Forecasting Methods

Master in MQDEE

Jorge Caiado
CEMAPRE/ISEG, University of Lisbon
Email: jcaiado@iseg.ulisboa.pt
Web: http://jcaiado100.wixsite.com/jorgecaiado

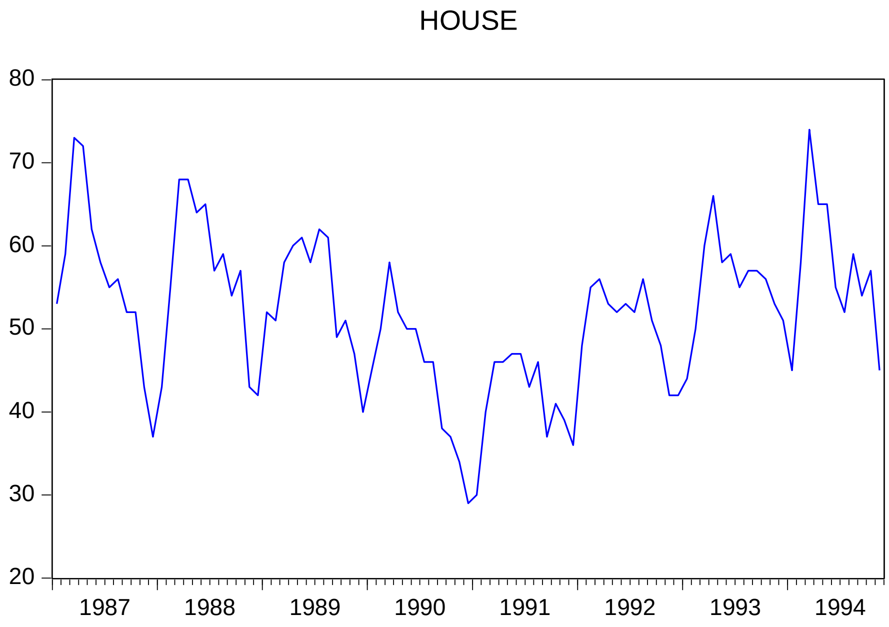
ISEG
LISBON
SCHOOL OF
ECONOMICS &
MANAGEMENT
UNIVERSIDADE DE LISBOA

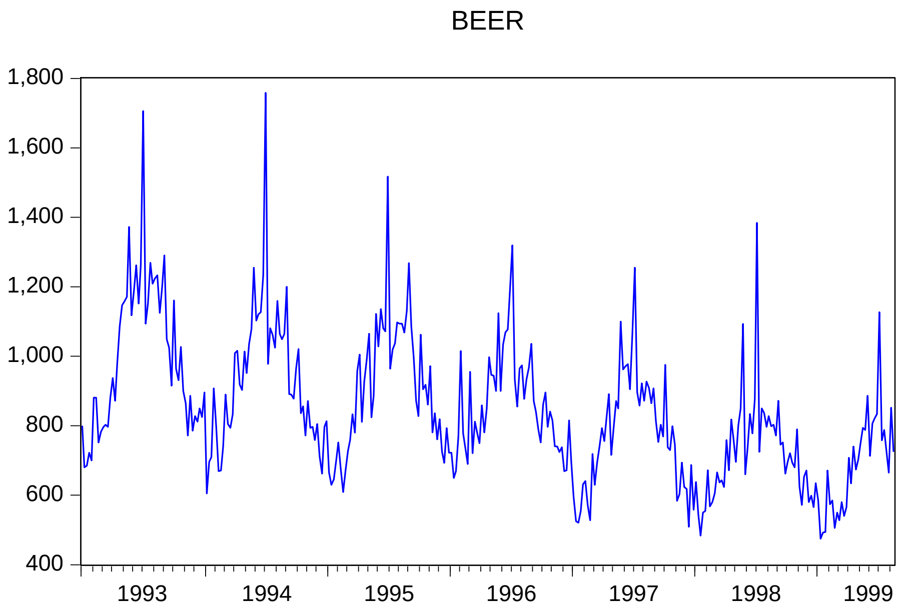
Introduction to Forecasting Methods

Time Series Analysis

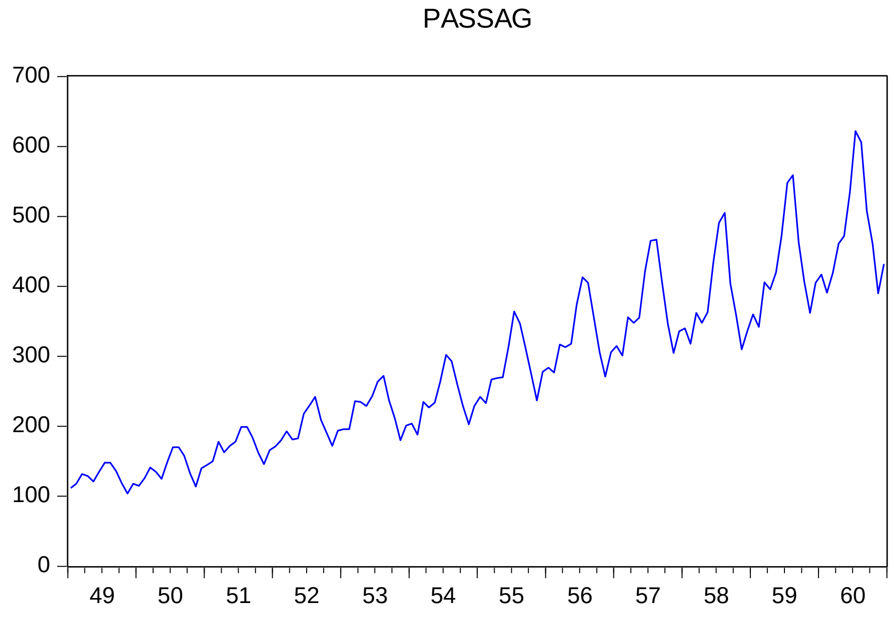
Definition of time series

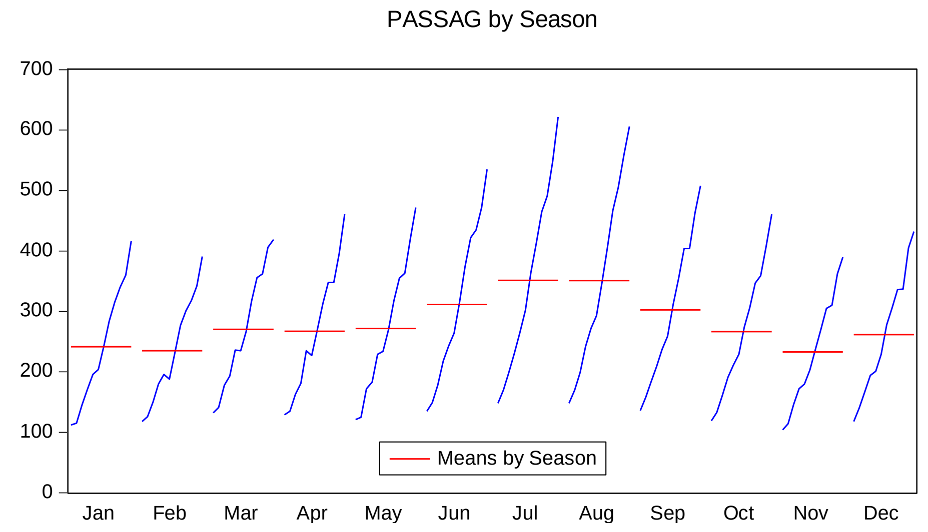
A **time series** is a sequence of observations over time. For example: monthly sales of new one-family houses; Daily stock indices; Weekly beer consumption; daily average temperature; Annual electricity production.





Time Series Analysis





Linear trend model

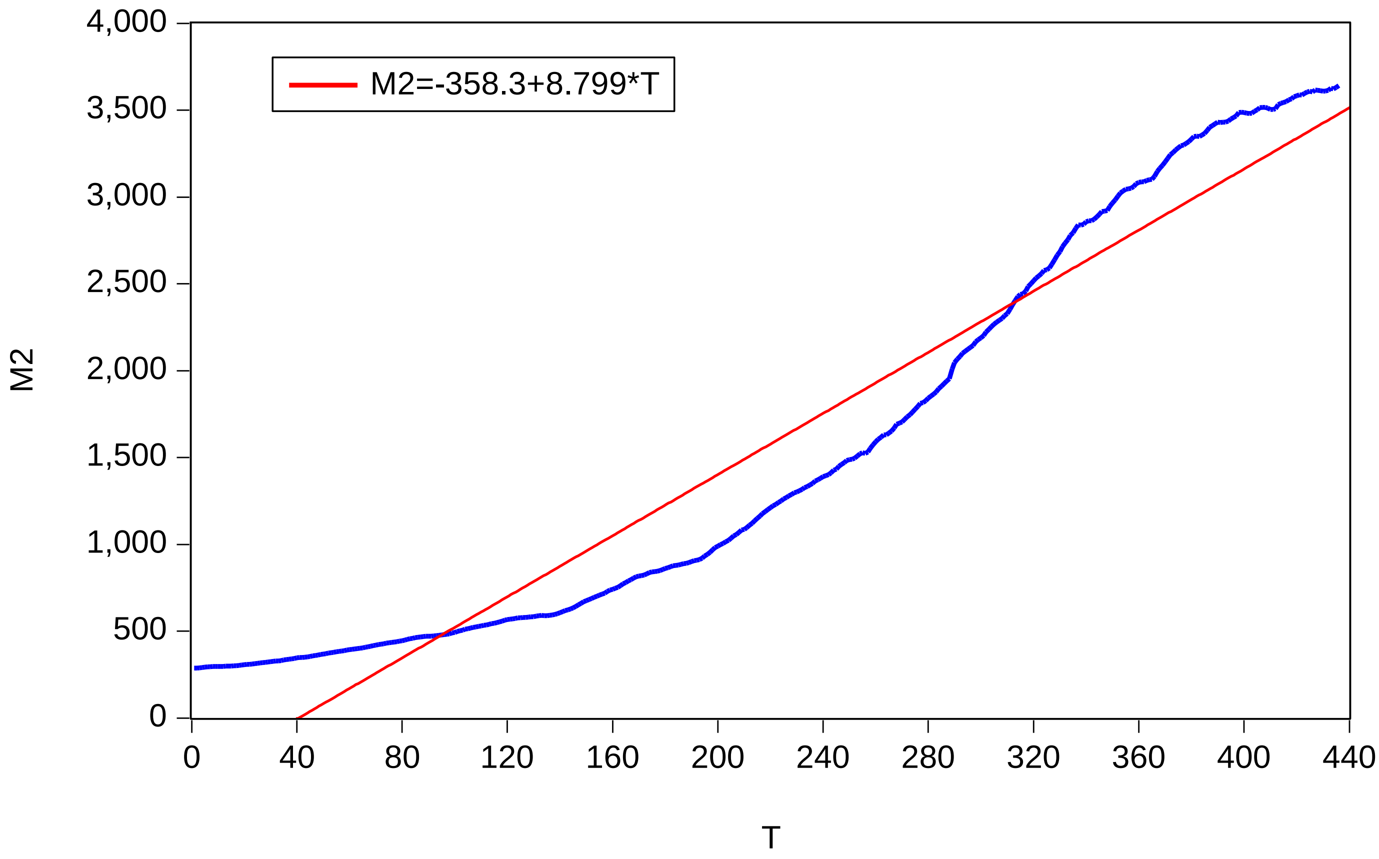
A common feature of time series data is a trend. We can model and forecast the trend in a time series data using the following regression model (called "linear trend model"):

$$Y_t = b_0 + b_1 t + \varepsilon_t$$

where $t = 1, 2, \ldots, T$ (time) is the explanatory or predictor variable.

Time Series Analysis

Example Figure below shows the estimated regression line of money stock M2 on time *T* (data from 1959:Q1 to 1995:Q4 in the US)



Time Series Analysis

Log-linear trend model

Suppose we want to find out the growth rate of consumption (Y_t) in Portugal from 2000Q1 to 2017Q3. Let Y_0 be the initial value of the consumption (i.e, the value in the end of 1999Q4).

We may use the following compound interest formula

$$Y_t = Y_0(1 + r)^t$$

where *r* is the compound rate of growth of Y. Taking the natural logarithm, we can write

$$\log Y_t = \log Y_0 + t \log(1 + r)$$

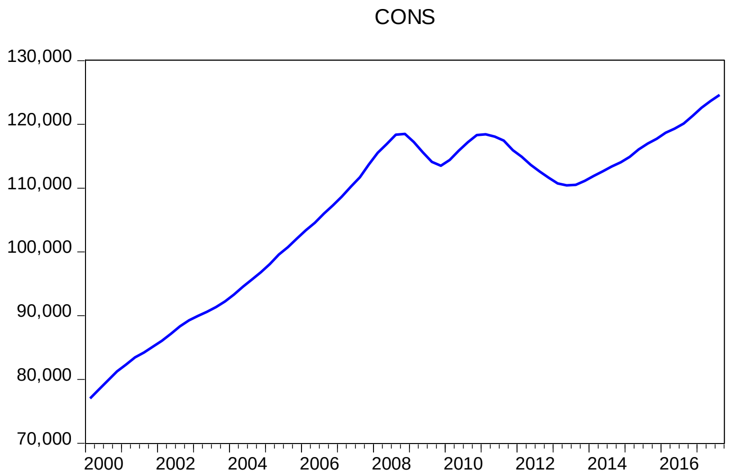
Now letting $b_0 = \log Y_0$ and $b_1 = \log(1 + r)$, we can write it as

$$\log Y_t = b_0 + b_1 t$$

Adding the disturbance term we obtain the so-called log-linear trend model:

$$\log Y_t = b_0 + b_1 t + \varepsilon_t$$

Time Series Analysis



Dependent Variable: LOG(CONS)
Method: Least Squares
Date: 07/21/18 Time: 11:54
Sample: 2000Q1 2017Q3
Included observations: 71

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	11.35948	0.014669	774.3708	0.0000
@TREND+1	0.005709	0.000354	16.12160	0.0000

R-squared	0.790214	Mean dependent var	11.56500
Adjusted R-squared	0.787173	S.D. dependent var	0.132554
S.E. of regression	0.061151	Akaike info criterion	-2.723172
Sum squared resid	0.258023	Schwarz criterion	-2.659434
Log likelihood	98.67260	Hannan-Quinn criter.	-2.697825
F-statistic	259.9059	Durbin-Watson stat	0.019370
Prob(F-statistic)	0.000000		

How to interpret this model?

Time Series Analysis

LISBON SCHOOL OF ECONOMICS & MANAGEMENT
UNIVERSIDADE DE LISBOA

Time series decomposition

We can think of a time series as containing four components: trend (T), cycle (C), seasonality (S) and noise or error (E).

For example, we may assume an additive model as follows:

$$Y_t = T_t + C_t + S_t + E_t$$

or

$$Y_t = TC_t + S_t + E_t$$

Alternatively, we can write a multiplicative model as

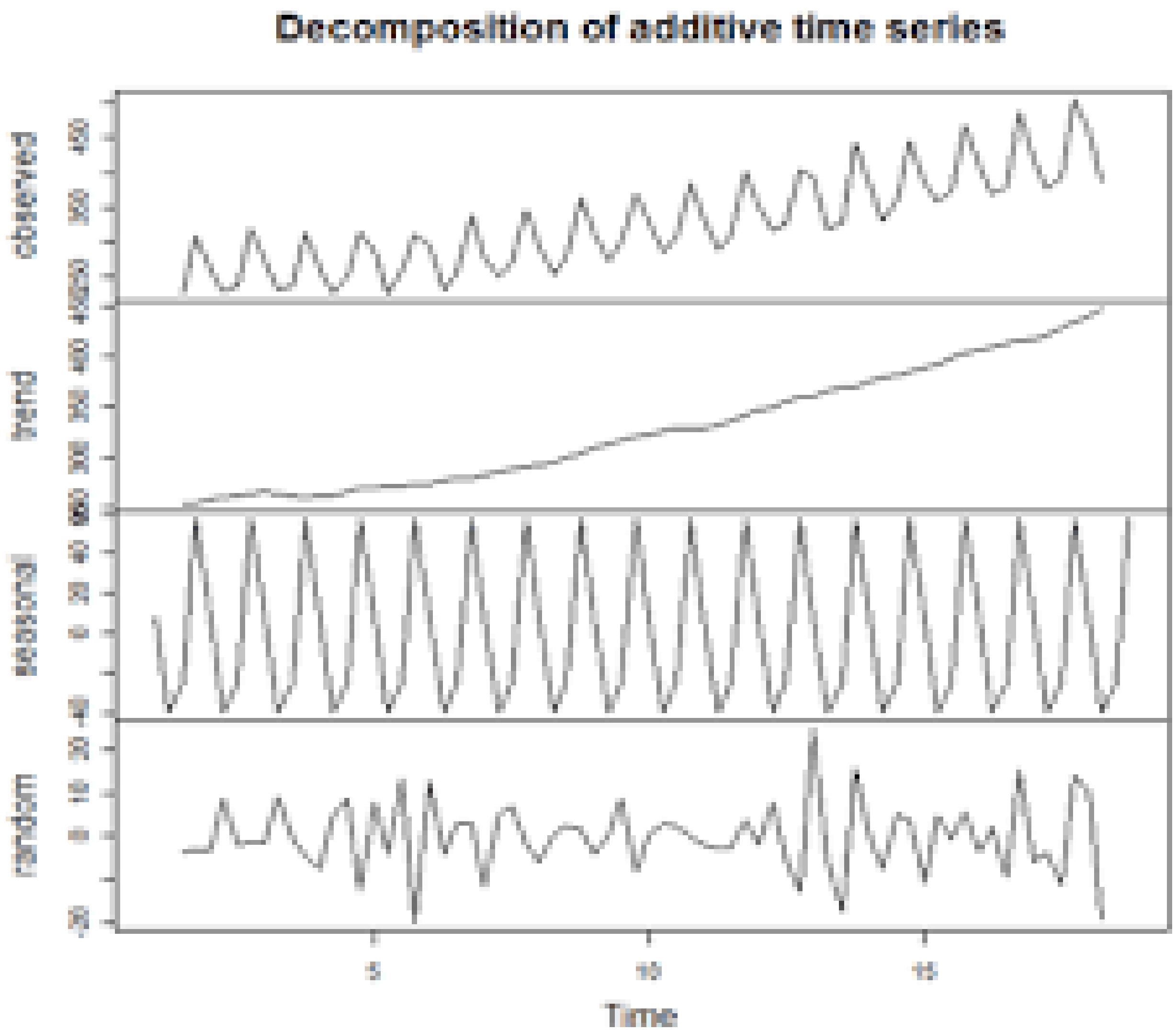
$$Y_t = T_t \times C_t \times S_t \times E_t$$

or

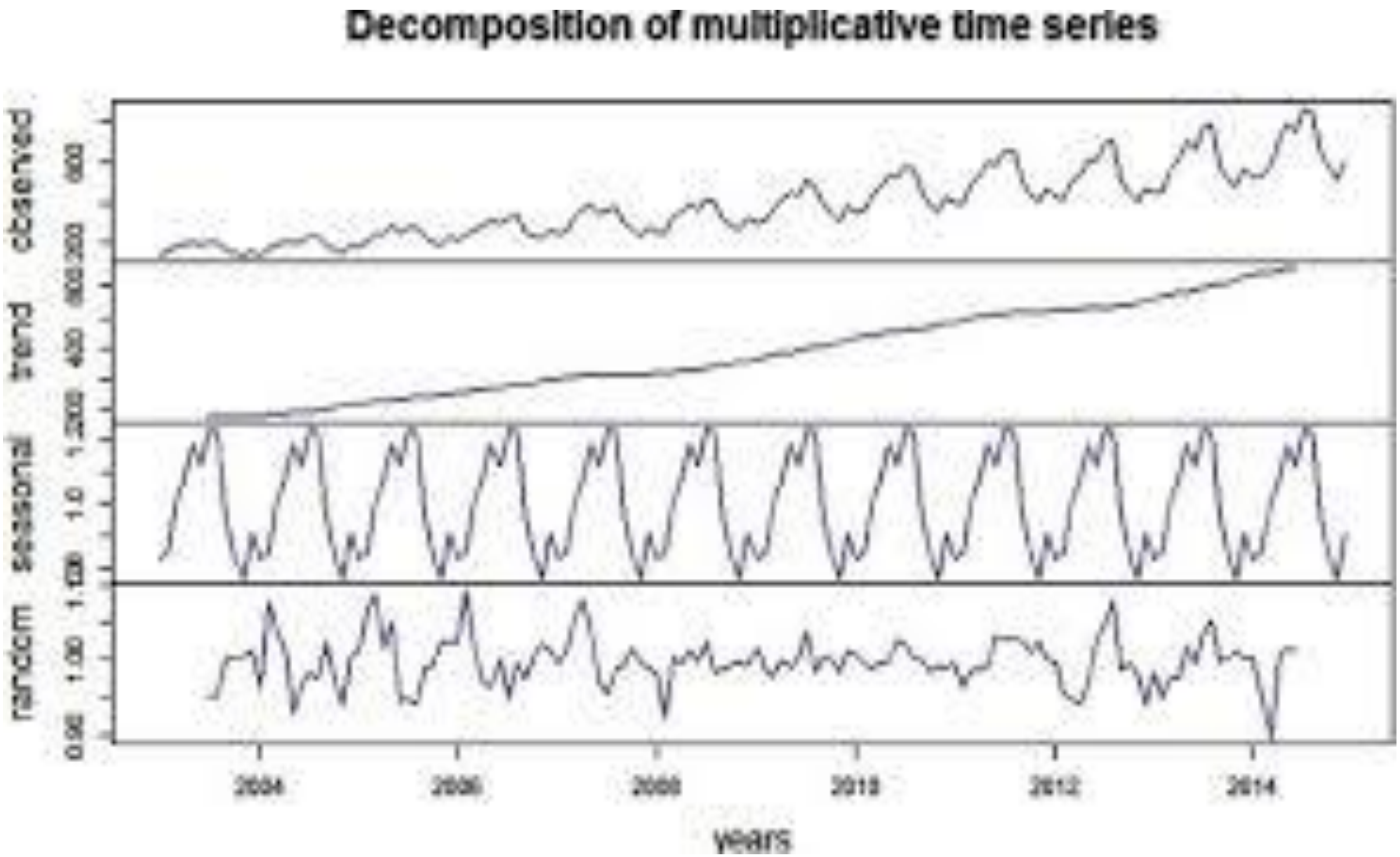
$$Y_t = TC_t \times S_t \times E_t$$

The additive model is most appropriate if the magnitude of the seasonal fluctuations or the variation around the trend-cycle does not vary with the level of the time series.

Time Series Analysis



Time Series Analysis



Time Series Analysis

LISBON SCHOOL OF ECONOMICS & MANAGEMENT
UNIVERSIDADE DE LISBOA

Seasonal Adjustment

Some economic time series observed at quarterly, monthly, weekly frequencies often exhibit cyclical seasonal movements that occur every quarter, month or week. For example, the monthly inflation rate in Angola reach a peak every December during Christmas period.

Seasonal adjustment → remove the cyclical seasonal movements from a series.

Moving average methods:

- Additive decomposition
- Multiplicative decomposition

The **seasonal period** is denoted by s (e.g., s=4 for quarterly data, s=12 for monthly data, s= 7 for daily data with a weekly pattern)

Time Series Analysis

LISBON SCHOOL OF ECONOMICS & MANAGEMENT
UNIVERSIDADE DE LISBOA

Multiplicative decomposition: $Y_t = TC_t \times S_t \times E_t$

Step1: Compute the trend-cycle component using a centered moving average as

$$\widehat{TC}_t = (0.5Y_{t-6} + \cdots + Y_t + \cdots + 0.5Y_{t+6})/12 \quad \text{if} \quad s = 12 \text{ (monthly)}$$

or

$$\widehat{TC}_t = (0.5Y_{t-2} + Y_{t-1} + Y_t + Y_{t+1} + 0.5Y_{t+2})/4 \quad \text{if} \quad s = 4 \text{ (quarterly)}$$

Step 2: Calculate the detrended series: $Y_t/\widehat{TC}_t$

Step 3: Estimate the seasonal components for each month or quarter, averaging the detrended values for that month or quarter. Then adjust the **seasonal indices so that they add to s**:

$$\hat{S}_m = i_m / \sqrt[12]{i_1 i_2 \cdots i_{12}} \quad \text{if} \quad s = 12 \text{ (monthly)}$$

or

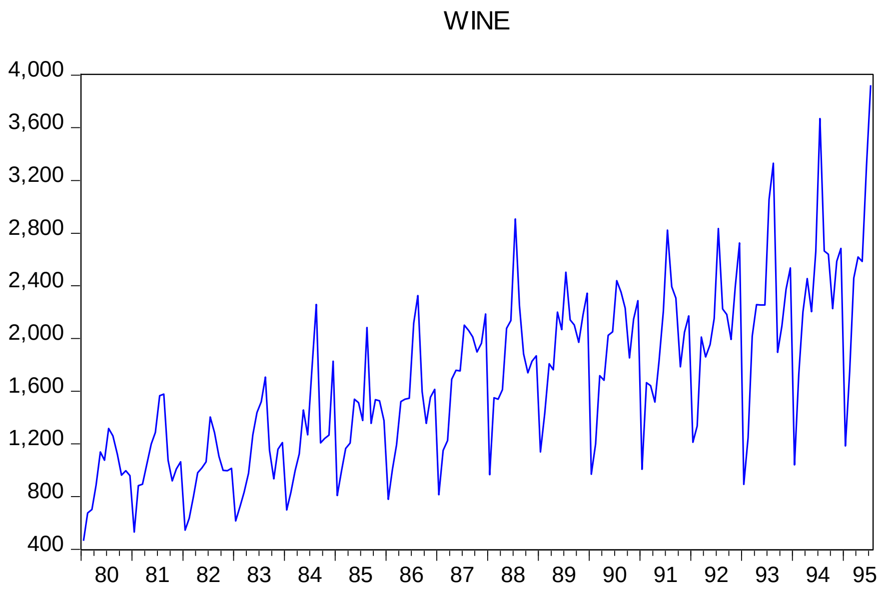
$$\hat{S}_q = i_q / \sqrt[12]{i_1 i_2 i_3 i_4} \quad \text{if} \quad s = 4 \text{ (quarterly)}$$

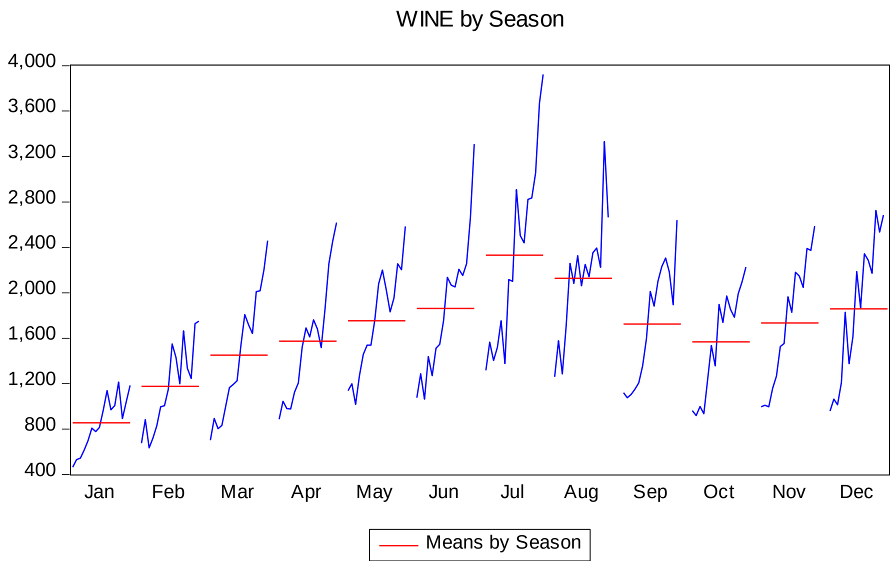
Time Series Analysis

Step 4: The seasonally adjusted series is obtained by **dividing** Y_t by the seasonal factors S_j. This gives Y_t^{SA}.

Step 5: The remainder component is calculated by **dividing** out the estimated seasonal and trend-cycle components: $\hat{E}_t = Y_t/(\widehat{TC}_t \times \hat{S}_t)$

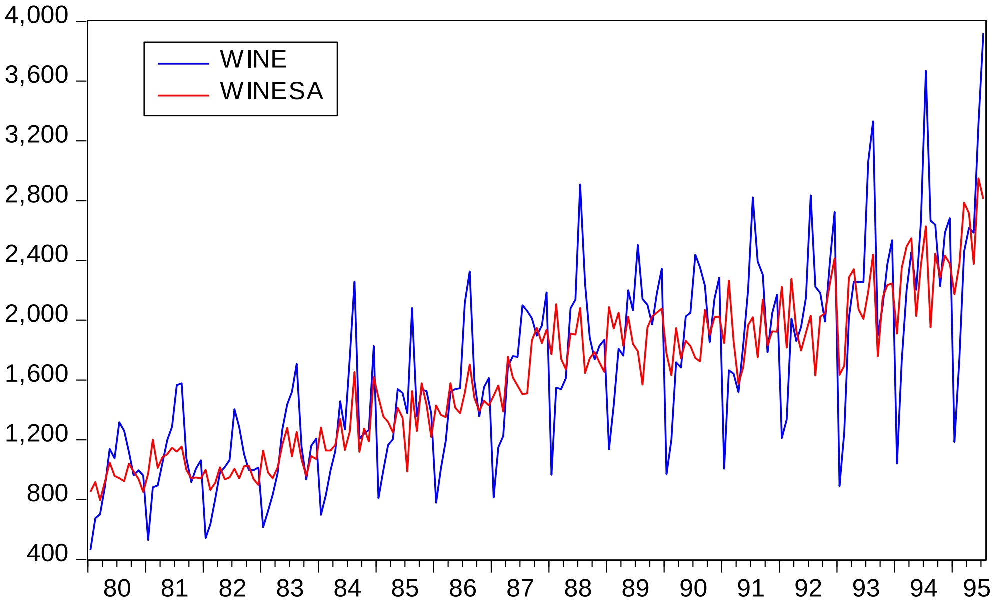
Example 17.1. Figure below shows the result of multiplicative decomposition the wine consumption in Australia from January 1980 to July 1995.





Time Series Analysis

LISBON SCHOOL OF ECONOMICS & MANAGEMENT
UNIVERSIDADE DE LISBOA



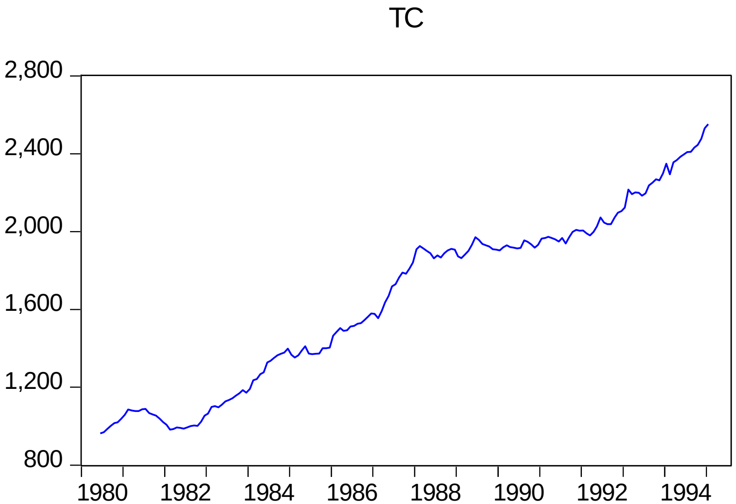
EVIEWS: Proc/Seasonal Adjustment/Moving average methods/Ratio to moving average

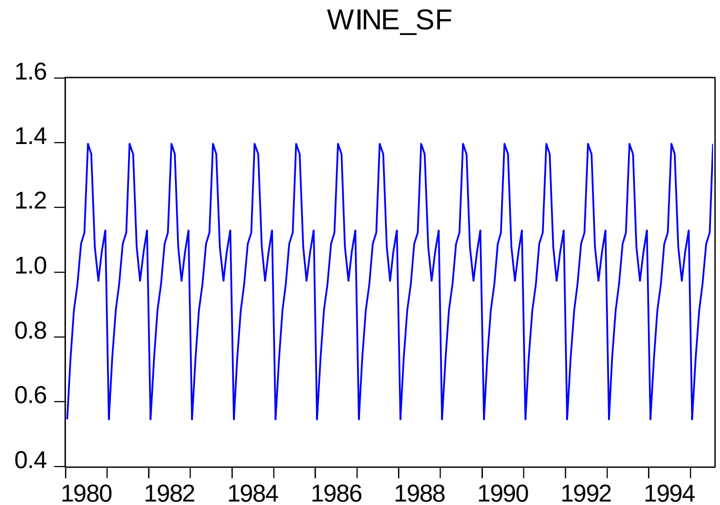
Date: 07/21/18 Time: 19:15
Sample: 1980M01 1995M07
Included observations: 187
Ratio to Moving Average
Original Series: WINE
Adjusted Series: WINESA

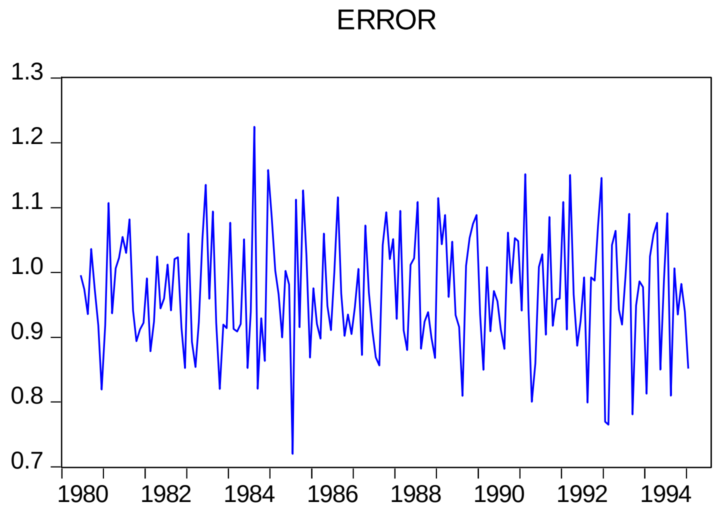
Scaling Factors:

1	0.545133
2	0.735193
3	0.882675
4	0.963865
5	1.087376
6	1.122619
7	1.396486
8	1.365222
9	1.078535
10	0.973587
11	1.063142
12	1.128828

Time Series Analysis







Time Series Analysis

Additive decomposition $Y_t = TC_t + S_t + E_t$

Step1: Compute the trend-cycle component using a centered moving average as

$$\widehat{TC}_t = (0.5Y_{t-6} + \cdots + Y_t + \cdots + 0.5Y_{t+6})/12 \quad \text{if} \quad s = 12 \text{ (monthly)}$$

or

$$\widehat{TC}_t = (0.5Y_{t-2} + Y_{t-1} + Y_t + Y_{t+1} + 0.5Y_{t+2})/4 \quad \text{if} \quad s = 4 \text{ (quarterly)}$$

Step 2: Calculate the detrended series: $Y_t - \widehat{TC}_t$

Step 3: Estimate the seasonal components for each month or quarter, averaging the detrended values for that month or quarter. Then adjust the **seasonal indices so that they add up to zero:**

$$\hat{S}_m = i_m - (i_1 + i_2 + \cdots + i_{12})/12 \quad \text{if} \quad s = 12 \text{ (monthly)}$$

or

$$\hat{S}_q = i_q - (i_1 + i_2 + i_3 + i_4)/4 \quad \text{if} \quad s = 4 \text{ (quarterly)}$$

Time Series Analysis

Step 4: The seasonally adjusted series is obtained by **subtracting** Y_t by the seasonal factors S_j. This gives Y_t^{SA}.

Step 5: The remainder component is calculated by **subtracting** the estimated seasonal and trend-cycle components: $\hat{E}_t = Y_t - \widehat{TC}_t - \hat{S}_t$

Other seasonal adjustment procedures: X12 ARIMA, STL and TRAMO/SEATS

Time Series Analysis

Example: The data below represent the monthly sales of houses in Ohio (US) from January 1987 to November July 1994 (EViews file: data-forecasting.wk1; page=house)

Plot the time series. Are there any seasonal fluctuations? Use additive decomposition to estimate the trend-cycle, seasonal indices and error component.

Date	Sales	Date	Sales	Date	Sales	Date	Sales
1987M01	53	1989M01	52	1991M01	30	1993M01	44
1987M02	59	1989M02	51	1991M02	40	1993M02	50
1987M03	73	1989M03	58	1991M03	46	1993M03	60
1987M04	72	1989M04	60	1991M04	46	1993M04	66
1987M05	62	1989M05	61	1991M05	47	1993M05	58
1987M06	58	1989M06	58	1991M06	47	1993M06	59
1987M07	55	1989M07	62	1991M07	43	1993M07	55
1987M08	56	1989M08	61	1991M08	46	1993M08	57
1987M09	52	1989M09	49	1991M09	37	1993M09	57
1987M10	52	1989M10	51	1991M10	41	1993M10	56
1987M11	43	1989M11	47	1991M11	39	1993M11	53
1987M12	37	1989M12	40	1991M12	36	1993M12	51
1988M01	43	1990M01	45	1992M01	48	1994M01	45
1988M02	55	1990M02	50	1992M02	55	1994M02	58
1988M03	68	1990M03	58	1992M03	56	1994M03	74
1988M04	68	1990M04	52	1992M04	53	1994M04	65
1988M05	64	1990M05	50	1992M05	52	1994M05	65
1988M06	65	1990M06	50	1992M06	53	1994M06	55
1988M07	57	1990M07	46	1992M07	52	1994M07	52
1988M08	59	1990M08	46	1992M08	56	1994M08	59
1988M09	54	1990M09	38	1992M09	51	1994M09	54
1988M10	57	1990M10	37	1992M10	48	1994M10	57
1988M11	43	1990M11	34	1992M11	42	1994M11	45
1988M12	42	1990M12	29	1992M12	42		

Forecast Evaluation

Suppose the forecast sample is $T+1, T+2, \ldots, T+k$ and denote the actual value in period t as Y_t and the forecasted value as $\hat{Y}_t$.

The three most commonly used forecast accuracy measures are:

$$RMSE = (1/k)\sum_{t=T+1}^{T+k}(Y_t - \hat{Y}_t)^2$$ (Root Mean Squared Error)

$$MAE = 1/k)\sum_{t=T+1}^{T+k}\left|Y_t - \hat{Y}_t\right|$$ (Mean Absolute Error)

$$MAPE = 1/k)\sum_{t=T+1}^{T+k}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \times 100$$ (Mean Absolute Percentual Error)

Time Series Analysis

LISBON SCHOOL OF ECONOMICS & MANAGEMENT
UNIVERSIDADE DE LISBOA

Exponential Smoothing

Exponential smoothing methods compute forecasts as weighted averages of past observations, with the weights decaying exponentially as the observation get older.

Single smoothing

The single exponential method is appropriate for forecasting series with no trend or seasonal pattern.

Forecast at time $t+1$: $\hat{Y}_{t+1} = \alpha Y_t + (1-\alpha)\hat{Y}_t$, where $0 \leq \alpha \leq 1$ is the damping or smoothing parameter.

By repeated substitutions, we obtain

$$\hat{Y}_{t+1} = \sum_{j=0}^{T-1} \alpha(1-\alpha)^j Y_{t-j}$$

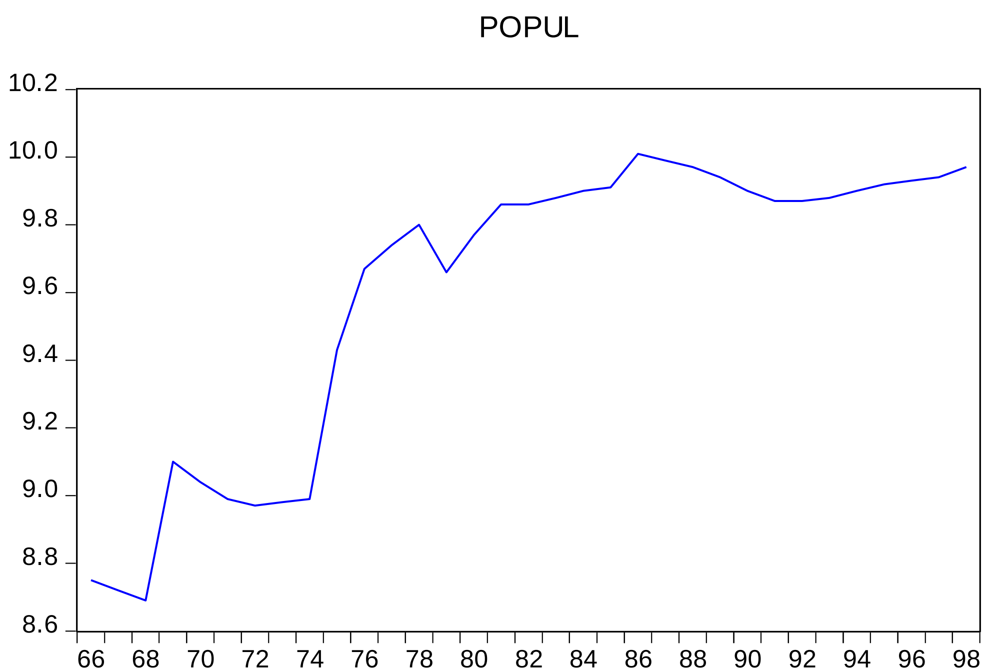
The forecast equation of single exponential method is given by:

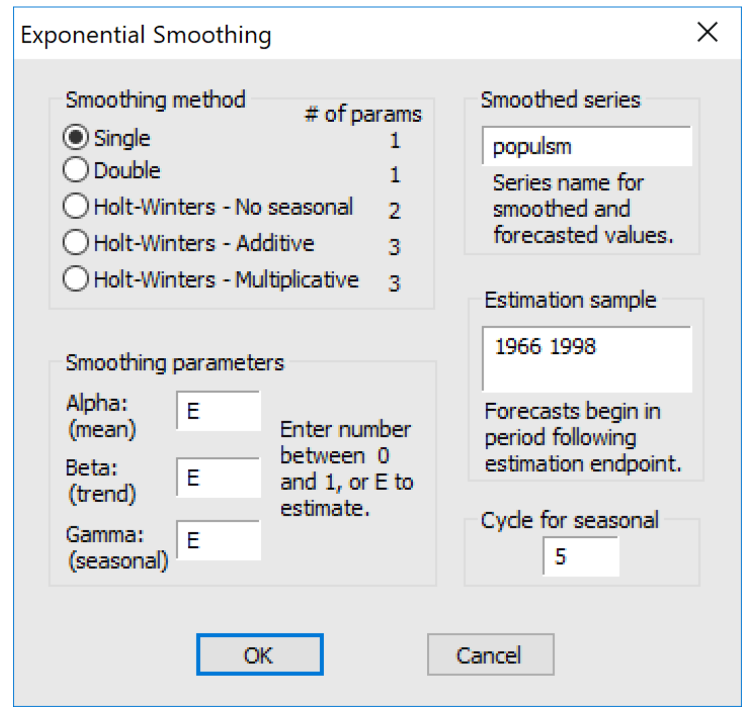
$$\hat{Y}_{T+k} = \hat{Y}_T \text{ for all } k > 0$$

Time Series Analysis

Initialization: We may use $\hat{Y}_2 = Y_1$ or the mean of the initial observations of Y_t. EViews uses the mean of the initial $(T+1)/2$ observations of Y_t to start the recursion.

Example: The figure below is a plot of the population in Portugal from 1966 to 1998 (Source: INE) .





Time Series Analysis

LISBON SCHOOL OF ECONOMICS & MANAGEMENT
UNIVERSIDADE DE LISBOA

Date: 08/02/18 Time: 21:44
Sample: 1966 1998
Included observations: 33
Method: Single Exponential
Original Series: POPUL
Forecast Series: POPULSM

Parameters:	Alpha	0.9990	
Sum of Squared Residuals			0.790758
Root Mean Squared Error		0.154798	
End of Period Levels:	Mean	9.969970	

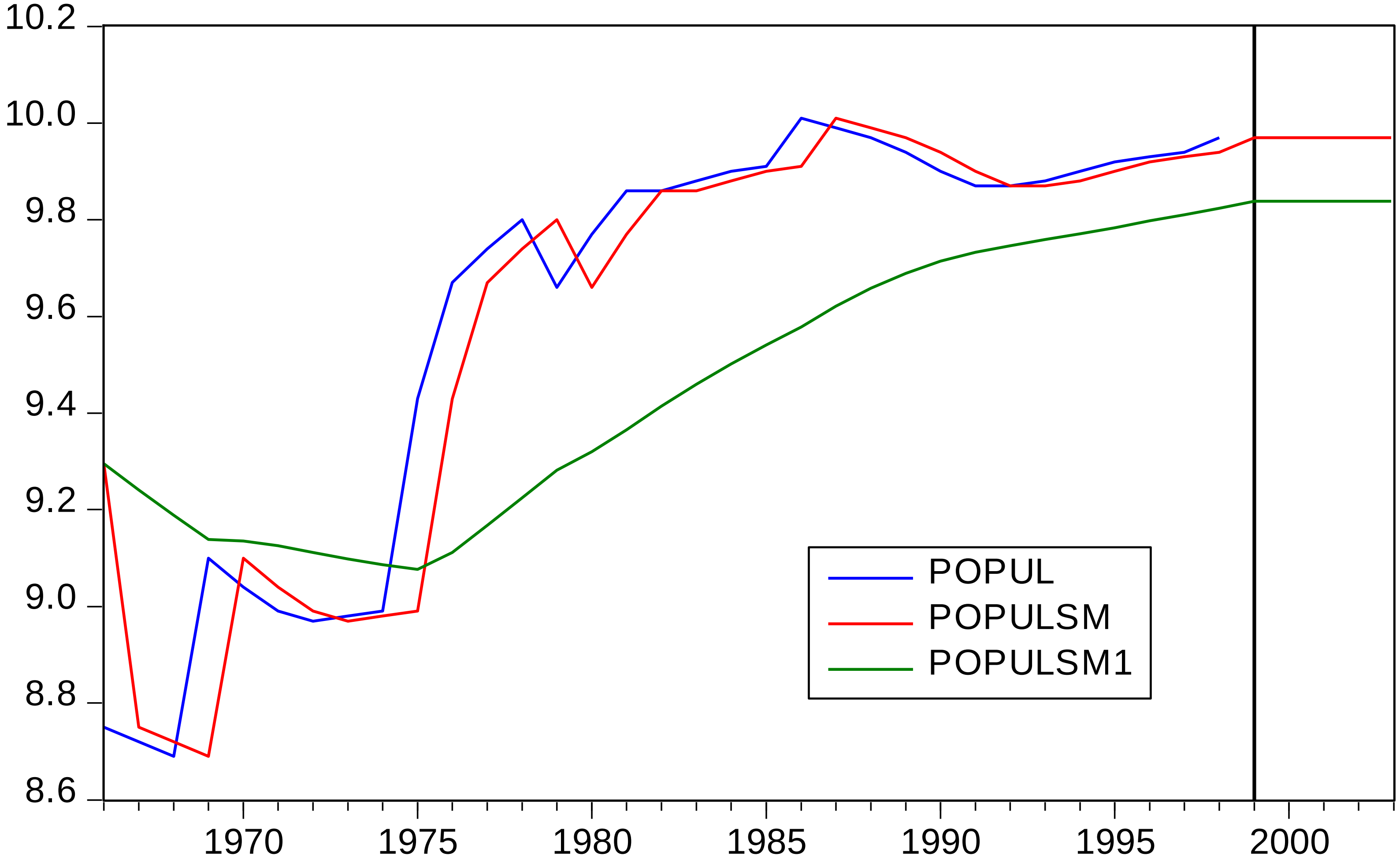
Optimal α
RMSE=0.155

Date: 08/02/18 Time: 21:46
Sample: 1966 1998
Included observations: 33
Method: Single Exponential
Original Series: POPUL
Forecast Series: POPULSM1

Parameters:	Alpha	0.1000	
Sum of Squared Residuals			3.902201
Root Mean Squared Error		0.343873	
End of Period Levels:	Mean	9.838343	

α=0.1
RMSE = 0.344

Time Series Analysis



Time Series Analysis

Double Smoothing

It is appropriate for series with a linear trend

The double smoothing method involves the following recursion equations:

$$S_t = \alpha Y_t + (1-\alpha) S_{t-1}$$
$$D_t = \alpha S_t + (1-\alpha) D_{t-1}$$

where S_t is the single smoothed series, D_t is the double smoothed series, and $0 \leq \alpha \leq 1$ is the smoothing parameter. Forecasts are computed as:

$$\hat{Y}_{T+k} = a_T + k b_T$$

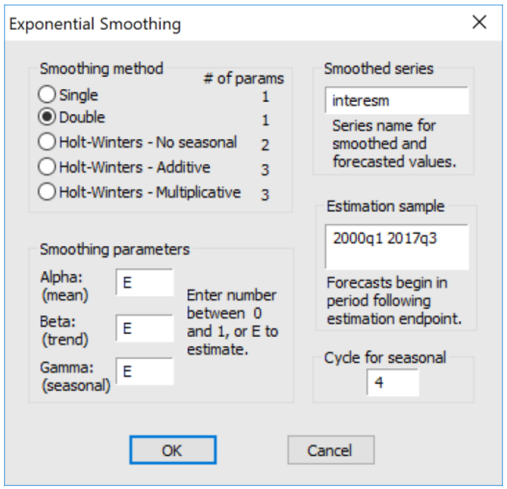
where $a_T = 2S_t - D_t$ and $b_T = (S_t - D_t)\alpha/(1-\alpha)$.

Initialization: $b_1 = (\sum_{t=m+1}^{2m} Y_t - \sum_{t=1}^{m} Y_t)/m^2$ (where m is an arbitrary number of observations) and $a_1 = (\sum_{t=1}^{m} Y_t)/m - b_1 \times (m+1)/2$.

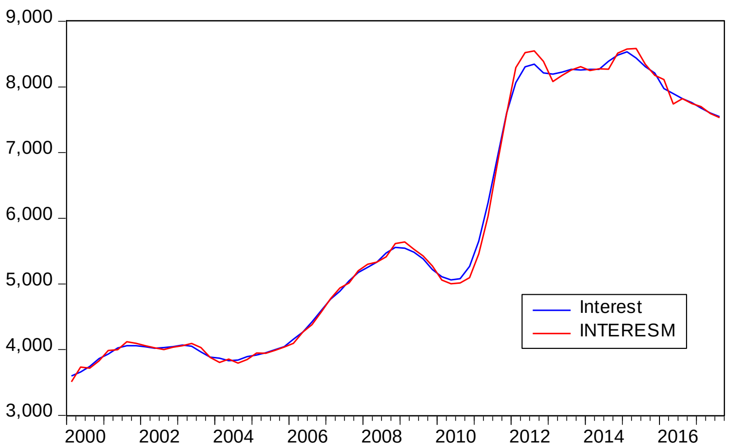
Optimization: We choose the smoothing parameter α by minimizing the sum of squares of one-step-ahead forecast errors.

Time Series Analysis

Example: To illustrate the application of double smoothing method, we forecast data on interest rate in Portugal from 2000q1 to 2017q3.



Sample: 2000Q1 2017Q3				
Included observations: 71				
Method: Double Exponential				
Original Series: INTEREST				
Forecast Series: INTERESM				
Parameters:	Alpha			0.9990
Sum of Squared Residuals				481412.8
Root Mean Squared Error				82.34356
End of Period Levels:		Mean		7548.200
		Trend		-56.83361



Time Series Analysis

LISBON SCHOOL OF ECONOMICS & MANAGEMENT
UNIVERSIDADE DE LISBOA

Holt's Linear Trend

Holt (1957) extended simple exponential method to allow forecasting of data with a linear time trend (and no seasonal variation).

$$a_t = \alpha Y_t + (1-\alpha)(a_{t-1} + b_{t-1})$$
$$b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1}$$

$$\hat{Y}_{t+k} = a_t + kb_t$$

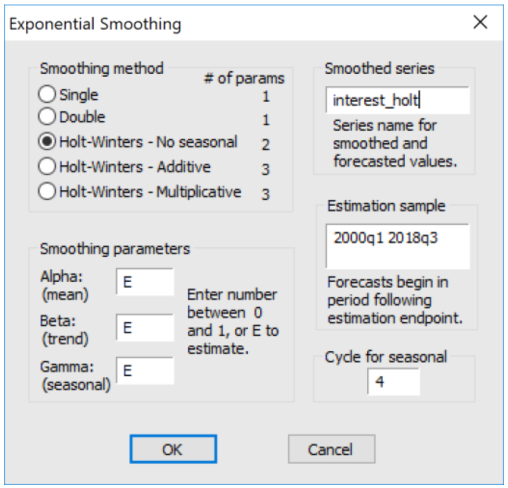
where a_t denotes the level of the series at time *t*; b_t denotes the trend (or slope) of the series at time *t*; $0 \leq \alpha \leq 1$ and $0 \leq \beta \leq 1$ are the smoothing parameters.

Initialization: $b_1 = (\sum_{t=m+1}^{2m} Y_t - \sum_{t=1}^{m} Y_t)/m^2$ (where m is an arbitrary number of observations) and $a_1 = (\sum_{t=1}^{m} Y_t)/m - b_1 \times (m+1)/2$.

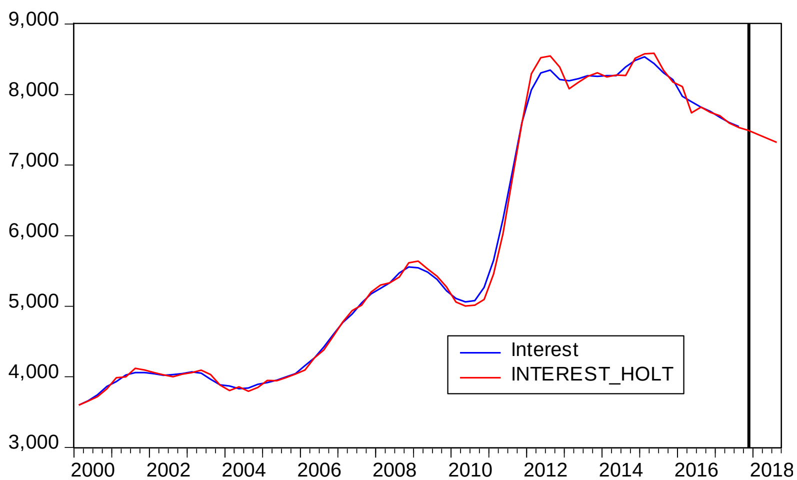
Optimization: We choose the smoothing parameter α by minimizing the sum of squares of one-step-ahead forecast errors.

Time Series Analysis

Example: To illustrate the application of Holt's linear trend method, we forecast again data on interest rate in Portugal from 2000q1 to 2017q3.



Sample: 2000Q1 2017Q3			
Included observations: 71			
Method: Holt-Winters No Seasonal			
Original Series: INTEREST			
Forecast Series: INTEREST_HOLT			
Parameters:	Alpha		1.0000
	Beta		1.0000
Sum of Squared Residuals			466952.1
Root Mean Squared Error			81.09741
End of Period Levels:		Mean	7548.200
		Trend	-56.80000



Time Series Analysis

LISBON SCHOOL OF ECONOMICS & MANAGEMENT
UNIVERSIDADE DE LISBOA

Holt-Winters additive method

This method is appropriate for series with a linear time trend and additive seasonal variation.

$$a_t = \alpha(Y_t - S_{t-s}) + (1-\alpha)(a_{t-1} + b_{t-1})$$
$$b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1}$$

$$S_t = \gamma(Y_t - a_t) + (1-\gamma)S_{t-s}$$

$$\hat{Y}_{t+k} = a_t + kb_t + S_{t+k-s}$$

where a_t denotes the level of the series at time t; b_t denotes of the trend (or slope) of the series at time t; S_t denotes the seasonal factor of the series, and s denotes the number of seasons in a year; $0 \leq \alpha \leq 1$, $0 \leq \beta \leq 1$ and $0 \leq \gamma \leq 1$ are the smoothing parameters.

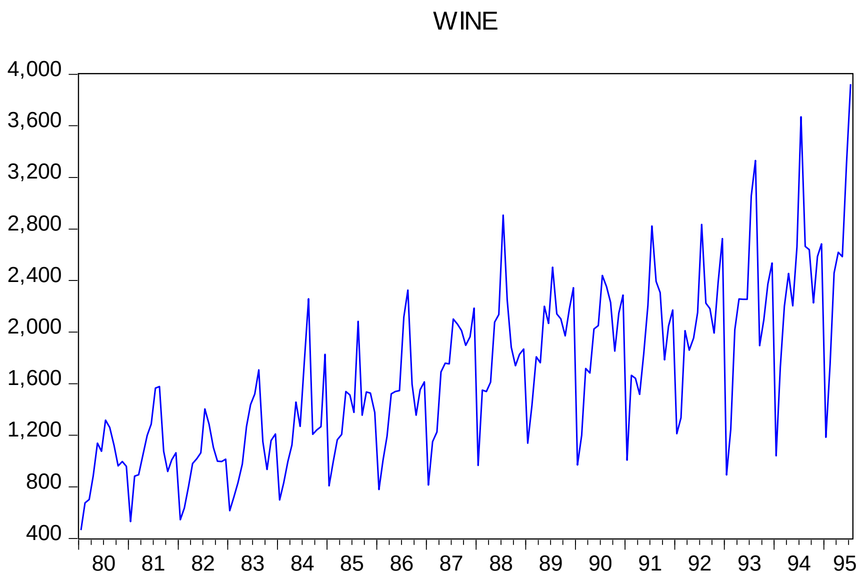
Initialization: A common approach is to set

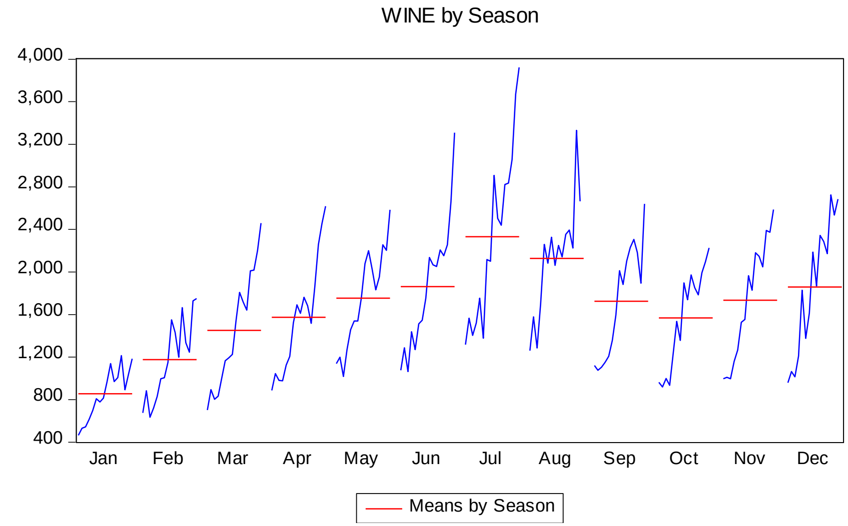
$b_s = (\sum_{t=s+1}^{2s} Y_t - \sum_{t=1}^{s} Y_t)/s^2$, $a_s = (\sum_{t=1}^{s} Y_t)/s$ and $S_i = Y_i - a_s, i = 1,2,\dots,s$

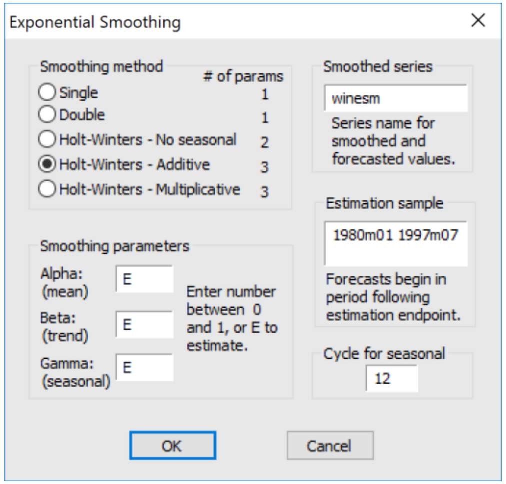
Optimization: We choose the smoothing parameters (α, β and γ) by minimizing the sum of squares of one-step-ahead forecast errors.

Time Series Analysis

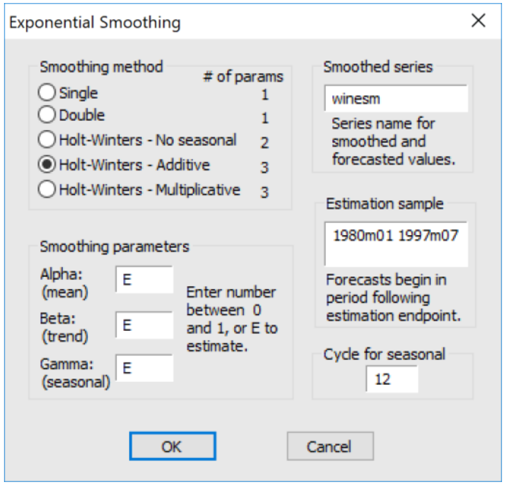
Example: We employ the Holt-Winters with additive seasonality to forecast wine consumption in Australia for the period 1980m1 to 1995m7.

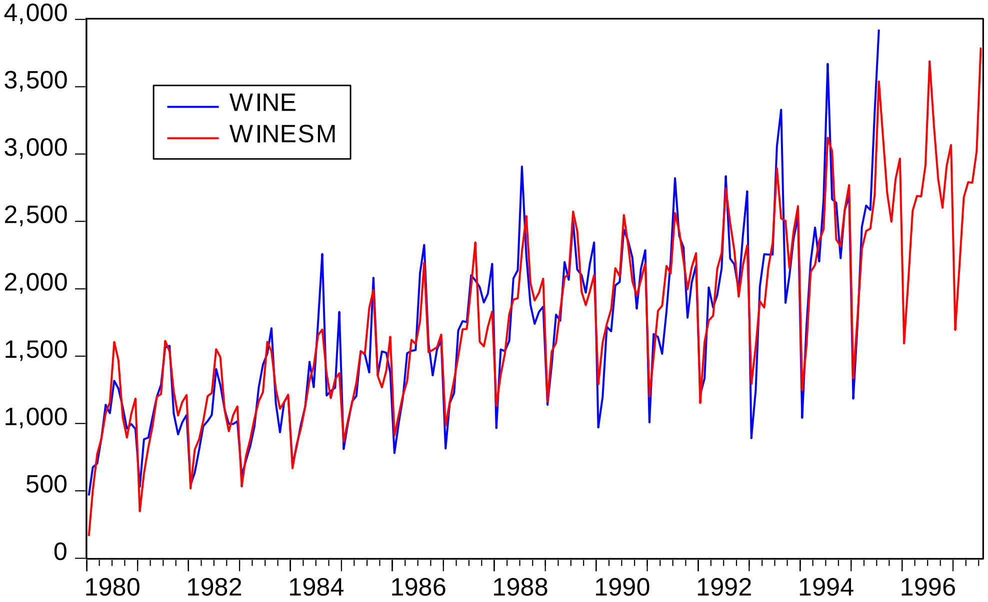






Time Series Analysis





Date: 10/25/18 Time: 22:17
Sample: 1980M01 1995M07
Included observations: 187
Method: Holt-Winters Additive Seasonal
Original Series: WINE
Forecast Series: WINESM

Parameters:	Alpha	0.1200
	Beta	0.0000
	Gamma	0.4001
Sum of Squared Residuals		8078709.
Root Mean Squared Error		207.8501

End of Period Levels:	Mean		2637.784
	Trend		8.530258
	Seasonals:	1994M08	455.5493
		1994M09	55.46994
		1994M10	-164.2915
		1994M11	141.5148
		1994M12	285.5265
		1995M01	-1095.073
		1995M02	-622.1864
		1995M03	-128.7721
		1995M04	-24.76443
		1995M05	-38.46030
		1995M06	187.3407
		1995M07	948.1462

Time Series Analysis

Holt-Winters multiplicative method

This method is appropriate for series with a linear time trend and multiplicative seasonal variation.

$$a_t = \alpha(Y_t/S_{t-s}) + (1-\alpha)(a_{t-1} + b_{t-1})$$
$$b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1}$$

$$S_t = \gamma(Y_t/a_t) + (1-\gamma)S_{t-s}$$

$$\hat{Y}_{t+k} = (a_t + kb_t)S_{t+k-s}$$

where a_t denotes the level of the series at time *t*; b_t denotes of the trend (or slope) of the series at time *t*; S_t denotes the seasonal factor of the series, and s denotes the number of seasons in a year; $0 \leq \alpha \leq 1$, $0 \leq \beta \leq 1$ and $0 \leq \gamma \leq 1$ are the smoothing parameters.

Initialization: A common approach is to set

$b_s = (\sum_{t=s+1}^{2s} Y_t - \sum_{t=1}^{s} Y_t)/s^2$, $a_s = (\sum_{t=1}^{s} Y_t)/s$ and $S_i = Y_i/a_s, i = 1,2, \ldots, s$

Optimization: We choose the smoothing parameters (α, β and γ) by minimizing the sum of squares of one-step-ahead forecast errors.

Time Series Analysis

Example: EViews file data-forecasting.wk1 contains the monthly air passengers in the US (page: passengers) for the period 1949m1 to 1960m12.

a) Plot the series and describe the main features of the series
b) Forecast the next two years using Holt-Winters multiplicative method.
c) Forecast the next two years using Holt-Winters additive method.
d) Report and compare the RMSE of the one-step ahead forecasts from the two approaches.

Linear Time Series Models

Stationarity

Definitions:

A **stochastic process** is a family of time indexed random variables, $Z(w,t)$: $t=0,\pm1,\pm2,\ldots$, where w is the sample space and t is the index set.

A time series is a realization (or sample function) from a certain stochastic process, Yt, $t=1,2,\ldots,n$.

A process Yt, $t=1,2,\ldots,n$ is said to be **weakly stationary** if it has constant mean, constant variance, and the covariance and the correlation between Yt and $Yt+k$ depend only on time difference k.

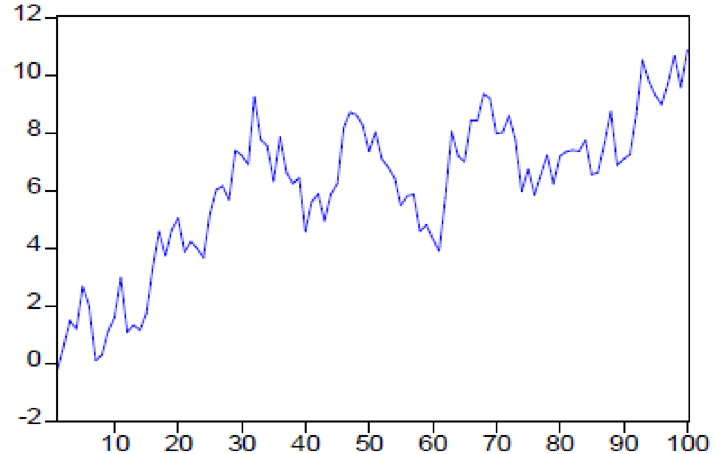
$$\mu_t = E(Y_t) = \mu \,,$$

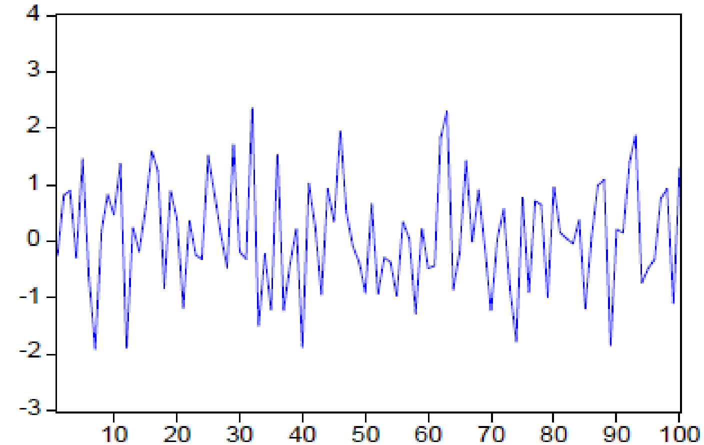
$$\sigma_t^2 = Var(Y_t) = E(Y_t - \mu_t)^2 = \sigma^2 \,,$$

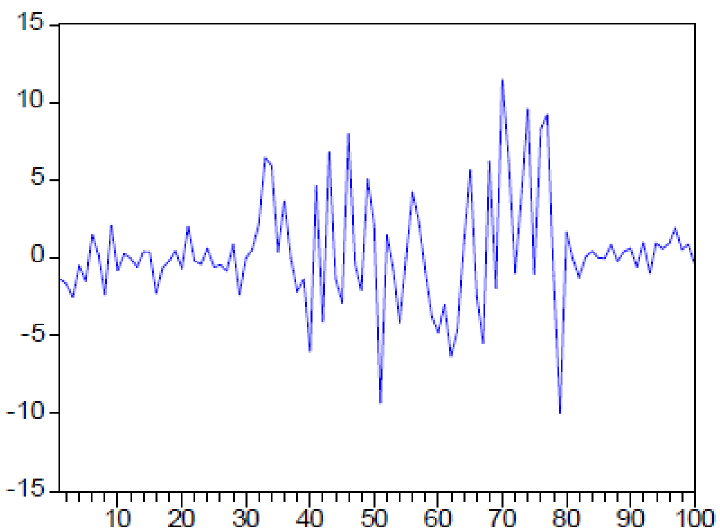
$$\gamma(t_1,t_2) = E(Y_{t_1} - \mu_{t_1})(Y_{t_2} - \mu_{t_2}) = \gamma(t_1+k,t_2+k)\,,\ \ \forall t_1,t_2,k$$

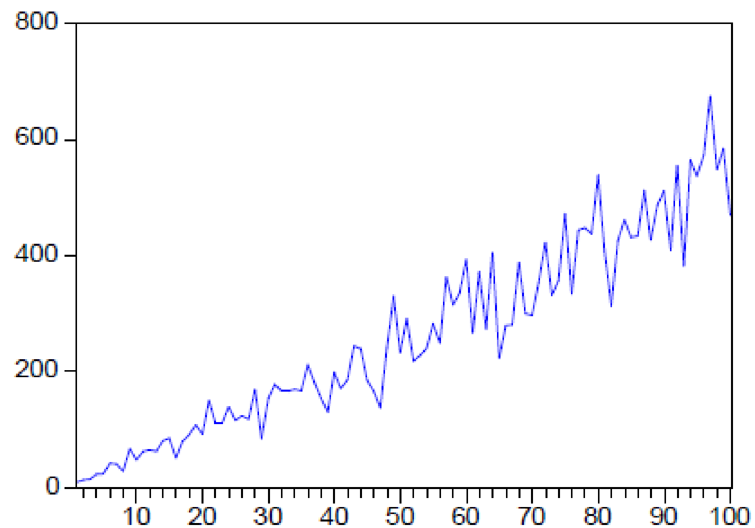
$$\rho(t_1,t_2) = \frac{\gamma(t_1,t_2)}{\sqrt{\sigma_{t_1}^2}\sqrt{\sigma_{t_2}^2}} = \rho(t_1+k,t_2+k)\,,\ \ \forall t_1,t_2,k \,,$$

Stationary?


12
10
8
6
4
2
0
-2
10 20 30 40 50 60 70 80 90 100


4
3
2
1
0
-1
-2
-3
10 20 30 40 50 60 70 80 90 100


15
10
5
0
-5
-10
-15
10 20 30 40 50 60 70 80 90 100


800
600
400
200
0
10 20 30 40 50 60 70 80 90 100

Autocorrelation

The **autocovariance function (ACOVF) and autocorrelation function (ACF)** represent the covariance and correlation between *Yt* and *Yt+k*, from the same process *Y* separated only by *k* time lags.

$$\gamma_k = Cov(Y_t, Y_{t+k}) = E[(Y_t - \mu)(Y_{t+k} - \mu)] \qquad \rho_k = \frac{Cov(Y_t, Y_{t+k})}{\sqrt{[Var(Y_t)][Var(Y_{t+k})]}} = \frac{\gamma_k}{\gamma_0}$$

The autocovariance function and the autocorrelation function have the following properties:

1) $\gamma_0 = Var(Y_t);\ \rho_0 = 1;$

2) $|\gamma_k| \leq \gamma_0;\ |\rho_k| \leq 1;$

3) $\gamma_k = \gamma_{-k};\ \rho_k = \rho_{-k}$ for all k;

Partial autocorrelation

The **partial autocorrelation function (PACF)** measures the correlation between Y_t and Y_{t-k}, when the effects of intervening variables Y_{t-1}, Y_{t-2}, ..., Y_{t-k+1} are removed. The partial autocorrelation coefficient of order k is denoted by ϕ_{kk} and can be derived by regressing Y_{t+k} against Y_{t+k-1}, Y_{t+k-2}, ..., Y_t:

$$Y_{t+k} = \phi_{k1} Y_{t+k-1} + \phi_{k2} Y_{t+k-2} + \cdots + \phi_{kk} Y_t + e_{t+k} .$$

Multiplying Y_{t+k-j} on both sides of the equation and taking expected values, we get

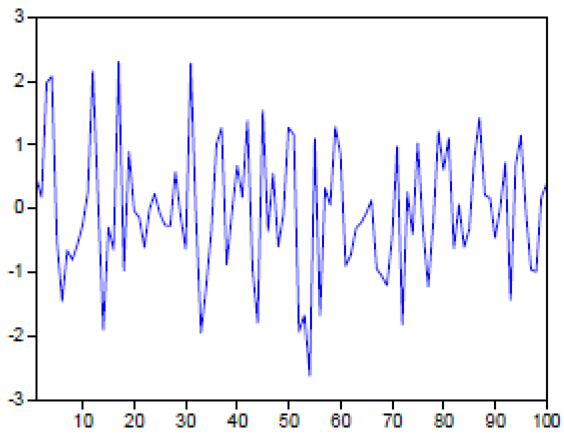
$$\phi_{11} = \rho_1, \quad \phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}, \quad \phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}, \ \ldots, \ \phi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & 1 \end{vmatrix}}$$

White noise

A process is called a "white noise" process if it is a sequence of uncorrelated random variables:

$$Y_t = \varepsilon_t ,$$

where ε_t has constant mean $E(\varepsilon_t) = \mu_\varepsilon$ (usually assumed to be 0) , constant variance $Var(\varepsilon_t) = \sigma_\varepsilon^2$ and null covariance $Cov(\varepsilon_t, \varepsilon_{t-k}) = 0$ for all $k \neq 0$. The ACF and PACF of a white noise process are null for all $k \neq 0$.



Simulation of a white noise process with zero mean and unit variance

Sample ACF and PACF

For a given observed time series, $Y_t, t = 1,2,\ldots,n$, the sample autocorrelation function (ACF) is defined as

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=k+1}^{n}\left(Y_t - \overline{Y}\right)\left(Y_{t-k} - \overline{Y}\right)}{\sum_{t=1}^{n}\left(Y_t - \overline{Y}\right)^2}, \quad k = 0,1,2,\ldots$$

The sample partial autocorrelation function (PACF) is obtained by a recursive method as follows:

$$\hat{\phi}_{kk} = \frac{\hat{\rho}_k - \sum_{j=1}^{k-1}\hat{\phi}_{k-1,j}\hat{\rho}_{k-j}}{1 - \sum_{j=1}^{k-1}\hat{\phi}_{k-1,j}\hat{\rho}_j},$$

with $\hat{\phi}_{11} = \hat{\rho}_1$ and $\hat{\phi}_{kj} = \hat{\phi}_{k-1,j} - \hat{\phi}_{kk}\hat{\phi}_{k-1,k-j}$, $j = 1,2,\ldots,k-1$.

Backshift notation

LISBON SCHOOL OF ECONOMICS & MANAGEMENT
UNIVERSIDADE DE LISBOA

A very useful notation in time series analysis is the backshift operator *B*, which is used as follows:

$$BY_t = Y_{t-1}.$$

In other words, *B* has the effect of shifting the data back one period.
For *k* periods, the notation is

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

For monthly data, B^{12} is used to shift attention to the same month last year, $B^{12}Y_t = Y_{t-12}$.

For quarterly data, the backshift operator is used as follows: $B^4 Y_t = Y_{t-4}$.

MA(∞) representation

The process Y_t can be expressed as a linear combination of a sequence of uncorrelated random variables:

$$Y_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},$$

where $\psi_0 = 1$, ε_t is a zero mean white noise with constant variance and $\sum_{j=0}^{\infty} \psi_j^2 < \infty$.

It can be shown that

$$E(Y_t) = 0, \quad Var(Y_t) = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_j^2, \quad E(\varepsilon_t Y_{t-k}) = \begin{cases} \sigma_\varepsilon^2, & k = 0 \\ 0, & k > 0, \end{cases}$$

$$\text{and } \rho_k = \frac{\gamma_k}{\gamma_0} = \frac{E(Y_t Y_{t+k})}{Var(Y_t)} = \frac{\sum_{j=0}^{\infty} \psi_j \psi_{j+k}}{\sum_{j=0}^{\infty} \psi_j^2}$$

AR(∞) representation

Another useful form is to write Y_t in an autoregressive representation, as follows:

$$Y_t = \pi_1 Y_{t-1} + \pi_2 Y_{t-2} + \cdots + \varepsilon_t = \sum_{j=1}^{\infty} \pi_j Y_{t-j} + \varepsilon_t,$$

or, equivalently,

$$\pi(B) Y_t = \varepsilon_t,$$

where $\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \cdots = 1 - \sum_{j=1}^{\infty} \pi_j B^j$ and $1 + \sum_{j=1}^{\infty} |\pi_j| < \infty$.

Autoregressive models

The finite-order representation of the autoregressive model described earlier, if only a finite number of π weights are nonzero, is given by

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t ,$$

where ε_t is a zero mean white noise series. Because $\sum_{j=1}^{\infty} |\pi_j| = \sum_{j=1}^{p} |\phi_j| < \infty$, the process is always invertible. To be stationaty, the roots of $(1 - \phi_1 B - \cdots - \phi_p B^p) = 0$ must be outside of the unit circle.

AR(1) model

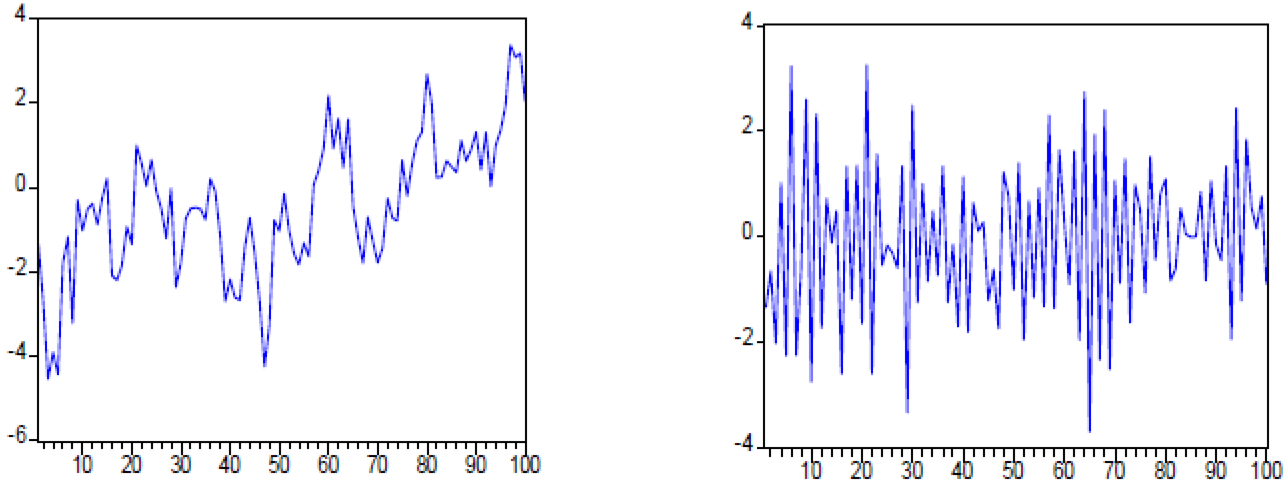
The first-order autoregressive model or AR(1) model is given by

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

where ε_t is a zero mean white noise series. The model is always invertible. To be stationary, the roots of $(1 - \phi B) = 0$ must be outside of the unit circle. Because the root $B = 1/\phi$, for a stationary model, we have $|\phi| < 1$.

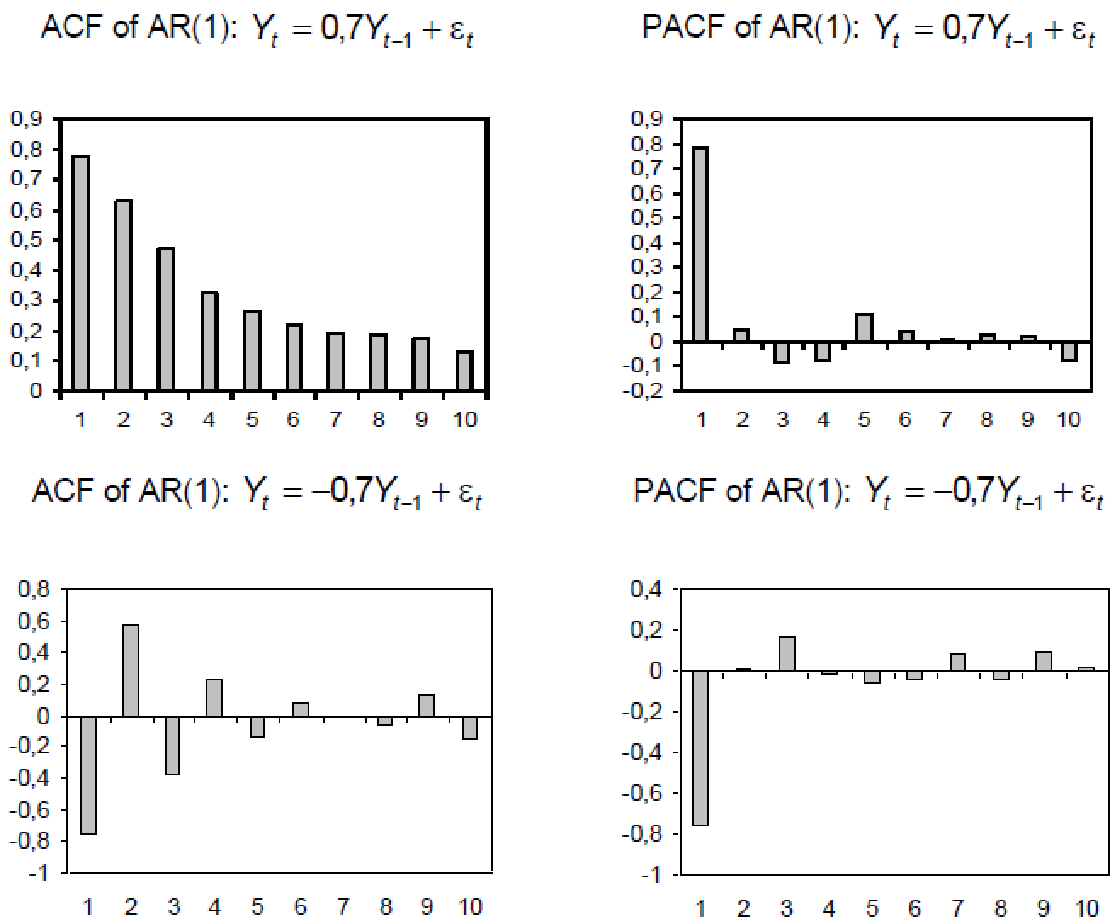
Autoregressive models

(a) $Y_t = 0{,}7Y_{t-1} + \varepsilon_t$ (b) $Y_t = -0{,}7Y_{t-1} + \varepsilon_t$



Simulated AR(1) models with $\phi = 0.7$ and $\phi = -0.7$

Autoregressive models



ACF and PACF of the simulated AR(1) models

Autoregressive models

AR(2) model
The second-order autoregressive AR(2) models is

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t ,$$

or

$$\phi_2(B) Y_t = \varepsilon_t ,$$

where ε_t is a zero mean white noise series. To be stationary, the roots of $\phi_2(B) = 1 - \phi_1 B - \phi_2 B^2 = 0$ must be outside of the unit circle. Thus, we have the following necessary and sufficient conditions for stationarity:

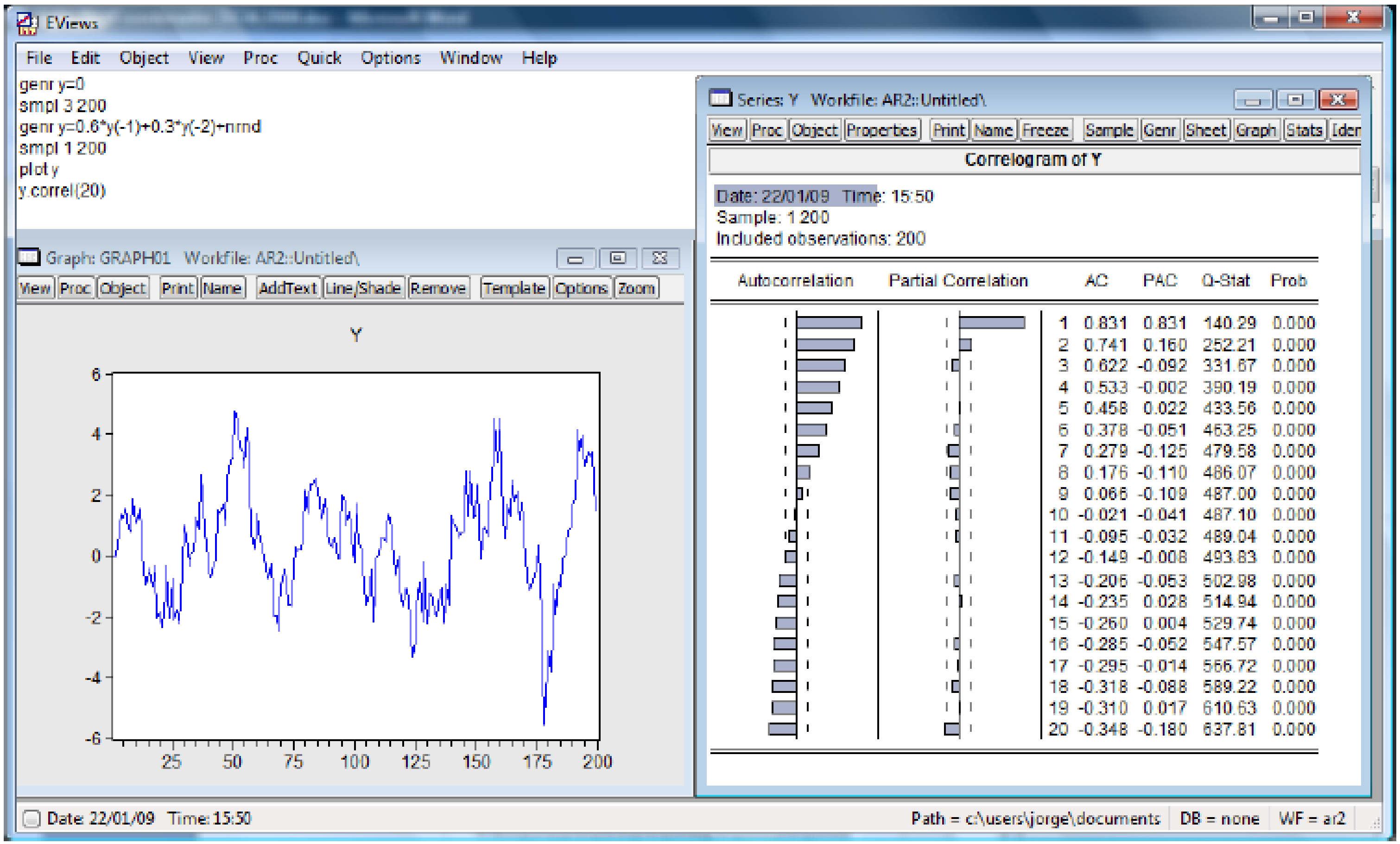
$$\phi_2 + \phi_1 < 1 \wedge \phi_2 - \phi_1 < 1 \wedge -1 < \phi_2 < 1.$$

The ACF tails off as an exponential decay or a damped sine waves depending on the roots of $\phi(B) = 0$, and the PACF cuts off after lag 2, $\phi_{kk} = 0$ for $k \geq 3$.

AR(p) model
More complicated conditions hold for AR(p) models with $p \geq 3$.
Econometric software (EViews among others) takes care of this.

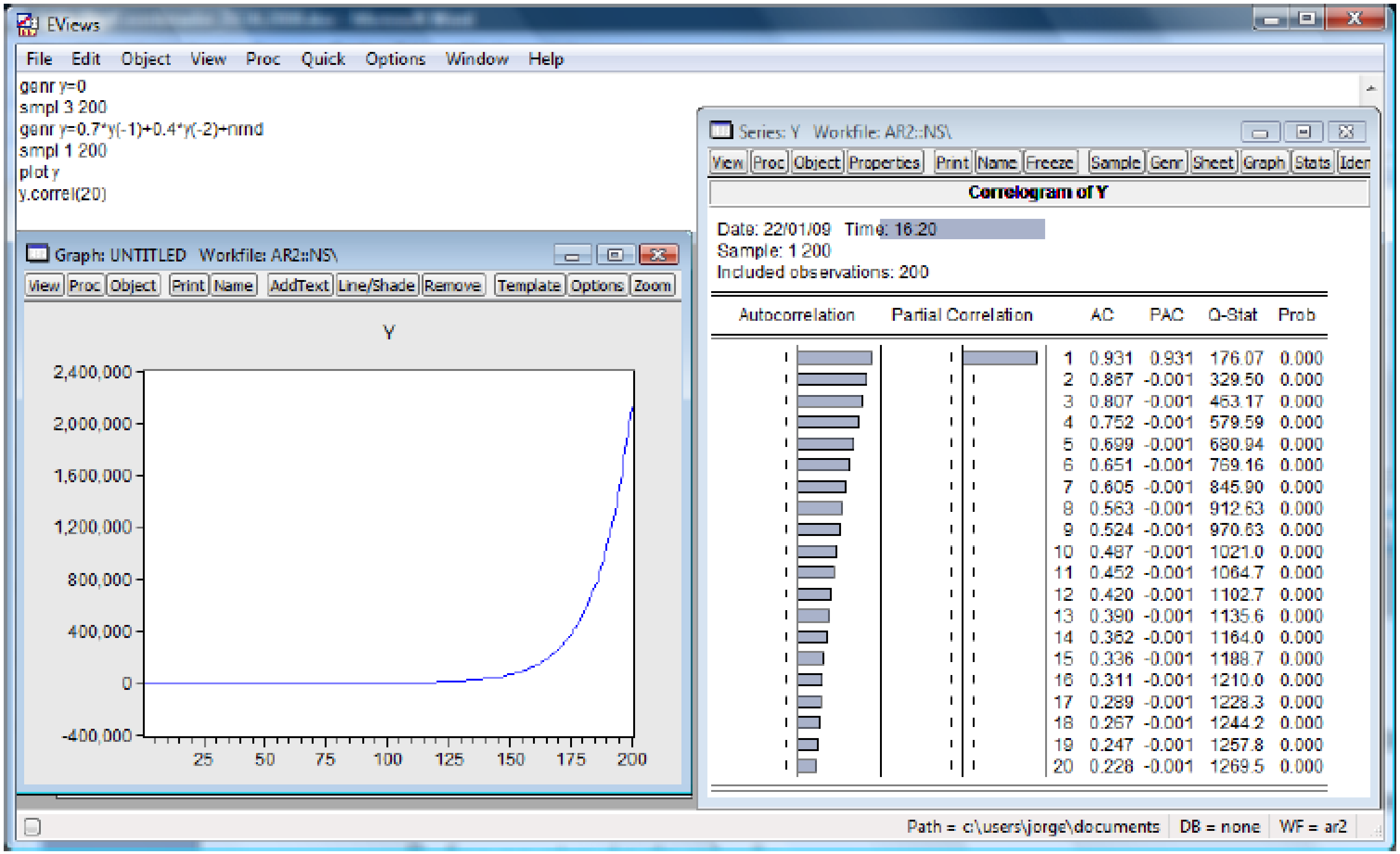
Autoregressive models

LISBON SCHOOL OF ECONOMICS & MANAGEMENT
UNIVERSIDADE DE LISBOA



ACF and PACF of a simulated stationary AR(2) model: $Y_t = 0{,}6Y_{t-1} + 0{,}3Y_{t-2} + \varepsilon_t$

Autoregressive models



ACF and PACF of a simulated nonstationary AR(2) model: $Y_t = 0{,}7Y_{t-1} + 0{,}4Y_{t-2} + \varepsilon_t$

Moving average models

The finite-order representation of the moving average model described earlier, if only a finite number of ψ weights are nonzero, is given by

$$Y_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q},$$

where ε_t is a zero mean white noise series. Because $1 + \theta_1^2 + \cdots + \theta_q^2 < \infty$, the process is always stationary. To be invertible, the roots of $(1 - \theta_1 B - \cdots - \theta_q B^q) = 0$ must be outside of the unit circle.

MA(1) model

The first-order moving average model or MA(1) model is

$$Y_t = \varepsilon_t - \theta_1\varepsilon_{t-1},$$

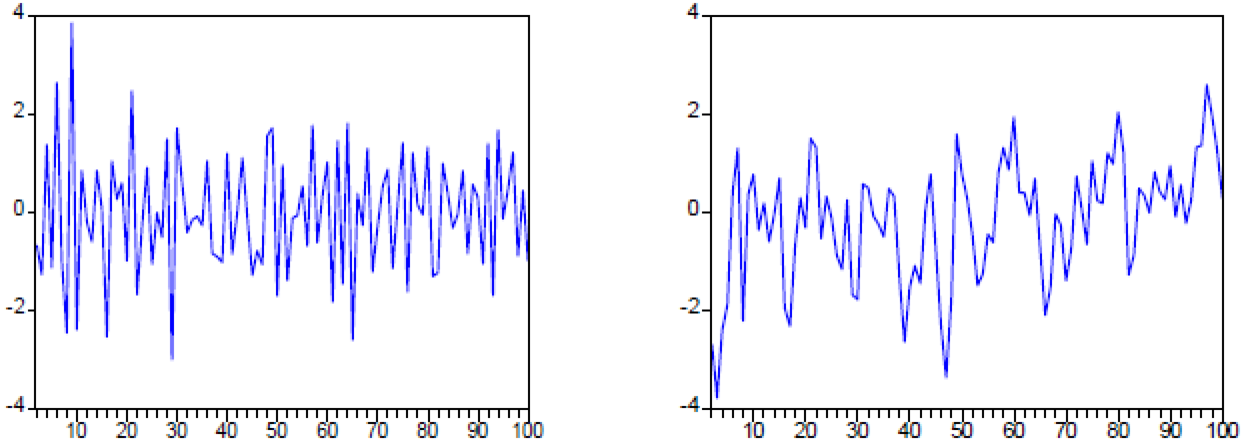
or

$$Y_t = \theta(B)\varepsilon_t,$$

where $\theta(B) = 1 - \theta_1 B$ and ε_t is white noise. To be invertible, the root of $\theta(B) = 0$ must lie outside the unit circle. Thus, we require $|\theta_1| < 1$.

Moving average models

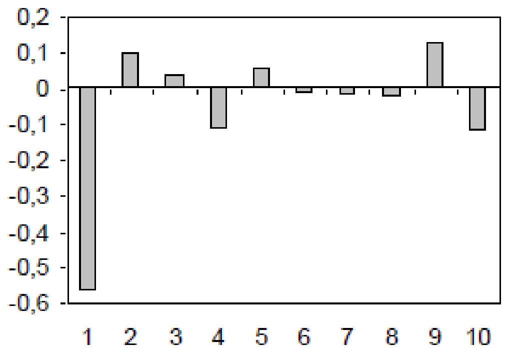
(a) $Y_t = \varepsilon_t - 0.75\varepsilon_{t-1}$ (b) $Y_t = \varepsilon_t + 0.75\varepsilon_{t-1}$



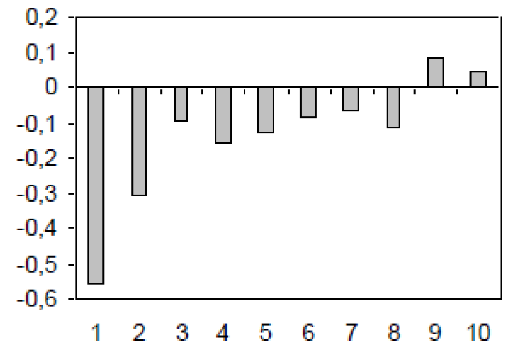
Simulated MA(1) models

Moving average models

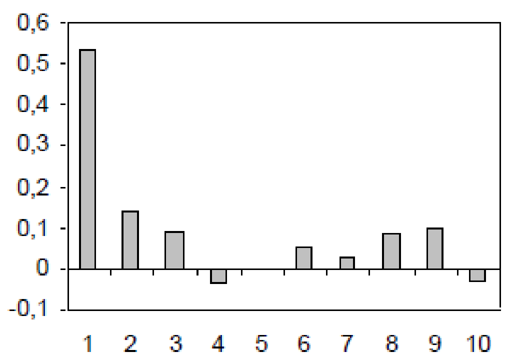
ACF of MA(1): $Y_t = \varepsilon_t - 0.75\varepsilon_{t-1}$



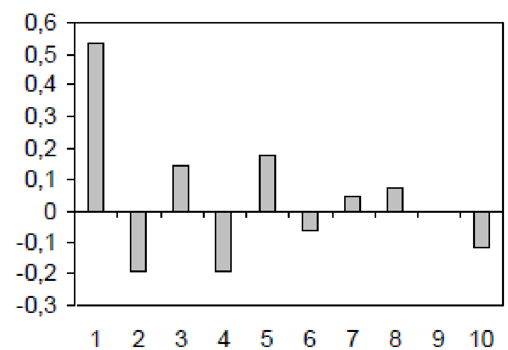
PACF of MA(1): $Y_t = \varepsilon_t - 0.75\varepsilon_{t-1}$



ACF of MA(1): $Y_t = \varepsilon_t + 0.75\varepsilon_{t-1}$



PACF of MA(1): $Y_t = \varepsilon_t + 0.75\varepsilon_{t-1}$



ACF and PACF of simulated MA(1) models

Moving average models

MA(2) model

The second-order moving average MA(2) model is given by

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2},$$

or

$$Y_t = \theta_2(B)\varepsilon_t,$$

where $\theta_2(B) = 1 - \theta_1 B - \theta_2 B^2$ and ε_t is white noise. To be invertible, the roots of $\theta_2(B) = 0$ must lie outside the unit circle. Hence, we have the following conditions:

$$\theta_2 + \theta_1 < 1 \wedge \theta_2 - \theta_1 < 1 \wedge -1 < \theta_2 < 1.$$

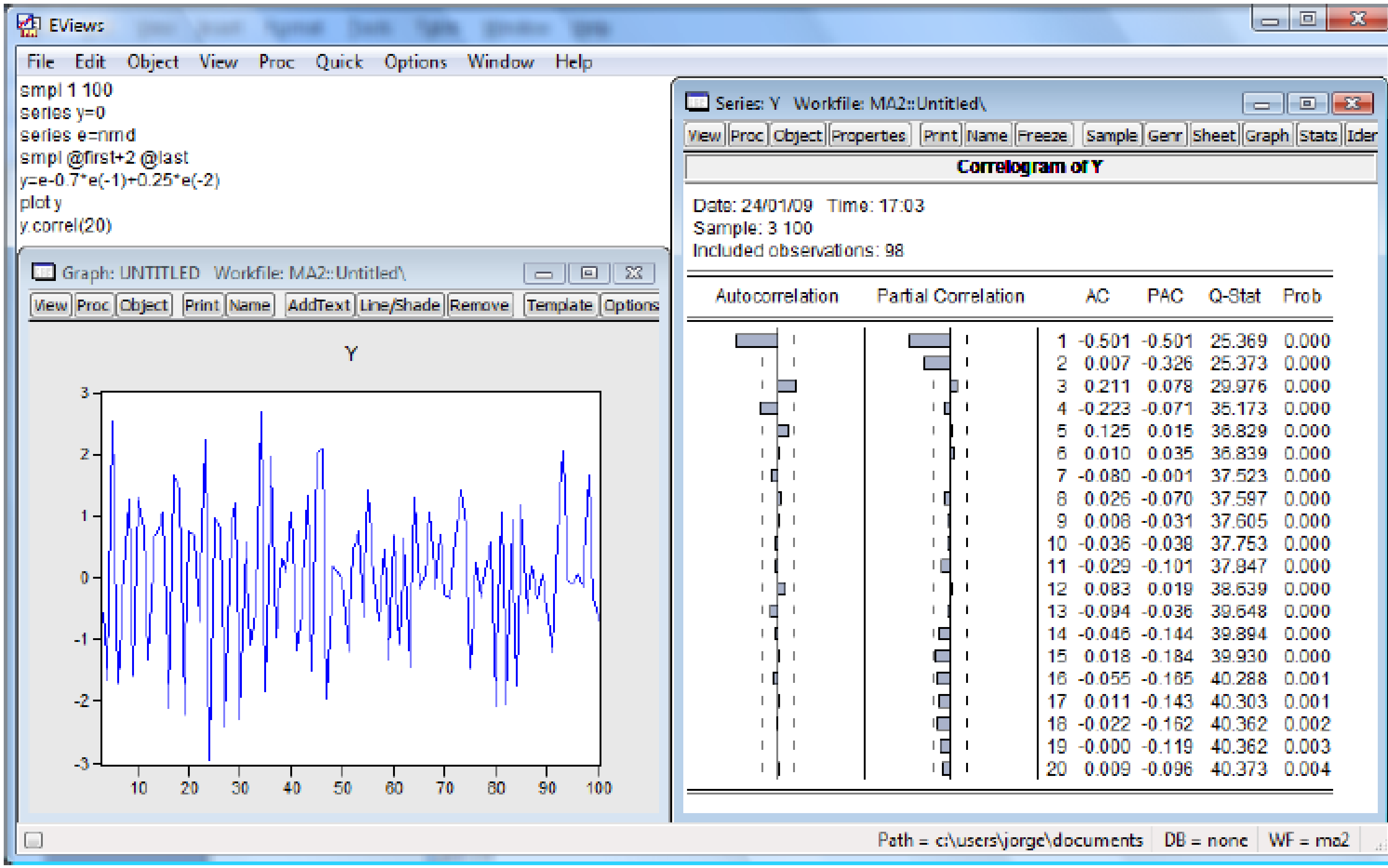
ACF of the MA(2) model cuts off after lag 2 and PACF tails off as an exponential decay or a damped sine wave depending on the roots of $\theta_2(B) = 0$.

MA(q) model

More complicated conditions hold for MA(q) models with $q \geq 3$.

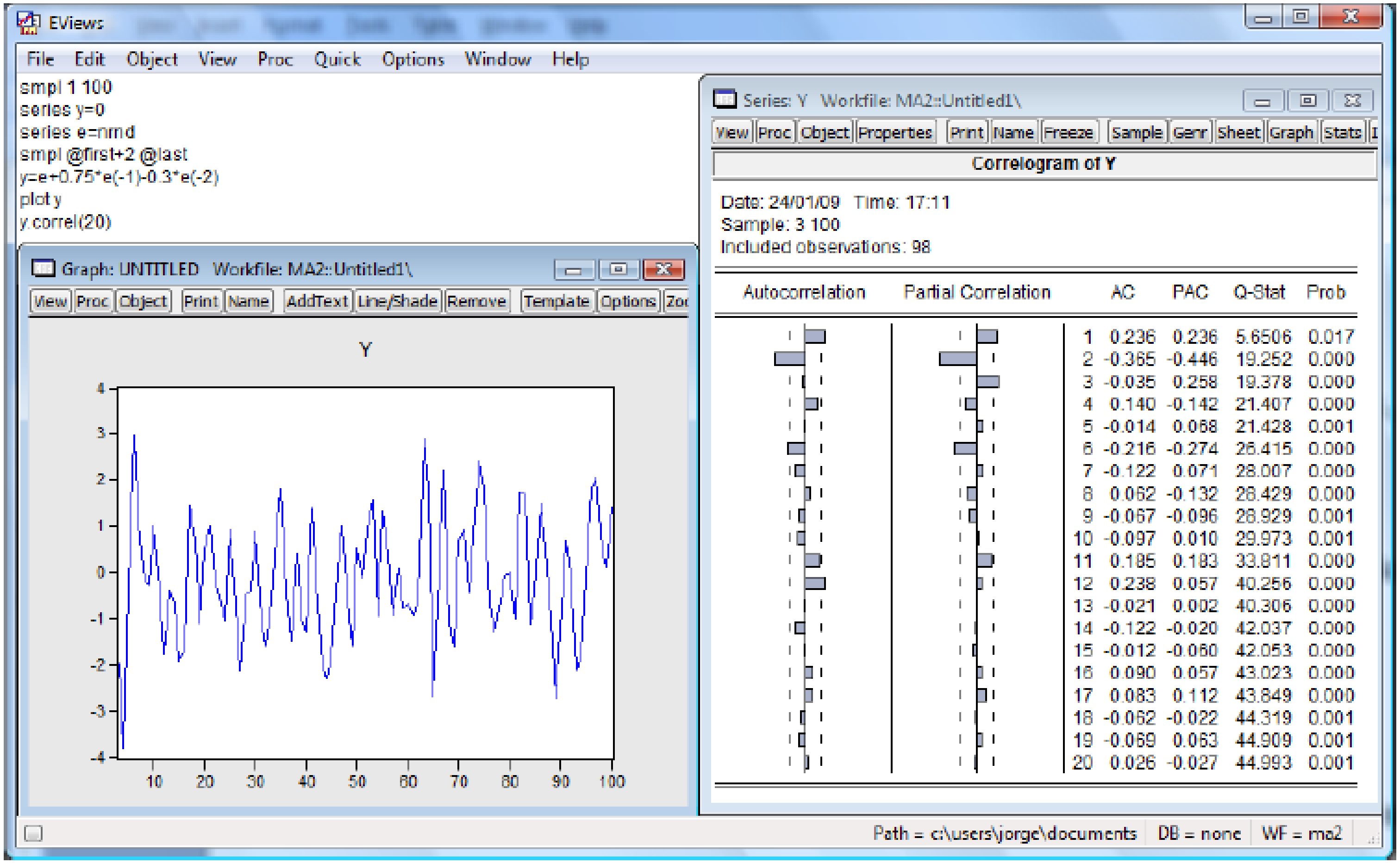
Econometric software (EViews among others) takes care of this.

Moving average models



ACF and PACF of the simulated MA(2) model: $Y_t = \varepsilon_t - 0{,}7\varepsilon_{t-1} + 0{,}25\varepsilon_{t-2}$

Moving average models



ACF and PACF of the simulated MA(2) model: $Y_t = \varepsilon_t + 0{,}75\varepsilon_{t-1} - 0{,}2\varepsilon_{t-2}$

Autoregressive and moving average models

ARMA(1,1) model

The mixed autoregressive and moving average ARMA(1,1) model includes the autoregressive AR(1) and moving average MA(1) models as special cases.

$$Y_t = \phi Y_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1},$$

or

$$\phi(B)Y_t = \theta(B)\varepsilon_t\,,$$

where $\phi(B) = 1 - \phi B$, $\theta(B) = 1 - \theta B$ and ε_t is white noise. To be stationary, the root of $\phi(B) = 0$ must lie outside the unit circle, i.e., $-1 < \phi < 1$. To be invertible, the root of $\theta(B) = 0$ must lie outside the unit circle, i.e., $-1 < \theta < 1$.

The ARMA(1,1) model can be written in a pure moving average representation as

$$Y_t = \psi(B)\varepsilon_t,$$

where

$$\psi(B) = (1 + \psi_1 B + \psi_2 B^2 + \cdots) = \frac{1 - \theta B}{1 - \phi B}.$$

Autoregressive and moving average models

LISBON SCHOOL OF ECONOMICS & MANAGEMENT
UNIVERSIDADE DE LISBOA

The ARMA(1,1) model can be written in a pure autoregressive representation as

$$\pi(B)Y_t = \varepsilon_t,$$

where

$$\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \cdots = \frac{1-\phi B}{1-\theta B}.$$

Both the ACF and PACF of a mixed ARMA(1,1) model tail off as k incresaes, with its shape depending on the signs and magnitudes of ϕ and θ.

ARMA(p,q) model

The general mixed autoregressive and moving average ARMA(p,q) model is given by

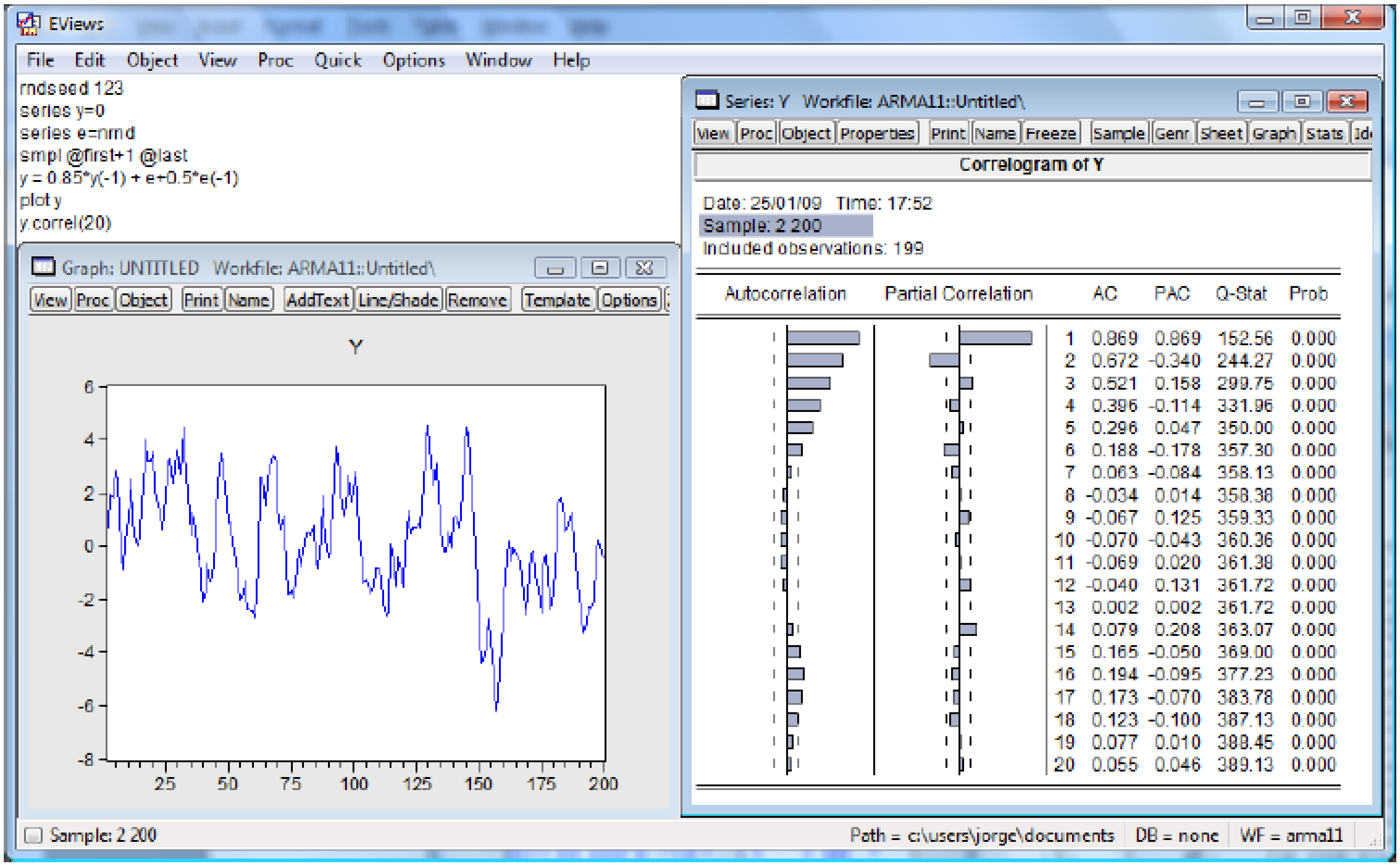
$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q},$$

or

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

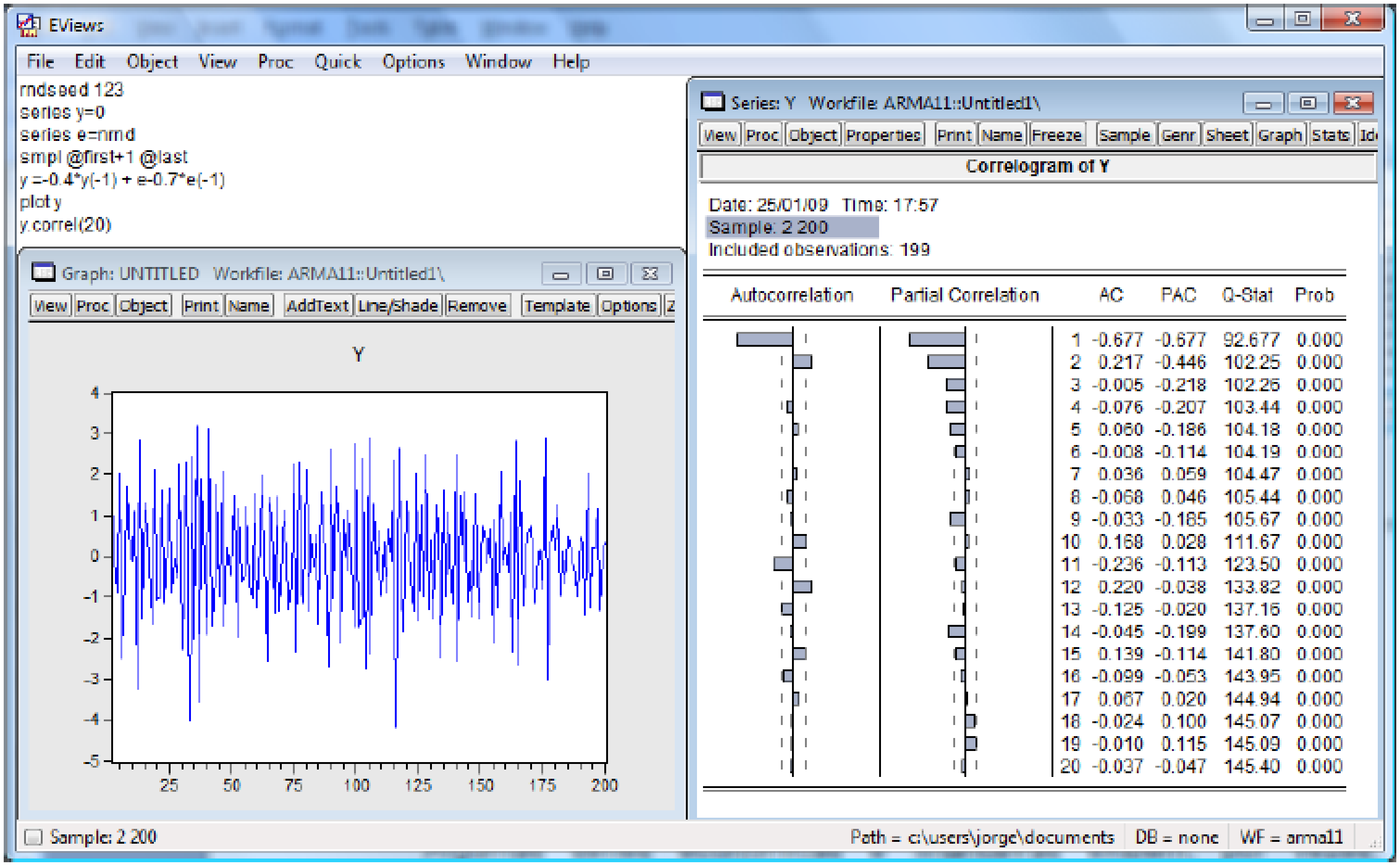
where $\phi_p(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$, $\theta_q(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$ and ε_t is white noise. To be stationary, the roots of $\phi_p(B) = 0$ must lie outside the unit circle. To be invertible, the roots of $\theta_q(B) = 0$ must lie outside the unit circle.

Autoregressive and moving average models



ACF and PACF of the ARMA(1,1) model: $Y_t = 0{,}85Y_{t-1} + \varepsilon_t + 0{,}5\varepsilon_{t-1}$

Autoregressive and moving average models



ACF and PACF of the ARMA(1,1) model: $Y_t = -0{,}4Y_{t-1} + \varepsilon_t - 0{,}7\varepsilon_{t-1}$

Seasonal ARMA models

LISBON SCHOOL OF ECONOMICS & MANAGEMENT
UNIVERSIDADE DE LISBOA

Seasonal autoregressive and moving average SARMA(P,Q)$_s$ model

The seasonal SARMA(P,Q)$_s$ model is represented by

$$Y_t = \Phi_1 Y_{t-s} + \cdots + \Phi_P Y_{t-Ps} + \varepsilon_t - \Theta_1 \varepsilon_{t-s} - \cdots - \Theta_Q \varepsilon_{t-Qs} ,$$

or

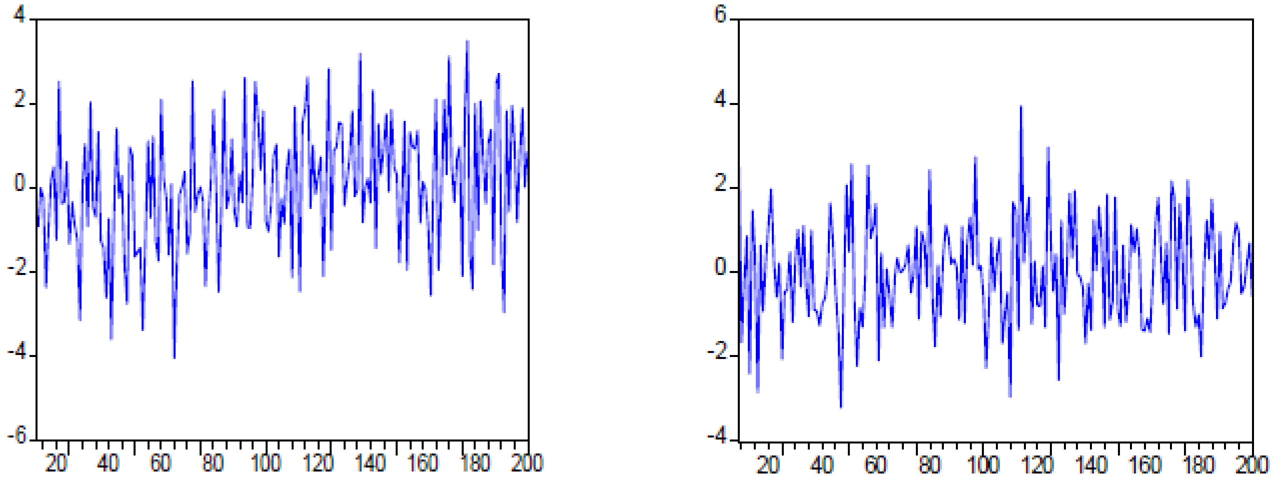
$$\Phi_P(B^s) Y_t = \Theta_Q(B^s) \varepsilon_t ,$$

where $\Phi_P(B^s) = 1 - \Phi_1 B^s - \cdots - \Phi_P B^{Ps}$, $\Theta_Q(B^s) = 1 - \Theta_1 B^s - \cdots - \Theta_Q B^{Qs}$ and ε_t is a zero mean white noise. To be stationary and invertible, the roots of $\Phi_P(B^s) = 0$ e $\Theta_Q(B^s) = 0$ must lie outside of the unit circle, respectively.

Both the ACF and PACF of the SARMA(P,Q)$_s$ model exhibit exponential decays and damped sine waves at the seasonal lags.

Seasonal ARMA models

(i) $(1-0.65B^{12})Y_t = (1+0.25B^{12})\varepsilon_t$ (ii) $(1-0.3B^{4})Y_t = (1-0.4B^{4}+0.15B^{8})\varepsilon_t$

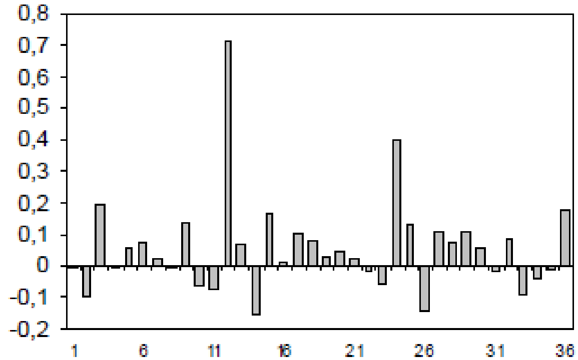


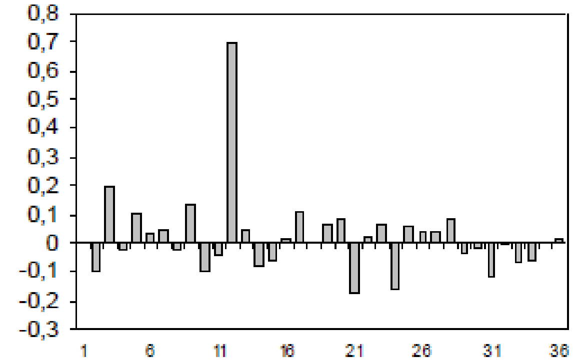
Simulated SARMA$(1,1)_{12}$ and SARMA$(1,2)_4$ models

Seasonal ARMA models

ACF of SARMA(1,1)$_{12}$ PACF of SARMA(1,1)$_{12}$

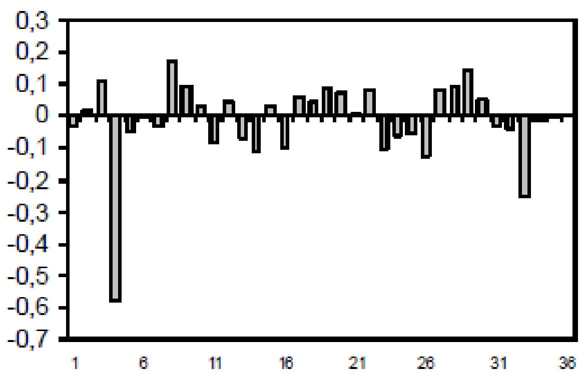
$$(1-0.65B^{12})Y_t = (1+0.25B^{12})\varepsilon_t$$

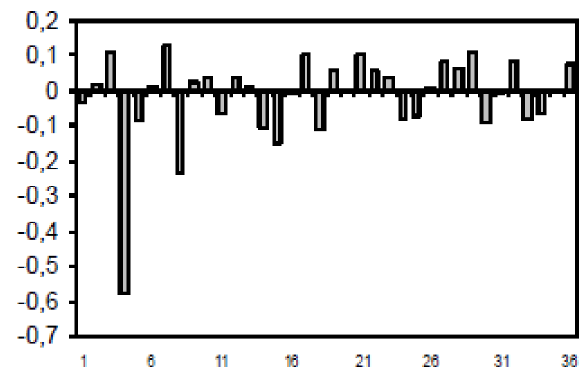




ACF of SARMA(1,2)$_4$ PACF of SARMA(1,2)$_4$

$$(1-0.3B^{4})Y_t = (1-0.4B^{4}+0.15B^{8})\varepsilon_t$$





ACF and PACF of simulated SARMA(1,1)$_{12}$ and SARMA(1,2)$_4$ models

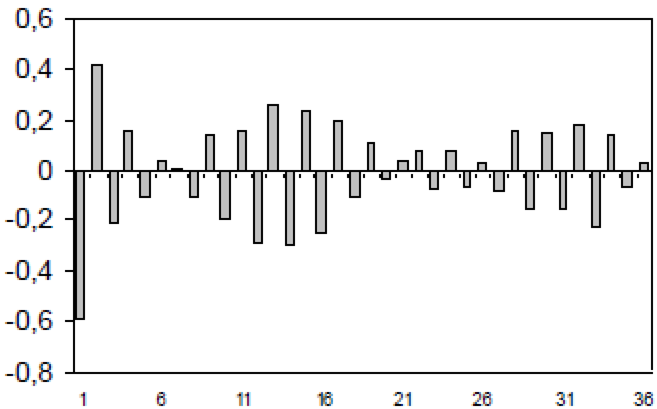
General multiplicative ARMA models

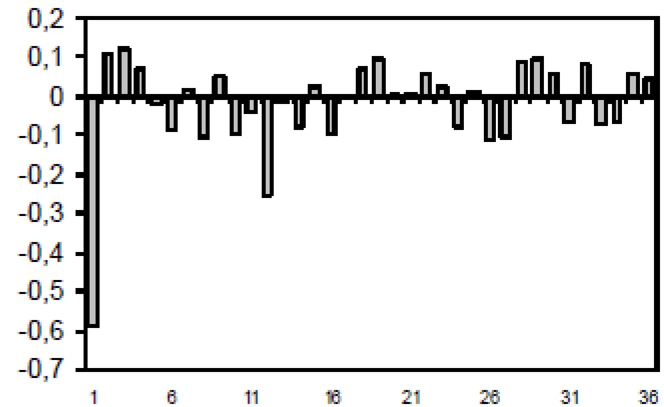
If we combine non-seasonal ARMA(p,q) and seasonal SARMA(P,Q)$_s$ models, we obtain a general multiplicative model of order $(p,q)\times(P,Q)_s$

$$(1-\phi_1 B-\cdots-\phi_p B^p)(1-\Phi_1 B^s-\cdots-\Phi_P B^{Ps})Y_t=(1-\theta_1 B-\cdots-\theta_q B^q)(1-\Theta_1 B^s-\cdots-\Theta_Q B^{Qs})\varepsilon_t,$$

or

$$\phi_p(B)\Phi_P(B^s)Y_t=\theta_q(B)\Theta_Q(B^s)\varepsilon_t.$$





ACF and PACF of a simulated SARMA(1,0)(1,0)$_{12}$ model: $(1-0.7B)(1+0.25B^{12})Y_t=\varepsilon_t$

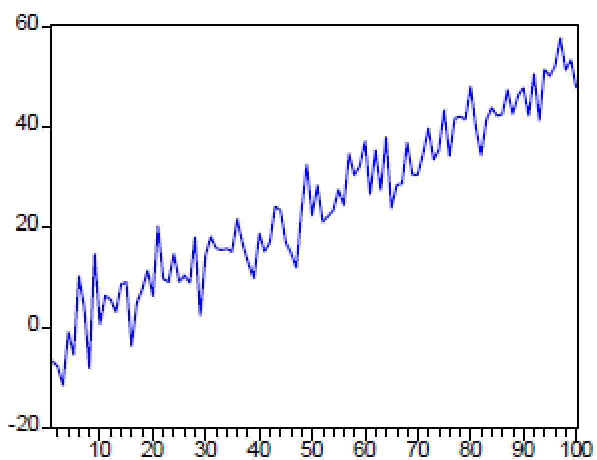
Linear nonstationary time series models

Nonstationary model in the mean

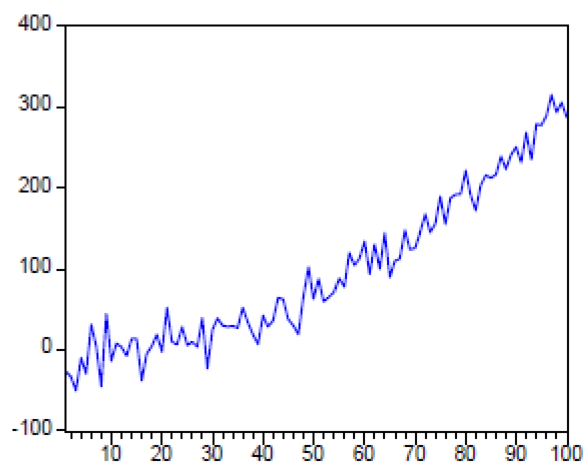
The mean function of a nonstationary model can be represented essentially by two models: deterministic trend models and stochastic trend models.

For a deterministic trend model, one can use the linear trend model, $Y_t = a + bt + \varepsilon_t$ or the quadratic trend model, $Y_t = a + bt + ct^2 + \varepsilon_t$.

Linear trend model



Quadratic trend model



Linear nonstationary time series models

Differencing and stochastic trend model

The dth differenced series, for some integer $d \geq 1$, is given by

$$\nabla^d Y_t = (1-B)^d Y_t .$$

For $d = 1$, we have first differences

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

For seasonal time series, we can use a sth seasonal differencing

$$\nabla^s Y_t = (1-B^s) Y_t = Y_t - Y_{t-s} .$$

Finally, a sth seasonal differencing of order D, for some integer $D \geq 1$ is given by

$$(\nabla^s)^D Y_t = (1-B^s)^D Y_t .$$

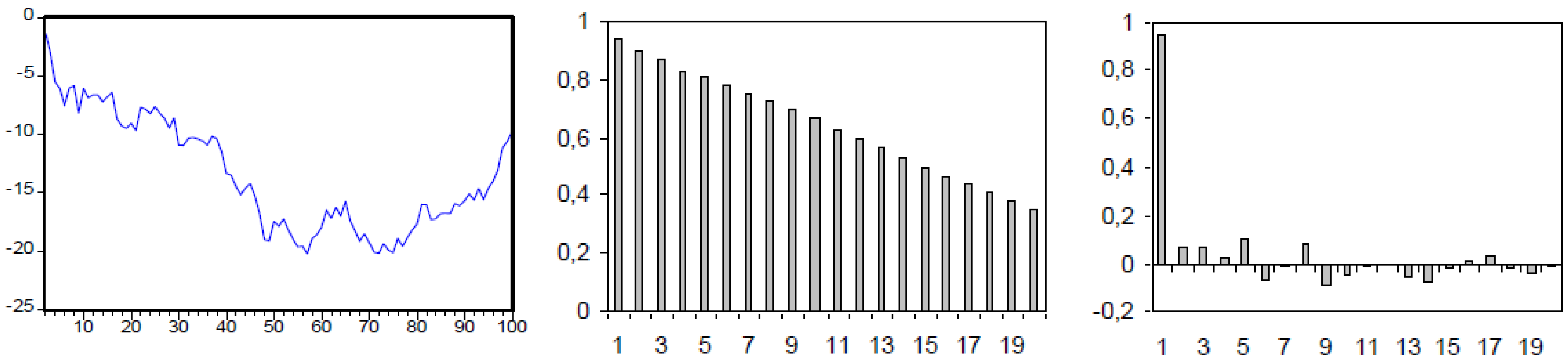
Usually D = 1,2 is sufficient to obtain seasonal stationarity.

Linear nonstationary time series models

A special case of the nonstationary models is the **stochastic trend model**,

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

where ε_t is white noise. This is the so-called "*random walk*" model.



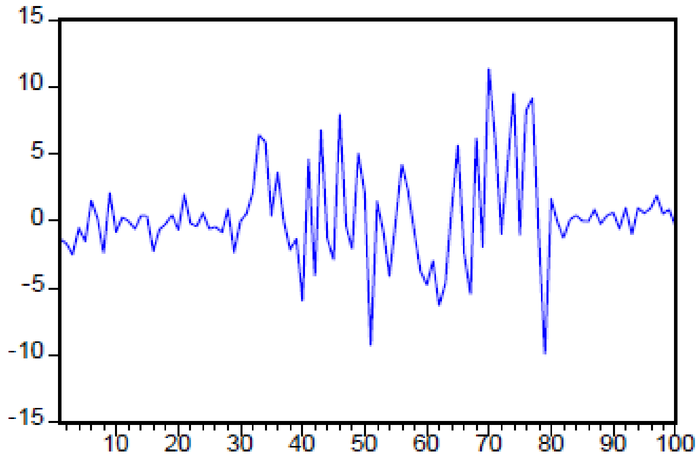
ACF and PACF of a simulated random walk model

Linear nonstationary time series models

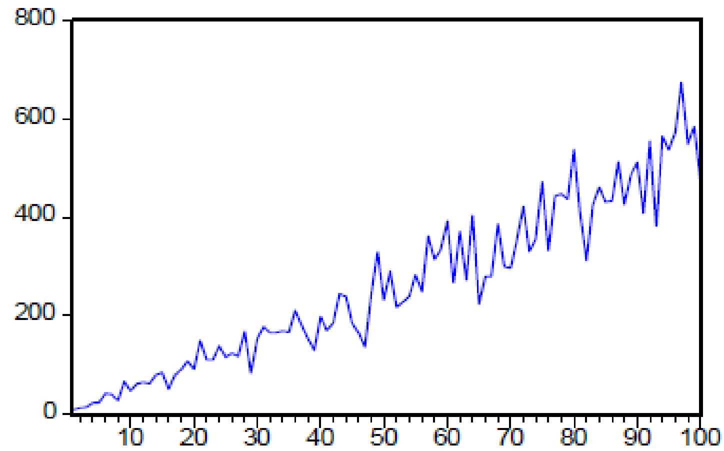
Nonstationarity in the variance

Many time series are stationary in the mean but are nonstationary in the variance. To reduce this type of nonstationarity, we need variance stabilizing transformations such as the power transformation of Box-Cox (1964),

$$X_t = T(Y_t) = \begin{cases} Y_t^{\lambda}, & \lambda \neq 0 \\ \log(Y_t), & \lambda = 0 \end{cases}$$



A simulated time series nonstationary in the variance but stationary in the mean



A simulated time series nonstationarity in both the mean and variance

Linear nonstationary time series models

In practice, we fit the model to $Y_t^{(\lambda)} = \dfrac{Y_t^{\lambda} - 1}{\lambda \tilde{Y}^{\lambda-1}}$, for various values of $\lambda \neq 0$, where $\tilde{Y}$ is the geometric mean of the series Y_t, and choose the value of λ that results in the smallest residual sum of squares. For $\lambda = 0$, we have $Y_t^{(0)} = \tilde{Y} \log(Y_t)$.

Autoregressive integrated moving average (ARIMA) models

A general model for representing nonstationary nonseasonal time series is given by the autoregressive integrated moving average ARIMA(p,d,q) model

$$(1 - \phi_1 B - \cdots - \phi_p B^p)(1 - B)^d Y_t = (1 - \theta_1 B - \cdots - \theta_q B^q)\varepsilon_t$$

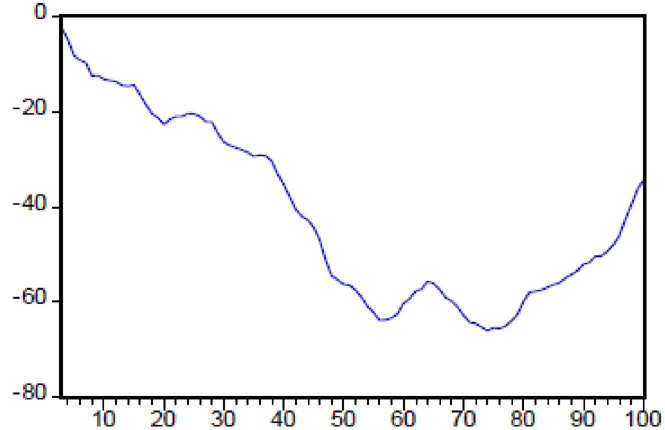
or

$$\phi_p(B)(1 - B)^d Y_t = \theta_q(B)\varepsilon_t ,$$

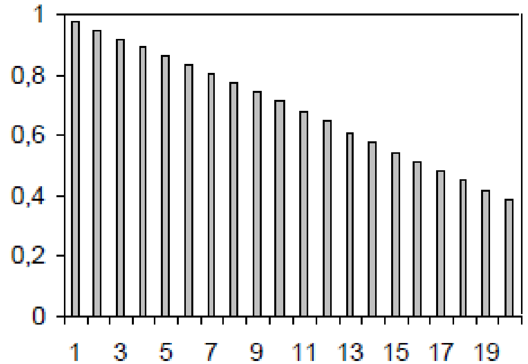
where $(1 - B)^d$ is the differencing operator of order d, for $d \geq 1$, $\phi_p(B)$ is a stationary autoregressive (AR) operator, $\theta_q(B)$ is an invertible moving average (MA) operator and ε_t is a zero mean white noise.

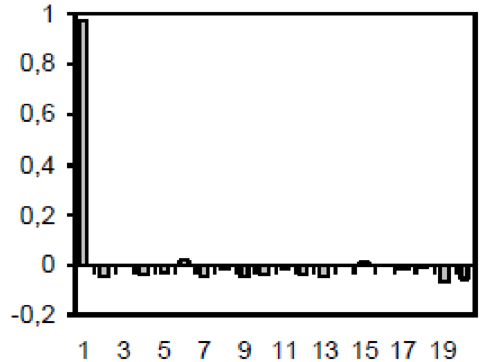
Some important special cases of the ARIMA model are ARIMA(0,1,0), ARIMA(1,1,0), ARIMA(0,1,1) and ARIMA (1,1,1) models.

Linear nonstationary time series models



A simulated series from ARIMA(1,1,0) model: $(1-0{,}75B)(1-B)Y_t = \varepsilon_t$





ACF and PACF of the ARIMA(1,1,0) model: $(1-0{,}75B)(1-B)Y_t = \varepsilon_t$

Linear nonstationary time series models

Multiplicative autoregressive integrated moving average models

The multiplicative seasonal ARIMA model is an extension of the nonseasonal ARIMA model, by adding seasonal autoregressive and moving average factors. The model, often denoted as SARIMA(p,d,q)(P,D,Q)$_s$, is represented by

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Y_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t\,,$$

where $\phi_p(B)$ and $\theta_q(B)$ are regular (nonseasonal) autoregressive and moving average factors, respectively, $\Phi_P(B^s)$ and $\Theta_Q(B^s)$ are seasonal autoregressive and moving average factors, respectively, and s is the seasonal period.

For example, consider the SARIMA(0,1,1)(0,1,1)$_{12}$ model

$$(1-B)(1-B^{12})Y_t = (1-\theta_1 B)(1-\Theta_1 B^{12})\varepsilon_t$$

Model identification

Steps for model identification

- Plot the time series and examine whether the series contains a trend, seasonality, ouliers, nonconstant variances and other nonstationary phenomena. Choose proper variance-stabilizing (Box-Cox's power transformation) and differencing transformations.

- Compute the sample ACF and the sample PACF of the original series and identify the degree of differencing d and D necessary to achieve stationarity. In practice, d and D are either 0, 1, or 2.

- Compute the sample ACF and the sample PACF of the transformed and differenced and identify the orders p and q for the regular autoregressive and moving average operators and the orders P and Q for the seasonal autoregressive and moving average operators, respectively. Usually, the needed orders of integers p, q, P and Q are less or equal to 3.

Model identification

Unit Root Tests

Statistical tests to determine the required order of differencing

- **Augmented Dickey-Fuller (ADF) test (most popular)**
  Null hypothesis: The data are non-stationary and non-seasonal (ϕ=0)
  $\nabla Y(t) = \phi Y(t-1) + b1\ \nabla Y(t-1) + \ldots + \nabla Y(t-p)$

- **Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test**
  Null hypothesis: The data are stationary and non-seasonal

- **Other tests:** Phillipis-Perron (PP) test; Seasonal tests

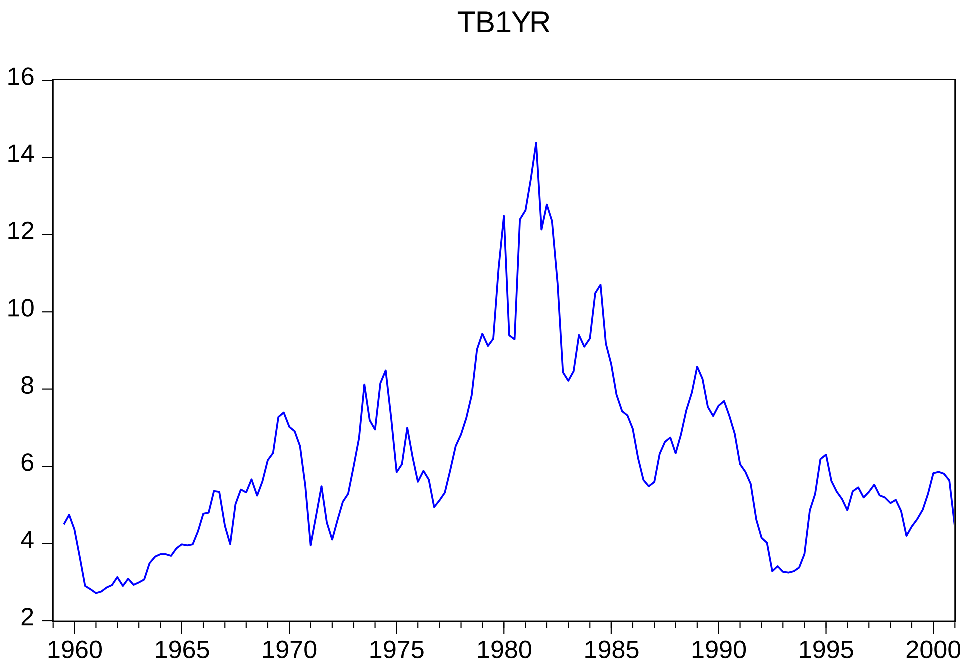
Model identification

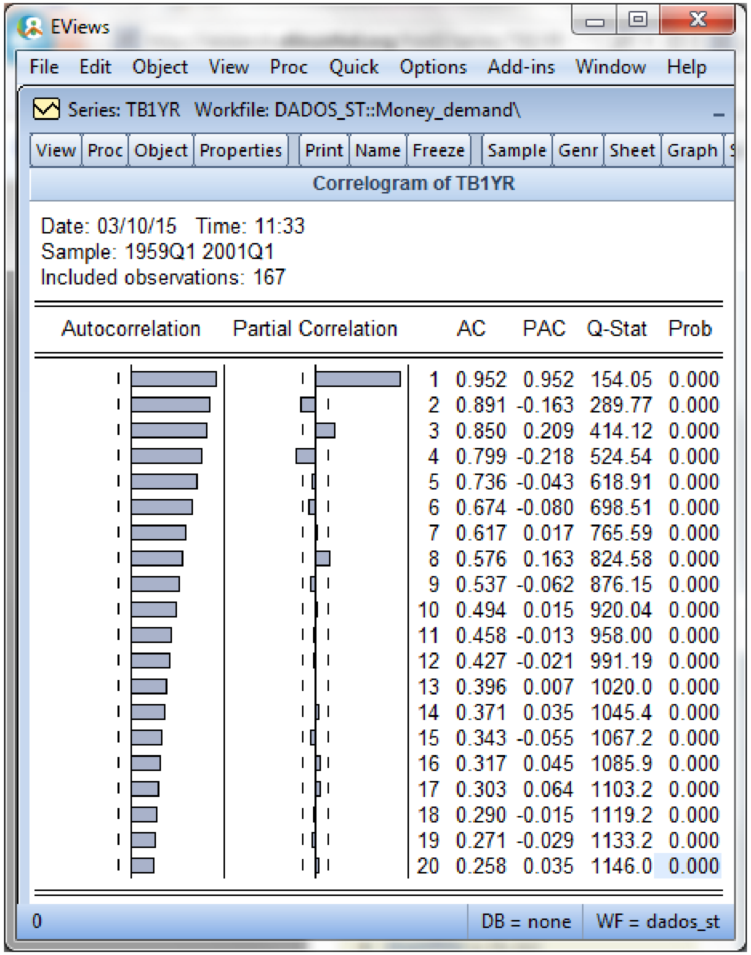
Theoretical ACF and PACF patterns for ARMA models

Model	**ACF**	**PACF**
AR(*p*)	Tails off as exponential decay or damped sine wave	Cuts off after lag *p*
MA(*q*)	Cuts off after lag *q*	Tails off as exponential decay or damped sine wave
ARMA(*p*,*q*)	Tails off after lag (*q*-*p*)	Tails off after lag (*q*-*p*)
SAR(*P*)	Tails off as exponential decay or damped sine wave at the seasonal lags *s*, 2*s*, ...	Cuts off after lag *P*×*s*
SMA(*Q*)	Cuts off after lag *Q*×*s*	Tails off as exponential decay or damped sine wave at the seasonal lags *s*, 2*s*, ...
SARMA(*P*,*Q*)	Tails off as exponential decay or damped sine wave at the seasonal lags *s*, 2*s*, ...	Tails off as exponential decay or damped sine wave at the seasonal lags *s*, 2*s*, ...
SARMA$(p,q)(P,Q)_s$	Tails off as exponential decay or damped sine wave at the seasonal and nonseasonal lags	Tails off as exponential decay or damped sine wave at the seasonal and nonseasonal lags

Model identification

Example: 1-Year US Treasury Bill: Secondary Market Rate




EViews

File Edit Object View Proc Quick Options Add-ins Window Help

Series: TB1YR Workfile: DADOS_ST::Money_demand\

View Proc Object Properties Print Name Freeze Sample Genr Sheet Graph

Correlogram of TB1YR

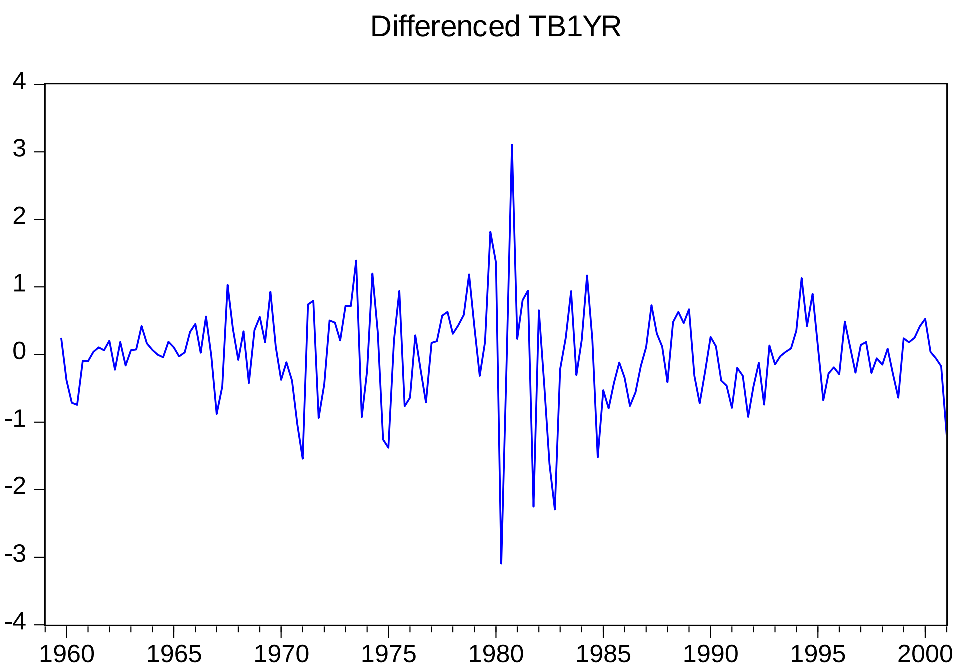
Date: 03/10/15 Time: 11:33
Sample: 1959Q1 2001Q1
Included observations: 167

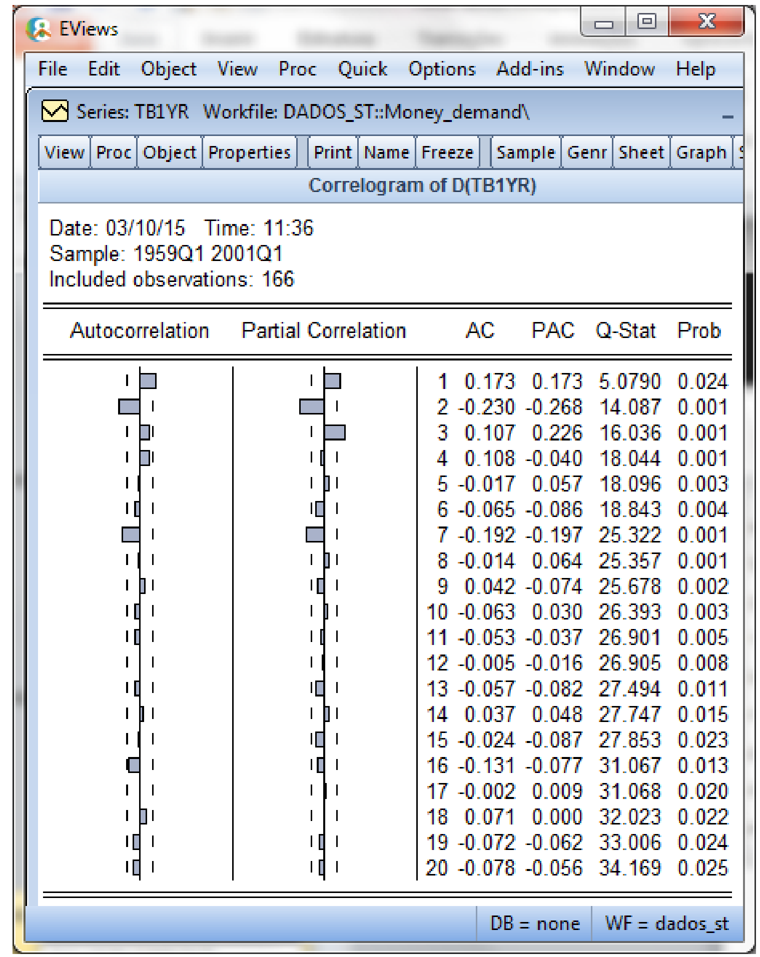
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.952	0.952	154.05	0.000
		2	0.891	-0.163	289.77	0.000
		3	0.850	0.209	414.12	0.000
		4	0.799	-0.218	524.54	0.000
		5	0.736	-0.043	618.91	0.000
		6	0.674	-0.080	698.51	0.000
		7	0.617	0.017	765.59	0.000
		8	0.576	0.163	824.58	0.000
		9	0.537	-0.062	876.15	0.000
		10	0.494	0.015	920.04	0.000
		11	0.458	-0.013	958.00	0.000
		12	0.427	-0.021	991.19	0.000
		13	0.396	0.007	1020.0	0.000
		14	0.371	0.035	1045.4	0.000
		15	0.343	-0.055	1067.2	0.000
		16	0.317	0.045	1085.9	0.000
		17	0.303	0.064	1103.2	0.000
		18	0.290	-0.015	1119.2	0.000
		19	0.271	-0.029	1133.2	0.000
		20	0.258	0.035	1146.0	0.000

0 DB = none WF = dados_st

Model identification

Example: 1-Year US Treasury Bill: Secondary Market Rate





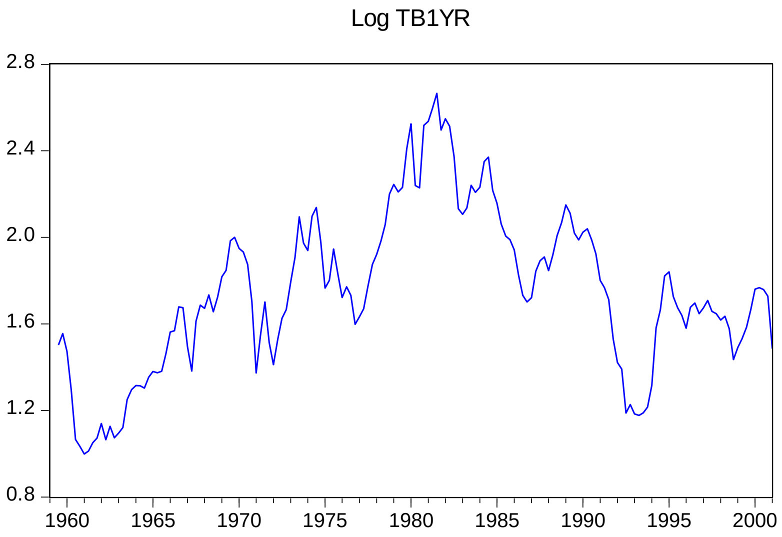
	AC	PAC	Q-Stat	Prob
1	0.173	0.173	5.0790	0.024
2	-0.230	-0.268	14.087	0.001
3	0.107	0.226	16.036	0.001
4	0.108	-0.040	18.044	0.001
5	-0.017	0.057	18.096	0.003
6	-0.065	-0.086	18.843	0.004
7	-0.192	-0.197	25.322	0.001
8	-0.014	0.064	25.357	0.001
9	0.042	-0.074	25.678	0.002
10	-0.063	0.030	26.393	0.003
11	-0.053	-0.037	26.901	0.005
12	-0.005	-0.016	26.905	0.008
13	-0.057	-0.082	27.494	0.011
14	0.037	0.048	27.747	0.015
15	-0.024	-0.087	27.853	0.023
16	-0.131	-0.077	31.067	0.013
17	-0.002	0.009	31.068	0.020
18	0.071	0.000	32.023	0.022
19	-0.072	-0.062	33.006	0.024
20	-0.078	-0.056	34.169	0.025

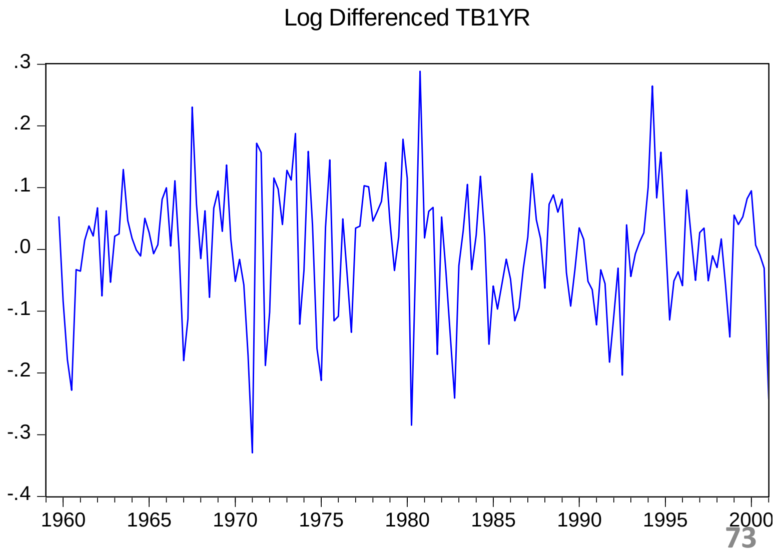
Model identification

Example: 1-Year US Treasury Bill - ARIMA(0,1,2) model

Box-Cox's power transformation on the 1-year Treasury Bill data

λ	Residual sum of squares
1	74.49
0.5	56.24
0	50.04
-0.5	52.04
-1	61.10





Model estimation

After identifying a tentative model, we need to estimate the parameters of the model.

We discuss two widely used estimation procedures:

- **Maximum likelihood estimators (MLE) method**
  The parameter values of the ARIMA model are obtained by minimizing the conditional log-likelihood function

$$\ln L_*(\phi,\theta,\sigma_\varepsilon) = -\frac{n}{2}\ln 2\pi\sigma_\varepsilon^2 - \frac{S_*(\phi,\theta)}{2\sigma_\varepsilon^2}$$

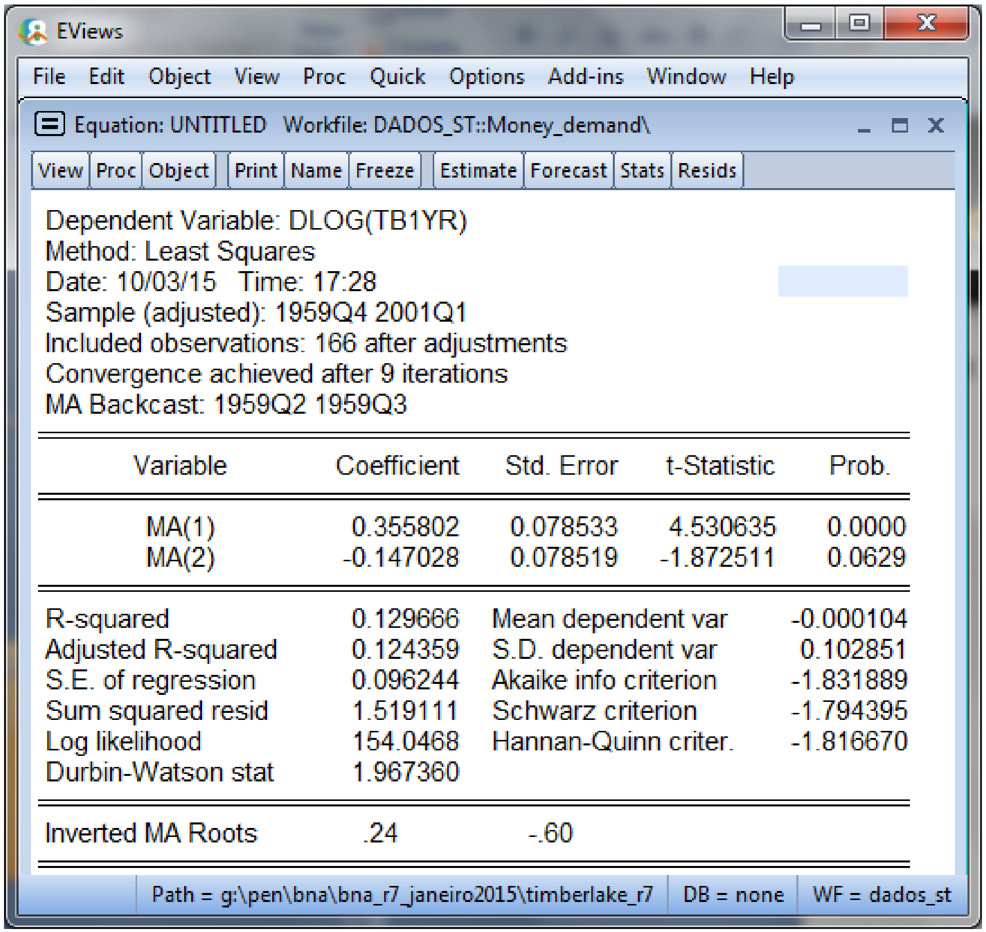
  where $S_*(\phi,\theta) = \sum_{t=p+1}^{n} \sigma_\varepsilon^2(\phi,\theta \mid \mathbf{Y})$ is the conditional sum of squares function.

- **Ordinary Least Squares (OLS) method**
  OLS is the most commonly used regression method in data analysis.
  However, for ARMA(p,q) models, the OLS estimator will be inconsistent unless we have q=0. For more details, see Wei (2006).

Different software will give different estimates. We use the EViews software.

Model estimation

Example: 1-Year US Treasury Bill - ARIMA(0,1,2) model



EViews

File Edit Object View Proc Quick Options Add-ins Window Help

Equation: UNTITLED Workfile: DADOS_ST::Money_demand\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

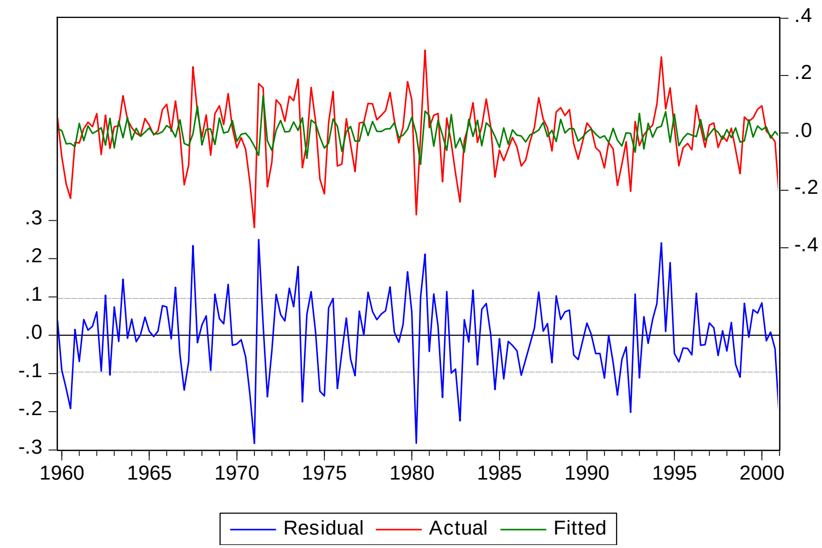
Dependent Variable: DLOG(TB1YR)
Method: Least Squares
Date: 10/03/15 Time: 17:28
Sample (adjusted): 1959Q4 2001Q1
Included observations: 166 after adjustments
Convergence achieved after 9 iterations
MA Backcast: 1959Q2 1959Q3

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	0.355802	0.078533	4.530635	0.0000
MA(2)	-0.147028	0.078519	-1.872511	0.0629

R-squared	0.129666	Mean dependent var	-0.000104
Adjusted R-squared	0.124359	S.D. dependent var	0.102851
S.E. of regression	0.096244	Akaike info criterion	-1.831889
Sum squared resid	1.519111	Schwarz criterion	-1.794395
Log likelihood	154.0468	Hannan-Quinn criter.	-1.816670
Durbin-Watson stat	1.967360		

Inverted MA Roots .24 -.60

Path = g:\pen\bna\bna_r7_janeiro2015\timberlake_r7 DB = none WF = dados_st



Coefficient covariance matrix

	MA(1)	MA(2)
MA(1)	0.006167	0.002640
MA(2)	0.002640	0.006165

Model diagnostic checking

Check on whether a particular model is adequate or not. This involves:

- **Analysis of the quality of parameter estimates**. Inspecting the statistical significance of individual parameter estimates provides some insight into the potential relative goodness of fit of the ARIMA model. To test the null hypothesis $H_0 : \beta_i = 0$, we use the test statistic:

$$|t| = \left| \frac{\hat{\beta}_i}{\sigma_{\hat{\beta}_i}} \right| > t_{(n-m)} \Rightarrow \text{Reject}\, H_0 : \beta_i = 0 .$$

- **Check whether the residuals are approximately white noise.** Compute the sample ACF and sample PACF of the residuals to check whether they are uncorrelated. Box and Pierce (1970) introduced a 'portmanteu' test to check the null hypothesis $H_0 : \rho_1 = \rho_2 = \cdots = \rho_k = 0$, with the test statistic

$$Q = n \sum_{j=1}^{k} \hat{\rho}_j^2 ,$$

which is asymptotically distributed as χ^2 with $k-m$ degrees of freedom, with m the number of estimated parameters.

Model diagnostic checking

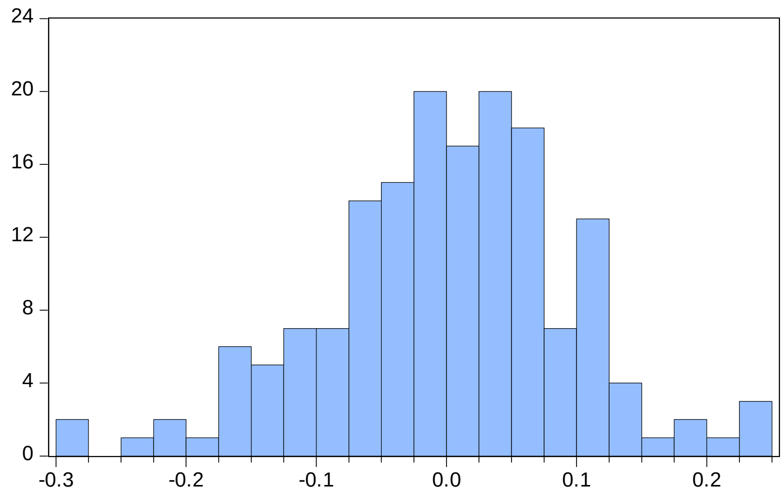
Ljung e Box (1978) proposed a modified version of the statistic Q,

$$Q^* = n(n+2)\sum_{j=1}^{k}\frac{\hat{\rho}_j^2}{n-j}.$$

This modified form of the 'portmanteu' test statistic is much closer to the $\chi^2(k-m)$ distribution for typical sample sizes n. Thus, if the calculated Q^* statistic exceeds the value $\chi^2(k-m)$ then the adequacy of the fitted ARMA model would be questionated.

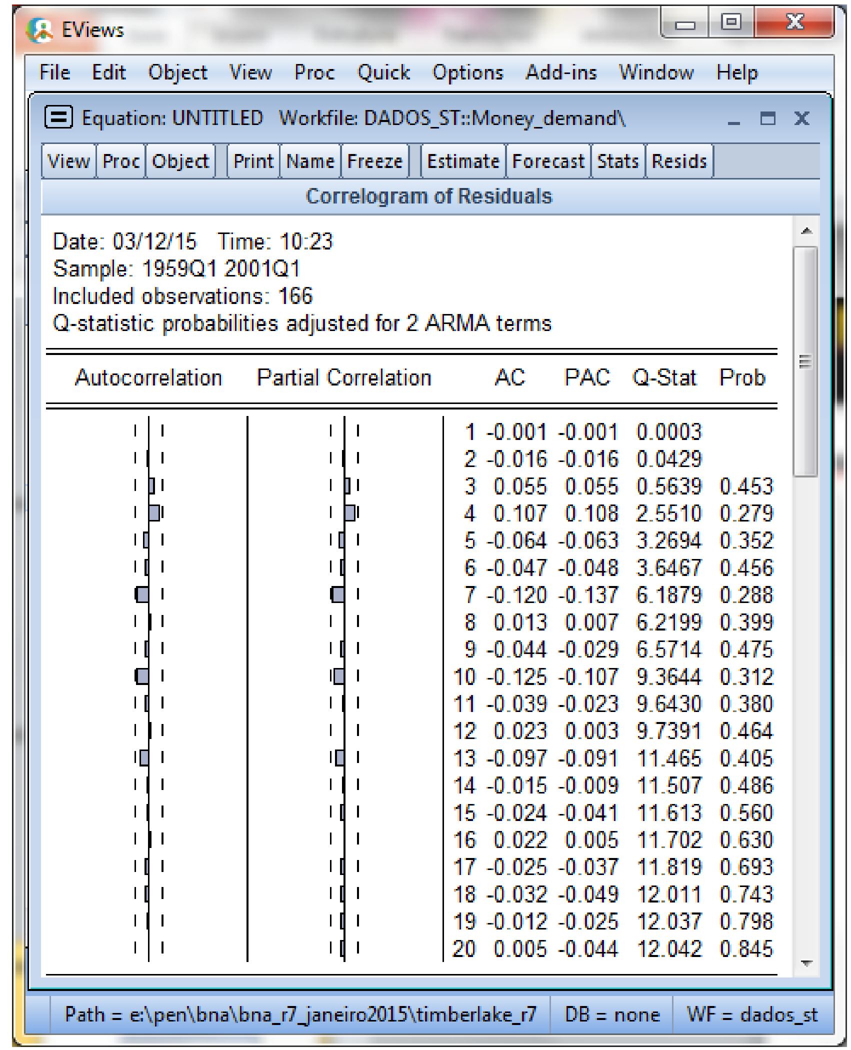
Model diagnostic checking

Example: 1-Year US Treasury Bill
ARIMA(0,1,2) model



Series: Residuals
Sample 1959Q4 2001Q1
Observations 166

Mean	-0.000345
Median	0.004791
Maximum	0.249859
Minimum	-0.282943
Std. Dev.	0.095951
Skewness	-0.229303
Kurtosis	3.464358
Jarque-Bera	2.946140
Probability	0.229221



Date: 03/12/15 Time: 10:23
Sample: 1959Q1 2001Q1
Included observations: 166
Q-statistic probabilities adjusted for 2 ARMA terms

Lag	AC	PAC	Q-Stat	Prob
1	-0.001	-0.001	0.0003	
2	-0.016	-0.016	0.0429	
3	0.055	0.055	0.5639	0.453
4	0.107	0.108	2.5510	0.279
5	-0.064	-0.063	3.2694	0.352
6	-0.047	-0.048	3.6467	0.456
7	-0.120	-0.137	6.1879	0.288
8	0.013	0.007	6.2199	0.399
9	-0.044	-0.029	6.5714	0.475
10	-0.125	-0.107	9.3644	0.312
11	-0.039	-0.023	9.6430	0.380
12	0.023	0.003	9.7391	0.464
13	-0.097	-0.091	11.465	0.405
14	-0.015	-0.009	11.507	0.486
15	-0.024	-0.041	11.613	0.560
16	0.022	0.005	11.702	0.630
17	-0.025	-0.037	11.819	0.693
18	-0.032	-0.049	12.011	0.743
19	-0.012	-0.025	12.037	0.798
20	0.005	-0.044	12.042	0.845

Model selection criteria

Selection criteria are based on summary statistics from residuals, computed from a fitted model (or on forecast errors calculated from out-of-sample forecasts).

- **Akaike Information Criteria (AIC)**
  Assume that a statistical model of m parameters is fitted to a given time series. Akaike (1974) introduced an information criterion defined as

  $$AIC = -2\ln L + 2m,$$

  where L is the maximum likelihood and n is the effective number of observations (or number of computed residuals from the series). The EViews software computes the AIC value as

  $$AIC = n\ln\hat{\sigma}_{\varepsilon}^{2} + 2m,$$

  where $\hat{\sigma}_{\varepsilon}^{2}$ is the residual variance for the fitted model.

- **Schwartz Bayesian criterion (SBC).** Schwartz (1978) introduced the following Bayesian criterion of model selection:

  $$SBC = n\ln\hat{\sigma}_{\varepsilon}^{2} + m\ln n,$$

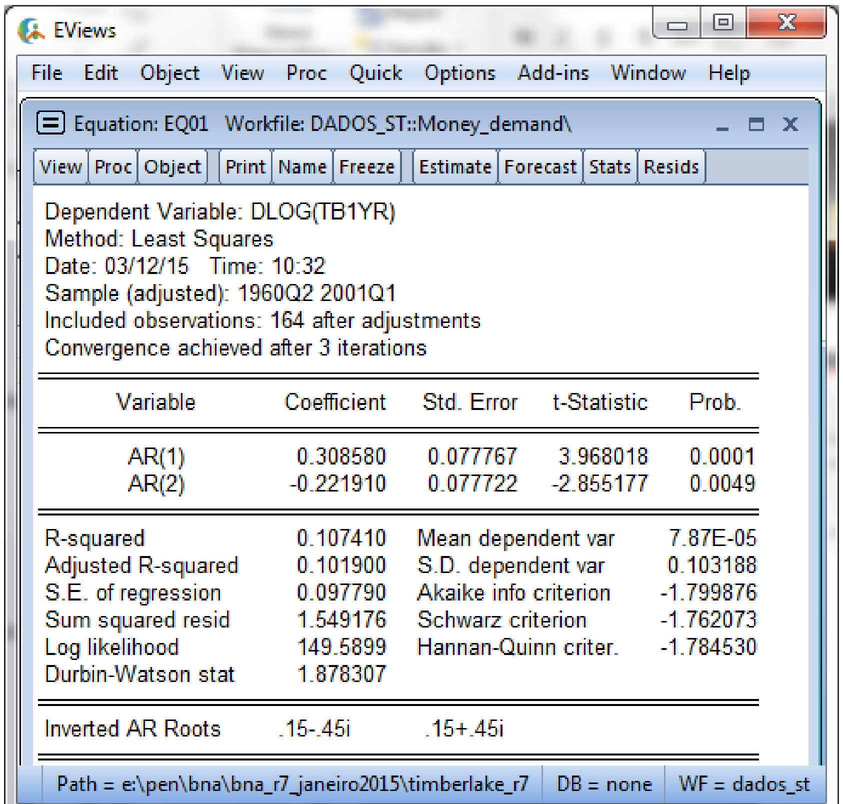
  where $\hat{\sigma}_{\varepsilon}^{2}$ is the residual variance for the fitted model, m is the number of parameters and n is the effective number of observations.

Model selection criteria

Example: 1-Year US Treasury Bill

AIC, BIC and HQ values for TB1YR models

	ARIMA(0,1,2)	ARIMA(2,1,0)
AIC	-1.832	-1.800
BIC	-1.794	-1.762
HQ	-1.817	-1.785



EViews

File Edit Object View Proc Quick Options Add-ins Window Help

Equation: EQ01 Workfile: DADOS_ST::Money_demand\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

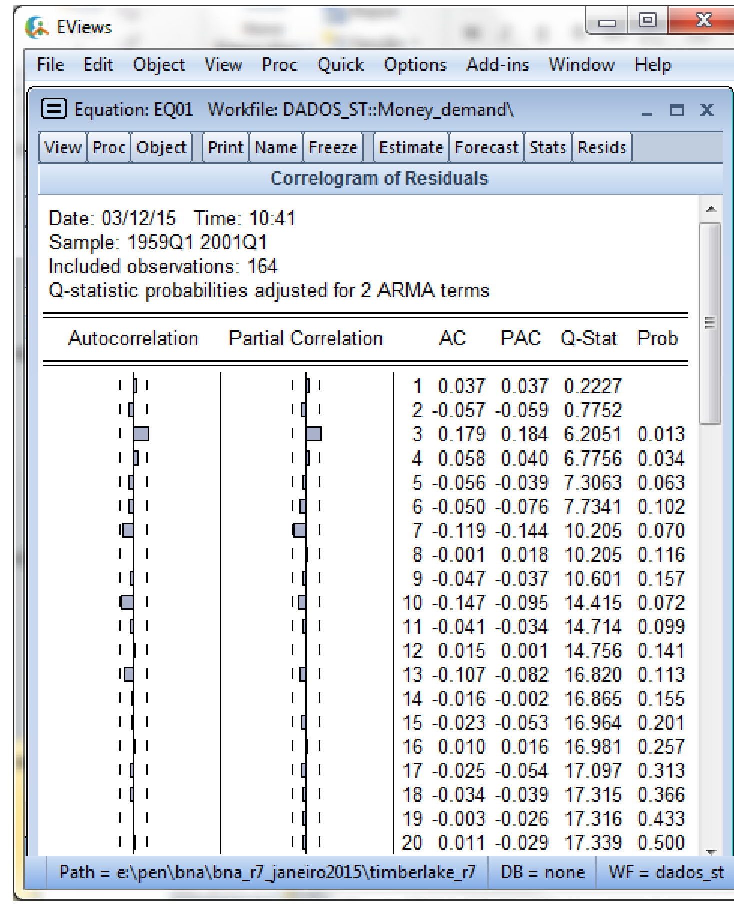
Dependent Variable: DLOG(TB1YR)
Method: Least Squares
Date: 03/12/15 Time: 10:32
Sample (adjusted): 1960Q2 2001Q1
Included observations: 164 after adjustments
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.308580	0.077767	3.968018	0.0001
AR(2)	-0.221910	0.077722	-2.855177	0.0049

R-squared	0.107410	Mean dependent var	7.87E-05
Adjusted R-squared	0.101900	S.D. dependent var	0.103188
S.E. of regression	0.097790	Akaike info criterion	-1.799876
Sum squared resid	1.549176	Schwarz criterion	-1.762073
Log likelihood	149.5899	Hannan-Quinn criter.	-1.784530
Durbin-Watson stat	1.878307		

Inverted AR Roots .15-.45i .15+.45i

Path = e:\pen\bna\bna_r7_janeiro2015\timberlake_r7 DB = none WF = dados_st



EViews

File Edit Object View Proc Quick Options Add-ins Window Help

Equation: EQ01 Workfile: DADOS_ST::Money_demand\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Correlogram of Residuals

Date: 03/12/15 Time: 10:41
Sample: 1959Q1 2001Q1
Included observations: 164
Q-statistic probabilities adjusted for 2 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.037	0.037	0.2227	
		2	-0.057	-0.059	0.7752	
		3	0.179	0.184	6.2051	0.013
		4	0.058	0.040	6.7756	0.034
		5	-0.056	-0.039	7.3063	0.063
		6	-0.050	-0.076	7.7341	0.102
		7	-0.119	-0.144	10.205	0.070
		8	-0.001	0.018	10.205	0.116
		9	-0.047	-0.037	10.601	0.157
		10	-0.147	-0.095	14.415	0.072
		11	-0.041	-0.034	14.714	0.099
		12	0.015	0.001	14.756	0.141
		13	-0.107	-0.082	16.820	0.113
		14	-0.016	-0.002	16.865	0.155
		15	-0.023	-0.053	16.964	0.201
		16	0.010	0.016	16.981	0.257
		17	-0.025	-0.054	17.097	0.313
		18	-0.034	-0.039	17.315	0.366
		19	-0.003	-0.026	17.316	0.433
		20	0.011	-0.029	17.339	0.500

Path = e:\pen\bna\bna_r7_janeiro2015\timberlake_r7 DB = none WF = dados_st

Forecasting

LISBON SCHOOL OF ECONOMICS & MANAGEMENT UNIVERSIDADE DE LISBOA

Suppose that at time $t = T$ we have the observations Y_T, Y_{T-1}, Y_{T-2}, ...

The minimum mean square error forecast of future value Y_{T+m} is defined in terms of the conditional expectation as a linear function of current and previous observations Y_T, Y_{T-1}, Y_{T-2}, ...

$$\hat{Y}_T(m) = E_T(Y_{T+m}) = E(Y_{T+m} \mid Y_T, Y_{T-1}, Y_{T-2}, \ldots),$$

where $\hat{Y}_T(m)$ is the m-step ahead forecast of Y_{T+m}, T is the forecast *origin* and m is the *lead time* (or forecast horizon).

Forecasts for ARMA models

Consider the general stationary ARMA(p,q) model:

$$\phi(B)Y_t = \theta(B)\varepsilon_t,$$

Because the model is stationary, it has an equivalent moving average representation

$$Y_t = \varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \cdots = \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j} = \psi(B)\varepsilon_t,$$

where $\psi_0 = 1$, ε_t is white noise and $\psi(B) = \sum_{j=0}^{\infty}\psi_j B^j = \dfrac{\theta(B)}{\phi(B)}$.

Forecasting

Suppose we have the observations $Y_1, Y_2, \ldots, Y_T$. For $t = T + m$, we have

$$Y_{T+m} = \sum_{j=0}^{\infty} \psi_j \varepsilon_{T+m-j}\ .$$

Standing at origin T, the forecast $\hat{Y}_T(m)$ of Y_{T+m} is defined as a linear combination of current and previous shocks $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots$

$$\hat{Y}_T(m) = \psi^*_m \varepsilon_T + \psi^*_{m+1} \varepsilon_{T-1} + \psi^*_{m+2} \varepsilon_{T-2} + \cdots$$

where the weights $\psi^*_m, \psi^*_{m+1}, \psi^*_{m+2}, \ldots$ are to be determined. Then, the mean square error of the forecast is

$$E\left[Y_{T+m} - \hat{Y}_T(m)\right]^2 = \sigma^2_\varepsilon \sum_{j=0}^{m-1} \psi_j^2 + \sigma^2_\varepsilon \sum_{j=0}^{\infty} \left[\psi_{m+j} - \psi^*_{m+j}\right]^2,$$

which is minimized when $\psi_{m+j} = \psi^*_{m+j}$. Hence,

$$\hat{Y}_T(m) = \psi_m \varepsilon_T + \psi_{m+1} \varepsilon_{T-1} + \psi_{m+2} \varepsilon_{T-2} + \cdots$$

Since $E(\varepsilon_{T+m} \mid Y_T, Y_{T-1}, Y_{T-2}, \ldots) = 0, j > 0$, then the minimum mean square error forecast of Y_{T+m} is the conditional expectation. That is

$$\hat{Y}_T(m) = \psi_m \varepsilon_T + \psi_{m+1} \varepsilon_{T-1} + \psi_{m+2} \varepsilon_{T-2} + \cdots = E_T(Y_{T+m})$$

Forecasting

LISBON SCHOOL OF ECONOMICS & MANAGEMENT
UNIVERSIDADE DE LISBOA

The forecast error for lead time *m* is

$$e_T(m) = Y_{T+m} - \hat{Y}_T(m) = \sum_{j=0}^{m-1} \psi_j \varepsilon_{T+m-j} \; .$$

Since $E_T[e_T(m)] = 0$, the variance of the forecast error is

$$Var[e_T(m)] = \sigma_\varepsilon^2 \sum_{j=0}^{m-1} \psi_j^2 \; .$$

Assuming the normality of ε's, the forecast limits are

$$Y_T(m) \pm z_{\alpha/2} \left[1 + \sum_{j=0}^{m-1} \psi_j^2 \right]^{1/2} \sigma_\varepsilon \, ,$$

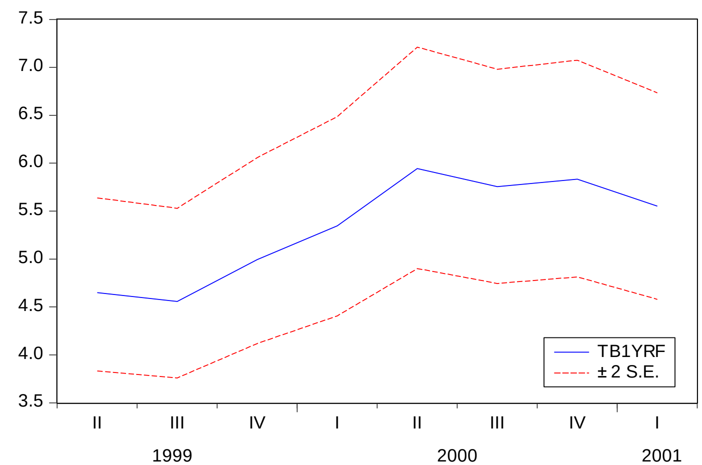
where $z_{\alpha/2}$ is the standard normal deviate such that $P(Z > z_{\alpha/2}) = \alpha/2$.

Find the *m*-step ahead forecast $\hat{Y}_T(m)$, the forecast error and the variance of the forecast error for AR(1), MA(1) and ARMA(1,1) models

Forecasting

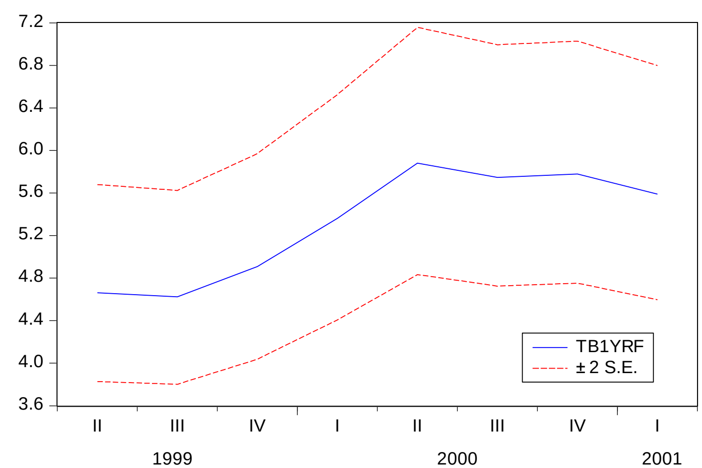
Example: 1-Year US Treasury Bill – **Static Forecasting**

ARIMA(0,1,2) model
log(TB1YR) series



Forecast: TB1YRF	
Actual: TB1YR	
Forecast sample: 1999Q2 2001Q1	
Included observations: 8	
Root Mean Squared Error	0.467643
Mean Absolute Error	0.322565
Mean Abs. Percent Error	6.540790
Theil Inequality Coefficient	0.043831
Bias Proportion	0.007367
Variance Proportion	0.006927
Covariance Proportion	0.985705

ARIMA(2,1,0) model
log(TB1YR) series

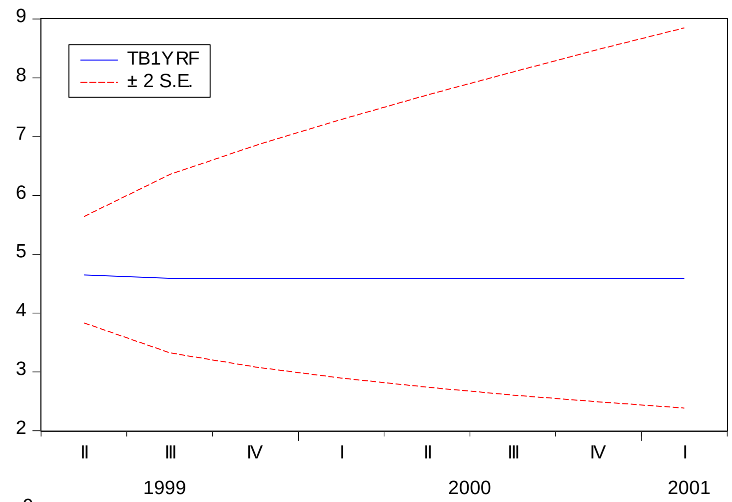


Forecast: TB1YRF	
Actual: TB1YR	
Forecast sample: 1999Q2 2001Q1	
Included observations: 8	
Root Mean Squared Error	0.476301
Mean Absolute Error	0.315519
Mean Abs. Percent Error	6.439636
Theil Inequality Coefficient	0.044692
Bias Proportion	0.004036
Variance Proportion	0.015674
Covariance Proportion	0.980290

Forecasting

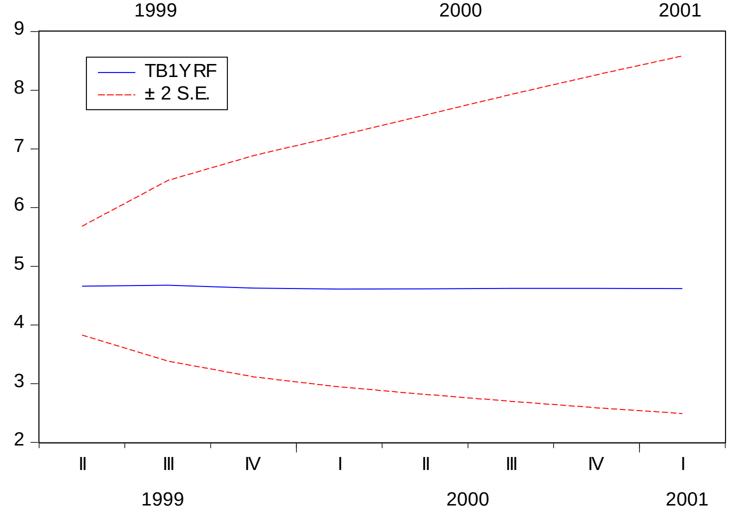
Example: 1-Year US Treasury Bill – **Dynamic Forecasting**

ARIMA(0,1,2) model
log(TB1YR) series



Forecast: TB1YRF	
Actual: TB1YR	
Forecast sample: 1999Q2 2001Q1	
Included observations: 8	
Root Mean Squared Error	0.883607
Mean Absolute Error	0.740036
Mean Abs. Percent Error	13.18370
Theil Inequality Coefficient	0.089116
Bias Proportion	0.609898
Variance Proportion	0.351946
Covariance Proportion	0.038155

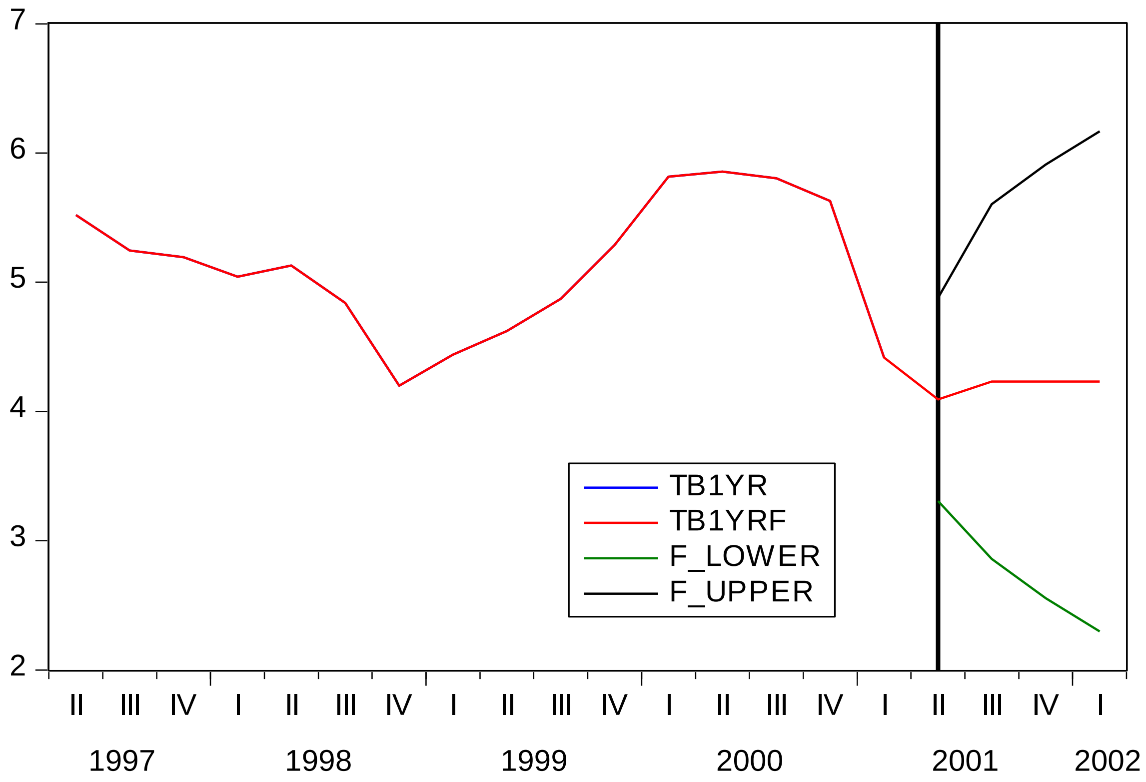
ARIMA(2,1,0) model
log(TB1YR) series



Forecast: TB1YRF	
Actual: TB1YR	
Forecast sample: 1999Q2 2001Q1	
Included observations: 8	
Root Mean Squared Error	0.861252
Mean Absolute Error	0.717839
Mean Abs. Percent Error	12.78505
Theil Inequality Coefficient	0.086577
Bias Proportion	0.582794
Variance Proportion	0.366424
Covariance Proportion	0.050782

Forecasting

Example: 1-Year US Treasury Bill
Forecasts for h=1,2,3 and 4 steps ahead and 95% forecast limits
ARIMA(0,1,2) model



Exercises

LISBON SCHOOL OF ECONOMICS & MANAGEMENT
UNIVERSIDADE DE LISBOA

1. Consider the quarterly unemployment rate (URATE) in U.S. between 1960:Q1 and 2008:Q1 (193 obs.) given in the EViews file "data_financial_econ.wk1" (Sheet 'Quarterly_US').

a) Describe the time plot. Do the data need transformation?

b) Identify a couple of ARIMA models that might be appropriate for the series.

c) Fit your best ARIMA model and carry out diagnostic checking on the residuals.

d) Produce forecasts for the next 4 periods using your preferred model.

e) Find the 95% forecast limits for forecasts in (d).

2. Consider the model

$$(1-B^4)(1-B)Y_t = (1-0.2B)(1-0.6B^4)\varepsilon_t$$

where ε_t is white noise. Find the eventual forecast function that generates the forecasts.

Exercises

3. Let Y_t be a stationary zero-mean process. Consider the models

$$X_t = (1 - 0.4B)Y_t \quad \text{and} \quad W_t = (1 - 2.5B)Y_t$$

a) Find the autocovariance generating functions of X_t and W_t.

b) Show that ACF of the above processes are identical.

4. Consider the ARIMA(0,2,3) model.

a) Write the model in terms of the backshift operator and without using the backshift operator.

b) Find the eventual forecast function.

5. Consider the ARIMA(0,1,1) model. Show that

$$Var[e_t(m)] = \sigma_\varepsilon^2 [1 + (m-1)(1-\theta)^2]$$

Exercises

6. Consider the model

$$(1-0.2B)(1-B)Y_t = (1-0.8B)\varepsilon_t$$

where $\sigma_\varepsilon^2 = 4$. Suppose we have the observations $Y_{49} = 30$, $Y_{48} = 25$ and $\varepsilon_{49} = -2$. Compute the forecast $\hat{Y}_{49}(m)$, for m = 50, 51, 52 and 53.

7. Consider the AR(2) model

$$(1-0.3B-0.6B^2)Y_t = \varepsilon_t$$

a) Find the MA representation of this model.

b) Find the PACF.

8. Consider the model

$$Y_t = 2 + 1.3Y_{t-1} - 0.4Y_{t-2} + \varepsilon_t + \varepsilon_{t-1}$$

a) Find the mean of Y_t.

b) Is the model invertible?

Exercises

9. Consider the model:

$$Y_t = 2 + \varepsilon_t - 0.6\varepsilon_{t-1}, \text{ with } \sigma_\varepsilon = 0.1$$

a) Find the eventual forecast function.

b) Find the variance of the forecast error.

10. Consider the ARMA(1,1) model. Show that

$$Var[e_t(m)] = \sigma_\varepsilon^2\left(1 + \sum_{j=1}^{m-1}\phi^{2(j-1)}(\phi-\theta)^2\right)$$

11. Consider the SARIMA(0,1,1)(0,1,1)$_{12}$ model

$$(1-B)(1-B^{12})Y_t = (1-\theta_1 B)(1-\Theta_1 B^{12})\varepsilon_t$$

a) Write the model without using the backshift operator.

b) Suppose that $\theta_1 = 0.33$ e $\Theta_1 = 0.82$. Find the eventual forecast function.