ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

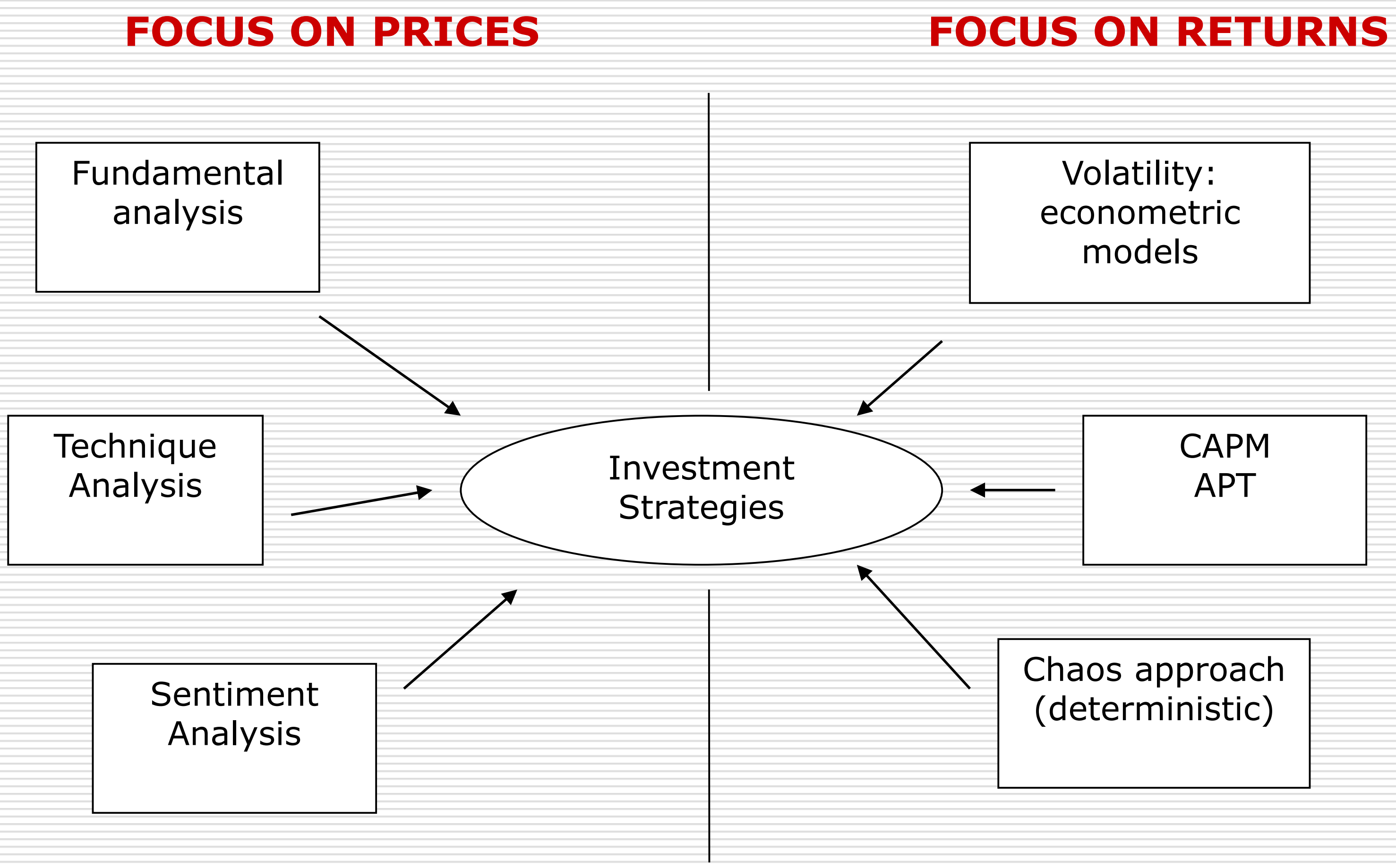
Trend modelling

Andrea Guizzardi
Department of Statistical Sciences
University of Bologna

Sources:

Robert D . Edwards "Technical Analysis of Stock Trends", Ninth Edition
Guizzardi A. (2002) « La previsione Economica » Guaraldi Ed. Rimini
Pring M. « introduction to techical analysis »

Some approach to Investment Strategies (trading)



The "classical" decomposition model.

$$P_t = signal_t + error_t$$

$$error_t \sim N(0, \sigma^2)$$

$$P_t = (trend_t + cycle_t + seasonality_t) + error_t$$

- Financial Markets prices are not predictable (RW theory)
- Trends are **the most persistent component.**
- Practitioners commonly identify with the word "trend" the more persistent components of the signal, sometimes also seasonality. Thus the price dynamics are broken down into a trend component and an erratic component.
- Identify (not predict) trends in prices, allows to build **probailized scenarios** on future prices dynamics

The trend.... Why?

- Because being the 'persistent' component of price, once identified can constitute the base to build a scenario about future stock prices that will have a "good chance” of realization.
- Compared to a Monte Carlo trading strategy, the trend knowledge will allow to set up a trading strategy with a "good chance” of lowering the risk / return ratio.

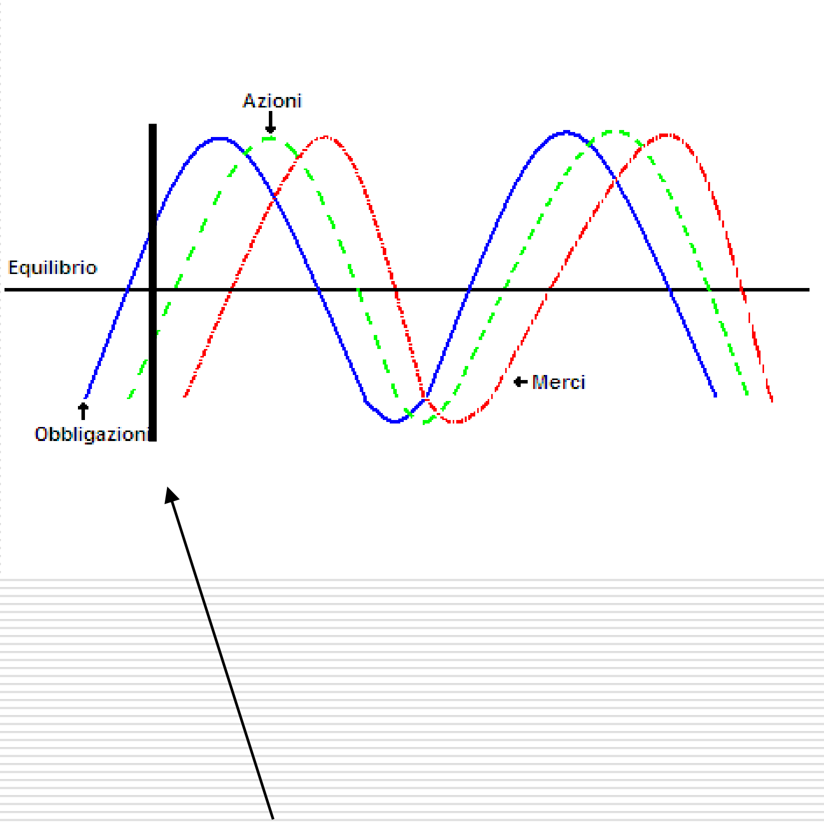
it is therefore important to define how the trend component can be extracted. In finance there are three broad classes:

1. **Calibration methods**
2. **Graphical models**
3. **Analytical models**

Note: some analytical methods can be seen as generalizations of special graphical methods. The consideration - developed later - is the basis for building automated trading systems

Trend Calibration:

the "bonds – stocks – commodities" trend-cycle sequence



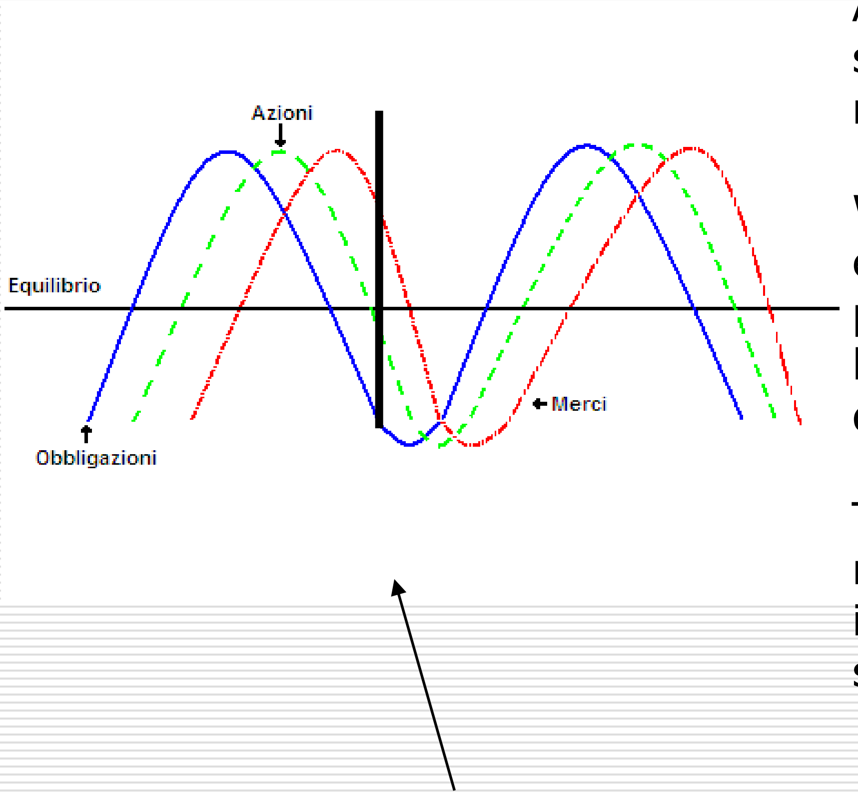
When a recession is going to start the Central bank starts to decrease the discount rate → the bond trend-cycle has a through **and** begins the growth phase (fixed rate bonds – BRDs, BTPs, OATs - increase their value).

During recession - with companies in trouble to financial statements and stock prices to a minimum - the stock market anticipates the subsequent recovery and begins its uptrend.

Nearby the end of recession the production starts to rise dragging the demand for commodities that supports their price.

At the beginning of the recovery phase of the business cycle all three markets are so on the upside.

Trend Calibration:
the “bonds – stocks – commodities” trend-cycle sequence



After the start of the expansion, interest rates stops decrease, the bond market reaches its maximum and reverses the first.

When economic growth is full and the expectations on the increase in corporate profits become moderate, the stock market keeps track of the maximum and starts descending.

To observe the downturn in the commodities market is necessary wait for a real slowdown in the expansion phase coinciding with the slowdown in demand for products.

At this point all three markets are in negative phase, a phase which continues until the beginning of the contraction of the economy, when the bond market reversed its trend starting a new growing phase

Trend Calibration:
the “bonds – stocks – commodities” trend-cycle sequence

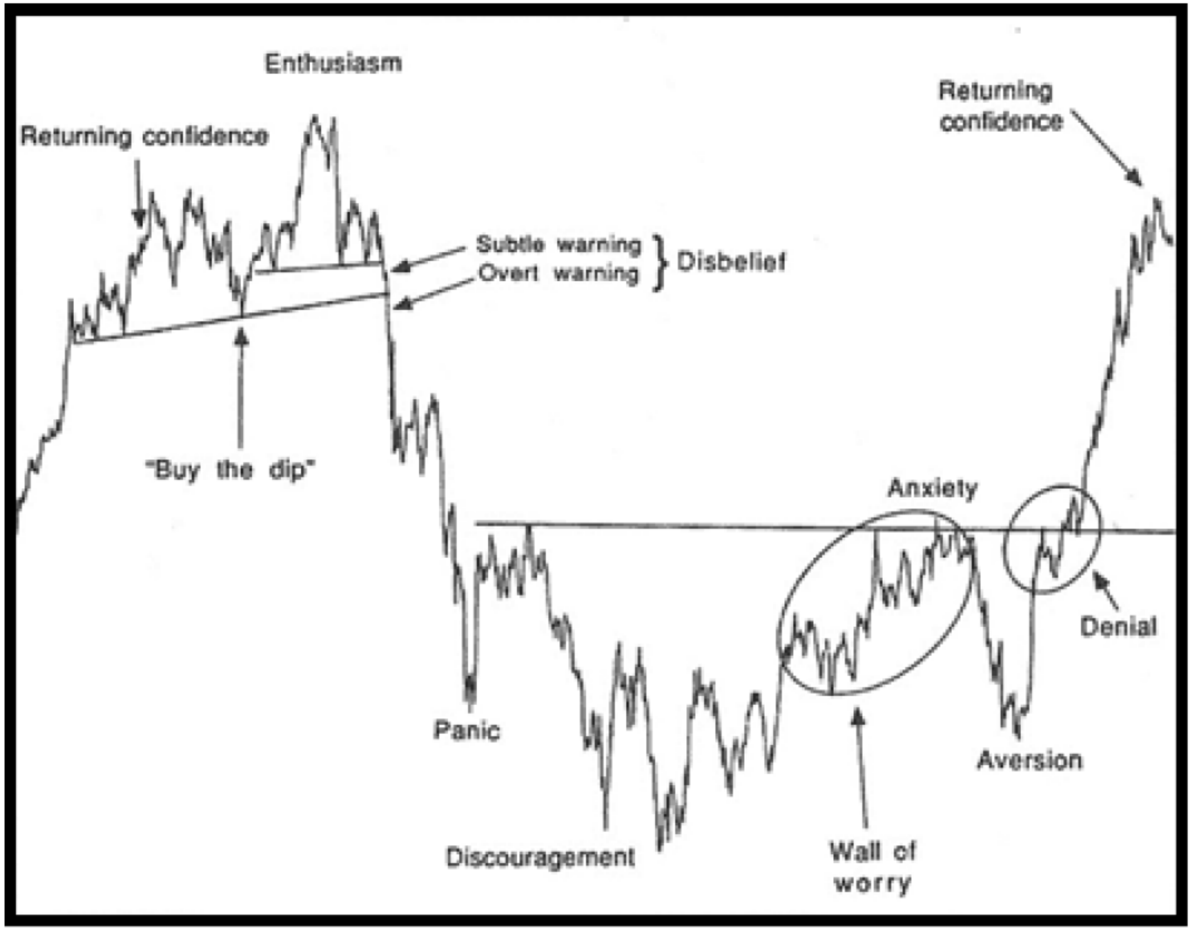
- The presented sequence is only an exemplification of the expected co-movements of prices in different markets.
- Prices of bonds and commodities, are more directly connected to the business cycle than equities.

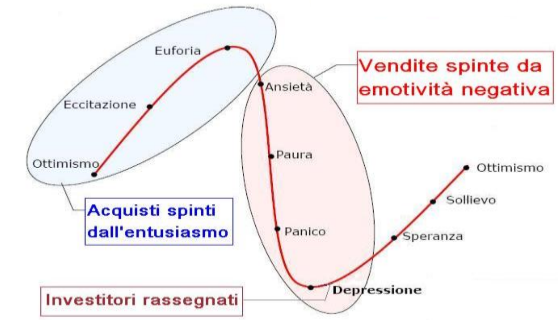
FUTURE PRICE ARE NOT PREDICTABLE but…

- The RW theory does not take into account the psychological motivations: the hopes, fears, knowledge, or the greed of investors.

- During the course of the business cycle, the evolution of the political situation, the speculative waves, exceptional events or doubts about persistence of the actual directional growth/lateral/decline phase, can influence prices, leading to important delays/advances in the rotation of the markets outlined above.

Trend Calibration: **the “sentiment” trend-cycle**





Calibration following a “contrarian” view

Il Panico dell’Ottobre 1998

09/10/1998

Il calo del dollaro e l'irrisolta emergenza degli hedge fund trascinano al ribasso tutti i listini

Borse sempre più nella tempesta

Le maggiori piazze europee perdono oltre il 4% - Bersagliati i titoli delle tlc e dell'auto - Milano annulla i guadagni '98

E la bufera non risparmia nemmeno l’obbligazionario

L'indice Dow Jones cede 270 punti per poi chiudere a -0,1% - Nasdaq in forte ribasso

Wall Street sfiora il crack ma si riprende

I fondi azionari registrano un saldo negativo di 2.295 miliardi per l'impatto dei crash di Borsa

I fondisti in fuga da Piazza Affari

Raccolta complessiva ai minimi da giugno '97 (11.215 miliardi)

07/10/1998

L’Euforia del 2000

Bolla speculativa? Le quotazioni in Borsa sono sempre più considerate il risultato d'un nuovo scenario economico, nel quale il maggior rischio ha preso invece la strada delle obbligazioni

Azioni, come guadagnare nella stagione di Internet

17/01/2000

GLI ANALISTI SONO SICURI, L'ENTUSIASMO VERSO I TECNOLOGICI È UN FENOMENO DESTINATO A DURARE

New economy, scommessa vincente

La grande maggioranza degli operatori punta sul Nuovo mercato

Ma suona un campanello d'allarme. Secondo le previsioni, il 95% degli start up del settore rischia di sparire entro tre anni

25/02/2000

PETER J CANELO, STRATEGIST DI MSDW, SPIEGA PERCHÉ NEI PROSSIMI ANNI CONVIENE PUNTARE SULLE BORSE USA

Il Dow Jones ha spazio per crescere fino a 36.000

Settembre 2000

Calibration following a “contrarian” view

La Negazione della realtà,
“questa volta è diverso!”

IL SONDAGGIO MENSILE MERRILL LYNCH GALLUP NON ESCLUDE LA POSSIBILITÀ DI UN RALLY AZIONARIO

Gestori, i mercati sono sottovalutati

Gli operatori approfittano dei ribassi per entrare nei mercati, ma a differenza del passato gli acquisti avvengono in modo selettivo L'impressione è che la corsa dei tassi sia arrivata al capolinea

18/11/2000

LO SPIEGA IL SONDAGGIO MERRILL LYNCH GALLUP SU UN CAMPIONE DI 241 MONEY MANAGER INTERNAZIONALI

Gestori, nel 2001 favorite le azioni

Due terzi degli investitori prevede che Bce e Banca d'Inghilterra taglieranno i tassi di interesse entro il primo semestre

Quest'anno la crescita economica dell'Europa sarà superiore a quella Usa

18/01/2001

GLI OPERATORI EUROPEI E AMERICANI CONSIDERANO I TITOLI TECNOLOGICI SEMPRE PIÙ SOTTOVALUTATI

Gestori pronti a ricomprare i tmt

Il sondaggio di Merrill Lynch e Gallup sostiene che è possibile un rally dei mercati azionari durante la prima parte del 2001, quando la Federal Reserve inizierà a tagliare il costo del denaro

13/12/2000

E' ora di investire in azioni un po' dei risparmi di riserva

Come nel '95, oggi è un momento adatto per questa strategia di investimento

28/11/2001

TISCALI company

Oct. 27 th 1999 → 46 euro

March 6 th 2000 → 1197 euro

May 14 th 2000 → 40 euro

Today → 0.04 euro

winning bet ???

This time is different ???

A “sentiment” index: the “Daily” Put/Call Ratio

The P/C Ratio is derived by dividing the total number (volume) of option put contracts purchased by the number of option calls, for a given trading day.

Calls, imply the buyer is optimistic (bullish),

whereas Puts imply the buyer is fearful of, or anticipates, a decline (bearish).

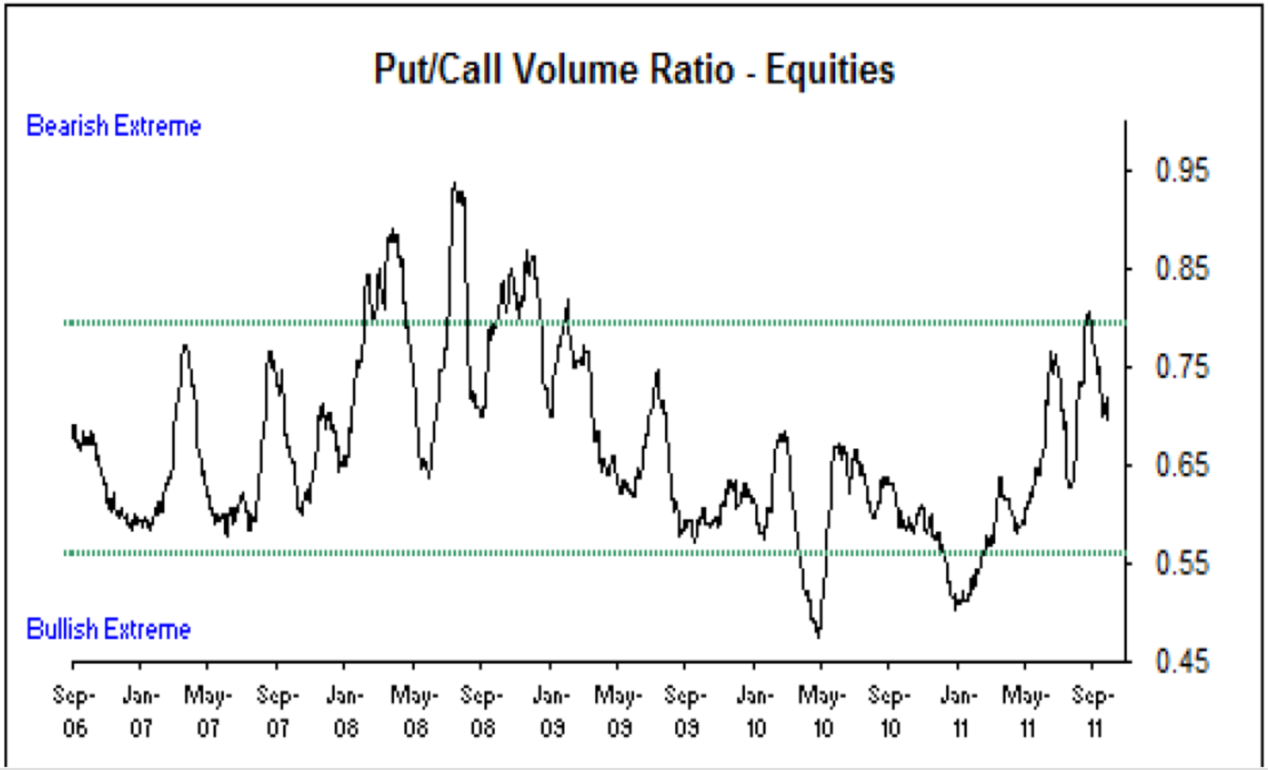
The Chicago Board of Options Exchange (CBOE) calculates the P/C Ratio for all equity and index options traded in their Daily Market Statistics.

Traditionally, the P/C Ratio has thresholds:

- .80 or greater is considered bearish. Readings above 1.00 over a number of trading days are considered strong signs of a market bottom.
- The readings are considered neutral in the .40-.50 range.
- Below .30 is considered bullish. The lower readings are considered strong signs of a topping market nearing a reversal.

Practitioners regard extreme high or low readings as contrarian signals.

A “sentiment” index: the “Daily” Put/Call Ratio



Graphical Methods:
The three Trends in the Dow theory

"… The "market," meaning the price of stocks in general, swings in trends, of which the most important are its *Major* or *Primary* Trends. These are the extensive up or down movements… … interrupted at intervals by *Secondary* Swings in the opposite direction — reactions or "corrections"…. Finally, the Secondary Trends are composed of *Minor* Trends or day-to-day fluctuations which are unimportant to Dow Theory. ..."

The Primary Trends

These are the broad, overall, up and down movements which usually (but not invariably) **last for more than a year**.

So long as each successive rally (price advance) reaches a higher level than the one before it, and each Secondary Reaction stops at a higher level than the previous reaction, the Primary Trend is *Up* (***Bull Market***).

Conversely, when each Intermediate Decline carries prices to successively lower levels and each intervening rally fails to bring them back up to the top level of the preceding rally, the Primary Trend is *Down* and that is called a ***Bear Market***

Long-term investor aim to buy stocks just as soon as he belives the bull marked has started — and then hold them until it becomes evident that it has ended. **Traders**, however, may well concern himself also with the Secondary Swings.

Graphical Methods:
The three Trends in the Dow theory

The Secondary Trends

These are the important reactions that interrupt the progress of prices in the Primary Direction. They are the Intermediate Declines or “corrections” which occur during Bull Markets, the Intermediate Rallies or “recoveries” which occur in Bear Markets. Normally, **they last for 3 weeks to many months**, and rarely longer. Normally (that is frequently), they <u>**retrace from one third to two thirds of the gain**</u> (or loss, as the case may be) in prices registered in the preceding swing in the Primary Direction.

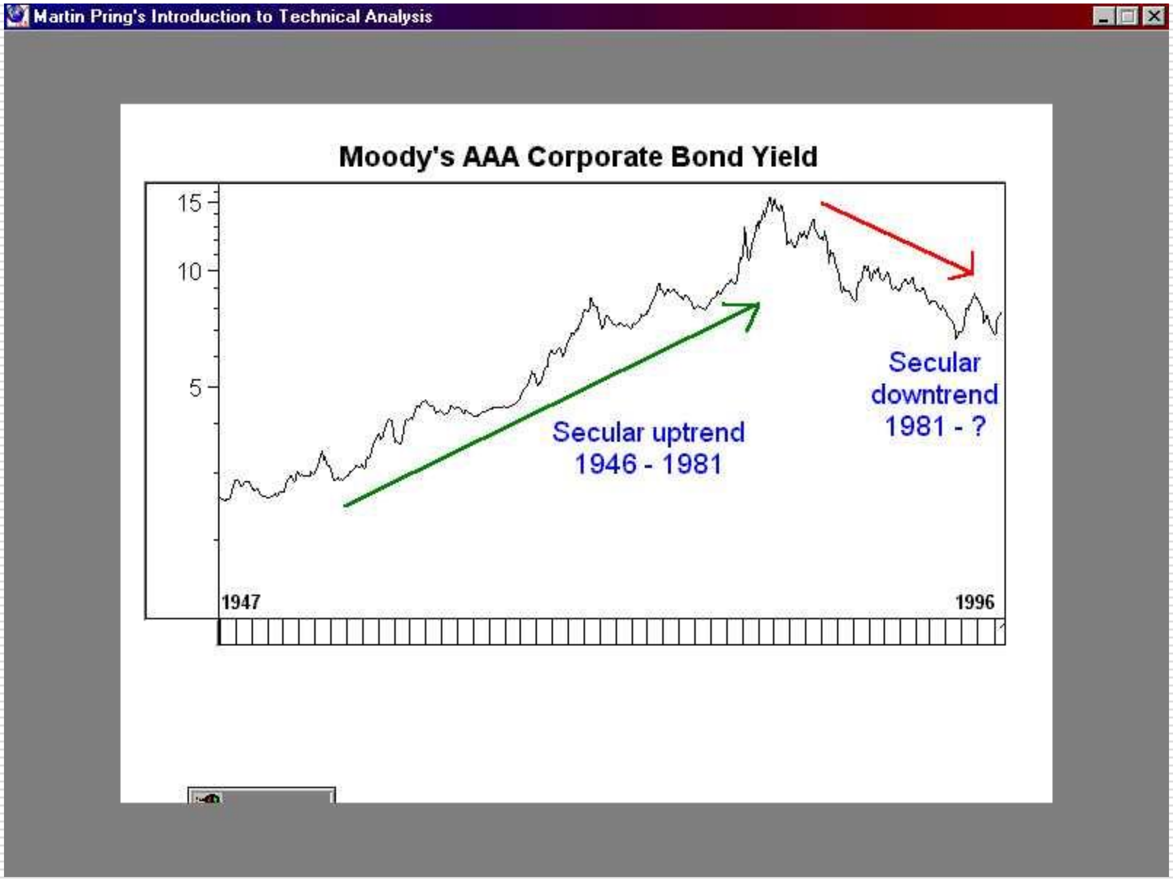
Many Secondary trends stop very close to the halfway mark, retracing 50% of the preceding Primary Swing.

> (Both Secondaries and the intervening segments of the Primary Trend are frequently lumped together as Intermediate Movements)

The Minor Trends

These are the brief (rarely as long as 3 weeks **usually less than 6 days**) fluctuations which are, so far as the Dow Theory is concerned, meaningless in themselves, but which, *in toto* make up the Intermediate Trends

Graphical Methods:
Secular Trends



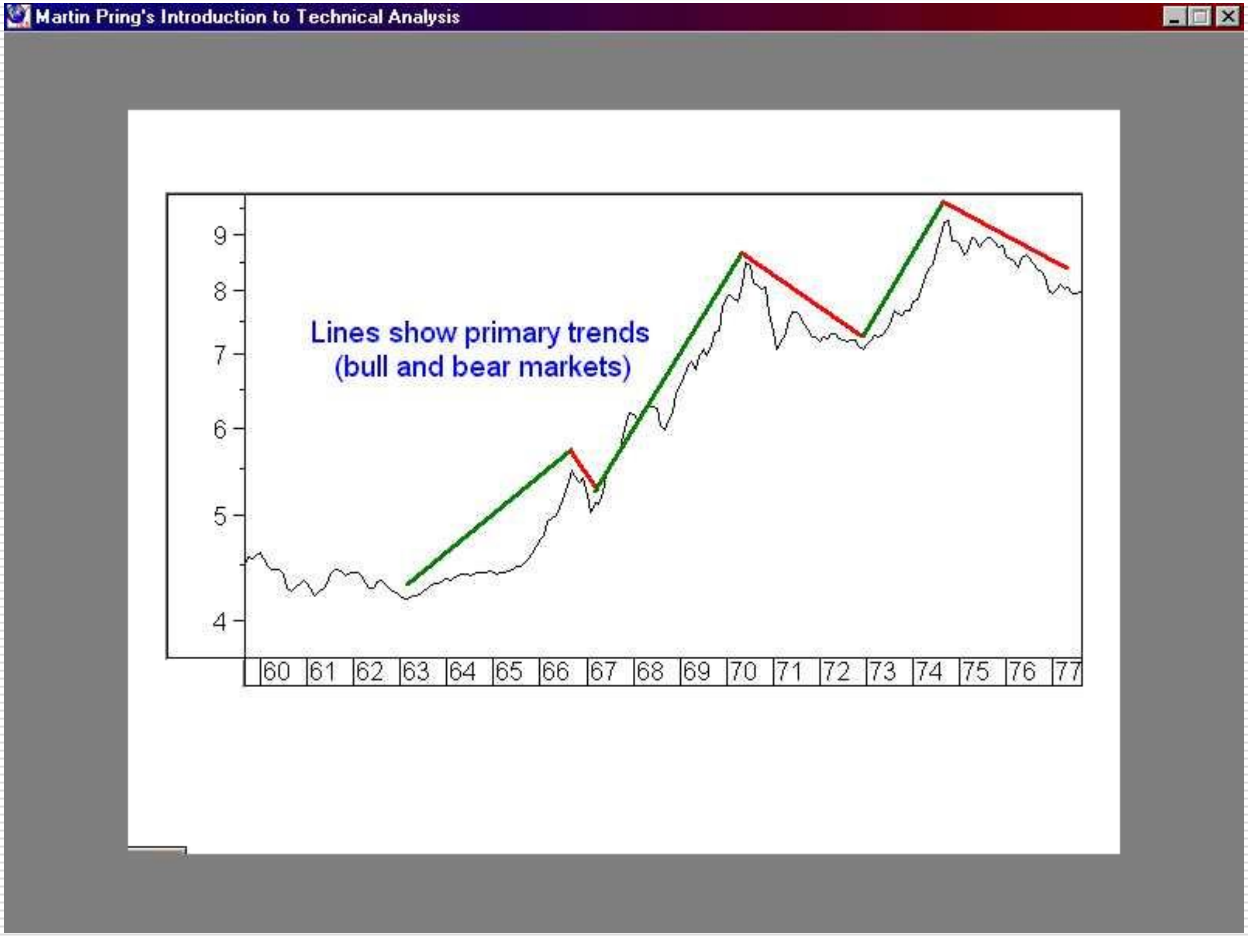
Practitioners:

«The longer a trend has been in existence, the greater the implication of its reversal once a signal has been given»

That is:

Long term returns are bigger than short term returns

Graphical Methods:
The three Trends in the Dow theory



Practitioners:

«financial market's secular trends are interrupted by primary trends»

That is:

(Real) Business Cycle exists and has a direct effect on financial markets

Analytical Methods:

Statistical induction (the trend «extraction»)

Everyone can specify and estimate (calibrate) statistical model from observation:

5; 7; **?**; 11

any doubt?

3; 5; 7; **?**; 11; 13; 15; 17

Two functional forms (model specification). Parameters in bold;

M1 → $Y=\mathbf{2}+Y(-1)$

M2 → $Y=\mathbf{1}+\mathbf{2}*t$ (with t the time)

And the forecast are:

$\hat{Y} = \mathbf{2}+7 = 9$

$\hat{Y} = \mathbf{1}+\mathbf{2}*4 = 9$

<u>With no error!</u> → (deduction not inference). But....

Problem: world is not deterministic:

Statistical induction
(trend «extraction»)

Starting from the same sequence (not rounded)

3,0; 5,3; 6,8; **?**; 10,5; 13,3; 15,1; 16,7

same model?
→ probably not!

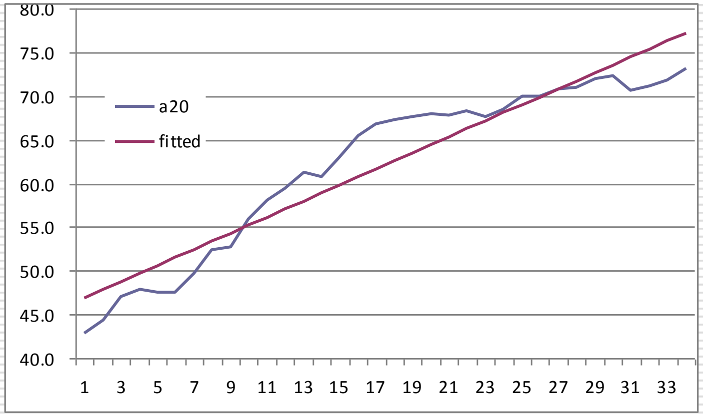
Think as a statistician can be useful in real world

Y=signal+error

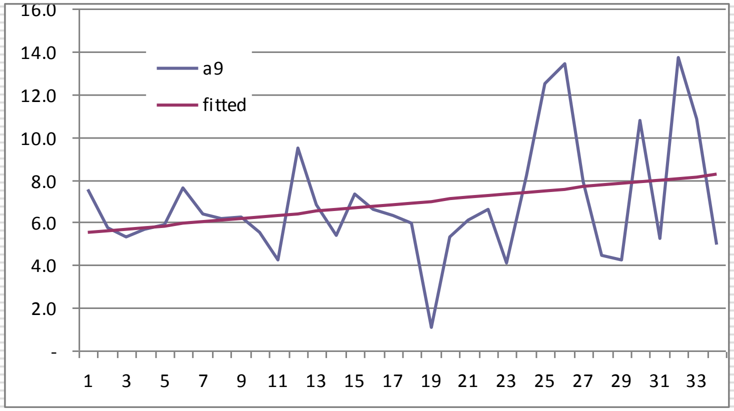
Error is unpredictable, has a distribution form.
Now (in the statistical perspective) the two models are not the same
The first is a high persistence model where shocks effects never end

Statistical induction
(trend «extraction»)

Y_t	t	M_1	M_2	e_1	e_2
3	1	3	3	0	0
5.3	2	5	5	0.3	0.3
6.8	3	7	7.3	-0.2	-0.5
8.9	4	9	8.8	-0.1	0.1
10.5	5	11	10.9	-0.5	-0.4
13.3	6	13	12.5	0.3	0.8
15.1	7	15	15.3	0.1	-0.2
16.7	8	17	17.1	-0.3	-0.4



- The basic idea is that what we observe (the time series) is generated bt an unknown DGP that could be approximatively decomposed (with error) in a trend component and in a cycle component
- The problem is to identify (extract) the best DGP, given a set of point of a time series.



Statistical induction
(trend «extraction»)

… considering the error component the two specification are not equivalent …

deterministic Trend vs stocastic Trend

$$\hat{y}_t = b_0 + b_1 t \qquad vs \qquad \hat{y}_t = b_0 + (1) \cdot y_{t-1}$$

- Key point:
  - Economic implication of the trend specification

Analytical Methods:
Modelling trend (with deterministic functions)

Deterministic trends evolve in a perfectly predictable way

$$P_t = \beta_0 + \beta_1 time_t$$

P = dependent **variable** (the object of interest)
time = independent **variable** (the object who explains the dinamics in dependent variable).
Is a time deterministic variable equal 1 in the first period of the sample, 2 in the second period and so on.
Beta$_0$ = the intercept **parameter** (the constant)
Is the value of trend at time 0.
Beta$_1$ = the slope. If it is positive the trend increase; if is negative the trend decrease.
The larger the absolut value of the **parameter** is the steeper the trend’s slope.

Analytical Methods:
Modelling trend (with deterministic functions)

- There are many deterministic function to explain dynamics in a time series:
  - Linear (previously seen)
  - Quadratic $P_t = \beta_0 + \beta_1 time_t + \beta_2 time_t^2$
  - Polinomial $P_t = \beta_0 + \beta_1 time_t + \beta_2 time_t^2 + ... + \beta_k time_t^k$
  - Exponential $P_t = \beta_0 \bullet \exp(\beta_1 time_t)$
  - Exponential trend is a non linear function of time in levels but in logarithms: $\ln(P_t) = \beta_0 + \beta_1 time_t$
- In polinomial trends the number of paramaters equals to the number of slope changes minus 1
- Trends do not change their dynamics by definition → better use simple functional form

Analytical Methods:
Modelling trend (with deterministic functions)

- **The logistic trend**: This trend is usually S-shaped, i.e., moving from an initial phase into a dynamic phase in order to then enter into a new (stationary) equilibrium phase again in the end. It is consistent with the formula:

$$P_t = \frac{1}{\beta_0 + \beta_1 \bullet r^{time_t}} \qquad 0 < r < 1$$

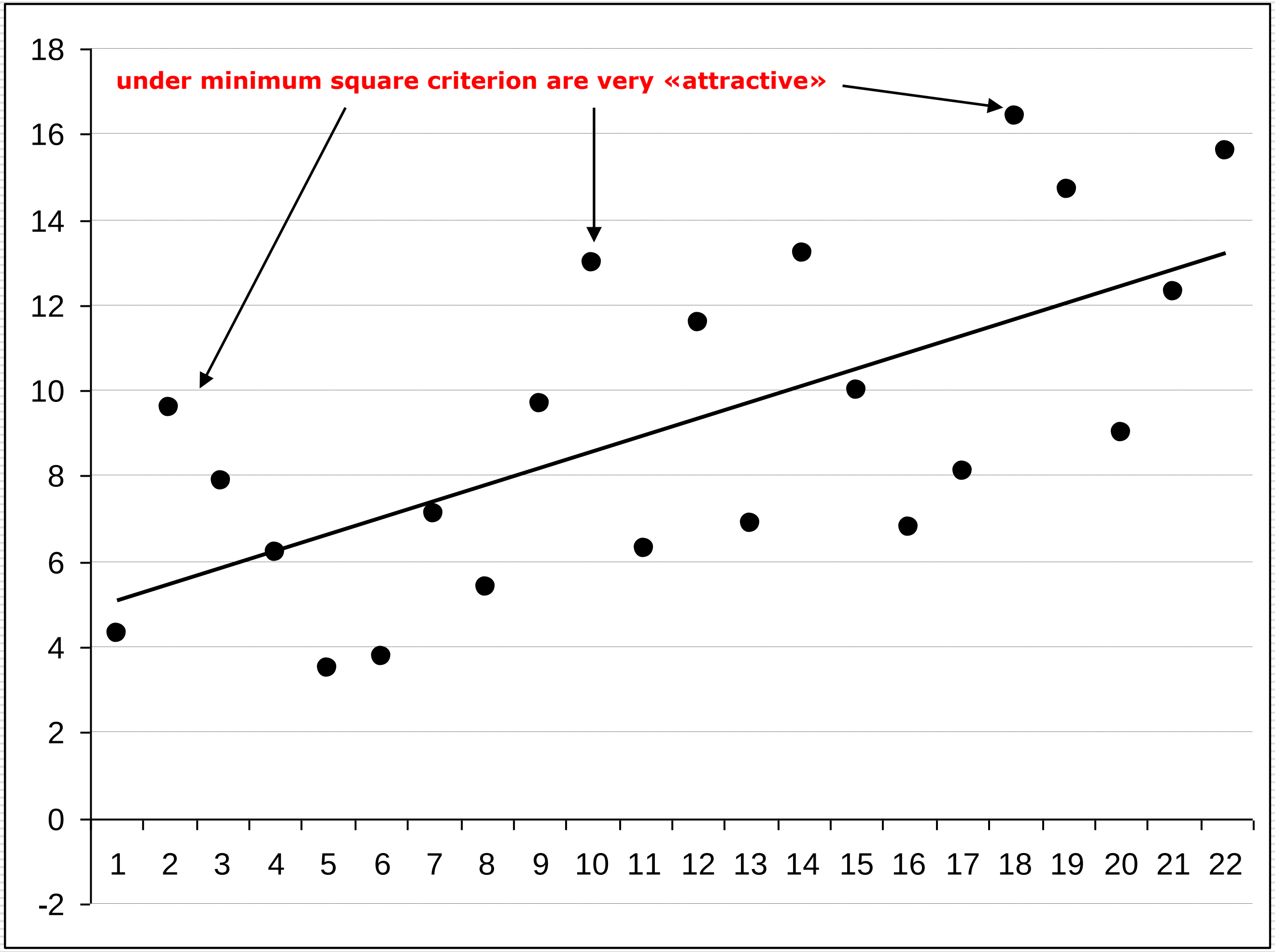
Deterministic functions
(analytical approach)

- General model: $$\mathbf{P}_{(T\times 1)} = \mathbf{X}_{(T\times k)}\boldsymbol{\beta}_{(k\times 1)} + \boldsymbol{\varepsilon}_{(T\times 1)}$$

- Hypotesis:
  1. Least Squares principle (loss function is quadratic)
  2. Residuales are Normally distributed, with zero mean and finite constant variance
  3. Residual are uncorrelated with the regressors
  4. The regressors *T* are error-free (non stochastic)
  5. The design matrix **T** must have full column rank *k*.

- Beta vector is BLUE $$\hat{\boldsymbol{\beta}} = \left(\mathbf{T}'\mathbf{T}\right)^{-1}\mathbf{T}'\mathbf{P}$$

Estimating (linear) deterministic trends



Modelling *(stocastic)* trend*

Unit Root in stocastic trends

- The random walk process (RW) can be represented as a sum of White Noise procesess

$$y_t = y_{t-1} + \varepsilon_t \quad \rightarrow \quad y_t = y_0 + \sum_{i=1}^{t} \varepsilon_{t-i} \qquad \varepsilon_t \sim WN\left(0, \sigma_\varepsilon^2\right)$$

Expected value and expected variance are:

$$E\left(y_t\right) = y_0 \quad ; \quad V\left(y_t\right) = E\left(y_0 + \sum_{i=1}^{t} \varepsilon_{t-i} - y_0\right)^2 = \sum_{i=1}^{t} \left(\varepsilon_{t-i}\right)^2 = t\sigma_\varepsilon^2 \xrightarrow[t\to\infty]{} \infty$$

So that the variance grows continously rather than converging to some **finite** unconditional variance.

- Same consideration for the RW with drift process: it is a non stationary process both in mean and variance.

$$y_t = \delta + y_{t-1} + \varepsilon_t \quad \rightarrow \quad y_t = y_0 + \delta \cdot t + \sum_{i=1}^{t} \varepsilon_{t-i} \qquad \varepsilon_t \sim WN\left(0, \sigma_\varepsilon^2\right)$$

$$E\left(y_t\right) = y_0 + \delta \cdot t \quad ; \quad V\left(y_t\right) = \sum_{i=1}^{t} \left(\varepsilon_{t-i}\right)^2 = t\sigma_\varepsilon^2 \xrightarrow[t\to\infty]{} \infty$$

* Some quick and incomplete references to ARMA modeling are listed in the appendix 1

Modelling *(stocastic)* trend

- More complex ARIMA(p,1,q) models behave like RW in certain key respect:
  - ARIMA(p,1,q) model are appropriately made stationary by differencing
  - Shocks to ARIMA(p,1,q) models have permanent effect even if – in contrast to RW – the long run effect of a unit shock may be greater ot less than unity depending on the parameter of the process.
  - The variance of a ARIMA(p,1,q) model grows without bound as time progress.

- The permanence of shocks means that optimal point forecast, even at long horizon, don't complitely revert to a mean or a trend.
- The fact that variance approaches to infinity as time progress imply that interval and density forecasts growth without bound in width and spread as the forecast horizon grows

EXAMPLE: $y_t = a_1 \cdot y_{t-1} + \varepsilon_t \qquad |a_1| = 1;\ \varepsilon_t \sim WN\left(0, \sigma_\varepsilon^2\right)$

The expected value h steps ahead (the optimal forecast in case of square loss) is:

$$\hat{y}_{t+h} = y_t \quad ; \quad if\ |a_1| < 1 \quad \hat{y}_{t+h} = a_1^h \cdot y_t$$

The variance of the forecast h steps ahead is:

$$V\left(\hat{y}_{t+h}\right) = (h-1) \cdot \sigma_\varepsilon^2 \qquad if\ |a_1| < 1 \quad V\left(\hat{y}_{t+h}\right) = \sigma_\varepsilon^2 \cdot \sum_{i=0}^{h-1} a_1^{2i}$$

Modelling *(stocastic)* trend

Unit roots estimation and testing

- Consider the following AR(1) process: (RW is used for illustration but the results carry over to general ARIMA(p,1,q) models)

$$y_t = a_1 \cdot y_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right)$$

- If the autoregressive coefficient is 1 (is a RW), OLS estimate are
  - **Superconsistent:** the difference between the LS estimate of the parameter and its true value 1, vanished quikly as the sample size grows respect to the case where $|a_1| < 1$ (it shrinks like 1/T while in the stationary case $\sqrt{T}\left(\hat{a}_1^{(LS)} - a_1\right)$ converge to a non degenerate random variable).
  - **Biased:** The LS estimator is biased downward. The larger the true value of the autoregressive parameter the larger the bias, so the bias is worst in the unit root case. Bias is larger if an intercept is included, and larger still if a trend is included. The bias vanished as sample size grows, as the estimate converge to the true population value.

- In testing unit root we can not use a t-statistics computed under the null of zero coefficient. **BUT** is possibile to coax standard software into printing t-statistics automatically for the null of autoregressive coefficient =1 simply rewriting the first order auto regression:

$$y_t - y_{t-1} = \left(a_1 - 1\right) \cdot y_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right)$$

Modelling *(stocastic)* trend

Unit roots estimation and testing

- Unfortunately in the unit root case the standard t-test don't follows the t-distribution. Instead it follows a non standard (Dikey-Fuller) distribution.
- Distribution changes for different alternative hypotesis.
  - Zero mean process
  - Non zero Mean
  - Linear trend

→ augmented Dickey Fuller statistics (**see Econometric courses**)

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \cdots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t,$$

- When the null hypotheses $\gamma = (a_1 - 1) = 0$ is rejected at certain significance level, a unit root exist.
- Imposing constraints on intercept and slope we obtain the three main versions of the test.
- By including lags of the order *p* the ADF formulation allows for higher-order autoregressive processes

Modelling Cycle

deterministic Cycle vs stocastic Cycle

$$\hat{y}_t = b_0 + b_1 \sin(t) \quad vs \quad \hat{y}_t = b_0 + b_1 \cdot y_{t-1}$$

- The cycle could be generated by a limited class of stocastic processes that are stationary, ergodic and asymptotically independent.

- **Modelling deterministic cycle**
- **Modelling stochastic cycle** (ARMA models)

Modelling *(stocastic)* Cycle

An example: the ARMA (1,1) model

$$y_t = c + a_1 \cdot y_{t-1} + b_1 \cdot \varepsilon_{t-1} + \varepsilon_t \qquad |a_1| < 1;\ \varepsilon_t \sim WN\left(0, \sigma_\varepsilon^2\right)$$

The expected value h steps ahead (the optimal forecast in case of square loss) is:

$$\hat{y}_{t+h} = c \cdot \left(\sum_{i=0}^{h-1} a_1^i\right) + a_1^h \cdot y_t + b_1 \cdot \varepsilon_t$$

The variance of the forecast h steps ahead is:

$$V\left(\hat{y}_{t+h}\right) = \sigma_\varepsilon^2 \cdot \left[\sum_{i=0}^{h-1} \left(\left(a_1 + b_1\right) \cdot a_1^{i-1}\right)^2\right]$$

Smoothing for trend estimation

(moving averages in finance)

The statistical approach to study (and use) moving averages in finance sees them as tools for the approximation of the underlying trend in the price series.
Known the trend, it will be possible to construct trading strategies by buying at the beginning of a period of upward trend and selling, after inversion, at the start of downward trend.

Assumption is that is possibile to break down the price of a financial asset in a trend component with a stable dynamic (predictable to some extent from the price serie) and in an unpredictable, irregular component (the error).

The error, measure everything that has to be considered "erratic‟, that is unpredictable given the available information (the price series).

$$P_t = T_t + \varepsilon_t$$

.

Smoothing for trend estimation

(moving averages in finance)

The moving average is a filter that reduces the erratic component (zero mean) allowing to approximate (display) the undergone trend

$$P_t = T_t + \varepsilon_t \qquad \varepsilon_t \sim i.i.d\left(0, \sigma_{\varepsilon_t}\right)$$

A moving average **will never be able to anticipate the trend inversions….** (as it is based on past price values).
A moving average is expected to reduce the error component but **there's no chance to completely eliminate it**.

Thus any average will only be able to

1. follow the trend - with a delay that is a **direct function** of average's amplitude
2. highlighting the trend with a precision that is an **inverse function** of the same amplitude.

Smoothing for trend estimation
(moving averages in finance)

The trade off between the ability to filter out the error component and the ability to react quickly to changes in trend direction, became evident by rewriting the moving average of width N:

$$P_t = T_t + \varepsilon_t = \frac{1}{N}\sum_{i=0}^{N-1} T_{t-i} + \frac{1}{N}\sum_{i=0}^{N-1} \varepsilon_{t-i}$$

Only for $N = 1$ the current trend component will coincide with the result of the current "average“, but – in that case - no reduction of the error is expected.
The higher is the amplitude the higer is the filtering power respect to the unpredictable (erratic) zero mean component.
However increasing N, the estimates of the current trend will be more and more underestimated (in bull market) or overestimated (in bear market).

Smoothing for trend estimation
(moving averages in finance)

The systematic underestimation / overestimation of the unobservable trend, is graphically displayed as a delay - of the averaged prices - in following the unobservable changes in the trend's slope.

This feature determines the jargon used by practitioners to defined moving averages.
When the amplitude is small the moving averages are called "**Fast**” that is they are able to “capture” a reversing trend with a short delay (i.e. fast averages follow the prices more closely).
Moving averages with medium or high amplitude are called “**Slow**" because they react “slowly” to trend’s reversals, even if they better filter erratic components.

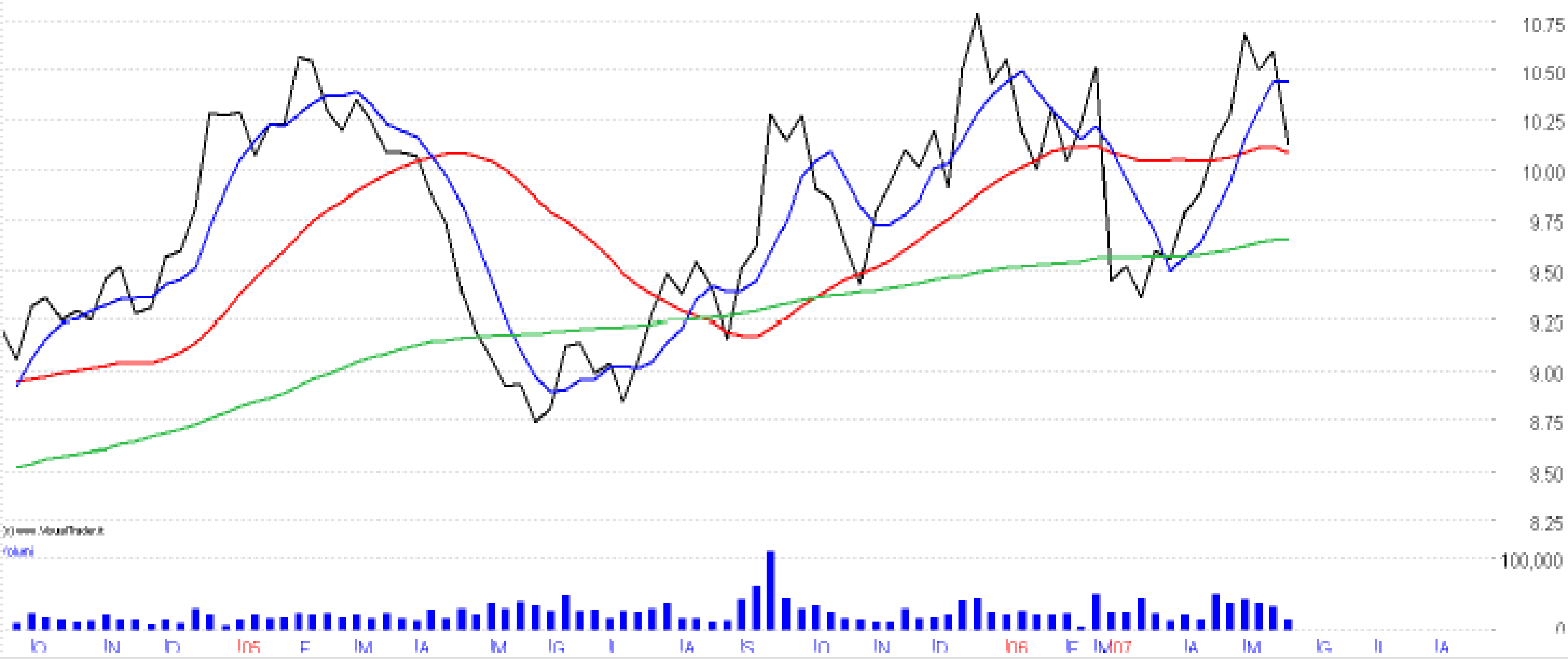
Charting moving averages results in a smoothed line that visually approximates the trend. That’s why moving averages are also defined "curvilinear trendline‟ by many practitioners

Smoothing for trend estimation

(moving averages in finance)

One side moving averages of weekly closing prices

Amplitudes: 100 (green); 21 (red); 5 (blue)



Smoothing for trend estimation

(weighted moving averages)

- Analysts also use other types of moving averages, assigning different weight to each observation. The ratio is that the value of the first averaged observation (maybe even far away in time) is not considered as "informative" as the value of the more recent observation

$$MA(N)_t = \frac{\sum_{n=0}^{N-1} w_n \cdot P_{t-n}}{\sum_{n=0}^{N-1} w_n}$$

(were *Wn are the weights)*

- Among the different structures of weight that is possible to assign the most frequently used is linear.
  Thus, for example, when a length of 10 is taken, the first observation is multiplied by 1, the following by 2, and so on up to give a weight of 10 to the tenth (the last) observation. The total is then divided by the sum of the weights, in this case 1+2 +3 +...+ 10 = 55.
- With this scheme the last observation weighs the 18% while the most distant one only 1.8%.

Smoothing for trend estimation
(weighted moving averages)

- Is thus evident, that a **weighted average is more effective in "following" the last observation** even when their dynamics are in contrast with the series "historical" dynamics.
  In contrast, in the case of a simple average, the slope of the moving average will take several periods (observations) before reporting that a trend has been reversed, "ballast" as is, from past observations.
- Anyway, it should be noted that the **weighted averages are expected to be much less effective in filtering the error component** (which by definition has zero mean).

- Other weighted moving averages are the exponential, the geometric, the triangular…. There are even methods that will vary the moving average length based on the historic volatility of the prices known as "adaptive" moving averages. The simple moving average is easier to construct and suffices for our purposes
- We only introduce the exponential moving average to show that is a particular case of ARIMA model.

Smoothing for trend estimation
(exponential moving averages)

- The **exponential moving average** is a particular weighted average where observations are weighted with coefficients increasing exponentially as a function of time.

  The average can takes into account the complete "history" of a price series because as the time interval grows, the weight of the initial data is reduced to become infinitesimal.
- It could be “faster” than others moving averages formulae in capturing changes in unobservable trend slope.

- The basic framework, in which we derive the exponential moving averages, is the RW with measurement errors ε:

$$c_{0,t} = T_t + \eta_t \quad \rightarrow \quad c_{0,t} = c_{0,t-1} + \eta_t \quad ; \quad \eta_t \sim WN(0, \sigma_\eta^2)$$

  Supposing that we can only observe a variable $y_t = c_{0,t} + \varepsilon_t$

  where ε_t and η_t are uncorrelated at all leads and lags.

Smoothing for trend estimation
(exponential moving averages)

- A smoother for $c_{0,t}$ defined $Sm_t(N)$ could be the one side aritmetic mean of its *N* (*unknown*) previous values. A **feasible** smoother at time *t* have to considers the observable *yt* and the "measurement error":

$$Sm_t(N) = \frac{1}{N}\left(\sum_{n=0}^{N-1} y_{t-n} + \sum_{n=0}^{N-1} \varepsilon_{t-n}\right)$$

  (where the residual sum goes to zero as *N* increase)

- At time *t+1* the feasible smoother is the aritmetic mean of *N+1* observable *yt* that could be expressed with an iterative formula as:

$$Sm_{t+1}(N+1) = \frac{1}{N+1}\left(\sum_{n=0}^{N} y_{t+1-n} + \sum_{n=1}^{N} \varepsilon_{t+1-n}\right) = \frac{1}{N+1}\left(\sum_{n=1}^{N} y_{t+1-n} + \sum_{n=1}^{N} \varepsilon_{t+1-n}\right) + \frac{1}{N+1}\left(y_{t+1} + \varepsilon_{t+1}\right)$$

$$Sm_{t+1}(N+1) = \frac{N}{N+1} Sm_t(N) + \left(1 - \frac{N}{N+1}\right)\left(y_{t+1} + \varepsilon_{t+1}\right)$$

- Is possible to generalize the smoother letting $\frac{N}{N+1}$ to be a parameter λ varying from 0 to 1.

Smoothing for trend estimation

(exponential moving averages)

- The smoothing algorithm could be stated as an iterative procedure:
  1. Inizialize at t=N so that the starting values of the recursion are obtained by simple arithmetic average of past prices.

$$Sm_t(N) = \frac{1}{N}\sum_{n=0}^{N-1} y_{t-n}$$

  2. *Update the previous smothed value* by a constrained linear combination of the current y value and the smoother calculated in the previous iteration.

$$Sm_t = (1-\lambda)\cdot y_t + \lambda \cdot Sm_{t-1} \qquad [\text{in some book} : \lambda = (1-\alpha)]$$

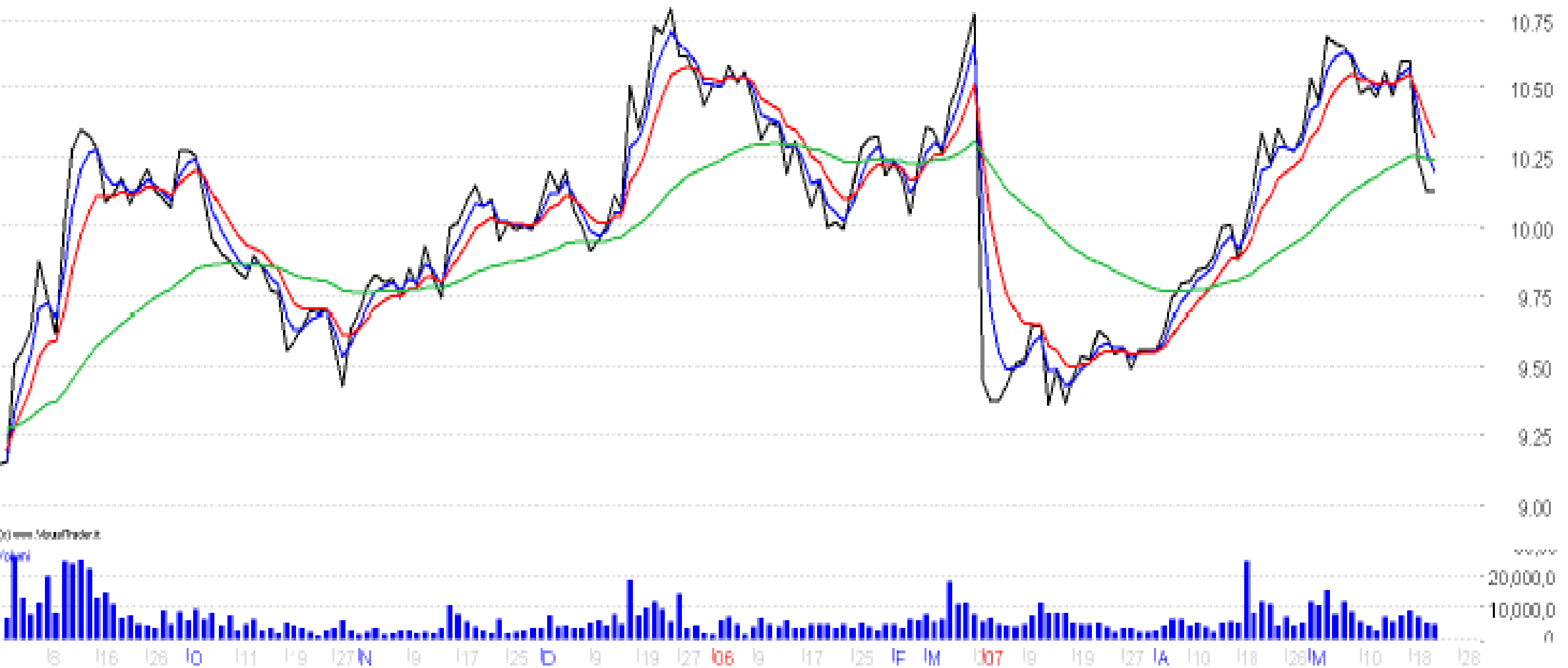
(Note: if lambda is equal to 1 (alpha =0), the filter coincides with the simple average
if lambda tends to 0 (alpha → 1) the filter tends to the RW)

When the DGP is descripted as a RW with measurement errors this exponential moving average represent the optimal estimate of *Co* and hence the optimal forecast for any future values of *y*.

Smoothing for trend estimation
(exponential moving averages)

Exponential moving averages of weekly closing prices:

Alpha=0.5; Alpha=0.25; Alpha=0.05



Smoothing for trend estimation
(exponential smoothing)

Previous smoother:

$$Sm_t = (1-\lambda)\cdot y_t + \lambda \cdot Sm_{t-1} \qquad [\alpha = (1-\lambda)]$$

That is a one side moving average with exponentially declining weights as is clear substituting backward

$$ES_t = (1-\lambda)y_t + \lambda \cdot ES_{t-1}$$
$$ES_t = (1-\lambda)y_t + \lambda \cdot [(1-\lambda)y_{t-1} + \lambda \cdot ES_{t-2}]$$
$$ES_t = (1-\lambda)y_t + \lambda \cdot \{(1-\lambda)y_{t-1} + \lambda \cdot [(1-\lambda)y_{t-2} + \lambda \cdot ES_{t-3}]\}\ldots$$
$$ES_t = (1-\lambda)y_t + (1-\lambda)\lambda y_{t-1} + (1-\lambda)\lambda^2 y_{t-2} + \ldots$$

$$ES_t = (1-\lambda)\sum_{i=0}^{\infty}\lambda^i y_{t-i} \qquad = \qquad \frac{\sum_{i=0}^{\infty}\lambda^i y_{t-i}}{\sum_{i=0}^{\infty}\lambda^i}$$

Smoothing for trend estimation

(exponential smoothing)

- For $y_t \equiv \text{Price at time } t$ $\quad for\ \lambda \in]0,1]_t \qquad (1-\lambda) = \left[\sum_{i=0}^{\infty} \lambda^i\right]^{-1}$

$$ES_t \cong \frac{\sum_{i=0}^{t} \lambda^i P_{t-i}}{\sum_{i=0}^{t} \lambda^i}$$

- The **exponential smoothing** approach is a particular weighted average where observations are weighted with coefficients declining exponentially as a function of time.

- ES can be seen as a generalization of the arithmetic mean, since the latter is obtained when the smoothing parameter lambda is equal to 1.
- Decrease the lambda parameter, increases the focus on recent observations and thus the ability to "follow" the unobserved trend loosing at the same time the ability to filter the erratic component.
- For lamba → 0, ES coincides with the value of the price at time t.

A generalization toward forecasting

(Exponential Smoothing)

- In order to make statistical inference on the estimated trend extracted (estimated) by ES, is usefull to show ES as an ARIMA(0.1.1) DGP.

- Calculation of smoother's expected value and variance allow to build confidence intervals for the price fluctuation and thus derive a method that - under the usual parametric statistics' assumptions - helps in discriminating between erratic fluctuations (false signals) and trend movements (operating signals).
  The optimal forecast and its variability are thus two basic steps allowing to design (probabilistic) trading strategies, whatever modelling approach you decide to follow.

- Under the assumption that the price (P_t) is described as a sum of exponential type signal **+** an error component type White Noise, for t very large (∞ limit), the unknown P_{t+1} will be:

$$P_{t+1} = (1-\lambda)\sum_{i=0}^{t} \lambda^{i} P_{t-i} + \varepsilon_{t+1}$$

A generalization toward forecasting
(Exponential Smoothing)

- Premultiplicating the whole expression by *(1 - λL)* with *L* the lag operator such that *L(Pt)=L(Pt-1)* we have:

$$(1-\lambda L)(P_{t+1}) = (1-\lambda)(1-\lambda L)\sum_{i=0}^{t}\lambda^{i}P_{t-i} + (1-\lambda L)\varepsilon_{t+1}$$

$$P_{t+1} - \lambda P_t = (1-\lambda)(\lambda^{0}P_t - \lambda^{t}P_0) + \varepsilon_{t+1} - \lambda\varepsilon_t$$

that given *λ ranges between 0 and 1* simplifies (for t → ∞) in the following ARIMA(0,1,1) process:

$$P_{t+1} = P_t - \lambda\varepsilon_t + \varepsilon_{t+1}$$

- The value *λ* is the coefficient of the MA(1) DGP. Its determination is possible by applying OLS to prices first differences (better in log).

$$\Delta P_{t+1} = -\lambda\varepsilon_t + \varepsilon_{t+1}$$

A generalization toward forecasting
(Exponential Smoothing)

- Note that in this last formulation,

$$P_{t+1} = P_t - \lambda\varepsilon_t + \varepsilon_{t+1}$$

  the exponential smoothing curve is determined by adding to the former price a share of the past forecast error. At time *T+h* it will be obtained by adding to the forecast at time *T+h-1* the corresponding forecasting error.

  If λ tends to 0, then the forecasting error is negligible and the determination of future values will depend solely on the present price (assumption of Random Walk and efficient markets).

  Conversely, if λ tends to 1, the ES forecaster coincides with the arithmetic mean, and the whole price history is considered.

- **In any case, prior to obtain the best punctual predictor, a loss function must be defined.**

A generalization toward forecasting
(Exponential Smoothing)

- Assuming that the cost function is of quadratic type, the “optimal predictor" for the price is the conditional expectation:

$$E(P_{T+1}) = E(P_T + \varepsilon_{T+1} - \lambda\varepsilon_T) = P_T - \lambda\varepsilon_T$$

As in the ARIMA (0,1,1) procesess, the optimal predictor variance is (for h greater than 0):

$$Var(P_{T+h}) = \sigma_\varepsilon^2 \left[1 + (h-1)(\lambda+1)^2 \right]$$

with σ_ε^2 the variance of the error component prices.

The knowledge of the expected value and variance permits (under usual assumptions on error term distribution) to build confidence intervals. Chosen a first type error level (ɑ), the corrispondig interval is build by "bands" within which price movements do not constitute trading signals at ɑ confidence level (they could be “false signals”).

Smoothing for trend estimation
(Holt-Winters "double exponential smoothing")

- The **Holt-Winters Smooting** algorithm considers that the observed time series is generated by a DGP that has both an «evolving» local level and «evolving» slope:

$$y_t = c_{0,t} + c_{1,t} \cdot t + \varepsilon_t \quad \text{where} \quad c_{0,t} = c_{0,t-1} + \eta_t \quad \text{and} \quad c_{1,t} = c_{1,t-1} + \nu_t$$

$$\eta_t \sim WN(0, \sigma_\eta^2) \ , \ \nu_t \sim WN(0, \sigma_\nu^2)$$

Where all the disturbance are orthogonal at all leads and lags

- The previous exponential moving averages recursive formula is “augmented” by a slope parameter that increase exponentially as a function of time.

Smoothing for trend estimation
(Holt-Winters "double exponential smoothing")

- The smoothing algorithm could be stated as an iterative procedure:
  1. Inizialize at *t=2*

$$HW_2 = y_2 \quad ; \quad F_2 = y_2 - y_1$$

  2. *Update the previous "smothed" values* by two constrained linear combination of local level and local slope calculated in the previous iteration.

$$HW_t = \alpha \cdot y_t + (1-\alpha) \cdot (HW_{t-1} - F_{t-1}) \quad \alpha \in]0,1[$$
$$F_t = \beta \cdot (HW_t - HW_{t-1}) + (1-\beta) \cdot F_{t-1} \quad \beta \in]0,1[$$

- The weight of the combination could be estimated by OLS.

- Although we have displayed the DGP below HW modelling, when we apply HW we generally don't assume that the data are actually generated by that process. We hope however that the actual DGP is close to the one for which HW is optimal.

What common forecasting models can be seen as special cases of ARIMA models?

Mean = ARIMA(0,0,0) with constant

RW = ARIMA(0,1,0)

RW with Drift = ARIMA(0,1,0) with deterministic trend (even constant)

Simple Exponential Smoothing = ARIMA(0,1,1)

What common forecasting models can be seen as special cases of ARIMA models?

	Expected value	Variance
Mean = ARIMA(0,0,0) with constant	$\hat{y}_{t+h} = c$	$V(\hat{y}_{t+h}) = \sigma_\varepsilon^2$
RW = ARIMA(0,1,0)	$\hat{y}_{t+h} = y_t$	$V(\hat{y}_{t+h}) = \sigma_\varepsilon^2 \cdot h$
RW with Drift = ARIMA(0,1,0) with constant	$\hat{y}_{t+h} = c \cdot h + y_t$	$V(\hat{y}_{t+h}) = \sigma_\varepsilon^2 \cdot h$
Simple Exponential Smoothing = ARIMA(0,1,1)	$\hat{y}_{t+h} = c \cdot h + y_t + b_1 \cdot \varepsilon_t$	$V(\hat{y}_{t+h}) = \sigma_\varepsilon^2 \cdot \left[1 + (h-1)(1+b_1)^2\right]$

Advanced topics

Volatility smoothing

An exponential smoother for volatility could be based on squared differences between prices and their expected value (the average).

$$Sm_t = (1-\lambda)\cdot y_t + \lambda \cdot Sm_{t-1} \qquad [\alpha = (1-\lambda)]$$

$$S_t^2(N) = (1-\lambda)\cdot (P_t - \overline{P}_t)^2 + \lambda \cdot S_{t-1}^2(N)$$

Lambda indicates the smoothing constant. The more it approaches to one, the more the volatility is persistent (it is not given any weight to the last observation).

The lambda value can be calibrated or estimated exactly as with the standard exponential smoothing setting. The same considerations apply.

Advanced topics
VaR in the RiskMetrics methodology

The standard exponential smoother is also known as the RiskMetrics methodology (J.P. Morgan (early 1990s), for the purposes of daily measuring and explaining risks)

RiskMetrics uses the Exponential Weighted MA (EWMA) of historical observations in order to capture the dynamic features of volatility.
Since the recent observations could weigh more than past observations, volatility reacts quicker to shocks respect to the equally weighted case also "recovering quickly" after a shock.

Based on a conditional Gaussian distribution, RiskMetrics uses the following updating mechanism for 1-day-ahead conditional variance

$$\sigma^2_{t+1|t} = (1-\lambda)\sum_{i=0}^{\infty} \lambda^i \cdot R^2_{t-i} \rightarrow (1-\lambda)R^2_t + \lambda\sigma^2_{t|t-1}$$

exponential average

with R_t, represent a time series of financial returns (t = 1, ..., T), with a time-varying conditional distribution.

$$p(R_t \mid I_{t-1};\theta) \sim N(0;\sigma_t)$$

Advanced topics

fat-tailed squared returns

$$Sm_t = (1-\lambda)\cdot y_t + \lambda \cdot Sm_{t-1} \qquad [\alpha = (1-\lambda)]$$

Previous EWMA model corresponds to a zero-intercept IGARCH(1,1) model

$$\sigma_{t+1}^2 = 0 + (1-\lambda)R_t^2 + \lambda\sigma_t^2 =$$
$$= (1-\lambda)R_t^2 + \left[-(1-\lambda)\sigma_t^2 + (1-\lambda)\sigma_t^2\right] + \lambda\sigma_t^2 = (1-\lambda)\cdot\left(R_t^2 - \sigma_t^2\right) + (1-\lambda+\lambda)\cdot\sigma_t^2$$

This model is often symplified (considering that variance is generally not observable and estimating it with squared returns). That way we assume that variance follows a random walk without a drift. The h-period forecasts take the form:

$$\sigma_{t+h}^2 = \left(\lambda + 1 - \lambda\right)\cdot\sigma_{t+h-1}^2 = \sigma_{t+h-1}^2 = \sigma_t^2$$

So that tomorrow’s variance is a function of today’s variance (recursively).

In Lucas et.al (2014) is pointed out that the use of squared return is “not optimal” when squared returns are conditionally fat-tailed, which is due to regular occurrence of large realizations, even though the variance has not changed substantially.
This may lead to highly biased volatility estimates, which translates into a persisting bias in the VaR due to the long memory properties of the IGARCH. The solution suggested by the authors is an EWMA scheme in which the shape of the conditional distribution is accounted for within Generalized Autoregressive Score (GAS) framework.

Lucas, A., Zhang, X. (2014) “Score Driven Exponentially Weighted Moving Averages and Value-at-Risk Forecasting”, Tinbergen Institute Discussion Paper, 2014,

Advanced topics

Score Driven Exponential smoothing (Generalized Autoregressive Score (GAS) framework).

Observation driven model → 0 intercept IGARCH(1,1) $\sigma_{t+1}^2 = (1-\lambda)R_t^2 + \lambda\sigma_t^2$

Parameters are functions of lagged dependent variables as well as current and lagged independent variables. Despite the stochastic nature of parameters, they are thought as perfectly predictable given the past information

Parameter driven model → 0 intercept IGAS(1,1) $f_{t+1} = A_1 s_t + f_t$

Parameters are stochastic processes with their own error source Creal et al. [2013] show that reduces precisely to the standard ES scheme setting $A_1 = 1 - \lambda$ and $f_t = \sigma_t$

f_t is time varying parameter (i.e $f_t = \sigma_t$; s_t is a function of the past data (financial returns). Note: usually

$$R_t = \sigma_t \varepsilon_t \quad \varepsilon_t \sim N(0;1)$$

If the conditional distribution is Student's *t*, the score of the *t*-distribution causes the volatility dynamics not to react too fiercely to large values of the time-varying parameter.

$$\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\,\Gamma(\frac{n}{2})}\left(1+\frac{t^2}{n}\right)^{-(\frac{n+1}{2})}$$

Advanced topics

Score Driven Exponential smoothing

Instead of use squared lagged observations $\sigma^2_{t+1|t} = (1-\lambda)\sum_{i=0}^{\infty} \lambda^i \cdot R^2_{t-i} \to (1-\lambda)R^2_t + \lambda\sigma^2_{t|t-1}$ is possible to update parameters using the score of the forecasting distribution. This allows the parameter dynamics to adapt automatically to any non-normal data features and robustifies the subsequent volatility estimates.

Lucas et.al (2014) shows that considering a Student's t distributions and a time varying degrees of freedom parameter the Score Driven Exponential smoothing techniques is expected to improve the coincidence between the number of violations the technique produces, and the number of time the losses have exceeded the level indicated or “promised” by the confidence level.

For any degrees of fredom >2:

$$\sigma_{t+1} = \sigma_t + A_1 s_t \quad \lambda = A_1(1+3\nu^{-1})$$

$$\sigma^2_{t+1} = \sigma^2_t + A(1+3\nu^{-1})\left(\frac{\nu+1}{\nu-2+y_t^2/f_t}y_t^2 - f_t\right)$$

$$= (1-\lambda)\sigma^2_t + \lambda\frac{\nu+1}{\nu-2+y_t^2/f_t}y_t^2$$

Appendix 1
Some quick and incomplete references to ARMA model

A **stochastic** or **random process** is a mathematical object usually defined as a collection of random variables.

Historically, the random variables were associated with or indexed by a set of numbers, usually viewed as points in time, giving the interpretation of a stochastic process representing numerical values of some system randomly changing over time,

- Is a random function of the time t and the event ϖ

$\{Y(\omega,t):\ \omega \in \Omega,\ t \in T\}$ (is a function from $(\Omega \times T)$ to R)

If: $t \in T\ e\ \omega \in \Omega$, $\{Y(\omega,t)\}$ is a stocastic process

$t = t_0\ e\ \omega \in \Omega$, $\{Y(\omega,t_0)\}$ is a random variable

$t = t_0\ e\ \omega = \omega_0$, $\{Y(\omega_0,t_0)\}$ is a real number

A time series $\{y_t;\ t = 1,\dots,T\}$ is a finite realization of a stochastic process, or as a set (a sequence) of sample values relating to different random variables $Y(\omega, t = t_0)$.

A SP is "known" when all the k-dimensional cumulative distribution function are known:

$$F(y_1) = P(Y_t \le y_1)\ (\forall y_1 \text{ e } \forall t \in T);$$

$$F(y_1, y_2) = P(Y_t \le y_1, Y_s \le y_2)\ (\forall y_1, y_2 \text{ e } \forall t,s \in T);$$

$$\dots$$

$$F(y_1, y_2, \dots, y_k) = P(Y_{t1} \le y_1, Y_{t2} \le y_2, \dots, Y_{tk} \le y_k)$$

- That is never!

Appendix 1
Some quick and incomplete references to ARMA model

It is therefore necessary:

1. synthesize distributions through position and variability moments. assuming that first and second order moment are sufficient to describe the distribution of the distribution of the random variables;

2. Restrict the set of possible process constraining the memory to k finite lag assuming that the process is homogeneous, in probabilistic terms, over time (stationarity).

Moments:

Expected value $E(Y_t) = \mu_t \qquad t \in T$

Variance $V(Y_t) = E\left[(Y_t - \mu_t)^2\right] = \sigma_t^2 \qquad t \in T$

Autocovariance $Cov(Y_t, Y_{t-k}) = E\left[(Y_t - \mu_t)(Y_{t-k} - \mu_{t-k})\right] = \gamma_k(t) \qquad t \in T \quad k = 0, \pm 1, \pm 2, \ldots$

Autocorrelation $Cor(Y_t, Y_{t-k}) = \dfrac{Cov(Y_t, Y_{t-k})}{\sqrt{V(Y_t)V(Y_{t-k})}} = \dfrac{\gamma_k(t)}{\sqrt{\sigma_t^2 \sigma_{t-k}^2}} \qquad t \in T \quad k = 0, \pm 1, \pm 2, \ldots$

Appendix 1
Some quick and incomplete references to ARMA model

Restrictions properties:

Stationarity (in the mean) $E(Y_t) = \mu < \infty \qquad t \in T$

Stationarity (in the variance) $V(Y_t) = E\left[(Y_t - \mu)^2\right] = \sigma^2 < \infty \qquad t \in T$

Then the process is also covariance and correlation stationary

$$Cov(Y_t, Y_{t-k}) = E\left[(Y_t - \mu)(Y_{t-k} - \mu)\right] = \gamma_k \qquad t \in T \quad k = 0, \pm 1, \pm 2, \dots$$

$$Cor(Y_t, Y_{t-k}) = \frac{Cov(Y_t, Y_{t-k})}{\sqrt{V(Y_t)V(Y_{t-k})}} = \frac{\gamma_k}{\sigma^2} = \frac{\gamma_k}{\gamma_0} \qquad t \in T \quad k = 0, \pm 1, \pm 2, \dots$$

Autocovariance function $\gamma : I \to R$

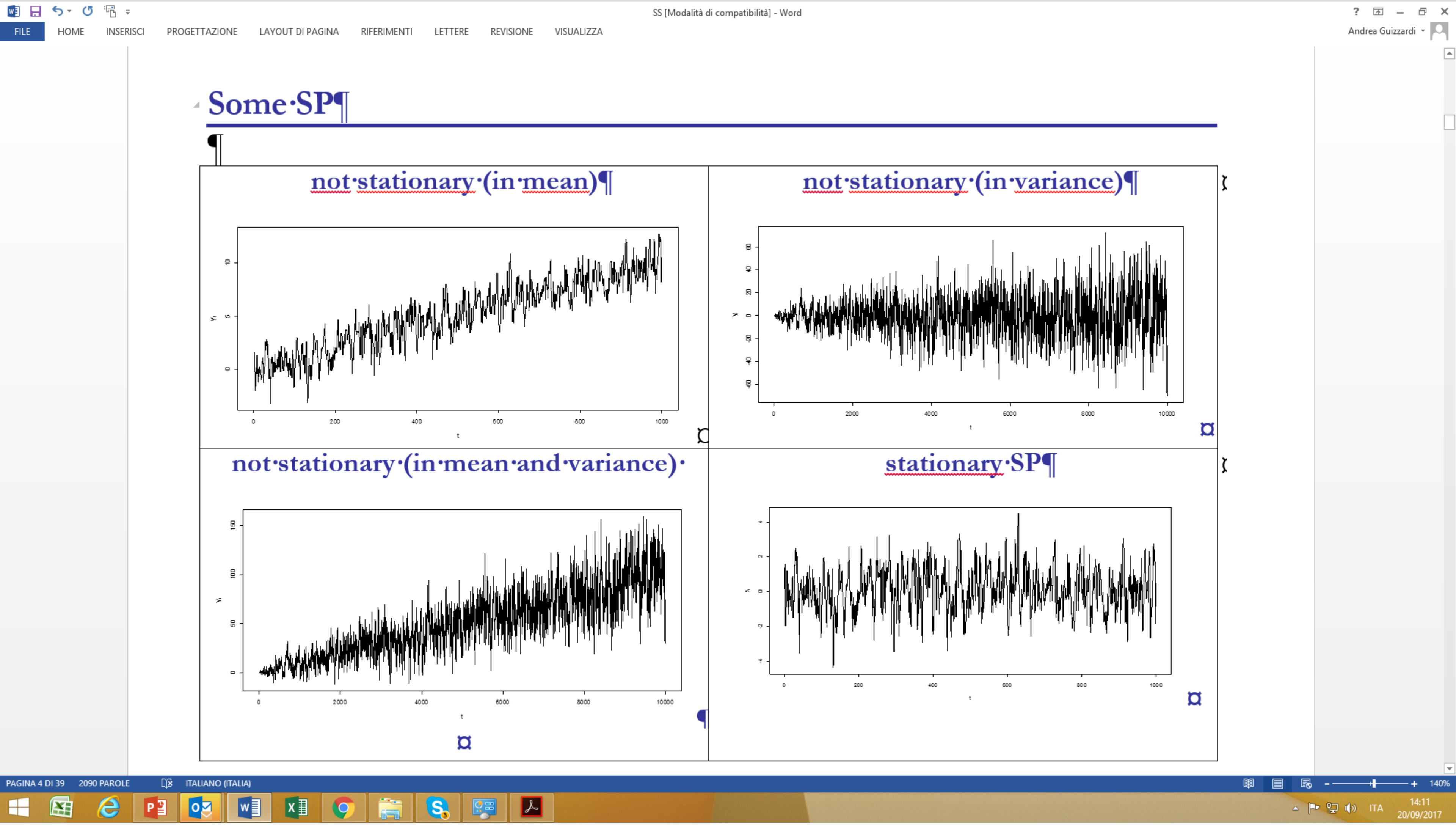
$$\gamma_k = Cov(Y_t, Y_{t-k}) \qquad \gamma_0 = Cov(Y_t, Y_t) = Var(Y_t) = \sigma^2$$

$$\gamma_k = \gamma_{-k}$$

Autocorrelation function (ACF) $\rho : I \to [-1, 1]$

$$\rho_k = \frac{Cov(Y_t, Y_{t-k})}{\sqrt{V(Y_t)V(Y_{t-k})}} = \frac{\gamma_k}{\gamma_0} \dots \qquad \rho_k = \rho_{-k} \qquad |\rho_k| \le 1$$

Appendix 1
Some quick and incomplete references to ARMA model



Appendix 1

Some quick and incomplete references to ARMA model

The White Noise (stationary process): $\{\varepsilon_t\}_{t\in T} \sim WN(0,\sigma_\varepsilon^2)$.

$$E(\varepsilon_t)=0; \quad V(\varepsilon_t)=\sigma_\varepsilon^2; \quad Cov(\varepsilon_t,\varepsilon_{t-k})=\gamma_k=0 \qquad \forall k\neq 0$$

A SP **is invertible** if it can be expressed as a function of the past values an a random component

$$Y_t = f(Y_{t-1},Y_{t-2},\ldots,\varepsilon_t) \qquad \{\varepsilon_t\}_{t\in T} \sim WN(0,\sigma_\varepsilon^2)$$

is a property used in process identification, very useful in forecasting.

A SP **is ergodic** if its statistical properties can be deduced from a single, sufficiently long, random sample of the process (i.e. limited memory: $\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} Cov(Y_t,Y_{t-k})=0$)

Thus: $\bar{y}=\frac{1}{T}\sum_{t=1}^{T} y_t$; $\hat{\sigma}^2=\frac{1}{T}\sum_{t=1}^{T}(y_t-\bar{y})^2$; $\hat{\gamma}_k=\frac{1}{T}\sum_{t=1}^{T}(y_t-\bar{y})(y_{t-k}-\bar{y})$ are consistent estimators for μ, σ^2 e γ_k.

Also imposing linearity in the time-dependence structure → ARMA process

Appendix 1
Some quick and incomplete references to ARMA model

Moving average process (zero mean: $E(Y_t)=0$)

The output variable depends linearly on the current and various past values of a stochastic term

$$Y_t=\varepsilon_t-\theta_1\varepsilon_{t-1}-\theta_2\varepsilon_{t-2}-\ldots-\theta_q\varepsilon_{t-q} \qquad \{\varepsilon_t\}_{t\in T}\sim WN\left(0,\sigma_\varepsilon^2\right)$$

In compact form $\quad Y_t=B^0\varepsilon_t-\theta_1B^1\varepsilon_t-\theta_2B^2\varepsilon_t-\ldots-\theta_qB^q\varepsilon_t=\theta_q(B)\varepsilon_t$

with $\theta_q(B)=\left(1-\theta_1B-\theta_2B^2-\ldots-\theta_qB^q\right)$

Autoregressive process (zero mean: $E(Y_t)=0$)

The output variable depends linearly on its own previous values and on a stochastic term

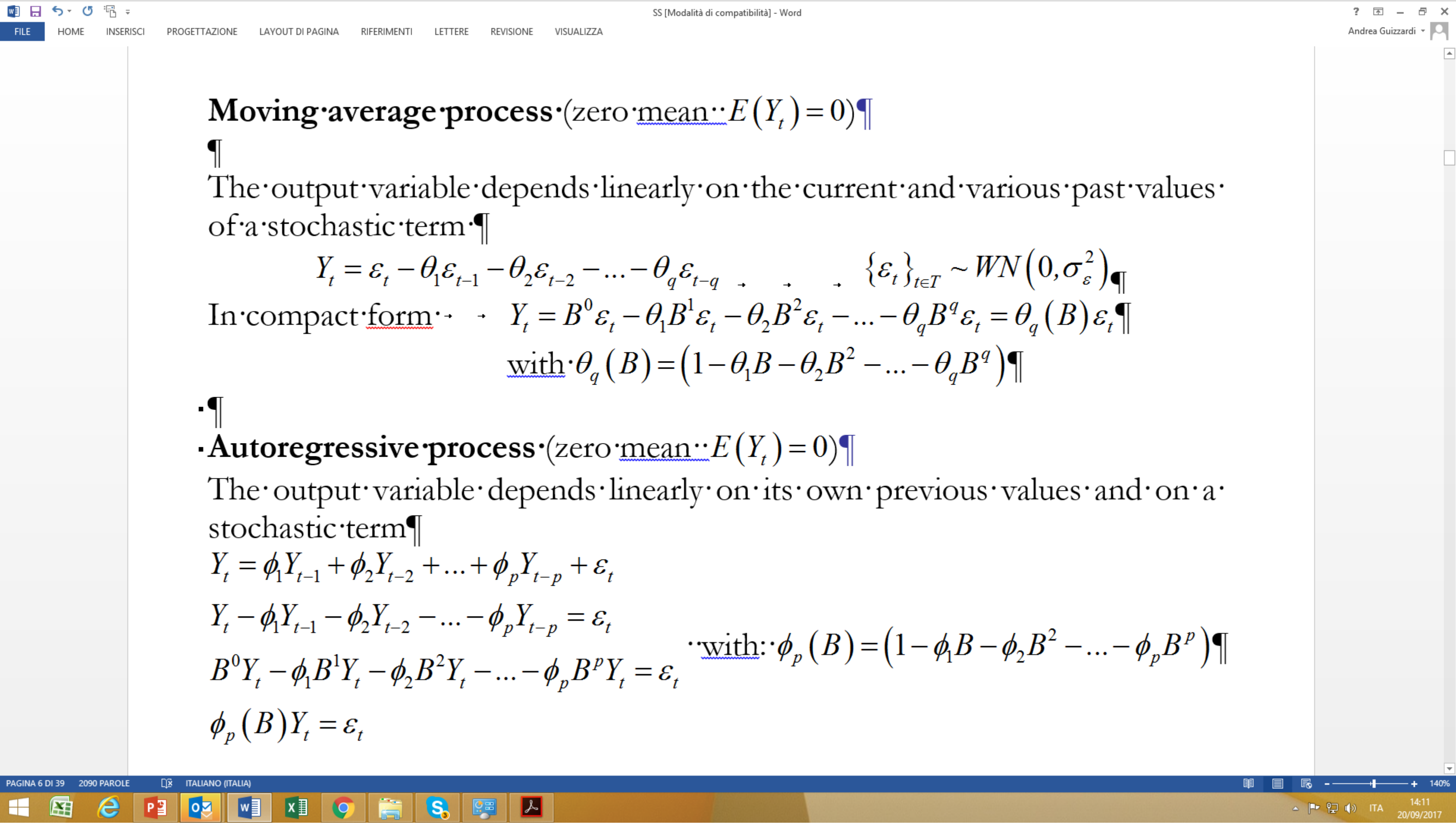
$$Y_t=\phi_1Y_{t-1}+\phi_2Y_{t-2}+\ldots+\phi_pY_{t-p}+\varepsilon_t$$

$$Y_t-\phi_1Y_{t-1}-\phi_2Y_{t-2}-\ldots-\phi_pY_{t-p}=\varepsilon_t$$

$$B^0Y_t-\phi_1B^1Y_t-\phi_2B^2Y_t-\ldots-\phi_pB^pY_t=\varepsilon_t \qquad \text{with: } \phi_p(B)=\left(1-\phi_1B-\phi_2B^2-\ldots-\phi_pB^p\right)$$

$$\phi_p(B)Y_t=\varepsilon_t$$

Some quick and incomplete references to ARMA model



Moving average process (zero mean: $E(Y_t)=0$)

The output variable depends linearly on the current and various past values of a stochastic term

$$Y_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \dots - \theta_q\varepsilon_{t-q} \qquad \{\varepsilon_t\}_{t\in T} \sim WN\left(0,\sigma_\varepsilon^2\right)$$

In compact form: $Y_t = B^0\varepsilon_t - \theta_1 B^1\varepsilon_t - \theta_2 B^2\varepsilon_t - \dots - \theta_q B^q\varepsilon_t = \theta_q(B)\varepsilon_t$

with $\theta_q(B) = \left(1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q\right)$

Autoregressive process (zero mean: $E(Y_t)=0$)

The output variable depends linearly on its own previous values and on a stochastic term

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

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

$$B^0 Y_t - \phi_1 B^1 Y_t - \phi_2 B^2 Y_t - \dots - \phi_p B^p Y_t = \varepsilon_t \qquad \text{with: } \phi_p(B) = \left(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p\right)$$

$$\phi_p(B) Y_t = \varepsilon_t$$

Appendix 1
Some quick and incomplete references to ARMA model

MA(1) $Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$ $\quad Y_t = (1-\theta_1 B)\varepsilon_t$ $\quad \{\varepsilon_t\}_{t\in T} \sim WN(0, \sigma_\varepsilon^2)$

Expected value: $E(Y_t) = 0$

Variance $\gamma_0 = V(Y_t) = E(Y_t^2) - [E(Y_t)]^2 = E(Y_t^2) - 0$

$$E(Y_t^2) = E\left[(\varepsilon_t - \theta_1\varepsilon_{t-1})^2\right] = E(\varepsilon_t^2 + \theta_1^2\varepsilon_{t-1}^2 - 2\theta_1\varepsilon_t\varepsilon_{t-1}) =$$
$$= E(\varepsilon_t^2) + E(\theta_1^2\varepsilon_{t-1}^2) - E(2\theta_1\varepsilon_t\varepsilon_{t-1}) = \sigma_\varepsilon^2 + \sigma_\varepsilon^2\theta_1^2 + 0 =$$
$$= (1+\theta_1^2)\sigma_\varepsilon^2$$

Autocovariance:

$$\gamma_1 = Cov(Y_t, Y_{t-1}) = E(Y_t Y_{t-1}) - E(Y_t)E(Y_{t-1}) = E(Y_t Y_{t-1}) =$$

lag1 $= E[(\varepsilon_t - \theta_1\varepsilon_{t-1})(\varepsilon_{t-1} - \theta_1\varepsilon_{t-2})] = E\left[\varepsilon_t\varepsilon_{t-1} - \theta_1\varepsilon_{t-1}^2 - \theta_1\varepsilon_t\varepsilon_{t-2} + \theta_1^2\varepsilon_{t-1}\varepsilon_{t-2}\right] =$

$$= 0 - \theta_1\sigma_\varepsilon^2 - 0 + 0 = -\theta_1\sigma_\varepsilon^2$$

$$\gamma_2 = Cov(Y_t, Y_{t-2}) = E(Y_t Y_{t-2}) - E(Y_t)E(Y_{t-2}) = E(Y_t Y_{t-2}) =$$

lag2 $= E[(\varepsilon_t - \theta_1\varepsilon_{t-1})(\varepsilon_{t-2} - \theta_1\varepsilon_{t-3})] = E\left[\varepsilon_t\varepsilon_{t-2} - \theta_1\varepsilon_{t-1}\varepsilon_{t-2} - \theta_1\varepsilon_t\varepsilon_{t-3} + \theta_1^2\varepsilon_{t-1}\varepsilon_{t-3}\right] =$

$$= 0 - 0 - 0 + 0 = 0$$

Appendix 1

Some quick and incomplete references to ARMA model

Autocorrelation: $\rho_1 = \frac{\gamma_1}{\gamma_0} = -\frac{\theta_1 \sigma_\varepsilon^2}{(1+\theta_1)\sigma_\varepsilon^2} = -\frac{\theta_1}{1+\theta_1^2}$

$$\rho_k = \frac{\gamma_k}{\gamma_0} = 0 \quad \text{if } k \geq 2$$

Is stationary but invertible only if it is algebraically equivalent to a converging infinite order AR model. (… *that is: the θ coefficients have values such that the characteristic equation* $(1-\theta_1 B) = 0$ *has solution that fall outside the unit circle*).

$|\theta| < 1$ is the invertibility condition for a MA(1))

Premultpying $Y_t = (1-\theta_1 B)\varepsilon_t$ by $(1-\theta_1 B)^{-1}$, the equality is possible only if $(1-\theta_1 B)^{-1}$ is not infinite. This happens if $(1-\theta_1 B)^{-1}$ can be seen as the result of e convergent series $\Sigma(\theta_1 B)^i$ requiring that $|-\theta_1|<1$. Thus if $|-\theta_1|<1$ the MA(1) can be represented as a AR(∞) process (invertible)

In the MA(q) frame, we can consider the factorization θ(B)= $(1-r_1B) \cdot (1-r_2B) \cdot \ldots \cdot (1-r_qB)$. Thus invertibility requires that the all the *q* solutions fall outside the unit circle

Appendix 1
Some quick and incomplete references to ARMA model

AR(1) $Y_t = \phi_1 Y_{t-1} + \varepsilon_t$ $\quad Y_t - \phi_1 Y_{t-1} = (1 - \phi_1 B) Y_t = \varepsilon_t \quad \{\varepsilon_t\}_{t \in T} \sim WN(0, \sigma_\varepsilon^2)$

Is invertible but stationary only if it is algebraically equivalent to a infinite order MA model. (… *advanced: the* ϕ *coefficients have values such that the characteristic equation* $(1 - \phi_1 B) = 0$ *has solution that fall outside the unit circle*). (i.e. if $|\phi_1| < 1$)

Expected value: $E(Y_t) = \phi_1 E(Y_{t-1}) = 0$ (for $|\phi_1| < 1$ given that: $\mu = \phi_1 \mu$)

Variance $\gamma_0 = V(Y_t) = E(Y_t^2) - [E(Y_t)]^2 = E(Y_t^2)$

$$E(Y_t^2) = E\left[(\phi_1 Y_{t-1} + \varepsilon_t)^2\right] = E(\phi_1^2 Y_{t-1}^2 + \varepsilon_t^2 + 2\phi_1 Y_{t-1}\varepsilon_t) =$$

$$= \phi_1^2 E(Y_{t-1}^2) + E(\varepsilon_{t-1}^2) + 2\phi_1 E(Y_{t-1}\varepsilon_t) = \qquad \rightarrow \gamma_0 = \frac{\sigma_\varepsilon^2}{1 - \phi_1^2}$$

$$= \phi_1^2 \gamma_0 + \sigma_\varepsilon^2$$

Autocovariance:

$$\gamma_1 = Cov(Y_t, Y_{t-1}) = E(Y_t Y_{t-1}) - E(Y_t)E(Y_{t-1}) = E(Y_t Y_{t-1}) =$$

lag1

$$= E[(\phi_1 Y_{t-1} + \varepsilon_t) Y_{t-1}] = \phi_1 \frac{\sigma_\varepsilon^2}{1 - \phi_1^2}$$

Appendix 1
Some quick and incomplete references to ARMA model

$$\gamma_2 = Cov(Y_t, Y_{t-2}) = E(Y_t Y_{t-2}) - E(Y_t)E(Y_{t-2}) = E(Y_t Y_{t-2}) =$$

l2g1 $$= E[(\phi_1 Y_{t-1} + \varepsilon_t) Y_{t-2}] = \phi_1 E(Y_{t-1} Y_{t-2}) + E[\varepsilon_t Y_{t-2}] =$$

$$= \phi_1 \gamma_1 = \phi_1^2 \gamma_0 = \phi_1^2 \frac{\sigma_\varepsilon^2}{1-\phi_1^2}$$

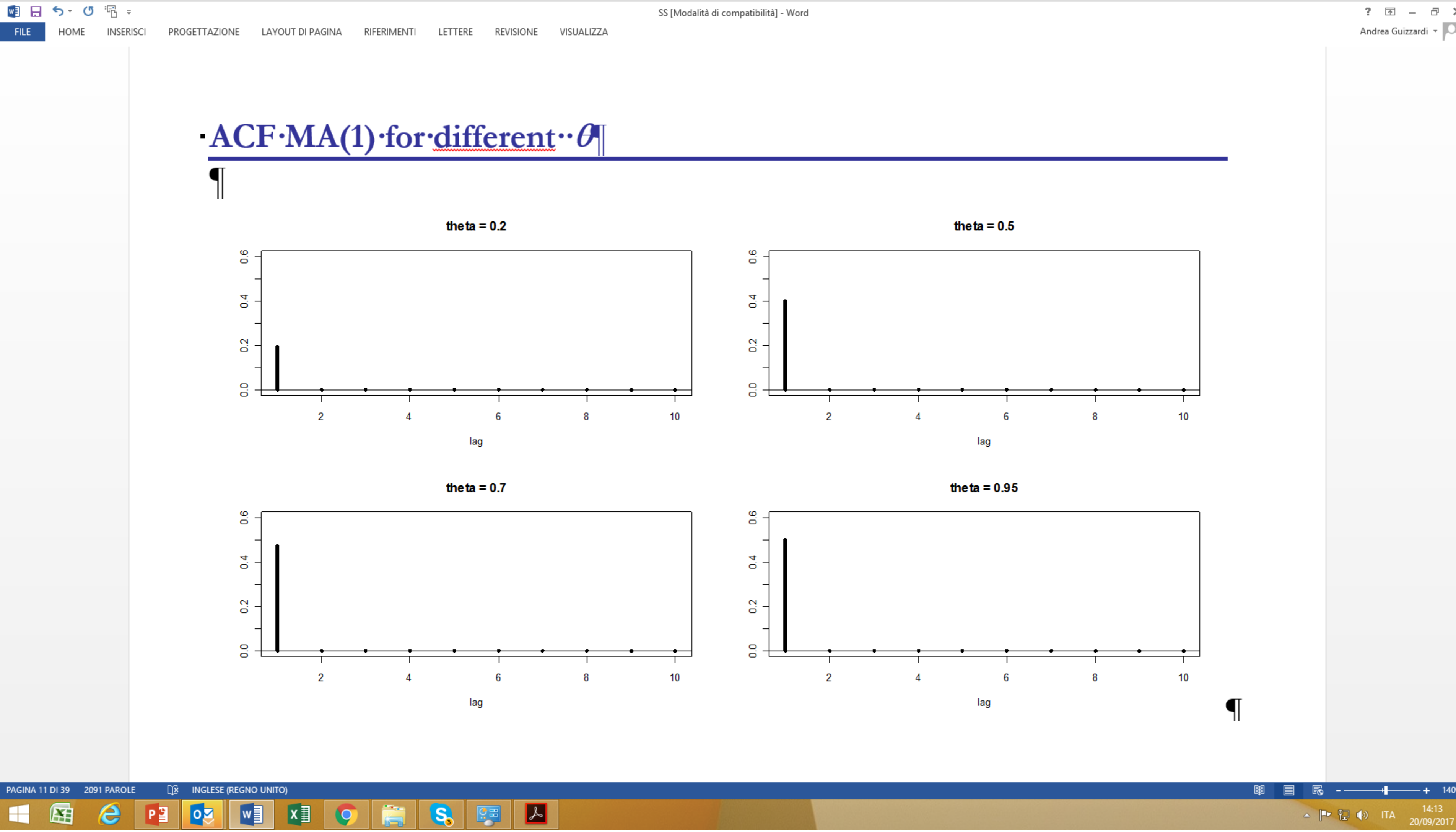
lag k $$\gamma_k = \phi_1^k \gamma_0 = \phi_1^k \frac{\sigma_\varepsilon^2}{1-\phi_1^2}$$

ACF $$\rho_k = \frac{\phi_1^k \gamma_0}{\gamma_0} = \phi_1^k \qquad \forall k \geq 0$$

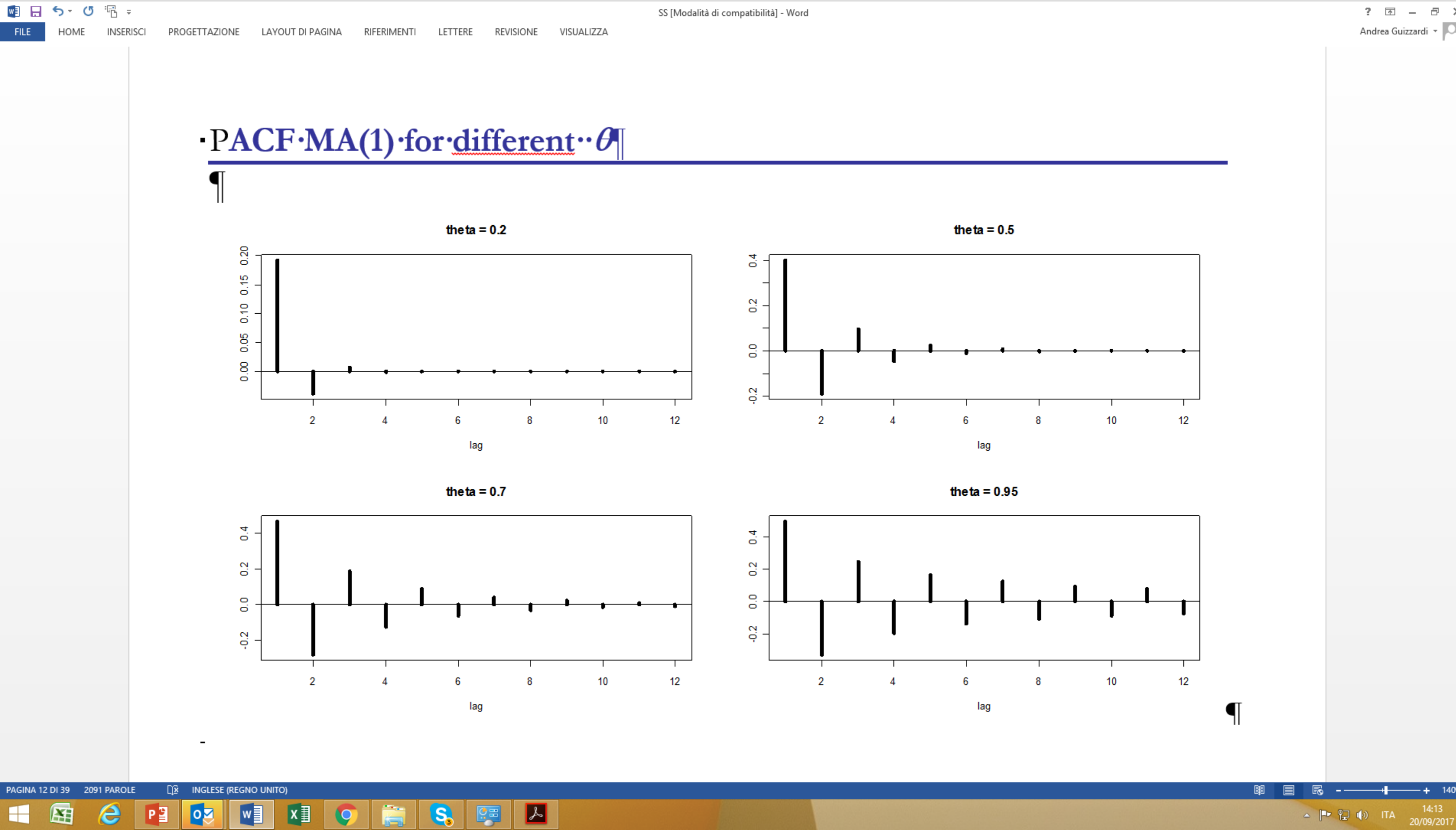
PACF $P_k = Corr(Y_t, Y_{t-k} | Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1})$ it measures the correlation between Y_t e Y_{t-k} once explained the linear dependence between Y_t e Y_{t-k} and all the variables $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}$.

- Complex to calculate, but it is equal 0 for *k>1*

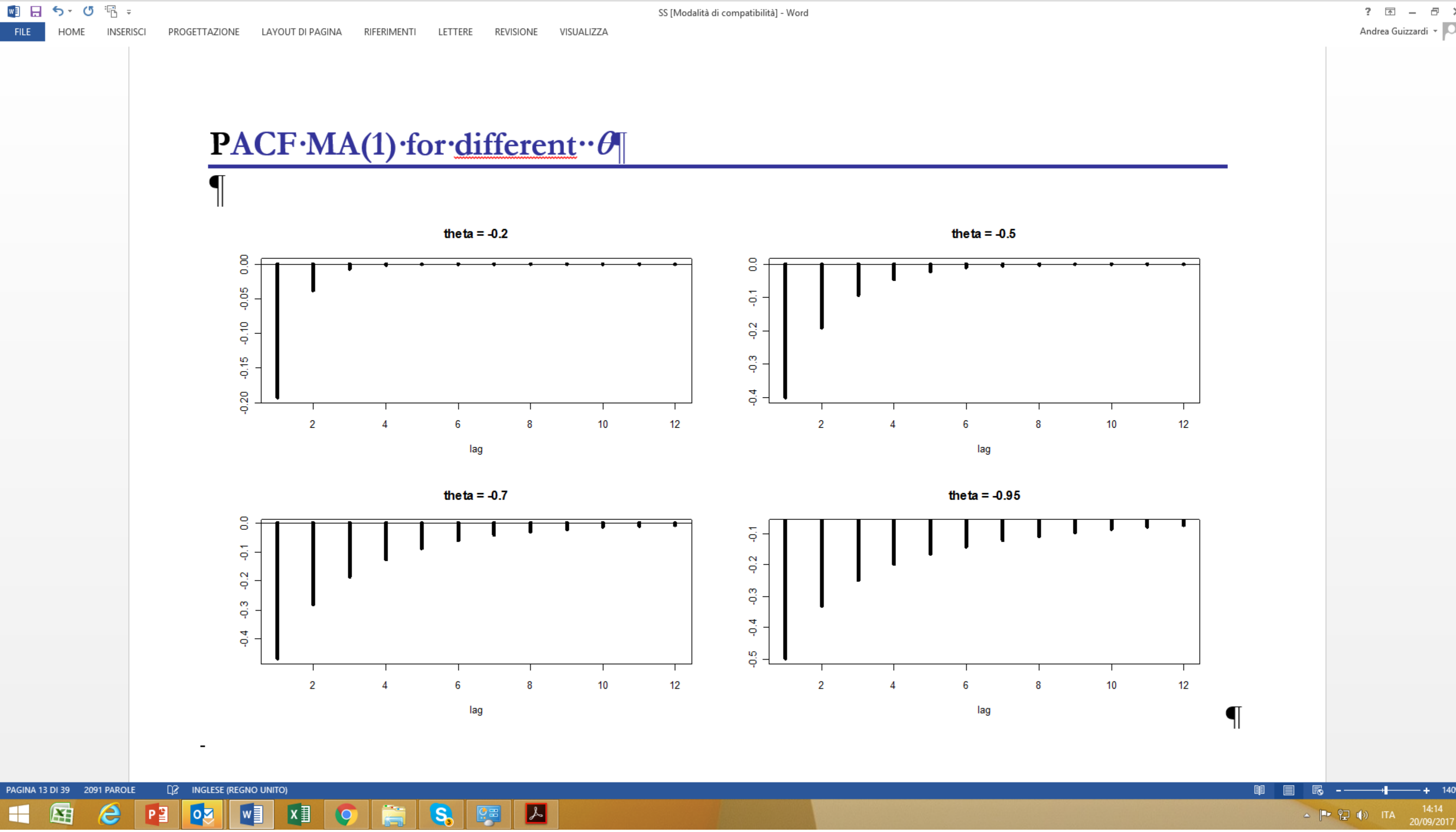
Appendix 1
Some quick and incomplete references to ARMA model



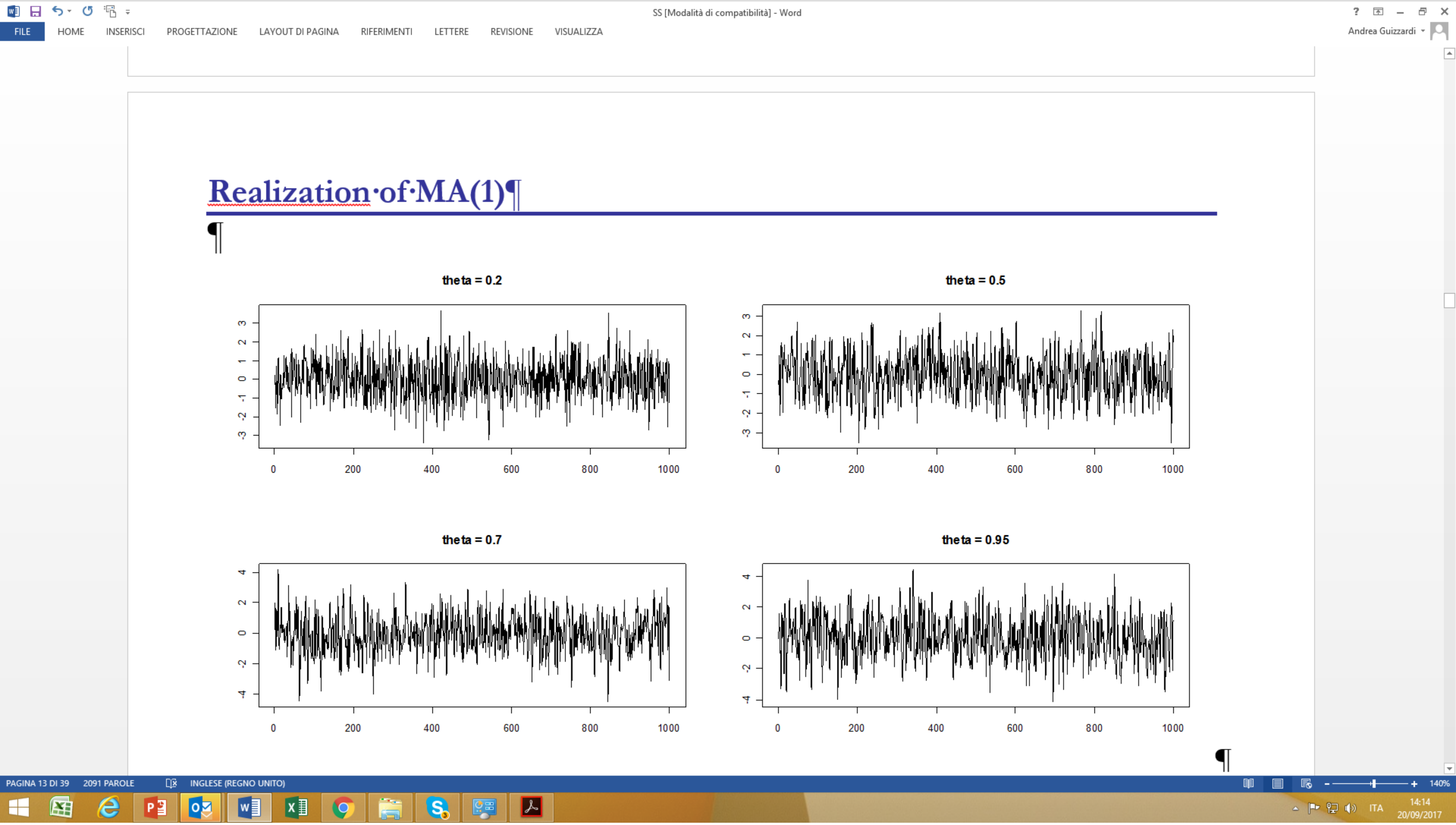
Appendix 1
Some quick and incomplete references to ARMA model



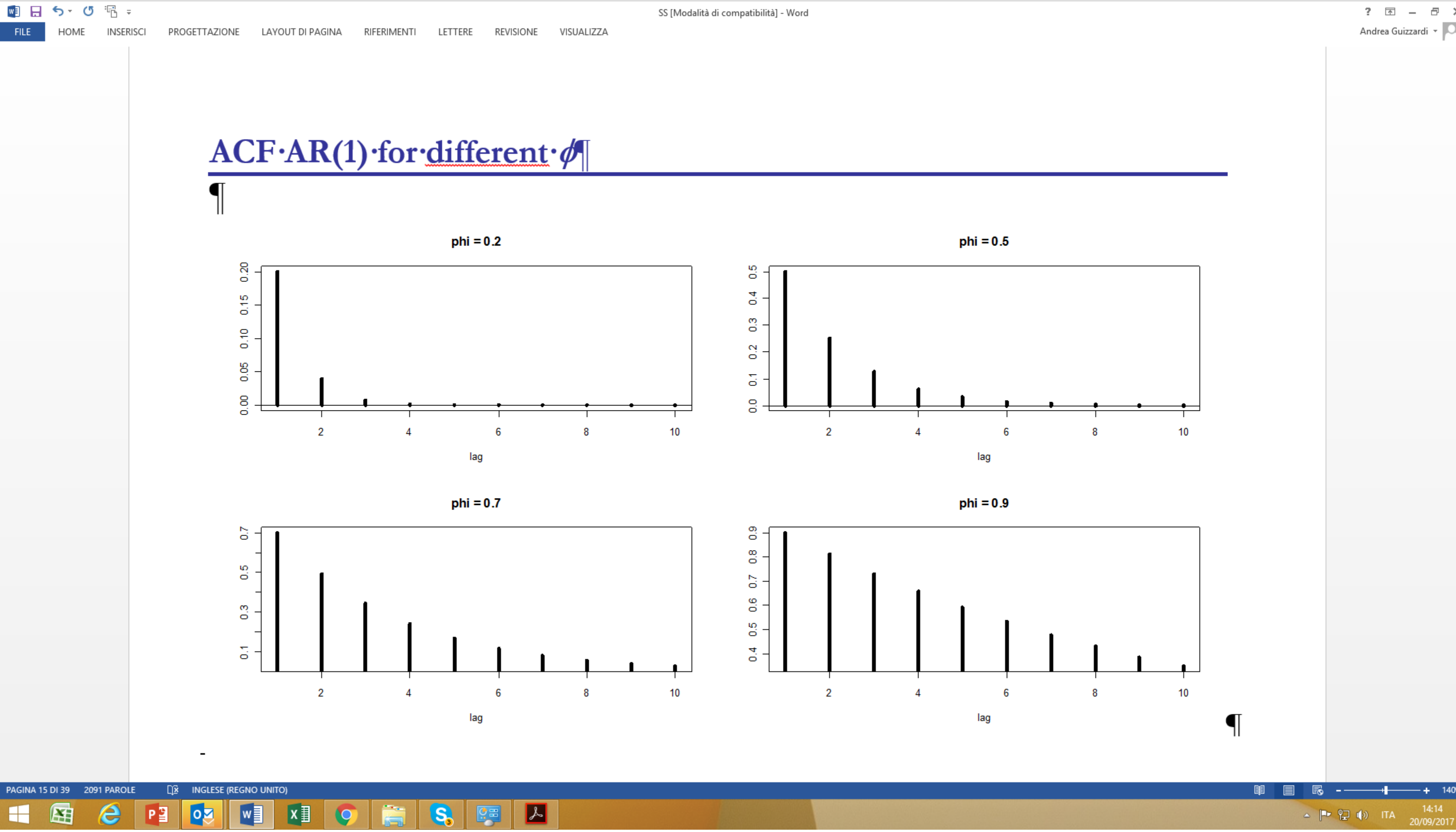
Appendix 1
Some quick and incomplete references to ARMA model



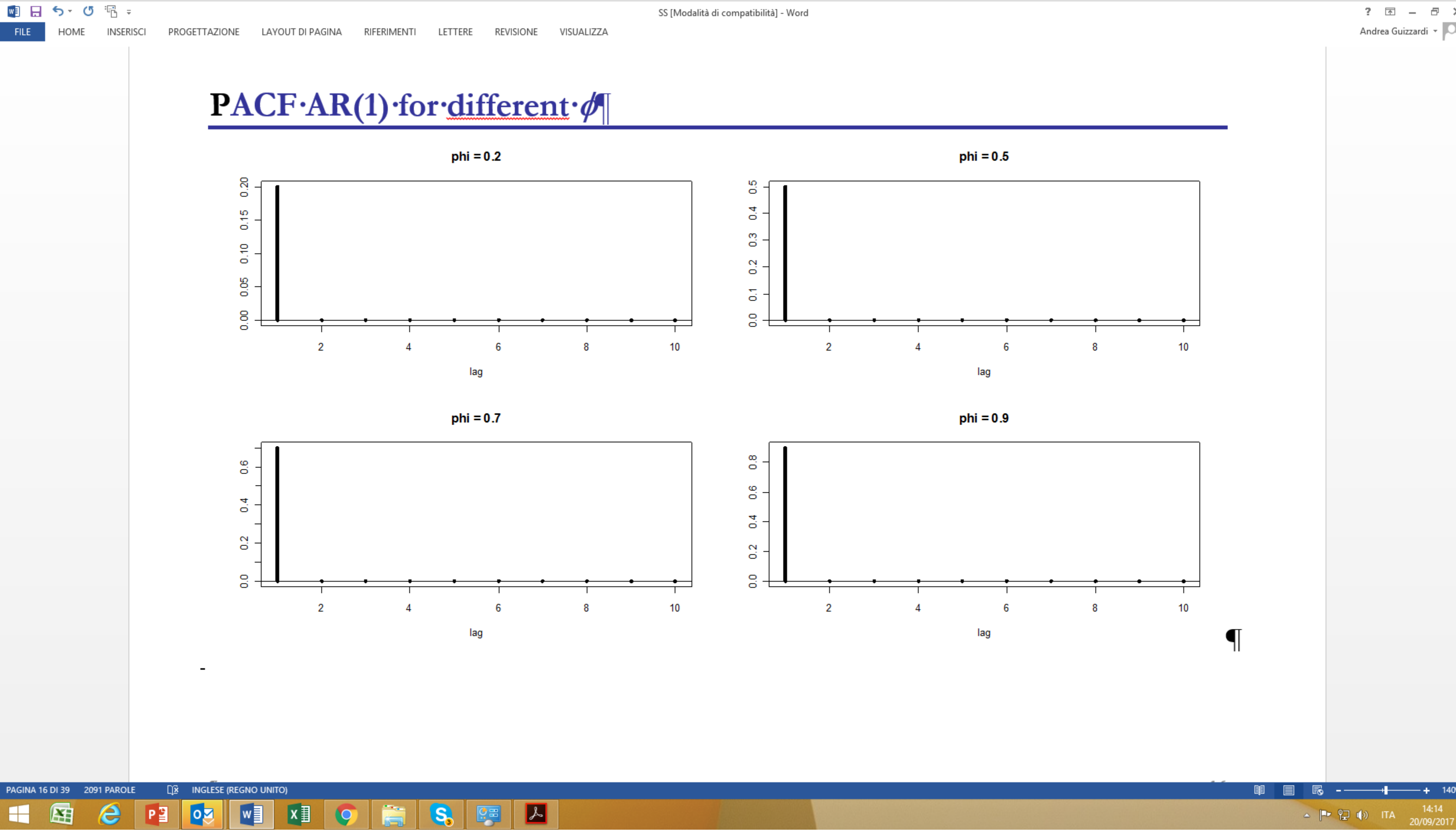
Appendix 1
Some quick and incomplete references to ARMA model



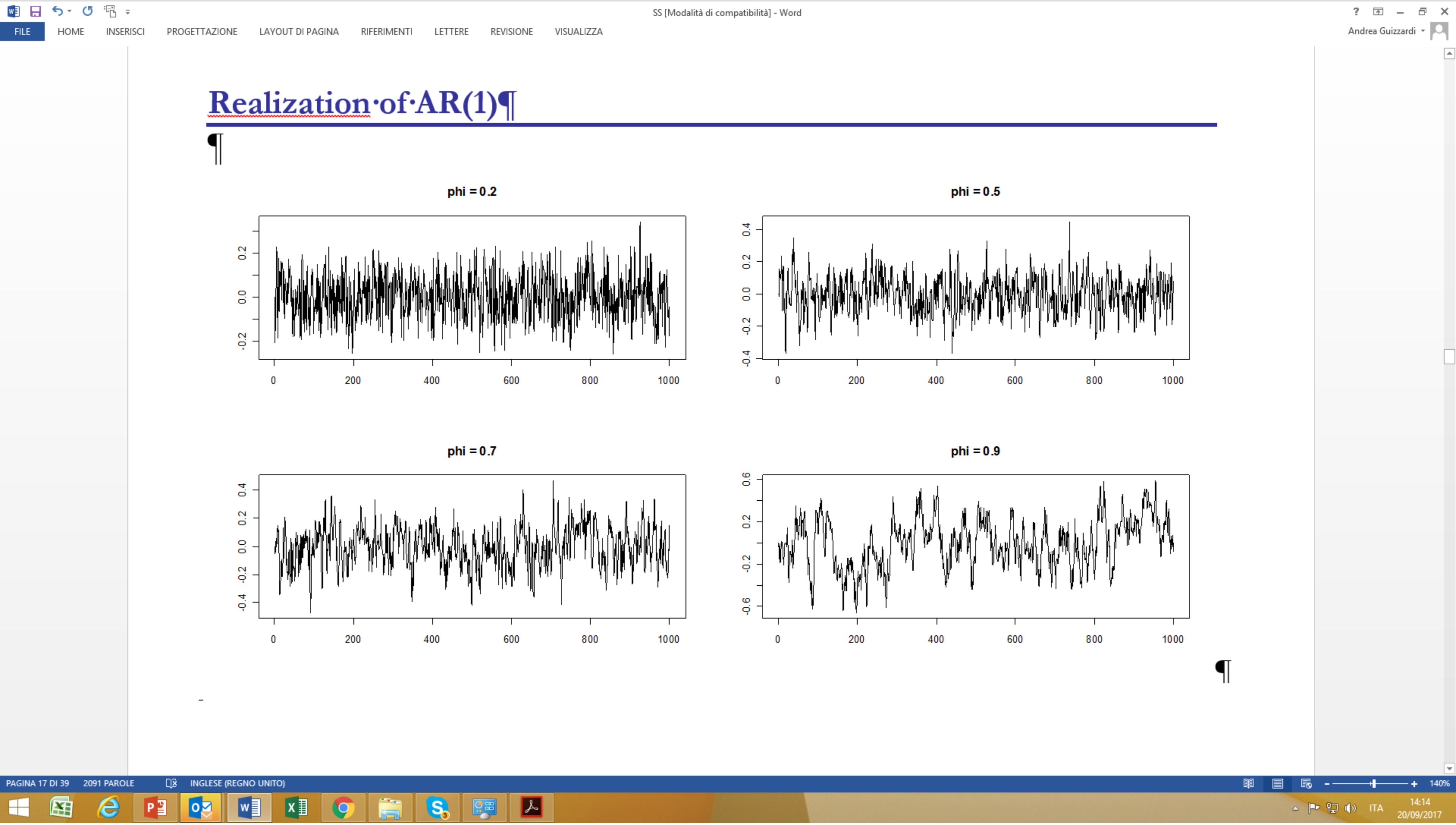
Appendix 1
Some quick and incomplete references to ARMA model



Appendix 1
Some quick and incomplete references to ARMA model



Appendix 1
Some quick and incomplete references to ARMA model



Appendix 1
Some quick and incomplete references to ARMA model

Mixed Autoregressive Moving Average Processes

The general ARMA(p,q) process is $\phi(L)y_t = \theta(L)\varepsilon_t$ where

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p$$
$$\theta(L) = 1 + \theta_1 L + \theta_2 L^2 - \ldots + \theta_q L^q.$$

Stationarity requires the **AR roots** of $\phi(z) = 0$ to **lie outside the unit circle**, and **invertibility** requires the same for the **MA roots** of $\theta(z) = 0$.

Given these conditions, the **mixed** ARMA(p,q) process may alternatively be expressed as a **pure** AR(∞) process or as a **pure** MA(∞) process (“Wold representation”) of infinite order, namely

$$\pi(L)y_t = \theta^{-1}(L)\phi(L)y_t = \varepsilon_t \quad \text{or} \quad y_t = \phi^{-1}(L)\theta(L)\varepsilon_t = \psi(L)y_t$$

Retrived at: http://www.feweb.vu.nl/econometriclinks/orientatie/econometricsii20073.pdf
basic knowledge of linear time series models for stationary data are assumed.

Identification of ARMA model orders, ACF

Method 1 for Identification: use ACF and PACF
This method helps to understand the data, but can only lead to an 'educated guess'.

Take variance of SACF r_k and SPACF ϕ_{kk} into account when trying to identify a model from SACF and SPACF. For $r_1 = \hat{\phi}_{11}$ use 'rule of thumb' variance of $\frac{1}{n}$ under H_0 that process is white noise. Use the following known properties:

1. The ACF of a **pure** MA(q) process "cuts off" after lag q.
2. The PACF of a **pure** AR(p) process "cuts off" after lag p.
3. ACF of pure AR(p) and mixed ARMA(p,q) ($p > 0$) processes **die out exponentially** (**after lag** q for mixed ARMA), **oscillating** in case of **negative or complex roots** of $\phi(z) = 0$. The AR order, p, is not easily derived from the ACF.

Retrived at: http://www.feweb.vu.nl/econometriclinks/orientatie/econometricsii20073.pdf
basic knowledge of linear time series models for stationary data are assumed.

Appendix 1
Some quick and incomplete references to ARMA model

Identification of ARMA model orders, AIC

Method 2: Minimize AIC or SIC

This is a general statistical procedure, not confined to time series analysis.

1. Estimate a collection models. Do not include models with p **and** q (too) large, ($p > 4$ and $q > 4$) unless you have a compelling reason (e.g. seasonal patterns).
2. Select model with the best trade-off between fit (residual sum of squared one-step-ahead forecasting errors) and number of parameters, according to AIC or SIC. Let p^* be $p + q$, One minimizes -2 $\times$ the loglikelihood plus a penalty depending on p^*.

AIC: $-2 * l(p^*) + 2p^*$ SIC: $-2 * l(p^*) + p^* \log n$

Note: in normal regression models:
$-2 * l(p^*) = c + n \log(\sum e_t^2/n)$, c.f. §4.3.2 and §5.2.1.

Retrived at: http://www.feweb.vu.nl/econometriclinks/orientatie/econometricsii20073.pdf
basic knowledge of linear time series models for stationary data are assumed.

Practical dangers for ARMA estimation

- $\phi(z) = 0$ and $\theta(z) = 0$ should **not** be so flexible as to have common roots. How to avoid? Do not overspecify AR and MA part simultaneously. Simplest example: fit ARMA(1,1) model to white noise ($\phi = -\theta$): inference on parameter estimates completely unreliable!
- NLS (and Eviews-) estimator and inference bad when $\theta(1) \approx 0$. The arbitrary assumptions about presample ε_t really hurt in this case. How to avoid? Do not "overdifference" the data. NLS not so problematic when $\phi(1) = 0$.
- Exact ML estimator and inference tricky when $\phi(1) \approx 0$. How to avoid? Do not "underdifference" the data.

$y_2 = \phi\, y_1 + \theta \varepsilon_1 + \varepsilon_2$	$\phi = -\theta$	
$y_2 = \phi\, y_1 - \phi\, \varepsilon_1 + \varepsilon_3$		
$y_3 = \phi\, y_2 - \phi\, \varepsilon_2 + \varepsilon_3$	$= \phi^2\, y_1 - \phi^2 \varepsilon_1 + \phi\, \varepsilon_2 - \phi\, \varepsilon_2 + \varepsilon_3$	$= \phi^2\, y_1 - \phi^2 \varepsilon_1 + \varepsilon_3$
$y_4 = \phi\, y_3 - \phi\, \varepsilon_3 + \varepsilon_4$	$= \phi^3\, y_1 - \phi^3 \varepsilon_1 + \phi\, \varepsilon_3 - \phi\, \varepsilon_3 + \varepsilon_4$	$= \phi^3\, y_1 - \phi^3 \varepsilon_1 + \varepsilon_4$
$y_5 = \phi\, y_4 - \phi\, \varepsilon_4 + \varepsilon_5$	$= \phi^4\, y_1 - \phi^4 \varepsilon_1 + \phi\, \varepsilon_4 - \phi\, \varepsilon_4 + \varepsilon_5$	$= \phi^4\, y_1 - \phi^4 \varepsilon_1 + \varepsilon_5$
.....	......	
$y_T =$	$= \phi^T\, y_1 - \phi^T \varepsilon_1 + \varepsilon_T \qquad T \to \infty$	$= \varepsilon_T$

Retrived at: http://www.feweb.vu.nl/econometriclinks/orientatie/econometricsii20073.pdf
basic knowledge of linear time series models for stationary data are assumed.

Appendix 1
Some quick and incomplete references to ARMA model

Testing and Evaluating ARMA models

- Check white noise assumption residuals, Ljung-Box test, LM test (Breusch Godfrey). If H_0 rejected: add AR or MA parameter, or add regressors §7.3.
- Check other assumptions using residuals (homoskedasticity, stationarity). If H_0 rejected allow for a changing mean §7.3, or variance §7.4, or both.
- Forecast performance out-of-sample. Assess empirical coverage of theoretical confidence intervals: count no. of observations outside confidence interval. Compare Root Mean Squared Prediction Error (RMSE), or Mean Absolute Prediction Error (MAE) with forecasts of benchmark models. If performance unsatisfactory: simplify model or allow for changing mean or variance in the model: model nonstationarity.

Retrived at: http://www.feweb.vu.nl/econometriclinks/orientatie/econometricsii20073.pdf
basic knowledge of linear time series models for stationary data are assumed.