac{\partial V}{\partial t} dt + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} dt$$

Now we can use a non-arbitrage argument to set

$$d\Pi = r\Pi dt$$

We finally substitute $d\Pi$ and Π in the above equation to end up with

$$\frac{\partial V}{\partial t} dt + \frac{1}{2} \sigma S^2 \frac{\partial^2 V}{\partial S^2} dt = r\left(V - \frac{\partial V}{\partial S} S\right) dt.$$

$$\boxed{\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS \frac{\partial V}{\partial S} - rV = 0}$$

Black - Scholes Formula
worth Nobel Prize

This is not a random equation is a PDE with some initial condition on the function V (which is the payoff function at expiry).

i Just exactly as we did in the binomial model!

WEEK 7

MONTE CARLO SIMULATION & OPTION PRICING

As we did in the binomial model we have computed a "close-form" (Black-Scholes PDE) to solve the price of an option under a stochastic model for the evolution of prices which follows

$$dS_t = \mu S_t dt + \sigma S_t dW_t$$

But in the binomial model we also saw that the price of the option could be computed as the expectation of the pay-off function, given we could compute this "magic" risk-neutral probability

> **Risk neutral Probability**
> Thanks to the Girsanov-theorem the risk neutral world in the stochastic environment is achieved when we use the risk-free rate for the trend of the stock evolution

$$\boxed{dS_t = r S_t dt + \sigma S_t dW_t}$$

Under the above assumptions, equivalently as done in the binomial model, we can compute the option value as

$$V(t,S) = e^{-r(T-t)}\, \mathbb{E}_{Q}[V(T,S)]$$

The steps to perform a Monte-Carlo algorithm evaluation of the price of the option are:

1) Simulate the risk-neutral random walk

2) Evaluate the payoff function on the previous realization of the algorithm

3) Repeat n-times steps 1) & 2)

4) Average the results obtained in 3)

5) Take the present value of the average

$\Downarrow$ This is the option price!!

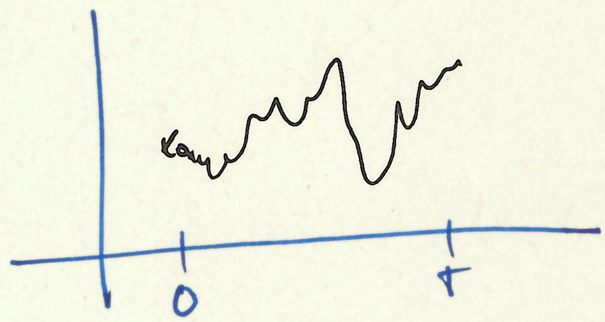
The only step we do not know how to perform at the moment is 1)

EULER DISCRETIZATION SCHEME

Let us consider the general stochastic differential equation:

$$dX_t = a(X_t)\,dt + b(X_t)\,dW_t \qquad (*)$$

for some functions $a(\cdot)$ and $b(\cdot)$. The objective is to simulate the path of the process from $[0,T]$



We first discretize the time interval

$$0 \le t_0 < t_1 \ldots < t_{n-1} \le t_n = T \;; \quad t_j = j\Delta T, \quad \Delta T = T/n$$

The discretization of $(*)$ is

$$X_{t+\Delta T} = X_t + \int_t^{t+\Delta T} a(X_s)\,ds + \int_t^{t+\Delta T} b(X_s)\,dW_s$$

We now analyse each term:

$$\int_t^{t+\Delta T} a(X_s)\,ds \simeq a(X_t)\int_t^{t+\Delta T} dt = a(X_t)\,\Delta T$$

$$\int_t^{t+\Delta T} b(X_s)\,dW_s \simeq b(X_t)\int_t^{t+\Delta T} dW_s = b(X_t)\underbrace{\left[W_{t+\Delta T} - W_t\right]}$$

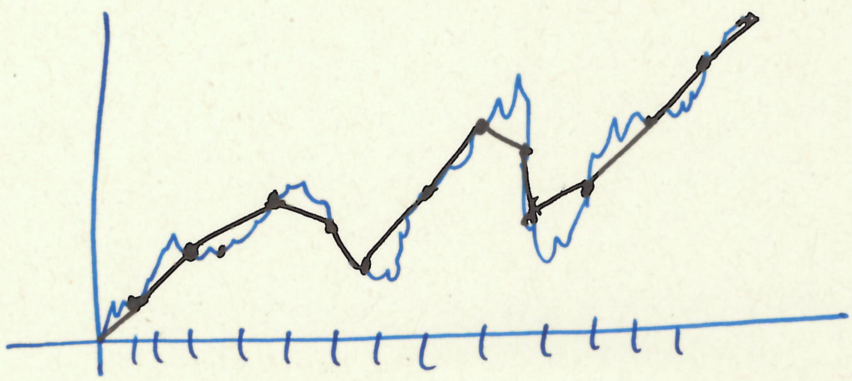
We know that increments of Brownian Motion are normally distributed with 0 mean and $\sqrt{\Delta T}$ as standard deviation.

Putting all together we get the Euler's discretization

$$\begin{cases} X_0 = X(0) \\ X_{t+\Delta T} = X_t + a(X_t)\,\Delta T + b(X_t)\sqrt{\Delta T}\,\varepsilon_t \end{cases}$$

where $\varepsilon_t \sim N(0,1)$

What we did ~~is~~ to build up a skeleton for the realization of the path



As an example, lets consider we want to price a call option, under the model

$$dS_t = rS_t dt + \sigma S_t dW_t$$

Monte Carlo method is as follow:

1. Simulate a path for S applying the Euler scheme to

   $S(0) = S_0$
   $a(S) = rS$
   $b(S) = \sigma S$

   for a given value N (discretization steps)

2. Apply the payoff function to S_N

3. Repeat 1) & 2) multiple times (n) an average it

4. The price of the option is

$$\left(\frac{1}{n}\sum_{i=1}^{n} f(S_N^i)\right)e^{-r(T-t)}$$

The advantage of Monte Carlo methods has now become clear, since it is much easier than to solve an PDE. Also higher dimensional problems are easy to generalize and minor changes in the model are ~~easi~~ easily fit onto the simulation.

(4) MATHEMATICAL OPTIMIZATION: PORTFOLIO MANAGEMENT

References

Argimiro Arratia, Ch 8	"Computational Finance"	
Ruppert, Ch 11 & 16	"Statistics and data analysis for financial engineering"	

Markowitz presented in 1952 the basic idea of portfolio selection: to find a combination of assets that in a given period of time produces the highest possible return at ~~a~~ least possible risk.

He received the Nobel price in economics in 1990.

This idea laid the basic foundation of portfolio management

- Determine the different combination of assets (portfolios)
- Choose the portfolio that best suits a particular investor.

TOY EXAMPLE

Imagine you are in a tropical island and have only USD 100 to invest. Investing opportunities are very limited, we can only invest in umbrellas and icecream.

The payoff of the investment depends on the weather as

weather	ice cream	umbrella
sunny	120	90
rainy	90	120

Assume the probability of the weather beeing sunny or rainy is the same.

Lets analyse the following two options:

a) Invest everything you have in either product

b) Split the investment equally between products.

a) The expected return is equivalent in each product

$$\mathbb{E}[R] = \frac{0.5 * 120 + 0.5 * 90}{100} - 1 = 5\%$$

but there is a chance of you earning more or less than that.

b) If we split the investment, the outcome does not depend on the evolution of the weather, you will earn \$45 in one product and \$60 in the other, hence

This outcome is for sure.

$$\mathbb{E}[R] = \frac{45 + 60}{100} - 1 = 5\%$$

Even though the outcome is the same, the difference between option a) and b) is that b) is riskless

The main concept of portfolio optimization is captured in this example. Based on the correlation between investment products we can reduce the risk of the portfolio and still get the desired return

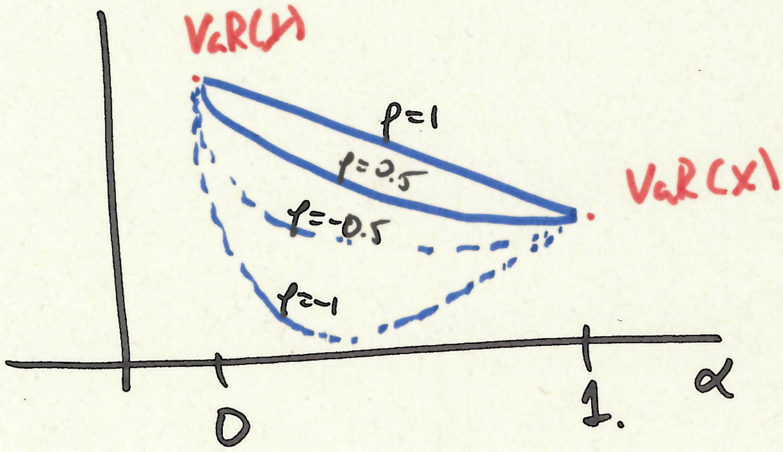
Diversification reduces risk

To be mathematically more precise. Let X, Y be random variables with variances σ_x^2 and σ_y^2. The variance of the combination is:

$$VaR(\alpha X + (1-\alpha) Y) = \alpha^2 \sigma_x^2 + (1-\alpha)^2 \sigma_y^2 + 2\alpha(1-\alpha) Cov(X, Y)$$

For different values of their correlation the variance looks like

Here, the risk is measured as variance!



The variance can be completely eliminated if and only if the correlation between X and Y is -1 or 1 and the variance of X and Y are not the same.

Portfolio optimization is the theory that studies the correlation between assets to build up a portfolio with low risk and high return

MEAN-VARIANCE MODEL

The starting hypothesis of Markowitz theory is that

- investors should consider expected return as desirable
- avoid the variance of returns (risk)

This rule of thumb has a first consequence known by investors as the importance of diversification.

Let P be a portfolio of N risky assets (stocks)

$$R_p = \sum_{i=1}^{N} w_i R_i \quad , \text{ where } \sum_{i=1}^{N} w_i = 1$$

are the weights and R_i the return of the asset i-th.

From the above equation the expected return is

$$\mu = \mathbb{E}[R_p] = \sum_{i=1}^{N} w_i \, \mathbb{E}[R_i]$$

and the variance

$$\sigma^2 = Var(R_p) = \sum_{i}^{N} \sum_{j}^{N} w_i w_j \sigma_{ij}$$

where $\sigma_{ij} = Cov(R_i, R_j)$

Consider the following extreme cases:

Option A The return of the assets are pairwise uncorrelated, ie

$$\mathrm{Cov}(R_i, R_j) = \begin{cases} 0 & i \neq j \\ \sigma_i^2 & i = j \end{cases}$$

Therefore

$$\mathrm{Var}(R_p) = \sum_i^N w_i^2 \sigma_i^2$$

Furthermore, assume the portfolio is equally weighted $w_i = 1/N$. Then

$$\mathrm{Var}(R_p) = \frac{1}{N^2} \sum_{i=1}^N \sigma_i^2 \leq \frac{[\max(\sigma_i \mid i=1..N)]^2}{N}$$

As $N \to \infty$, the variance of the portfolio decreases to zero

Option B The return of the portfolio are similar correlated. Let us consider an equally weighted portfolio with $\mathrm{Var}(R_i) = \sigma$ an correlation $0 \leq c \leq 1$. Then

$$\mathrm{Var}(R_p) = \sum_i \sum_j w_i w_j \sigma_{ij} = \frac{1}{N^2} \sum_i \sum_j \sigma_{ij}$$

$$= \frac{1}{N^2} \Big(\sum_{1 \leq i \leq N} \sigma_{ii} + \sum_{i \neq j} \sigma_{ij} \Big)$$

$$= \frac{\sigma^2}{N} + \Big(1 - \frac{1}{N}\Big) c\sigma^2 = \frac{1-c}{N} \sigma^2 + c\sigma^2$$

This shows that no matter how large is N, it is impossible to reduce the variance of the portfolio below σ.

The conclusion from the above example is:

- The greater the number of pairwise uncorrelated assets, the smaller the risk of the portfolio.
- The greater the presence of similarly correlated assets, the closer is the risk of the portfolio to a risk common to all assets.

Therefore, diversification in a mean-variance world is accomplished by considering highly uncorrelated assets in some reasonable number.

MINIMUM RISK MEAN-VARIANCE PORTFOLIO

Given a portfolio of N risky assets, consider the following matrices:

$w = (w_1, \ldots, w_N) \longrightarrow$ vector of weights

$C = [\sigma_{ij}]_{1 \le i, j \le N} \longrightarrow$ matrix of covariances $\sigma_{ij} = Cov(R_i, R_j)$

$\mu = (\mu_1, \ldots, \mu_N) \longrightarrow$ vector of expected returns $\mu_i = \mathbb{E}[R_i]$

with this vector notation we can write the following expressions for the expected return of a portfolio and its variance

$$\mathbb{E}[R_p] = w^t \mu$$

$$Var(R_p) = w^t C w$$

According to Markowitz's rationale the investor's objective is to obtain a certain level of return with the smallest amount of risk.

The minimum variance optimization problem is formulated as:

Find w such that

$$\min_w \; w^t C w$$

subject to: $w^t \mu = r^*$

and $\sum_i w_i = 1$

for a given level of expected return r^*

This is a quadratic programming problem ~~wit~~ which can be solved using "Lagrange multipliers". Before we give the analytic solution, let us observe:

- There is no restriction on the sign of the weights w_i, therefore we might have long and short positions

- we use all our money to invest, ie the sum of the weights is 1.

The Lagrange multipliers consists in defining the Lagrangian

$$\mathcal{L} = \underbrace{w^t C w}_{\text{objective optimization}} - \underbrace{\lambda_1 (w^t \mu - r^*) - \lambda_2 (w^t \mathbb{1} - 1)}_{\text{restrictions}}$$

$$= \sum_{1 \le i,j \le N} w_i w_j \sigma_{ij} - \lambda_1 \left(\sum_{i=1}^{N} w_i \mu_i - r^* \right) - \lambda_2 \left(\sum_{i=1}^{N} w_i - 1 \right)$$

Then we look for a local minimum of the function with respect to λ_1, λ_2

$$\frac{\partial L}{\partial \lambda_1} = 0$$

$$\frac{\partial L}{\partial \lambda_2} = 0$$

Imposing these equations we ensure the restrictions are met

$$\frac{\partial L}{\partial w_i} = 0 \qquad i = 1, \ldots, N$$

$\Rightarrow$ Optimizes the objective to find a local extreme for w_i

We end up with $N+2$ equations

$$\begin{cases} 2\sum_{i=1}^{N} w_i \sigma_{ij} - \lambda_1 \mu_j - \lambda_2 = 0 & j = 1, \ldots, N \\ \sum_{i=1}^{N} w_i \mu_i = r^* \\ \sum_{i=1}^{N} w_i = 1. \end{cases}$$

Which in matrix formulation is reduced to:

$$\begin{pmatrix} C & \mu & \mathbb{1} \\ \mu^t & 0 & 0 \\ \mathbb{1}^t & 0 & 0 \end{pmatrix} \begin{pmatrix} w \\ \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} 0 \\ r^* \\ \mathbb{1} \end{pmatrix}$$

where $\mathbb{1} = (1, \ldots, 1)$.

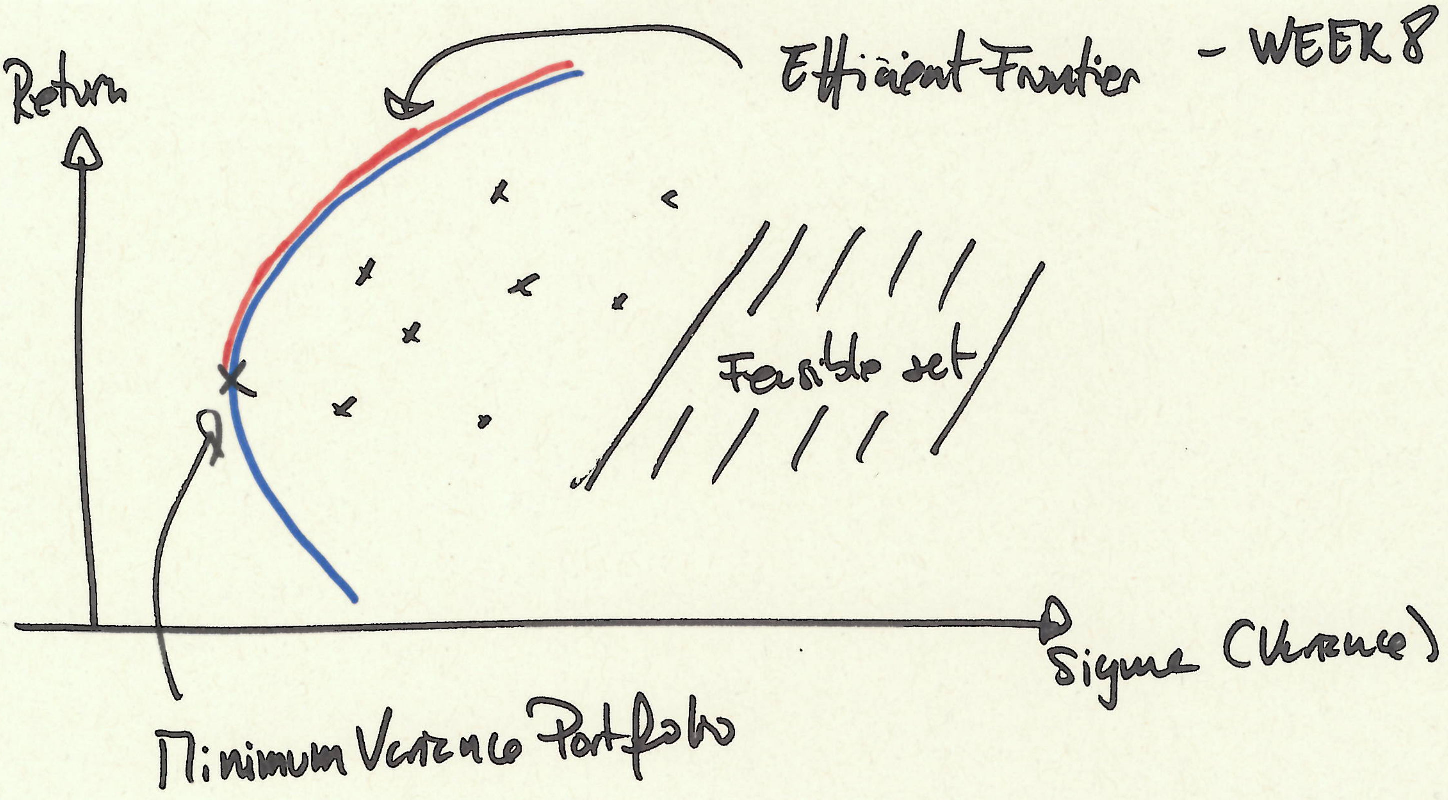
The above sistema is easily solve by any linear method.

This solution is termed efficient in the sense of being the portfolio with that expected return (r^*) and minimum variance. Any other portfolio with that return (r^*) must have higher variance.

THE EFFICIENT FRONTIER

For a set of N assets, we consider a given objective return r^*, then the Lagrange optimization problem finds the optimal weights $\{w_i\}_i$ for the minimal variance portfolio σ^*_i.

If we plot the pair (σ^*, r^*) for different values of r^* we get the efficient frontier.



- The region inside the hyperbola is the feasible set, consisting in all possible combination of the N-risky assets, ie all possible existing portfolios.

- The upper branch of the hyperbola is the efficient frontier, ie for a given level of risk (σ) it gives you the highest return

- The locus of the hyperbola corresponds to the Minimum Variance Portfolio.

Anything outside the hyperbola is not possible to achieve.

- The lower branch of the hyperbola is clearly not efficient and hence not desirable from the point of view of the investor.

Matrix formulation substituting page 10 of 08_0_Portfolio_optimization.pdf:

$$\begin{pmatrix} C & \mu & \mathbb{1} \\ \mu^t & 0 & 0 \\ \mathbb{1}^t & 0 & 0 \end{pmatrix} \begin{pmatrix} \hat{\omega} \\ \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} \hat{0} \\ r^* \\ 1 \end{pmatrix}$$

where $\mathbb{1} = (1, 1, \cdots, 1)^t$, $\hat{0} = (0, 0, \cdots, 0)^t$ and $\hat{\omega} = (\omega_1, \omega_2, \cdots, \omega_N)^t$

PORTFOLIO OPTIMIZATION WITH ONE RISK-FREE ASSET

We are going to complete the theory of portfolio optimization with the inclusion of a risk-free asset.
We start by assuming one risky asset and one risk-free asset

	expected return	standard deviation
risky	0.15	0.25
risk-free	0.06	0.0

Assume that the fraction w of our wealth is invested on the risky asset, then

→ $\mathbb{E}[R_p] = w\, 0.15 + (1-w)\, 0.06 = 0.06 + w\, 0.09$

→ $\sigma_p^2 = w^2 (0.25)^2 + (1-w)^2\, 0^2 = 0.25^2\, w^2$.

To decide which portion w to invest in the risky asset one chooses:

→ The level of expected return $\mathbb{E}[R_p]$

→ The amount of risk, σ_p^2, willing to hold.

Assume there is some "optimal" risky portfolio, then the choice to invest between this "optimal" portfolio and in a risk-free asset is made as in the example ~~above~~ above. This optimal risky portfolio is called Tangency portf

TWO RISKY ASSETS & THE TANGENCY PORTFOLIO

The mathematics of considering N-risky assets and one risk-free asset is most easily understood when the are only two risky asset.

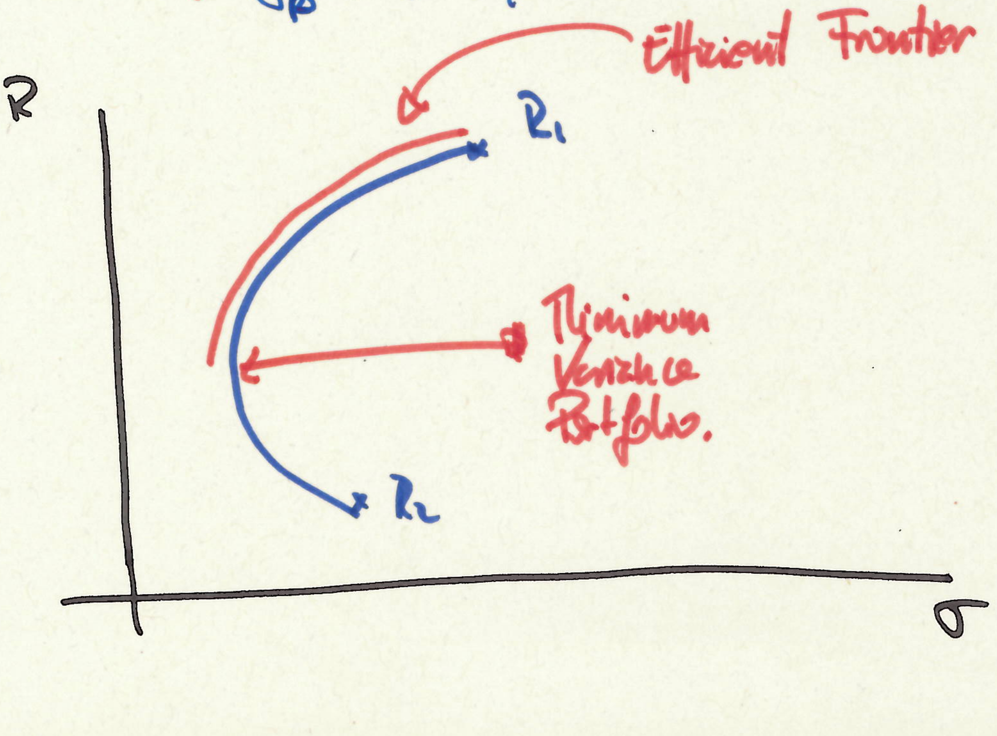
Let R_1, R_2 be the returns of risky asset one and risky asset 2 respectively, and let w the amount owned by R_1. (hence $1-w$ owned for R_2)

The return of the portfolio is

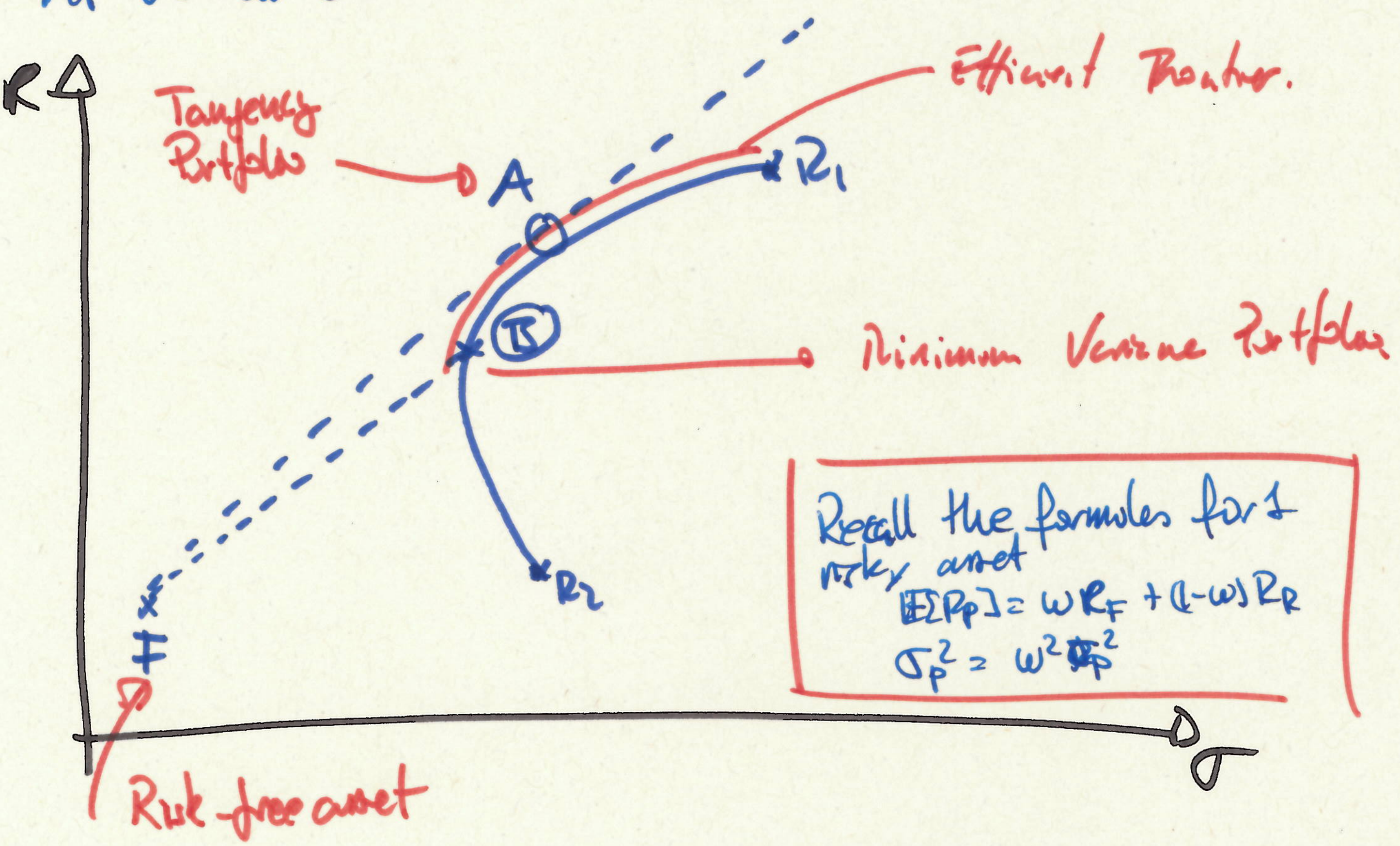
- $R_p = wR_1 + (1-w)R_2$
- $\mathbb{E}[R_p] = w\mu_1 + (1-w)\mu_2$

Let ρ_{12} be the correlation between the return on the risky assets. The variance of the return is

- $\sigma_p^2 = w^2\sigma_1^2 + (1-w)^2\sigma_2^2 + 2w(1-w)\rho_{12}\sigma_1\sigma_2$

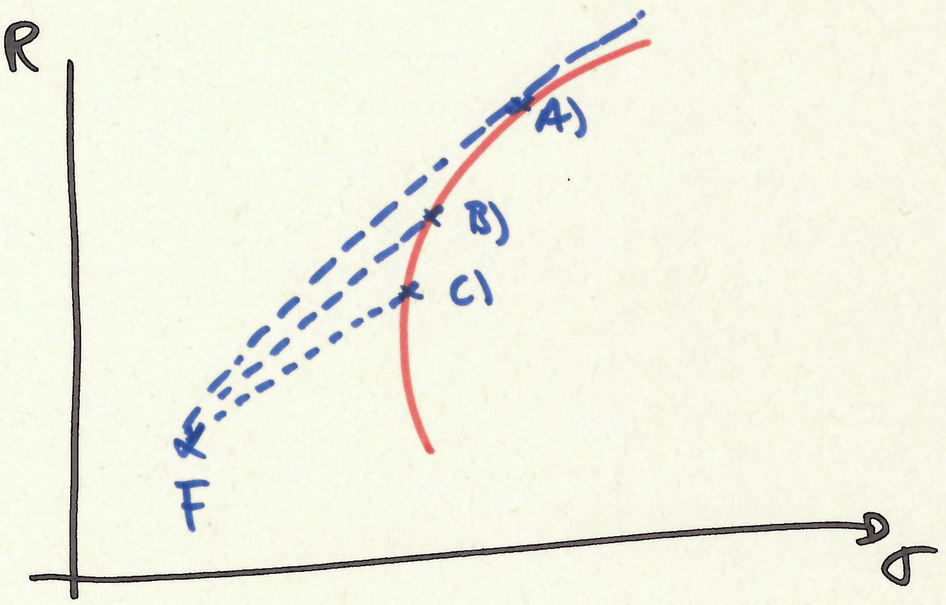


Let us now enhance the plot with the risk-free asset



The point F in the plot represents the risk-free asset. Each line from F to any portfolio on the efficient frontier is a possible allocation of wealth between the risk-free asset and the combination of the two risky assets.

Although not all combinations of the risk-free asset and a point on the efficient frontier seem efficient or optimal.



For each point in line B), the corresponding vertical intersection in line A) is more efficient as it has the same risk for a greater expected return.

Only the combination of the risk-free asset with the tangency portfolio is optimal, in the sense that no other line would have larger slope

The slope of this set of investment is called Sharpe's ratio

$$\text{Sharpe Ratio} = \frac{\mathbb{E}[R_p] - \mu_f}{\sigma_{R_p}}$$

where

$\mathbb{E}[R_p]$ is the expected return of the portfolio in the efficient frontier

σ_{R_p} is the standard deviation of the portfolio in the efficient frontier

μ_f is the return of the risk-free asset.

Sharpe's ratio can be thought as "risk-reward ratio"; re the amount of reward for unit of risk. A line with a larger slope gives more reward for a unit of risk.

Optimal Portfolios:

Note that with the inclusion of one risk-free asset all optimal portfolios are a combination of the risk-free asset and the tangency portfolio.

Therefore all optimal portfolios have the same combination of the risky assets. The only parameter that can vary is the amount of wealth allocated to the risk-free asset or ~~the~~ to the tangency portfolio.

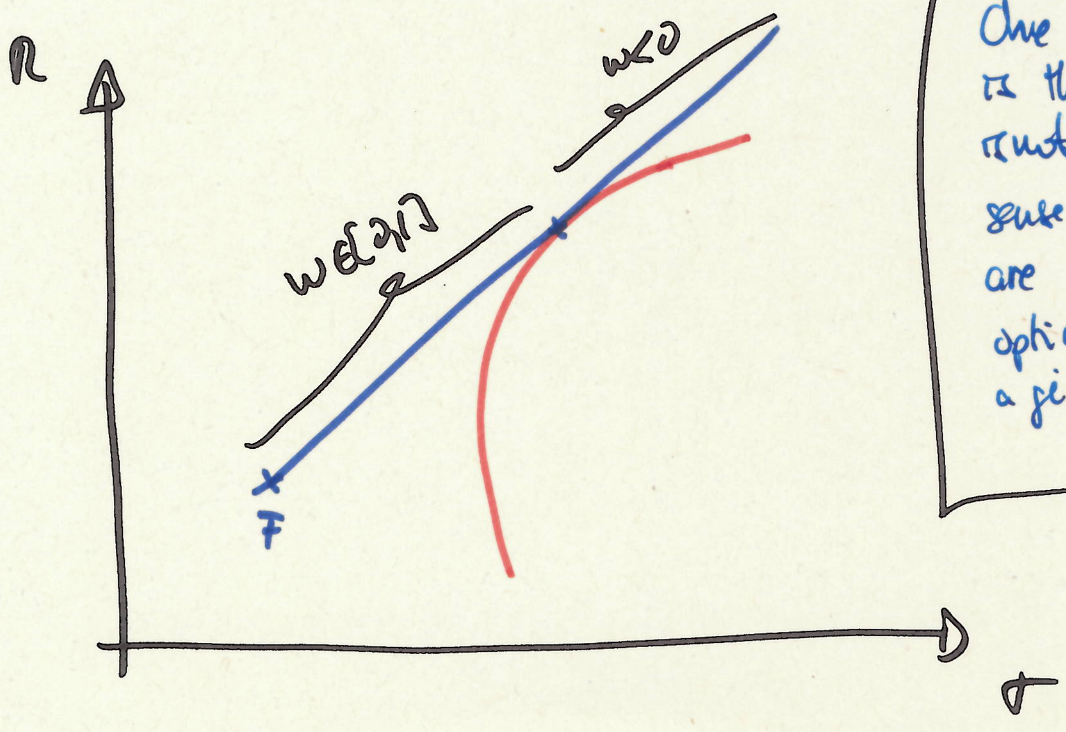
The mix of the tangency portfolio with the risk-free asset yields the following rules:

- $\mathbb{E}[R_P] = \mu_f + w(\mathbb{E}[R_T] - \mu_f)$
- $\sigma_P = w\,\sigma_T$

LEVERAGE POSITION

Sometimes the expected return of the portfolio asked by the investor is not accomplished by any point between the risk-free asset and the tangency portfolio. In such cases one need to borrow money to invest more in the risky asset.

$$P = w R_f + (1-w) R_T$$



One final consideration is that the model is not dynamic, in th sense that all computati are performed to the optimal portfolio for a given time horizon

CAPITAL ASSET PRICING MODEL

CAPM starts trying to answer the following question:

What would be the risk premium on securities if the following assumptions were true?

i) The market prices are in "equilibrium". For each asset demands = supply.

ii) Everyone has the same forecast of expected returns and risk

iii) Everyone ~~has the~~ choose portfolios optimally, as discussed earlier.

iv) The market only rewards unavoidable risk.

① CAPITAL MARKET LINE (CML)

The CML relates the excess return on an efficient portfolio to its risk. The excess expected return is also known as the risk premium

$$\mu_R = \mu_f + \frac{\mu_M - \mu_f}{\sigma_M}\sigma_R$$

R is an efficient portfolio.
μ_f is the risk free rate
σ_R is the risk of the efficient portfolio.
μ_M is the return on the tangent portfolio.
σ_M is the risk of the tangent portfolio.

The slope of CLM is

$$\frac{\mu_M - \mu_f}{\sigma_M}$$

which is equal to $\quad \dfrac{\mu_R - \mu_f}{\sigma_R} = \dfrac{\mu_M - \mu_f}{\sigma_M}$

which says that the risk-to-reward (Sharpe's ratio) for any efficient portfolio equals that of the tangent portfolio.

⑩ SECURITY MARKET LINE

The SML relates the excess return on an asset to the slope of its regression on the tangent portfolio. The SML differs from the CML in that SML applies to all assets while CML only to efficient portfolios.

Assume stocks are indexed by j-th. Define

σ_{jn} = covariance between the returns on the j-th asset and the tangent portfolio

$$\beta_j = \frac{\sigma_{jn}}{\sigma_n^2}$$

It follows that β_j is the slope of the best linear predictor of j-th stock's return using the returns of the tangent portfolio as predictor variable, i.e.

$$\mu_j - \mu_f = \beta_j (\mu_n - \mu_f)$$

This equation states that the risk-premium for a security is the product of β_j and the risk premium of the tangent portfolio. Consequently, β_j is a measure of how "aggressive" the j-th asset is.

By this definition the β of the tangent portfolio is 1, hence

$\beta_i > 1 \Rightarrow$ "aggresive"

$\beta_j = 1 \Rightarrow$ "risk-averse"

$\beta_i < 1 \Rightarrow$ "defensive"

Assets with $\beta > 1$ tend to overreact the market, and assets with $\beta < 1$ tend to be more conservative. This is a highly monitored parameter in the industry

As simple as it may seem, the computation of β is highly complex. For instance, ask yourself the following questions:

i) Shall I use historical returns to obtain β, or forecast it?

ii) If historical data is used, how long should the period be?

iii) Shall I compute β with the tangent portfolio or should I restrict to a sector index?

iv) Is my β estimation very sensitive? One would prefer to have a robust estimation of β.

5 PROBABILITY CALCULUS: RISK ESTIMATION

We have spent some time now forcasting macroeconomic time series, computing prices of derivatives and optimizing portfolios with the end objective to earn as much money as possible with a given level of risk.

Now we might ask ourselves how to measure risk. The tasks carried out in the risk department have the objective to ensure that the portfolio managers and front office officers operate within the limits stablished by top management.

Therefore their techniques are very similar to those employed in those tasks.

Risk department are usually denoted as second line of defense. The first line of defense would be the same front office managers and the third line of defense would be the internal auditors.

RISK MAP

→ **Market Risk**: Related to prices and evolution of them it is present in any operation involving equities, FX, rates and fixed income

→ **Credit Risk**: It is related to the ability of the counterparty to meet his obligations in a financial contract.

→ **Balance Sheet Risk**: Is the risk coming from the asimitries of the balansheet due to market movements.
For example short deposits vs long loans.
(ALM, Cash flow matching)

→ **Capital Risk**: Regulatory, Economic or Capital allocation

→ **Operational Risk**: Present in every aspect of any operation and due to human errors.

RISK ON MARKET VALUE OR CASHFLOW

Once we have a portfolio of risky assets we need to determine which risk factors drive their evolution. The objective is to:

a) Quantify the variability of those risk factors into the value of my portfolio

b) Analysing the cashflow model of the portfolio.

The main risk factors are related with market variables (prices, curves, etc...) or macroeconomic variables. All those factors generate the risk market.

We also need to fix the ~~horizontal~~ time horizon for the analysis.

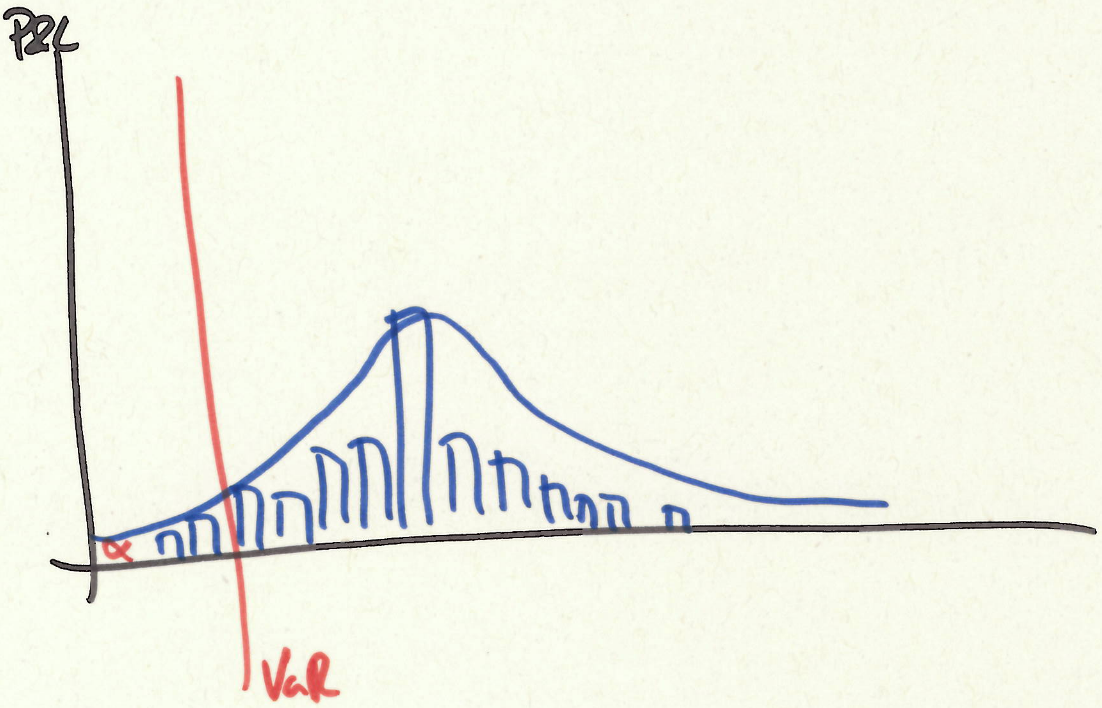
Generally speaking once we have targeted the main driving risk factors we need to built up a model to sistemize the aggregate behaviour.

VALUE AT RISK

By far the most common risk measure is value at risk (VaR)

VaR

For a given time horizon T and level of confidence α, VaR is the maximum loss that a given portfolio can have within T time at $(1-\alpha)$%.



Another variation of the VaR measure is the tail VaR or conditional VaR

Tail VaR

For a given time horizon T and level of confidence α, tail VaR is the expected loss once VaR is overcome.

There exists different methodologies to compute VaR.

→ **Historical Simulation**: For a given historical set of data, the idea is to simulate historical values of our portfolio to derive the simulated histogram. It does not assume any particular distribution and the idea is to compute an "empirical" VaR.

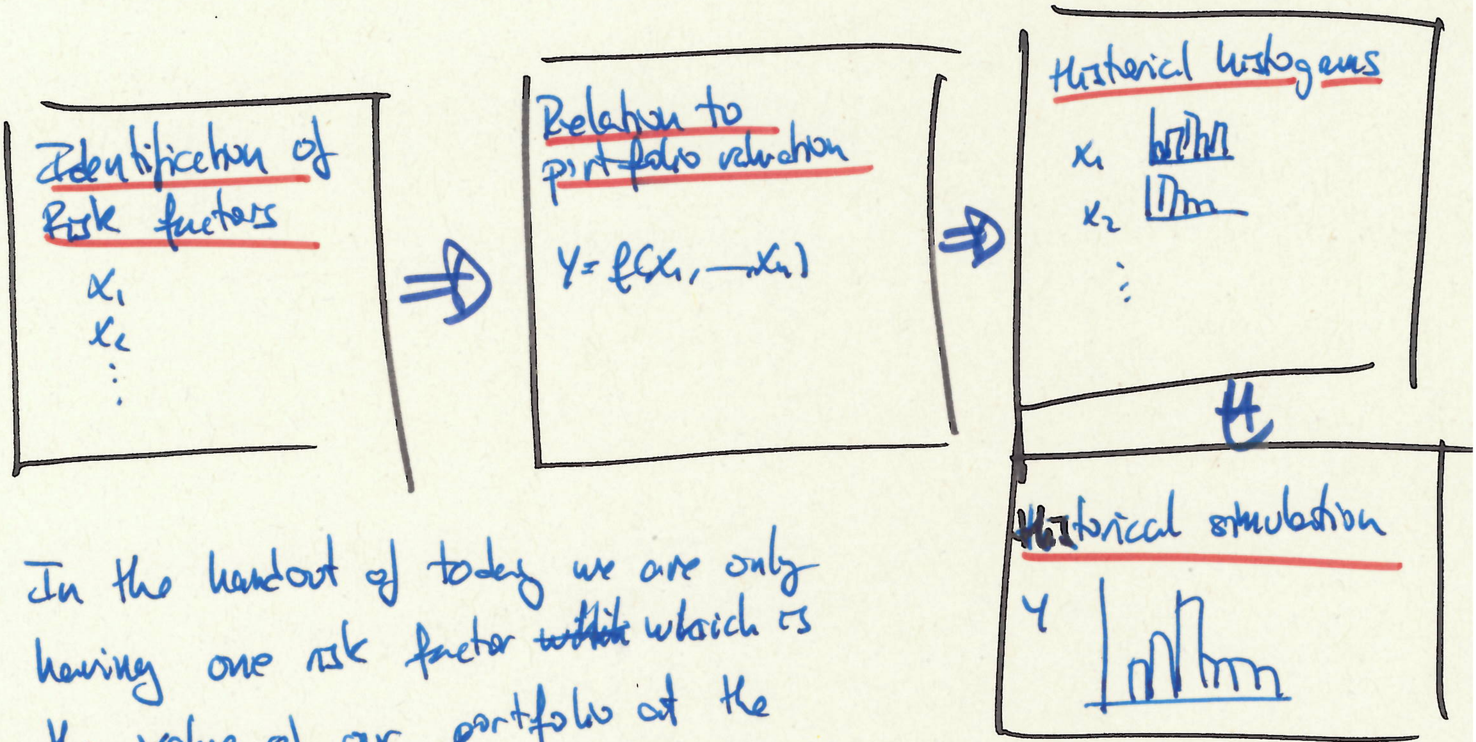
→ **Monte Carlo simulation**: Assumes a given model for the evolution of the underlying assets and computes the VaR using a monte carlo simulation.

HISTORICAL SIMULATION

① For a given portfolio of risky assets we identify the main risk factors and the interaction of those factors with the value of our portfolio.

② We compute empirical historical distribution of such the risk factors and combine them to produce the historical simulated empirical distribution of our portfolio.

③ We finally compute the risk metrics on the histogram.



In the handout of today we are only having one risk factor ~~which~~ which is the ~~value~~ value of our portfolio at the same time so we can go directly to Step 3

MONTE CARLO SIMULATION

Essentially one needs to follow the same steps as for the historical simulation but instead of deriving the historical histograms for the risk factors we simulate under some given model by means of a Monte Carlo simulation.

For today's exercise we are asked to simulate the outcome of a put option and built ~~the~~ a Monte Carlo algorithm to compute the Risk metrics for a put option.

In this case the value or payoff of a put option will only have one risk factor (the price of the underlying stock) and we will simulate it through a stochastic differential equation.

Identification of Risk factors ⇒ Relation to portfolio Valuation ⇒ Monte Carlo simulation of Risk factors

⇓

Monte Carlo simulation of portfolio value.