Time Series Analysis - Part 1

Dr. Esam Mahdi

Islamic University of Gaza - Department of Mathematics

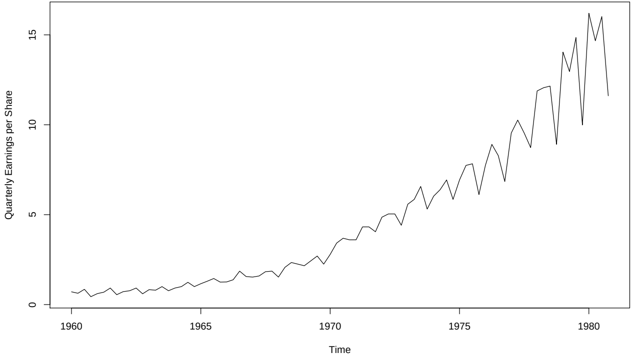
April 9, 2017

What is a Time Series?

- A collection of data $\{X_t\}$ recorded over a period of time (daily, weekly, monthly, quarterly, yearly, etc.), analyzed to understand the past, in order to predict the future (forecast), helping managers and policy makers to make well-informed and sound decisions.
- An important feature of most time series is that observations close together in time tend to be correlated (serially dependent).
- Time could be discrete: $t = 1, 2, ...$, or continous: $t > 0$.

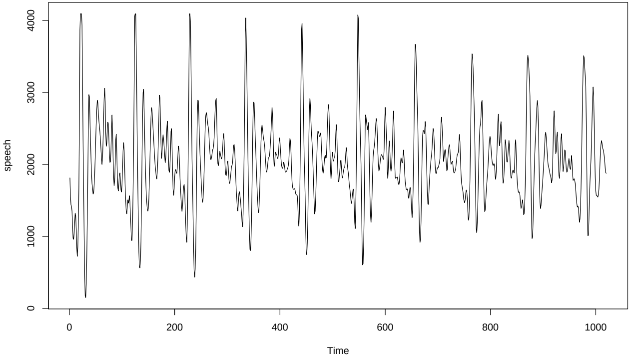
The Nature of Time Series Data - Example 1

Quarterly earnings per share for 1960Q1 to 1980Q4 of the U.S. company, Johnson & Johnson, Inc. (Upward trend).



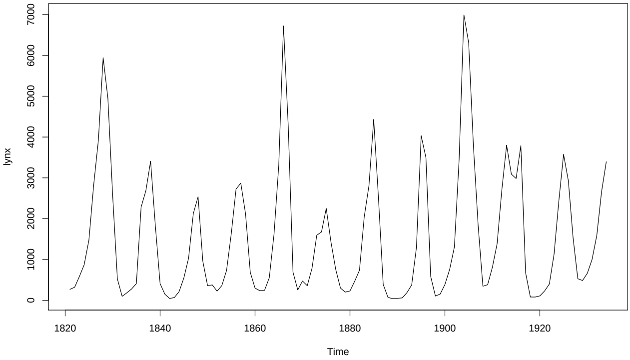
The Nature of Time Series Data - Example 2

A small .1 second (1000 points) sample of recorded speech for the phrase "aaa...hhh". (Regular repetition of small wavelets).



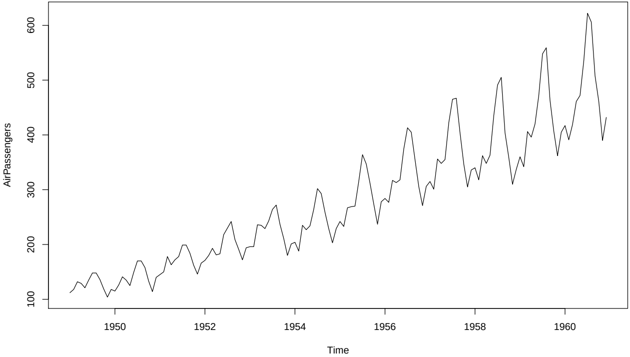
The Nature of Time Series Data - Example 3

Annual numbers of lynx trappings in McKenzie river in Northwest Territories of Canada over the years 1821-1934. (Aperiodic cycles of approximately 10 years).



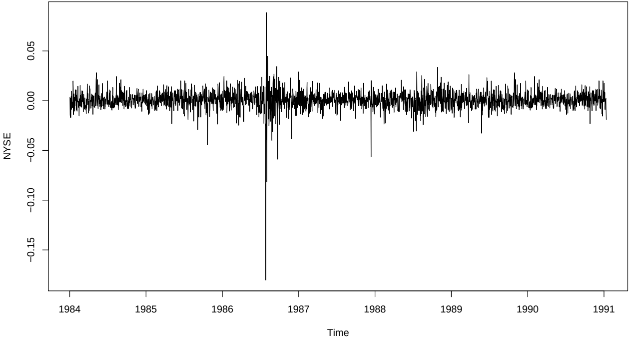
The Nature of Time Series Data - Example 4

Monthly Airline Passenger Numbers 1949-1960. (Seasonality appears to increase with the general trend).



The Nature of Time Series Data - Example 5

Returns of the New York Stock Exchange (NYSE) from February 2, 1984 to December 31, 1991. (Average return of approximately zero, however, volatility (or variability) of data changes over time).



Components of a Time Series

In general, the time series can be decompose into 4 components: **Secular Trend, Seasonal Variation, Cyclical Variation, and Irregular Variation** that can be modelled deterministically with mathematical functions of time (see the next slide).

- General Trend: The smooth long term direction (upward or downward) of a time series.
- Seasonal Variation: Patterns of change in a time series within a year which tend to repeat each year.
- Cyclical Variation: The rise and fall of a time series over periods longer than one year.
- Irregular Variation: Random and follow no regularity in the occurrence pattern.

Three Types of Time Series Decomposition

1. **The additive decomposition model:**

$$X_t = T_t + S_t + C_t + I_t$$

   - X_t = Original data at time t
   - T_t = Trend value at time t
   - S_t = Seasonal fluctuation at time t,
   - C_t = Cyclical fluctuation at time t,
   - I_t = Irregular variation at time t.

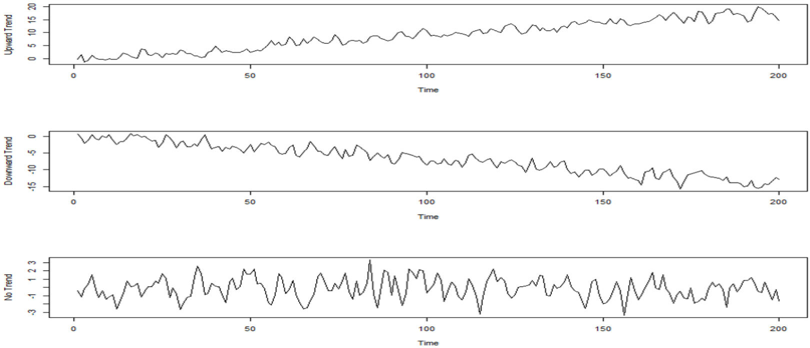
2. **The multiplicative decomposition model:**

$$X_t = T_t \times S_t \times C_t \times I_t$$

3. **The Mixed decomposition model:** For example, if T_t and C_t are correlated to each others, but they are independent from S_t and I_t, the model will be $X_t = T_t \times C_t + S_t + I_t$

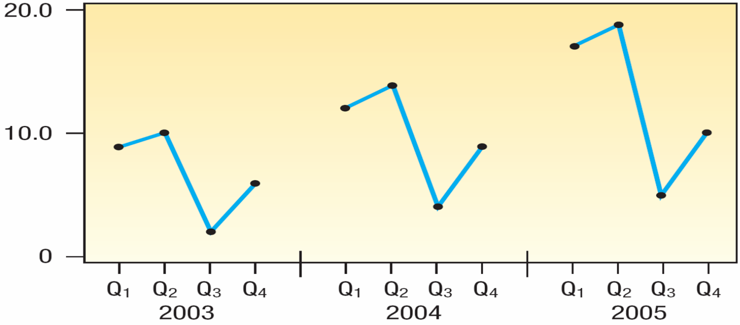
Secular (General) Trend: Sample Chart

The increase or decrease in the movements of a time series that does not appear to be periodic is known as a trend.



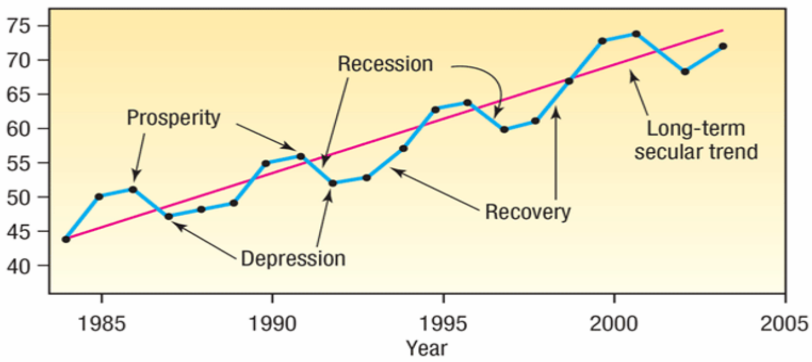
Seasonal Variation: Sample Chart

Short-term fluctuation in a time series which occur periodically in a year. This continues to repeat year after year, although the term is applied more generally to repeating patterns within any fixed period, such as restaurant bookings on different days of the week.



Cyclical Variation: Sample Chart

Recurrent upward or downward movements in a time series where the period of cycle is greater than a year. Also these variations are not regular as seasonal variation.



Time Series Decomposition in R

In R, the function `decompose()` estimates trend, seasonal, and irregular effects using the Moving Averages method (MA). In this case, the series is decompose into 3 components: **trend-cycle component, seasonal component, and irregular component**.

1. **The additive decomposition model:**

$$X_t = T_t + S_t + I_t$$

2. **The multiplicative decomposition model:**

$$X_t = T_t \times S_t \times I_t$$

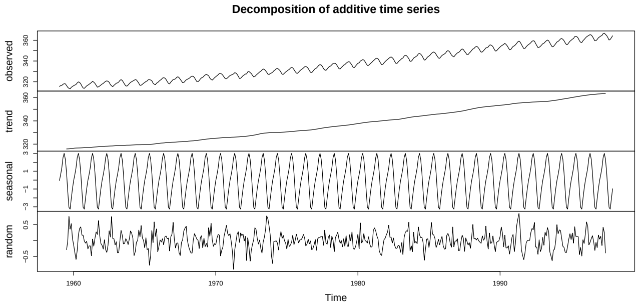
   where T_t is the trend-cycle component (containing both trend and cycle components).

Continue to next slide ...

The Function decompose() in R - Example

Atmospheric concentrations of CO2 (monthly from 1959 to 1997).

```
R> D <- decompose(co2)
R> season.term <- D$figure; trend.term <- D$trend
R> random.term <- D$random; plot(D)
```



Procedure for Decomposition

1. Decompose the series into its components (say multiplicative model: $X_t = T_t \times S_t \times C_t \times I_t$.
2. Create the centered moving averages series ($MA_t = T_t \times C_t$).
3. Deseasonalize the series (Deseasonalizing means: remove the seasonal effects when seasonal pattern of period m exists). This will be done as follows:
   - Find the ratio $X_t/MA_t = S_t \times I_t$.
   - Take the average of each corresponding season to eliminate the irregular term (randomness) and get the seasonal indices.
   - Normalize seasonal indices to get SI_t (divide each seasonal index's value by the sum of all indices' values and then multiply by m) to make sure they average to 1 (sum $= m$).
   - Find the ratio of X_t/SI_t (this is the deseasonalized series $X_t^\star$).
4. Estimate the trend for the deseasonalized series.
5. Forecast future values of each component ($\hat{X}_t^\star$).
6. Reseasonalize predictions ($ResP = SI_t \times \hat{X}_t^\star$).

Estimating the Seasonality: Moving Averages Method

- **Moving Averages method (MA)** is used to smooth out the short-term fluctuations and identify the long-term trend in the original series (i.e, to create a smooth series in order to reduce the random fluctuations in the original series and estimates the trend-cycle component). Moving averages of length $m \in \mathbb{Z}$ is obtained by taking the average of the initial subset of m consecutive observations (usually m denotes the seasonal pattern of m periods). Then the subset is modified by "shifting forward"; that is, excluding the first observation of the series and including the next observation following the initial subset in the series. This creates a new subset of m consecutive observations, which is averaged. Do the shifting forward method again and repeat the procedures until you get the moving averages of span of period m.

Continue to next slide ...

Centered Moving Averages (Centered MA)

- If m is odd, MA is termed a centered moving averages. For example: For the data of observations a, b, c, d, e, f, g, the moving averages with span 3 (or length 3) is

$$\{\frac{a+b+c}{3}, \frac{b+c+d}{3}, \frac{c+d+e}{3}, \frac{d+e+f}{3}, \frac{e+f+g}{3}\},$$

- If m is even, then apply a moving average of span 2 on the resulted series created by the moving average of even span, m, to get the centered moving averages series. For example: The moving averages with span of 4 is $m_1 = \frac{a+b+c+d}{4}, m_2 = \frac{b+c+d+e}{4}, m_3 = \frac{c+d+e+f}{4}, m_4 = \frac{d+e+f+g}{4}$, so that the centered moving averages will be

$$\{\frac{m_1+m_2}{2}, \frac{m_2+m_3}{2}, \frac{m_3+m_4}{2}\}.$$

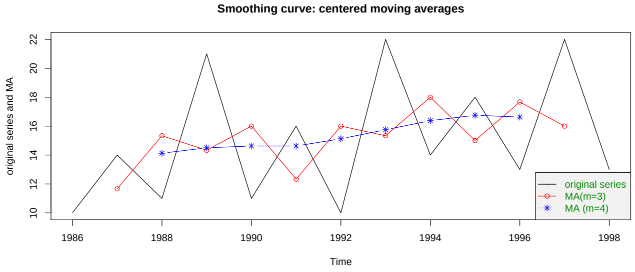
Example: Moving Averages Method

Time	X_t	three-years moving averages	four-years moving averages	
			four-years	two-years
1986	10	-	-	-
1987	14	$(10+14+11) \div 3 = 11.7$	-	-
			$(10+14+11+21) \div 4 = 14.00$	
1988	11	$(14+11+21) \div 3 = 15.3$		$(14.0+14.25) \div 2 = 14.125$
			$(14+11+21+11) \div 4 = 14.25$	
1989	21	$(11+21+11) \div 3 = 14.3$		$(14.25+14.75) \div 2 = 14.500$
			$(11+21+11+16) \div 4 = 14.75$	
1990	11	$(21+11+16) \div 3 = 16.0$		$(14.75+14.50) \div 2 = 14.625$
			$(21+11+16+10) \div 4 = 14.50$	
1991	16	$(11+16+10) \div 3 = 12.3$		$(14.50+14.75) \div 2 = 14.625$
			$(11+16+10+22) \div 4 = 14.75$	
1992	10	$(16+10+22) \div 3 = 16.0$		$(14.75+15.50) \div 2 = 15.125$
			$(16+10+22+14) \div 4 = 15.50$	
1993	22	$(10+22+14) \div 3 = 15.3$		$(15.50+16.00) \div 2 = 15.750$
			$(10+22+14+18) \div 4 = 16.00$	
1994	14	$(22+14+18) \div 3 = 18.0$		$(16.00+16.75) \div 2 = 16.375$
			$(22+14+18+13) \div 4 = 16.75$	
1995	18	$(14+18+13) \div 3 = 15.0$		$(16.75+16.75) \div 2 = 16.750$
			$(14+18+13+22) \div 4 = 16.75$	
1996	13	$(18+13+22) \div 3 = 17.7$		$(16.75+16.50) \div 2 = 16.625$
			$(18+13+22+13) \div 4 = 16.50$	
1997	22	$(13+22+13) \div 3 = 16.0$	-	-
1998	13	-	-	-

R-code: The Centered of Moving Averages Series

```
R> x <- c(10,14,11,21,11,16,10,22,14,18,13,22,13)
R> y <- numeric() ; z <- numeric(); w <- numeric()
R> ## Compute three-years moving averages (say: y)
R> for (i in 1: (length(x)-2)){y[i] <- sum(x[i:(i+2)])/3}
R> ## Compute four-years moving averages (say: z)
R> for (i in 1: (length(x)-3)){z[i] <- sum(x[i:(i+3)])/4}
R> ## Compute two-years moving averages (say: w)
R> for (i in 1: (length(z)-1)){w[i] <- sum(z[i:(i+1)])/2}
R> round(y,2); round(z,2) ; round(w,2)

 [1] 11.67 15.33 14.33 16.00 12.33 16.00 15.33 18.00 15.00

 [1] 14.00 14.25 14.75 14.50 14.75 15.50 16.00 16.75 16.75

[1] 14.12 14.50 14.62 14.62 15.12 15.75 16.38 16.75 16.62
```

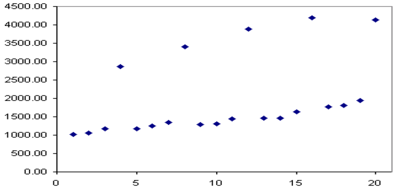
```
R> plot(ts(x,start=1986),ylab="original series and MA")
R> points(ts(y,start=1987),pch=1,type="o",col="red")
R> points(ts(w,start=1988),pch=8,type="b",col="blue")
R> legend(x="bottomright",c("original series","MA(m=3)",
+ "MA (m=4)"),col=c(1,2,4),text.col="green4",
+ pch=c(46,1,8),lty=c(1,19,15),merge=TRUE,bg='gray95')
R> title(main="Smoothing curve: centered moving averages")
```



Example - Deseasonalizing Using Moving Averages Method

Year	Qtr	Revenue
1992	1	1026.00
1992	2	1056.00
1992	3	1182.00
1992	4	2861.00
1993	1	1172.00
1993	2	1249.00
1993	3	1346.00
1993	4	3402.00
1994	1	1286.00
1994	2	1317.00
1994	3	1449.00
1994	4	3893.00
1995	1	1462.00
1995	2	1452.00
1995	3	1631.00
1995	4	4200.00
1996	1	1776.25
1996	2	1808.25
1996	3	1941.75
1996	4	4128.75

Toys R Us Revenue (millions $)



Example: Computing Moving Averages

Year	Qtr	Revenue	Moving Avg
1992	1	1026.00	
1992	2	1056.00	
1992	3	1182.00	**1531.3**
1992	4	2861.00	1567.8
1993	1	1172.00	1616
1993	2	1249.00	1657
1993	3	1346.00	1792.3
1993	4	3402.00	1820.8
1994	1	1286.00	1837.8
1994	2	1317.00	1863.5
1994	3	1449.00	1986.3
1994	4	3893.00	2030.3
1995	1	1462.00	2064
1995	2	1452.00	2109.5
1995	3	1631.00	2186.3
1995	4	4200.00	2264.8
1996	1	1776.25	2353.9
1996	2	1808.25	2431.6
1996	3	1941.75	**2413.8**
1996	4	4128.75	

Calculate Moving Average with span of 4

$$\frac{1026+1056+1182+2861}{4} = 1531.3$$

Example: Centered moving averages

Year	Qtr	Revenue	Moving Avg	Centered MA
1992	1	1026.00		
1992	2	1056.00		
1992	3	1182.00	1531.3	1549.5
1992	4	2861.00	1567.8	1591.9
1993	1	1172.00	1616.0	1636.5
1993	2	1249.00	1657.0	1724.6
1993	3	1346.00	1792.3	1806.5
1993	4	3402.00	1820.8	1829.3
1994	1	1286.00	1837.8	1850.6
1994	2	1317.00	1863.5	1924.9
1994	3	1449.00	1986.3	2008.3
1994	4	3893.00	2030.3	2047.1
1995	1	1462.00	2064.0	2086.8
1995	2	1452.00	2109.5	2147.9
1995	3	1631.00	2186.3	2225.5
1995	4	4200.00	2264.8	2309.3
1996	1	1776.25	2353.9	2392.7
1996	2	1808.25	2431.6	2422.7
1996	3	1941.75	2413.8	
1996	4	4128.75		

Center MA if even number

$$\frac{(1531.3 + 1567.8)}{2} = 1549.5$$

Example: Computing Seasonal Ratios

Year	Qtr	Revenue	Moving Avg	Centered MA	Ratio
1992	1	1026.00			
1992	2	1056.00			
1992	3	1182.00	1531.3	1549.5	**0.763**
1992	4	2861.00	1567.8	1591.9	1.797
1993	1	1172.00	1616	1636.5	0.716
1993	2	1249.00	1657.0	1724.6	0.724
1993	3	1346.00	1792.3	1806.5	0.745
1993	4	3402.00	1820.8	1829.3	1.86
1994	1	1286.00	1837.8	1850.6	0.695
1994	2	1317.00	1863.5	1924.9	0.684
1994	3	1449.00	1986.3	2008.3	0.722
1994	4	3893.00	2030.3	2047.1	1.902
1995	1	1462.00	2064.0	2086.8	0.701
1995	2	1452.00	2109.5	2147.9	0.676
1995	3	1631.00	2186.3	2225.5	0.733
1995	4	4200.00	2264.8	2309.3	1.819
1996	1	1776.25	2353.9	2392.7	0.742
1996	2	1808.25	2431.6	2422.7	0.746
1996	3	1941.75	2413.8		
1996	4	4128.75			

Calculate the ratio of the revenue to the centered moving average

1182÷ 1549.5= 0.7628

Example: Calculating Raw Seasonal Indices

Year	Qtr	Revenue	Moving Avg	Centered MA	Ratio	Avg Ratio
1992	1	1026.00				
1992	2	1056.00				
1992	3	1182.00	1531.30	1549.50	0.763	
1992	4	2861.00	1567.80	1591.90	1.797	
1993	1	1172.00	1616.00	1636.50	0.716	**0.714**
1993	2	1249.00	1657.00	1724.60	0.724	0.708
1993	3	1346.00	1792.30	1806.50	0.745	0.741
1993	4	3402.00	1820.80	1829.30	1.860	1.844
1994	1	1286.00	1837.80	1850.60	0.695	
1994	2	1317.00	1863.50	1924.90	0.684	
1994	3	1449.00	1986.30	2008.30	0.722	
1994	4	3893.00	2030.30	2047.10	1.902	
1995	1	1462.00	2064.00	2086.80	0.701	
1995	2	1452.00	2109.50	2147.90	0.676	
1995	3	1631.00	2186.30	2225.50	0.733	
1995	4	4200.00	2264.80	2309.30	1.819	
1996	1	1776.25	2353.90	2392.70	0.742	
1996	2	1808.25	2431.60	2422.70	0.746	
1996	3	1941.75	2413.80			
1996	4	4128.75				

Example: Normalizing Seasonal Indices (Make Sum SI =4)

$$\frac{4\times 0.7135}{0.7135+0.7077+0.7406+1.844} = 0.7124.$$

Year	Q	Revenue	Moving Avg	Centered MA	Ratio	Avg Ratio	SI
1992	1	1026.00					**0.7124**
1992	2	1056.00					0.7066
1992	3	1182.00	1531.30	1549.50	0.763		0.7394
1992	4	2861.00	1567.80	1591.90	1.797		1.8415
1993	1	1172.00	1616.00	1636.50	0.716	0.714	0.7124
1993	2	1249.00	1657.00	1724.60	0.724	0.708	0.7066
1993	3	1346.00	1792.30	1806.50	0.745	0.741	0.7394
1993	4	3402.00	1820.80	1829.30	1.860	1.844	1.8415
1994	1	1286.00	1837.80	1850.60	0.695		0.7124
1994	2	1317.00	1863.50	1924.90	0.684		0.7066
1994	3	1449.00	1986.30	2008.30	0.722		0.7394
1994	4	3893.00	2030.30	2047.10	1.902		1.8415
1995	1	1462.00	2064.00	2086.80	0.701		0.7124
1995	2	1452.00	2109.50	2147.90	0.676		0.7066
1995	3	1631.00	2186.30	2225.50	0.733		0.7394
1995	4	4200.00	2264.80	2309.30	1.819		1.8415
1996	1	1776.25	2353.90	2392.70	0.742		0.7124
1996	2	1808.25	2431.60	2422.70	0.746		0.7066
1996	3	1941.75	2413.80				0.7394
1996	4	4128.75					1.8415

Example: Deseasonalizing Raw Data

Year	Q	Revenue	Moving Avg	Centered MA	Ratio	Avg Ratio	SI	X^*_t
1992	1	1026.00					0.7124	**1440.2**
1992	2	1056.00					0.7066	1494.4
1992	3	1182.00	1531.3	1549.5	0.763		0.7394	1598.5
1992	4	2861.00	1567.8	1591.9	1.797		1.8415	1553.6
1993	1	1172.00	1616.0	1636.5	0.716	0.714	0.7124	1645.1
1993	2	1249.00	1657.0	1724.6	0.724	0.708	0.7066	1767.6
1993	3	1346.00	1792.3	1806.5	0.745	0.741	0.7394	1820.3
1993	4	3402.00	1820.8	1829.3	1.860	1.844	1.8415	1847.4
1994	1	1286.00	1837.8	1850.6	0.695		0.7124	1805.1
1994	2	1317.00	1863.5	1924.9	0.684		0.7066	1863.8
1994	3	1449.00	1986.3	2008.3	0.722		0.7394	1959.6
1994	4	3893.00	2030.3	2047.1	1.902		1.8415	2114.0
1995	1	1462.00	2064.0	2086.8	0.701		0.7124	2052.2
1995	2	1452.00	2109.5	2147.9	0.676		0.7066	2054.9
1995	3	1631.00	2186.3	2225.5	0.733		0.7394	2205.7
1995	4	4200.00	2264.8	2309.3	1.819		1.8415	2280.7
1996	1	1776.25	2353.9	2392.7	0.742		0.7124	2493.3
1996	2	1808.25	2431.6	2422.7	0.746		0.7066	2559.0
1996	3	1941.75	2413.8				0.7394	2626.0
1996	4	4128.75					1.8415	2242.0

Estimation the Trend

In the absence of the seasonal effect, the time series model may be written in the Wold representation:

$$X_t = \mu_t + \epsilon_t,$$

where $\mu_t = T_t$ is a deterministic component (trend), $\epsilon_t = I_t$ is an independent and identically distributed ($i.i.d.$) stochastic component, and μ_t & ϵ_t are uncorrelated, and μ_t can be modeled in many ways:

- Linear: $\mu_t = a + bt$.
- Quadratic: $\mu_t = a + bt + ct^2$.

There are two methods that can be used to estimate μ_t:

- Method 1: Least squares method.
- Method 2: Smoothing by means of moving averages.

Estimation the Trend - Least Squares Method

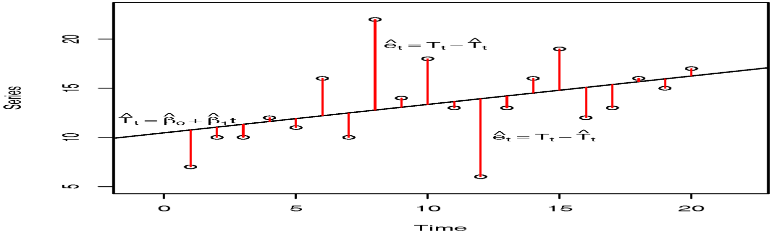
- The long term trend of many time series often approximates a straight line. If T_t is the dependent variable and t is the independent variable (represents the time index), the least squares fitted line (Simple Linear Regression):

$$\begin{aligned} \hat{T}_t &= \hat{\beta}_0 + \hat{\beta}_1 t, \\ \hat{\beta}_1 &= \frac{\sum_{i=1}^{n} t_i T_i - n\bar{t}\bar{T}}{\sum_{i=1}^{n} t_i^2 - n\bar{t}^2}, \\ \hat{\beta}_0 &= \bar{T} - \hat{\beta}_1 \bar{t}. \end{aligned}$$

- Residual is the difference between the true and predicted value:

$$\hat{e}_t = T_t - \hat{T}_t.$$

Least Squares Method - Root Mean Square Error



Root Mean Square Error is a goodness of fit measure defined as: the square root of the mean of the squares of the residuals

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} \hat{e}_t^2}$$

Year	Qtr	Revenue	$X_t^\star$	t	$tX_t^\star$	t^2
1992	1	1026.00	1440.2	1	1440.2	1
1992	2	1056.00	1494.4	2	2988.8	4
1992	3	1182.00	1598.5	3	4795.5	9
1992	4	2861.00	1553.6	4	6214.4	16
1993	1	1172.00	1645.1	5	8225.5	25
1993	2	1249.00	1767.6	6	10605.6	36
1993	3	1346.00	1820.3	7	12742.1	49
1993	4	3402.00	1847.4	8	14779.2	64
1994	1	1286.00	1805.1	9	16245.9	81
1994	2	1317.00	1863.8	10	18638.0	100
1994	3	1449.00	1959.6	11	21555.6	121
1994	4	3893.00	2114.0	12	25368.0	144
1995	1	1462.00	2052.2	13	26678.6	169
1995	2	1452.00	2054.9	14	28768.6	196
1995	3	1631.00	2205.7	15	33085.5	225
1995	4	4200.00	2280.7	16	36491.2	256
1996	1	1776.25	2493.3	17	42386.1	289
1996	2	1808.25	2559.0	18	46062.0	324
1996	3	1941.75	2626.0	19	49894.0	361
1996	4	4128.75	2242.0	20	44840.0	400
Sum	-	-	39423.4	210	451804.8	2870

$$\hat{\beta}_1 = \frac{451804.8 - 20(\frac{210}{20})(\frac{39423.4}{20})}{2870 - 20(\frac{210}{20})^2} = 56.93,$$

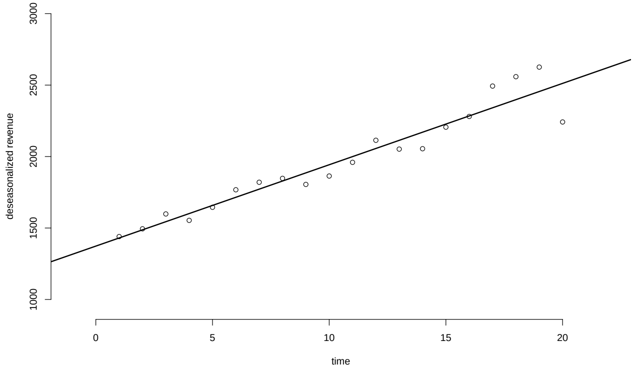
$$\hat{\beta}_0 = \frac{39423.4}{20} - 56.931(\frac{210}{20}) = 1373.39,$$

regression line to deseasonalized data:

$$\hat{X}_t^\star = 1373.39 + 56.93t$$

```
R> t <- 1:20
R> x<-c(1440.2,1494.4,1598.5,1553.6,1645.1,1767.6,1820.3,
+       1847.4,1805.1,1863.8,1959.6,2114.0,2052.2,2054.9,
+       2205.7,2280.7,2493.3,2559.0,2626.0,2242.0)
R> trend.line <- lm(x~t)
R> trend.line

Call:
lm(formula = x ~ t)

Coefficients:
(Intercept)            t
    1373.39        56.93
```

Regression Line for Deseasonalized Data



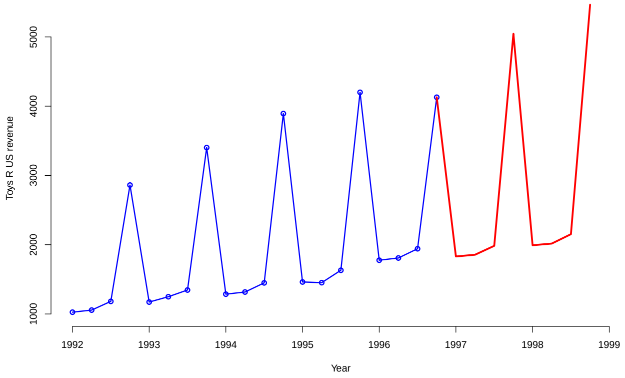
$$\hat{X}_t^\star = 1373.39 + 56.93t$$

Year	t	Revenue T_t	Seasonal Index SI	Forecast $\hat{X}_t^\star$	Reseasonalize Predictions $ResP = SI \times \hat{X}_t^\star$	Square Error $(T_t - ResP)^2$
1992	1	1026.00	0.7124	1430.3	1018.983	97.005
1992	2	1056.00	0.7066	1487.3	1050.927	51.550
1992	3	1182.00	0.7394	1544.2	1141.832	2950.947
1992	4	2861.00	1.8415	1601.1	2948.503	2257.815
1993	5	1172.00	0.7124	1658.0	1181.217	167.376
1993	6	1249.00	0.7066	1715.0	1211.841	2765.401
1993	7	1346.00	0.7394	1771.9	1310.219	2341.483
1993	8	3402.00	1.8415	1828.8	3367.862	343.651
1994	9	1286.00	0.7124	1885.8	1343.450	6503.070
1994	10	1317.00	0.7066	1942.7	1372.755	6225.828
1994	11	1449.00	0.7394	1999.6	1478.607	1603.193
1994	12	3893.00	1.8415	2056.6	3787.221	3299.437
1995	13	1462.00	0.7124	2113.5	1505.684	3759.864
1995	14	1452.00	0.7066	2170.4	1533.669	13358.150
1995	15	1631.00	0.7394	2227.4	1646.995	467.893
1995	16	4200.00	1.8415	2284.3	4206.581	12.770
1996	17	1776.25	0.7124	2341.2	1667.917	23123.700
1996	18	1808.25	0.7066	2398.2	1694.584	25875.590
1996	19	1941.75	0.7394	2455.1	1815.382	29205.720
1996	20	4128.75	1.8415	2512.0	4625.940	72893.410
Sum	-	-	-	-	-	197303.852

$$\hat{X}_t^\star = 1373.39 + 56.93t$$

$$\text{RMSE} = \sqrt{197303.852/20} = 99.32$$

Plot the Forecast - 2 Years (8 Quarters) Ahead



Differencing to detrend and deseasonalized a series

- Sometimes, the first difference operator, $\Delta X_t = X_t - X_{t-1}$, is used to remove the trend of the series (as we will discuss later).

$$\text{e.g., if } X_t = \beta_0 + \beta_1 t \text{ then } X_{t-1} = \beta_0 + \beta_1(t-1),$$

$$\text{so that } \Delta X_t = \beta_1 \text{ (Constant).}$$

- Similarly, a seasonal component with a span period $m \in \mathbb{Z}^+$ in the time series can be removed by differencing the series at lag m. That is $\Delta_m X_t = X_t - X_{t-m}$ (as we will discuss later).

Eliminating the Cycle Component (Multiplicative Models)

After creating the moving averages ($MA_t = T_t C_t$) and estimating the trend (T_t) series, we can estimate the cycle component from the following equation

$$C_t = \frac{MA_t}{T_t}.$$

Forecasts Using Exponential Smoothing Method

Definition

For a sequence of observations $\{X_1, X_2, \ldots\}$ (assuming there is no systematic trend or seasonal effects), the exponential smoothing series $\{\hat{X}_t : t \geq 1\}$ is given by the formula

$$\hat{X}_{t+1} = \hat{X}_t + \alpha(X_t - \hat{X}_t),$$

or equivalently,

$$\hat{X}_{t+1} = \alpha X_t + (1 - \alpha)\hat{X}_t,$$

where $0 < \alpha < 1$ is the exponentially weighted moving average (referred to as the smoothing parameter).

- X_t is the actual observed value at time t.
- $\hat{X}_{t+1}$ is the one-step-ahead forecast value, where $\hat{X}_1 = X_1$.

Exponential Smoothing Weighted MA (EWMA)

- Whereas the smoothing based on the moving averages method puts equal weights on the past observations, exponential smoothing method assign exponentially decreasing weights over time.
- The value of α determines the amount of smoothing.
- A value of α near 1 gives a little smoothing for estimating the mean level and $\hat{X}_{t+1} \approx X_t$.
- A value of α near 0 gives highly smoothed estimates of the mean level and takes little account of the most recent observation.
- A typical value for α is 0.2.
- In practice, we determine the value of α for which the mean squared errors $\text{MSE} = \sum_{t=2}^{n} \frac{(X_t - \hat{X}_t)^2}{n}$ is minimized.

Example - Forecasts Using Exponential Smoothing

For the following daily observations, use the exponential smoothing method with two different smoothing parameters $\alpha = 0.2$ and 0.4 to forecast the 1-day ahead. What is the optimal value of α?

Day	Sales (X_t)
1	39
2	44
3	40
4	45
5	38
6	43
7	39

Example - Forecasts Using Exponential Smoothing

Day	Sales	$\alpha = 0.2$		$\alpha = 0.4$	
t	X_t	Forecast ($\hat{X}_t$)	Square errors (e^2)	Forecast ($\hat{X}_t$)	Square errors (e^2)
1	39	39	-	39	-
2	44	39	25	39	25
3	40	40	0	41	1
4	45	40	25	40.6	19.36
5	38	41	9	42.36	19.0096
6	43	40.4	6.76	40.616	5.6834
7	39	40.92	3.69	41.5696	6.6028

For $\alpha = 0.2$: $\hat{X}_{t+1} = 0.2X_t + 0.8\hat{X}_t$.

- $\hat{X}_1 = X_1 = 39$
- $\hat{X}_2 = 0.2(39) + 0.8(39) = 39$
- $\hat{X}_3 = 0.2(44) + 0.8(39) = 40$
- $\hat{X}_4 = 0.2(40) + 0.8(40) = 40$
- $\hat{X}_5 = 0.2(45) + 0.8(40) = 41$
- $\hat{X}_6 = 0.2(38) + 0.8(41) = 40.4$
- $\hat{X}_7 = 0.2(43) + 0.8(40.4) = 40.92$
- $\hat{X}_8 = 0.2(39) + 0.8(40.92) = 40.54$

For $\alpha = 0.4$: $\hat{X}_{t+1} = 0.4X_t + 0.6\hat{X}_t$.

- $\hat{X}_1 = X_1 = 39$
- $\hat{X}_2 = 0.4(39) + 0.6(39) = 39$
- $\hat{X}_3 = 0.4(44) + 0.6(39) = 41$
- $\hat{X}_4 = 0.4(40) + 0.6(41) = 40.6$
- $\hat{X}_5 = 0.4(45) + 0.6(40.6) = 42.36$
- $\hat{X}_6 = 0.4(38) + 0.6(42.36) = 40.616$
- $\hat{X}_7 = 0.4(43) + 0.6(40.616) = 41.5696$
- $\hat{X}_8 = 0.4(39) + 0.6(41.5696) = 40.5418$

MSE $(\alpha = 0.2) = 11.58$ and MSE$(\alpha = 0.4) = 12.78$ so that $\alpha = 0.2$ is the optimal.

Exponential Smoothing in R

```
HoltWinters(x, alpha = NULL, beta = NULL, gamma =
NULL, seasonal = c("additive", "multiplicative"),
start.periods = 2, l.start = NULL, b.start = NULL,
s.start = NULL,optim.start = c(alpha = 0.3, beta =
0.1, gamma = 0.1), optim.control = list())
```

- Exponential smoothing is a special case of the Holt-Winters algorithm, implemented in the R function `HoltWinters()` where the additional parameters `beta` and `gamma` set to be `FALSE`.
- If we do not specify a value for `alpha`, the function `HoltWinters()` will estimates the value that minimises the MSE .

Example - Forecasts Using Exponential Smoothing in R

```
R> x <- c(27,34,31,24,18,19,17,12,26,14,18,33,31,31,19,
+ 17,10,25,26,18,18,10,4,20,28,21,18,23,19,16)
R> y <- ts(x)
R> ## Consider alpha=0.2
R> out1 <- HoltWinters(y,alpha=0.2,beta=FALSE,gamma=FALSE)
R> #out1$fitted##Output 1st column is the smoothing series
R> #
R> ## Let R estimates alpha
R> out2 <- HoltWinters(y,beta=FALSE,gamma=FALSE)

## The following code plot the original series and EWMA
R> plot(y)## Plot the original series and EWMA
R> lines(out1$fitted[,1],col="red")#alpha=0.2
R> lines(out2$fitted[,1],col="blue")#estimated alpha
```

Example - Plot the Forecasts with Confidence Intervals

```
R> plot(out1,xlim=c(2,38))
R> pred <- predict(out1,n.ahead=6,prediction.interval=T)
R> lines(pred[,1],col="blue")
R> lines(pred[,2],col="green")
R> lines(pred[,3],col="green")
```

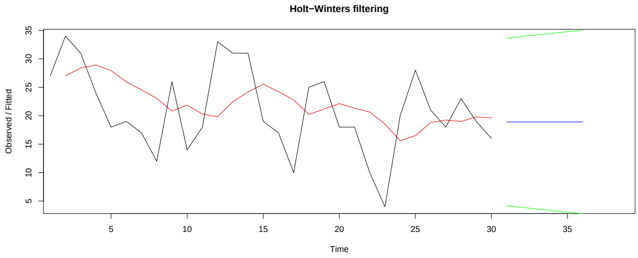


Figure : Exponentially weighted moving average.

The Mean and the Autocovariance Functions

Definition

The mean function of a time series $\{X_t\}$ is defined to be $\mu_t = \mathbb{E}(X_t)$, whereas the variance function is defined to be $\mathbb{V}ar(X_t) = \mathbb{E}[(X_t - \mu_t)^2]$.

The function μ_t specifies the first order properties of the time series.

Definition

The autocovariance function of a time series $\{X_t\}$ is defined to be $\gamma_X(s,t) = \mathbb{C}ov(X_s, X_t) = \mathbb{E}(X_s - \mu_s)(X_t - \mu_t)$, for any two time points t and s.

The function $\gamma(s,t)$ specifies the second order properties of the time series.

Remarks

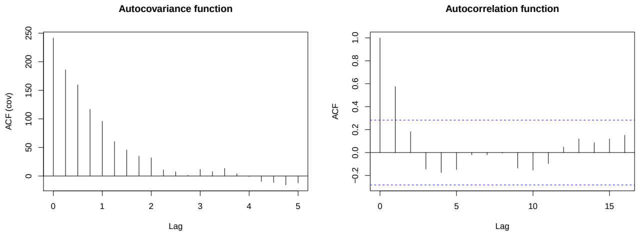
- If $h = t - s$, the parameter $\gamma_X(h)$ is called the h^{th} order or lag h autocovariance of $\{X_t\}$. Thus, $\gamma_X(0) = \mathbb{V}ar(X_t)$.
- The difference of two moments in time is called lag.
- $\forall h, \gamma_X(h) = \mathbb{C}ov(X_t, X_{t+h}) = \mathbb{C}ov(X_t, X_{t-h}) = \gamma_X(-h)$.
- $\forall h$, the autocorrelation $\rho_X(h) = \rho_X(-h)$.
- If a_i's constants, X_i's and Y_j's r.v.'s $i = 1, \ldots, n, j = 1, \ldots, m$

$$\mathbb{C}ov(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i^2 \mathbb{V}ar(X_i) + 2\sum\sum_{i<j} a_i a_j \mathbb{C}ov(X_i, X_j)$$

$$\mathbb{C}ov(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} \mathbb{C}ov(X_i, Y_j)$$

The Plot of the Autocovariance and Autocorrelation (Correlogram) Functions

- The plot of $\gamma(h)$ against the lag h values is called the autocovariance function (ACVF).
- The plot of $\rho(h)$ against the lag h values is called the autocorrelation function (ACF).
- Note that $\rho(0) = 1$ and $-1 \leq \rho(h) \leq 1$ for all h.



Sample Autocovariance and Autocorrelation

- The lag h sample autocovariance is defined as

$$\hat{\gamma}_X(h) = \frac{1}{n} \sum_{t=h+1}^{n} (x_t - \bar{x})(x_{t-h} - \bar{x}),$$

  or equivalently

$$\hat{\gamma}_X(h) = \frac{1}{n} \sum_{t=1}^{n-h} (x_t - \bar{x})(x_{t+h} - \bar{x}),$$

  where $\bar{x} = \frac{1}{n} \sum_{t=1}^{n} x_t$.
- The lag h sample autocorrelation is defined as

$$-1 \leq \hat{\rho}_X(h) = \frac{\hat{\gamma}_X(h)}{\hat{\gamma}_X(0)} = \frac{\sum_{t=1}^{n-h}(x_t - \bar{x})(x_{t+h} - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2} \leq 1$$

Example - Sample Autocorrelation

Find $\hat{\rho}(h)$ where $h = 1, 2,$ and 3

t	Z_t	Z_{t+1}	Z_{t+2}	Z_{t+3}	$\cdots;$	Z_{t-1}	Z_{t-2}
1	13	8	15	4			
2	8	15	4	4		13	
3	15	4	4	12		8	13
4	4	4	12	11		15	8
5	4	12	11	7		4	15
6	12	11	7	14		4	4
7	11	7	14	12		12	4
8	7	14	12			11	12
9	14	12				7	11
10	12					14	7

The sample mean of these ten values is $\bar{Z} = 10$. Thus,

$$\hat{\rho}_1 = \frac{(13-10)(8-10) + (8-10)(15-10) + \cdots + (7-10)(14-10) + (14-10)(12-10)}{(13-10)^2 + (8-10)^2 + \cdots + (14-10)^2 + (12-10)^2}$$

$$= \frac{-27}{144} = -.188$$

$$\hat{\rho}_2 = \frac{(13-10)(15-10) + (8-10)(4-10) + \cdots + (11-10)(14-10) + (7-10)(12-10)}{144}$$

$$= \frac{-29}{144} = -.201$$

$$\hat{\rho}_3 = \frac{(13-10)(4-10) + (8-10)(4-10) + \cdots + (12-10)(14-10) + (11-10)(12-10)}{144}$$

$$= \frac{26}{144} = .181$$

Example - Sample Autocovariance and Autocorrelation

Find $\hat{\gamma}_X(h)$ and $\hat{\rho}_X(h)$ where $h = 0, 1,$ and 3.

t	X_t	t	X_t
1990	2	1997	6
1991	1	1998	2
1992	3	1999	4
1993	5	2000	3
1994	1	2001	7
1995	6	2002	8
1996	2	2003	1

$$\hat{\gamma}_X(0) = \frac{1}{14}\sum_{t=1}^{14}(x_t - \bar{x})^2 = 5.23$$

$$\hat{\gamma}_X(1) = \frac{1}{14}\sum_{t=2}^{14}(x_t - \bar{x})(x_{t-1} - \bar{x}) = -1.152$$

$$\hat{\gamma}_X(3) = \frac{1}{14}\sum_{t=4}^{14}(x_t - \bar{x})(x_{t-3} - \bar{x}) = -0.961$$

$$\hat{\rho}_X(1) = \frac{\hat{\gamma}_X(1)}{\hat{\gamma}_X(0)} = -0.22$$

$$\hat{\rho}_X(3) = \frac{\hat{\gamma}_X(3)}{\hat{\gamma}_X(0)} = -0.184$$

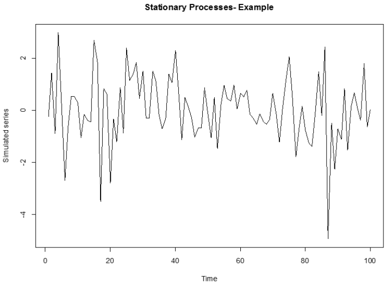
R-Code for the Previous Example

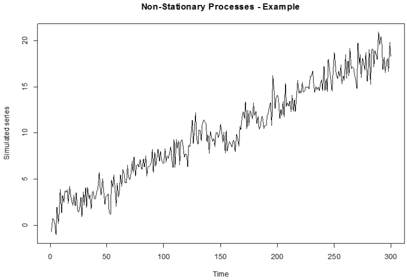
```
R> x <- c(2,1,3,5,1,6,2,6,2,4,3,7,8,1)
R> Cov <- acf(x, type ="covariance",lag.max=3,plot=F)
R> Cor <- acf(x, lag.max=3,plot=FALSE)
R> Cov

Autocovariances of series 'x', by lag

     0      1      2      3
 5.230 -1.152  0.456 -0.961

R> Cor

Autocorrelations of series 'x', by lag

     0      1      2      3
 1.000 -0.220  0.087 -0.184
```

Stationary Models

Definition

Stationary models: assume that the process remains in statistical equilibrium with probabilistic properties that do not change over time, in particular varying about a fixed constant mean level and with constant variance.





Why does a time series has to be stationary?

- Stationarity is defined uniquely, so there is only one way for time series data to be stationary, but lots of ways for it to be non-stationary. Thus stationarity is needed to model the dependence structure uniquely.
- It is preferred that the estimators of parameters such as the mean and variance, if they exist, to not be changed over time.
- In most cases, stationary data can be approximated with stationary ARMA model as we will discuss later.
- Stationary processes avoid the problem of spurious regression.

Strong (Strictly) Stationarity

Definition

The joint distribution of $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$ is defined as:
$F_{X_{t_1},X_{t_2},\ldots,X_{t_n}}(x_1, x_2, \ldots, x_n) = P(X_{t_1} \leq x_1, X_{t_2} \leq x_2, \ldots, X_{t_n} \leq x_n)$,
where $x_1, x_2, \ldots, x_n$ are any real numbers.

Definition

A time series $\{X_t\}$ is said to be Strong (or Strictly) Stationary if for any time points $t_1, t_2, \ldots, t_n \in \mathbb{Z}$, where $n \geq 1$ and any scaler shift (lag) $h \in \mathbb{Z}$, the joint distribution of $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$ is the same as the joint distribution of $X_{t_1+h}, X_{t_2+h}, \ldots, X_{t_n+h}$; i.e.,
$F_{X_{t_1},X_{t_2},\ldots,X_{t_n}}(x_1, x_2, \ldots, x_n) = F_{X_{t_1+h},X_{t_2+h},\ldots,X_{t_n+h}}(x_1, x_2, \ldots, x_n)$

Weak (Covariance) Stationarity

Definition

Time Invariant Process: A process has time invariant if it does not depend on time.

Definition

A time series $\{X_t\}$ is said to be Weak Stationary or Covariance Stationary or Second-order Stationary if

- The mean $\mu_t = \mathbb{E}(X_t) = \mu$ is independent of t.
- For all $t \& h, \gamma_X(h) = \mathbb{C}ov(X_t, X_{t+h}) = \mathbb{E}[(X_t - \mu)(X_{t+h} - \mu)]$ is time-invariant; i.e., the covariance function depends only on the time separation h and not the actual time t.

Remarks

- An independent and identically distributed (i.i.d) stochastic process satisfies $\gamma(h) = \rho(h) = 0 \ \forall h \neq 0$.
- As the joint distribution of X_t and X_{t+h} determines μ_t and $\gamma_X(h)$ if both exist,then the strict stationarity implies weak stationarity, while the converse is not true in general.
- If $\{X_t\}$ is a Gaussian stochastic process, then the weak stationarity is equivalent to strong stationarity (why ?).
- A stationary time series $\{X_t\}$ is **Ergodic** if sample moments converge in probability ($\xrightarrow{p}$) to population moments; i.e. if $\bar{x} \xrightarrow{p} \mu, \hat{\gamma}_X(h) \xrightarrow{p} \gamma_X(h)$, and $\hat{\rho}_X(h) \xrightarrow{p} \rho_X(h)$.

Some Useful Trigonometric Functions

$$
\begin{aligned}
cos(x \pm y) &= cos(x)cos(y) \mp sin(x)sin(y) \\
sin(x \pm y) &= sin(x)cos(y) \pm cos(x)sin(y) \\
sin(2x) &= 2sin(x)cos(x), \qquad cos(2x) = cos^2(x) - sin^2(x) \\
2sin(x)sin(y) &= cos(x-y) - cos(x+y) \\
2cos(x)cos(y) &= cos(x-y) + cos(x+y) \\
sin(2\pi k + x) &= sin(x), \text{ where } k = 0, \pm 1, \pm 2, \pm 3, \ldots \\
cos(2\pi k + x) &= cos(x), \text{ where } k = 0, \pm 1, \pm 2, \pm 3, \ldots \\
sin(\pi - x) &= +sin(x), \qquad sin(\pi + x) = -sin(x) \\
cos(\pi - x) &= -cos(x), \qquad cos(\pi + x) = -cos(x) \\
sin(\pi/2 - x) &= +cos(x), \qquad sin(\pi/2 + x) = +cos(x) \\
cos(\pi/2 - x) &= +sin(x), \qquad cos(\pi/2 + x) = -sin(x)
\end{aligned}
$$

Stationary and Non-stationary Examples

Example 1:

Which of the following time sequence is a weak stationary?

- $X_t = \epsilon_t$, where $\epsilon_t \sim i.i.d.(0,1)$ for all $t \in \mathbb{R}$.
  Solution:
  $\mathbb{E}(X_t) = \mathbb{E}(\epsilon_t) = 0$, which is independent of t; and $\gamma_X(h) = \mathbb{E}(\epsilon_t \epsilon_{t+h}) = 0 \ \forall h \neq 0$, & $\gamma_X(h) = 1$, for $h = 0$, which is also independent of t. Hence, the process is a weak stationary.
- $X_t = t + \epsilon_t$, where $\epsilon_t \sim i.i.d.(0,1)$ for all $t \in \mathbb{R}$.
  Solution:
  $\mathbb{E}(X_t) = \mathbb{E}(t + \epsilon_t) = t$, dependents of t. Hence, the process is not stationary. Note that $\gamma_X(h)$ is independent of t because $\gamma_X(h) = \mathbb{E}([t + \epsilon_t - t][t + h + \epsilon_{t+h} - t - h]) = 0 \ \forall h \neq 0$, & $\gamma_X(h) = 1$, for $h = 0$.

Stationary and Non-stationary Examples

Example 2:

Check whether the following model is a weak stationary or not?
$X_t = Asin(t + B)$, where A is a random variable with a zero mean and a unit variance and B is a random variable with a Uniform distribution $(-\pi, \pi)$ independent of A.

Solution:

$\mathbb{E}(X_t) = \mathbb{E}(Asin(t + B)) = \mathbb{E}(A)\mathbb{E}(sin(t + B)) = 0$, which is independent of t.

$$\begin{aligned}\gamma_X(h) &= \mathbb{E}(X_t X_{t+h}) = \mathbb{E}\Big\{Asin(t + B)Asin(t + h + B)\Big\} \\ &= \mathbb{E}(A^2)\mathbb{E}\Big\{\frac{1}{2}(cos(h) - cos(2t + 2B + h))\Big\}\end{aligned}$$

(see the next slide)

Stationary and Non-stationary Examples (Cont.)

Recall that $Y \sim Uniform(a, b)$, if its probability density function

$$f_Y(y) = \begin{cases} \frac{1}{b-a}, & a \leq y \leq b \\ 0, & \text{otherwise.} \end{cases}$$

where $\mathbb{E}(Y) = \frac{1}{a+b}$ and $\mathbb{V}ar(Y) = \frac{(b-a)^2}{12}$.

$$\begin{aligned} \gamma_X(h) &= \frac{1}{2}cos(h) - \frac{1}{2}\mathbb{E}\Big\{cos(2t + 2B + h)\Big\} \\ &= \frac{1}{2}cos(h) - \frac{1}{2}\int_{-\pi}^{\pi} cos(2t + 2B + h).\frac{1}{2\pi}dB \\ &= \frac{1}{2}cos(h) - \frac{1}{8\pi}[sin(2t + 2B + h)]_{-\pi}^{\pi} \\ &= \frac{1}{2}cos(h), \text{ which is independent of } t, \end{aligned}$$

Hence, the process is a second-order stationary.

White Noise Process

Definition

The stochastic process $\{\epsilon_t\}$ is said to be a strong White Noise process with mean zero and variance σ_ϵ^2 and written as $\epsilon_t \overset{i.i.d.}{\sim} WN(0, \sigma_\epsilon^2)$ if and only if it is independent and identically distributed (i.i.d.) with zero mean and covariance function

$$\gamma_\epsilon(h) = E(\epsilon_t \epsilon_{t+h}) = \begin{cases} \sigma_\epsilon^2, & h = 0; \\ 0, & h \neq 0. \end{cases}$$

Thus an i.i.d. white noise process has

- constant mean (usually this constant equals zero).
- constant variance.
- zero autocovariance, except at lag zero.

White Noise Process

Definition

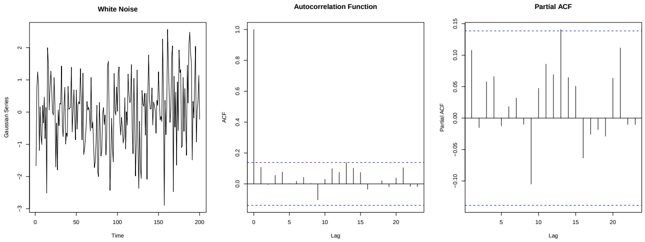
For a second-order properties, the assumption of i.i.d. is not required, only the absence of correlation is needed. In this case the process $\{\epsilon_t\}$ is said to be a weak White Noise process.

Remarks:

- If $\epsilon_t \sim N(0, \sigma_\epsilon^2)$, $\{\epsilon_t\}$ is a Gaussian White Noise process.
- Two random variables X and Y are uncorrelated when their covariance coefficient is zero.
- Two random variables X and Y are independent when their joint probability distribution is the product of their marginal probability distributions: $\forall x, y \Rightarrow P_{X,Y}(x, y) = P_X(x)P_Y(y)$.
- If X and Y are independent, then they are also uncorrelated. However, in general, the converse is not always true.

A Simulated Gaussian White Noise Series

```
R> set.seed(646)
R> white.noise <- rnorm(200)
R> ts.wn<- ts(white.noise)
R> par(mfrow=c(1,3))
R> plot(ts.wn,ylab="Gaussian Series",main="White Noise")
R> acf(ts.wn, main="Autocorrelation Function")
R> acf(ts.wn, type="partial", main="Partial ACF")
```



Backward (Lag), Forward (Lead) and Difference Operators

The lag, lead and difference notations are very convenient way to write linear time series models and to characterize their properties.

Definition

The lag (backshift) operator denotes by $\mathbf{B}(.)$ on an element of a time series is used to produce the previous element.

Definition

The lead (forward) operator denotes by $\mathbf{F}(.)$ on an element of a time series is used to shift the time index forward by one unit.

Definition

The difference operator denotes by Δ expresses the difference between two consecutive random variables (or their realizations).

- $\mathbf{B}X_t = X_{t-1}$,
- $\mathbf{B}^2 X_t = \mathbf{B}(\mathbf{B}X_t) = \mathbf{B}X_{t-1} = X_{t-2}$,
- More generally, $\mathbf{B}^h X_t = X_{t-h}$, where h is any integer.
- $\mathbf{F}X_t = X_{t+1}$,
- $\mathbf{F}^2 X_t = \mathbf{F}(\mathbf{F}X_t) = \mathbf{F}X_{t+1} = X_{t+2}$,
- More generally, $\mathbf{F}^h X_t = X_{t+h}$, where h is any integer.
- $\Delta X_t = X_t - X_{t-1} = (1 - \mathbf{B})X_t$,
- $\Delta^2 X_t = \Delta(\Delta X_t) = \Delta X_t - \Delta X_{t-1} = (1 - \mathbf{B})\Delta X_t = (1 - \mathbf{B})^2 X_t = X_t - 2X_{t-1} + X_{t-2}$,
- More generally, $\Delta^d X_t = (1 - \mathbf{B})^d X_t$, where d is any integer.

Note that

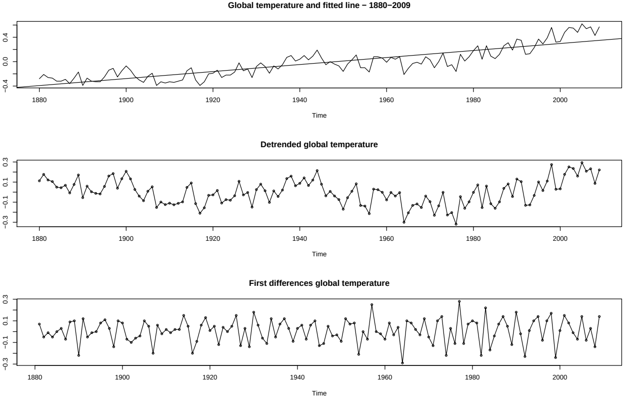
- Positive values of h define lags, while negative values define leads: $\mathbf{B}^{-h} X_t = \mathbf{F}^h X_t = X_{t+h}$,
- $\mathbf{B}^0 X_t = \mathbf{F}^0 X_t = \Delta^0 X_t = X_t$.

Backshift, Forward and Difference Operators in R

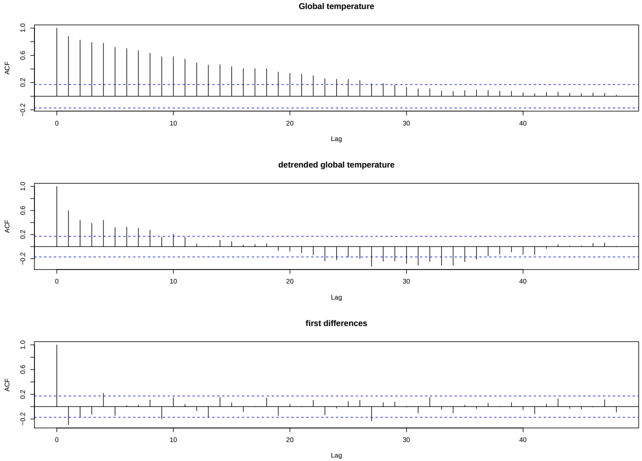
- To get $(1 - \mathbf{B}^h)X_t = X_t - X_{t-h}$ for all lag integer $1 \leq h < t$, type the R-code:
  ```
  > diff(X, lag=h)
  ```
- Note that the opposite of `diff()` is `cumsum()`
- To get $\Delta^d X_t = (1 - \mathbf{B})^d X_t$ for all order of integer difference $1 \leq d < t$, type the R-code:
  ```
  > diff(X, differences = d)
  ```
- In general, to get $(1 - \mathbf{B}^h)^d X_t$ for all lag and difference integers $1 \leq h \times d < t$, type the R-code:
  ```
  > diff(X, lag=h, differences = d)
  ```
- Note that the R-code:
  `> diff(diff(X))` produces the same result as
  ```
  > diff(X, differences = 2)
  ```

Transform Data to Stationary: Detrending vs. Differencing

Example: Global temperature data for the years 1880-2009. The model is $X_t = \beta_0 + \beta_1 t + \epsilon_t$. Detrending is $\hat{e}_t = X_t - \hat{\beta}_0 - \hat{\beta}_1 t$. Differencing is $\Delta X_t = X_t - X_{t-1}$.



Detrending vs. Differencing



Linear Processes

For a stationary time series without trends or seasonal effects. That is; if necessary, any trends, seasonal or cyclical effects have already been removed from the series, we might construct a linear model for a time series with autocorrelation. The most important special cases of linear processes are:

- Autoregressive Model (AR),
- Moving-Average Model (MA),
- Autoregressive Moving Average Model (ARMA),
- Autoregressive-Integrated-Moving Average Model (ARIMA).

These models can be used for predicting the future (forecasting) development of a time series.

Autoregressive (AR) Models With Zero Mean

Definition

A time series $\{X_t\}$, with zero mean, satisfies

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \epsilon_t$$

(ϵ_t is White Noise and ϕ_i are parameters, $i = 1, 2, \ldots, p$) is called Autoregressive Process of order p, denoted by AR(p).

- With backshift operator, the process is: $\Phi_p(\mathbf{B})X_t = \epsilon_t$, $\Phi_p(\mathbf{B}) = 1 - \phi_1 \mathbf{B} - \phi_2 \mathbf{B}^2 - \ldots - \phi_p \mathbf{B}^p$ polynomial of degree p
- The AR(p) model, $\forall p \geq 1$, is a stationary model if and only if the $|\text{roots}|$ of $1 - \phi_1 \mathbf{B} - \phi_2 \mathbf{B}^2 - \ldots - \phi_p \mathbf{B}^p = 0 > 1$.

Autoregressive (AR) Models With Mean μ

- If the mean, μ, of X_t is not zero, replace X_t by $X_t - \mu$ to get:

  $$X_t - \mu = \phi_1(X_{t-1} - \mu) + \phi_2(X_{t-2} - \mu) + \ldots + \phi_p(X_{t-p} - \mu) + \epsilon_t$$

  or write

  $$X_t = \alpha + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \epsilon_t,$$

  where $\alpha = \mu(1 - \phi_1 - \phi_2 - \ldots - \phi_p)$.
- For example, the AR(2) model

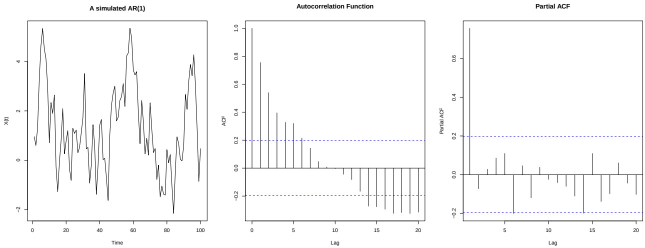
  $$X_t = 1.5 + 1.2X_{t-1} - 0.5X_{t-2} + \epsilon_t$$

  is $X_t - \mu = 1.2(X_{t-1} - \mu) - 0.5(X_{t-2} - \mu) + \epsilon_t$, where $1.5 = \mu(1 - 1.2 - (-0.5))$. Thus the model has mean $\mu = 5$

  $$X_t - 5 = 1.2(X_{t-1} - 5) - 0.5(X_{t-2} - 5) + \epsilon_t$$

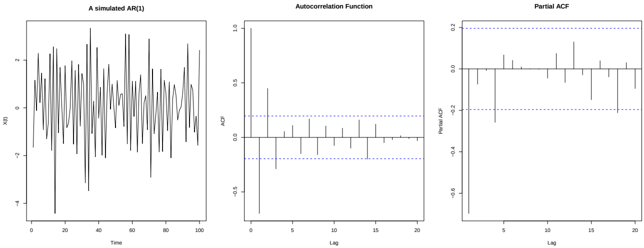
A Simulated AR (1) Series: $X_t = 0.8X_{t-1} + \epsilon_t$

```
R> set.seed(12345)
R> ar1.sim<-arima.sim(n=100,list(order=c(1,0,0),ar=0.8))
R> par(mfrow=c(1,3))
R> plot(ar1.sim, ylab="X(t)", main="A simulated AR(1)")
R> acf(ar1.sim, main="Autocorrelation Function")
R> acf(ar1.sim, type="partial", main="Partial ACF")
```



A Simulated AR (1) Series: $X_t = -0.8X_{t-1} + \epsilon_t$

```
R> set.seed(12345)
R> ar11.sim<-arima.sim(n=100,list(order=c(1,0,0),ar=-0.8))
R> par(mfrow=c(1,3))
R> plot(ar11.sim, ylab="X(t)", main="A simulated AR(1)")
R> acf(ar11.sim, main="Autocorrelation Function")
R> acf(ar11.sim, type="partial", main="Partial ACF")
```



A Simulated AR (2) Series

```
R> set.seed(1965)
R> ar2<-arima.sim(n=200,list(order=c(2,0,0),ar=c(1.1,-0.3))
R> par(mfrow=c(1,3))
R> plot(ar2, ylab="X(t)", main="A simulated AR(2)")
R> acf(ar2, main="Autocorrelation Function")
R> acf(ar2, type="partial", main="Partial ACF")
```

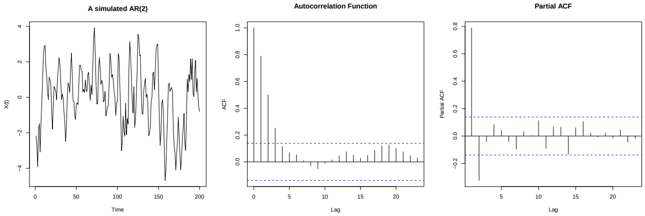


Figure : A simulated AR (2) process: $X_t = 1.1X_{t-1} - 0.3X_{t-2} + \epsilon_t$

Moving Average (MA) Models

Definition

A time series $\{X_t\}$, with zero mean, satisfies

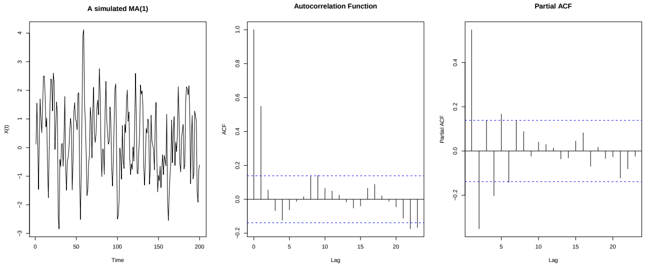
$$X_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}$$

(ϵ_t is White Noise and θ_j are parameters, $j = 1, 2, \ldots, q$) is called Moving Average Process of order q, denoted by MA (q).

- With backshift operator, the process is: $X_t = \Theta_q(\mathbf{B})\epsilon_t$, $\Theta_q(\mathbf{B}) = 1 + \theta_1 \mathbf{B} + \theta_2 \mathbf{B}^2 + \ldots + \theta_q \mathbf{B}^q$ polynomial of degree q
- The MA (q), $\forall q \geq 1$, process is always stationary (regardless of the values of $\theta_j, j = 1, 2, \ldots, q$).
- If the $|$roots$|$ of $1 + \theta_1 \mathbf{B} + \theta_2 \mathbf{B}^2 + \ldots + \theta_q \mathbf{B}^q = 0 > 1$, the MA (q) process is said to be invertible.

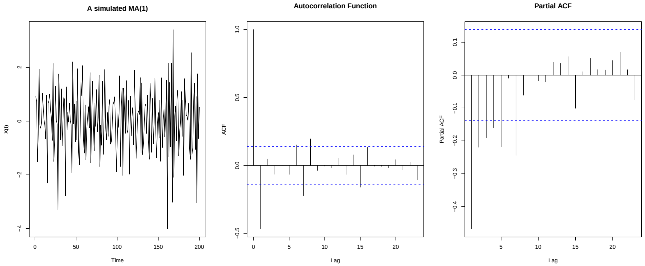
A Simulated MA (1) Series: $X_t = \epsilon_t + 0.8\epsilon_{t-1}$

```
R> set.seed(6436)
R> ma1.sim<-arima.sim(n=200,list(order=c(0,0,1),ma=0.8))
R> par(mfrow=c(1,3))
R> plot(ma1.sim, ylab="X(t)", main="A simulated MA(1)")
R> acf(ma1.sim, main="Autocorrelation Function")
R> acf(ma1.sim, type="partial", main="Partial ACF")
```



A Simulated MA (1) Series: $X_t = \epsilon_t - 0.8\epsilon_{t-1}$

```
R> set.seed(6436)
R> ma11.sim<-arima.sim(n=200,list(order=c(0,0,1),ma=-0.8))
R> par(mfrow=c(1,3))
R> plot(ma11.sim, ylab="X(t)", main="A simulated MA(1)")
R> acf(ma11.sim, main="Autocorrelation Function")
R> acf(ma11.sim, type="partial", main="Partial ACF")
```



A Simulated MA (2) Series

```
R> set.seed(14762)
R> ma2<-arima.sim(n=200,list(order=c(0,0,2),ma=c(0.8,0.6)))
R> par(mfrow=c(1,3))
R> plot(ma2, ylab="X(t)", main="A simulated MA(2)")
R> acf(ma2, main="Autocorrelation Function")
R> acf(ma2, type="partial", main="Partial ACF")
```

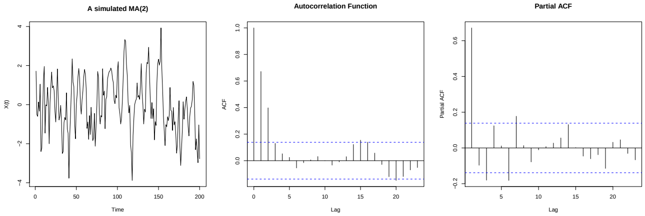


Figure : A simulated MA(2) process: $X_t = \epsilon_t + 0.8\epsilon_{t-1} + 0.6\epsilon_{t-2}$

Autoregressive-Moving Average (ARMA) Models

Definition

A time series $\{X_t\}$, with zero mean, is called Autoregressive-Moving Average of order (p, q) and denoted by ARMA (p, q) if it can be written as

$$X_t = \underbrace{\sum_{i=1}^{p} \phi_i X_{t-i}}_{\text{Autoregressive Part}} + \underbrace{\sum_{j=0}^{q} \theta_j \epsilon_{t-j}}_{\text{Moving Average Part}} ,$$

ϵ_t is White Noise, ϕ_i & θ_j for $i = 1, 2, \ldots, p$, $j = 0, 1, \ldots, q$, are the parameters of the Autoregressive & the Moving Average parts respectively, where $\theta_0 = 1$.

Autoregressive-Moving Average (ARMA) Models

- With backshift operator, the process can be rewritten as:

$$\underbrace{\Phi_p(\mathbf{B})X_t}_{\text{Autoregressive Part}} = \underbrace{\Theta_q(\mathbf{B})\epsilon_t}_{\text{Moving Average Part}},$$

where $\Phi_p(\mathbf{B}) = 1 - \sum_{i=1}^{p} \phi_i \mathbf{B}^i$ and $\Theta_q(\mathbf{B}) = 1 + \sum_{j=1}^{q} \theta_j \mathbf{B}^j$.

- If the mean, μ, of X_t is not zero, replace X_t by $X_t - \mu$ to get:

$$\Phi_p(\mathbf{B})(X_t - \mu) = \Theta_q(\mathbf{B})\epsilon_t.$$

can also be written as:

$$X_t = \alpha + \underbrace{\phi_1 X_{t-1} + \ldots + \phi_p X_{t-p}}_{\text{Autoregressive Part}} + \epsilon_t + \underbrace{\theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q}}_{\text{Moving Average Part}},$$

where $\alpha = \mu(1 - \phi_1 - \phi_2 - \ldots - \phi_p)$ and μ is called the intercept obtained from the output of the R function `arima()`.

Three Conditions for ARMA Models

The ARMA model is assumed to be stationary, invertible and identifiable, where:

- The condition for the stationarity is the same for the pure AR (p) process: i.e., the *roots* of $1-\phi_1\mathbf{B}-\phi_2\mathbf{B}^2-\ldots-\phi_p\mathbf{B}^p$ lie outside the unit circle.
- The condition for the invertibility is the same for the pure MA (q) process: i.e., the *roots* of $1+\theta_1\mathbf{B}+\theta_2\mathbf{B}^2+\ldots+\theta_q\mathbf{B}^q$ lie outside the unit circle.
- The identifiability condition means that the model is not redundant: i.e., $\Phi_p(\mathbf{B})=0$ and $\Theta_q(\mathbf{B})=0$ have no common roots. (see next page)

Example: Model Redundancy (Shared Common Roots in ARMA Models)

Consider the ARMA (1,2):

$$X_t = 0.2X_{t-1} + \epsilon_t - 1.1\epsilon_{t-1} + 0.18\epsilon_{t-2},$$

this model can be written as

$$(1 - 0.2\mathbf{B})X_t = (1 - 1.1\mathbf{B} + 0.18\mathbf{B}^2)\epsilon_t$$

or equivalently

$$(1 - 0.2\mathbf{B})X_t = (1 - 0.2\mathbf{B})(1 - 0.9\mathbf{B})\epsilon_t$$

cancelling $(1 - 0.2\mathbf{B})$ from both sides to get:

$$X_t = (1 - 0.9\mathbf{B})\epsilon_t$$

Thus, the process is not really an ARMA (1,2), but it is a MA (1)$\equiv$ARMA (0,1).

A Simulated ARMA (1,1) Series

```
R> set.seed(6436);par(mfrow=c(1,3))
R> arma1.sim<-arima.sim(n=200,list(order=c(1,0,1),
+ ar=0.8,ma=-0.6))
R> plot(arma1.sim,ylab="X(t)",main="A simulated ARMA(1,1)"
R> acf(arma1.sim, main="Autocorrelation Function")
R> acf(arma1.sim, type="partial", main="Partial ACF")
```

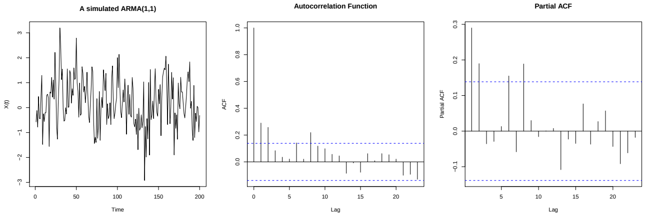


Figure : A simulated ARMA (1,1) process: $X_t = 0.8X_{t-1} + \epsilon_t - 0.6\epsilon_{t-1}$

A Simulated ARMA (2,1) Series

```
R> set.seed(6436);par(mfrow=c(1,3))
R> arma21.sim<-arima.sim(n=200,list(order=c(2,0,1),
+ ar=c(0.8,-0.4),ma=-0.6))
R> plot(arma21.sim,ylab="X(t)",main="A simulated ARMA(2,1)"
R> acf(arma21.sim, main="Autocorrelation Function")
R> acf(arma21.sim, type="partial", main="Partial ACF")
```

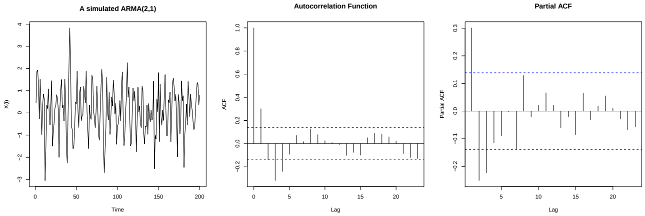


Figure : A simulated process: $X_t = 0.8X_{t-1} - 0.4X_{t-2} + \epsilon_t - 0.6\epsilon_{t-1}$

A Simulated ARMA (1,2) Series

```
R> set.seed(6436);par(mfrow=c(1,3))
R> arma12.sim<-arima.sim(n=200,list(order=c(1,0,2),
+ ar=0.8,ma=c(-0.6,0.4)))
R> plot(arma12.sim,ylab="X(t)",main="A simulated ARMA(1,2)"
R> acf(arma12.sim, main="Autocorrelation Function")
R> acf(arma12.sim, type="partial", main="Partial ACF")
```

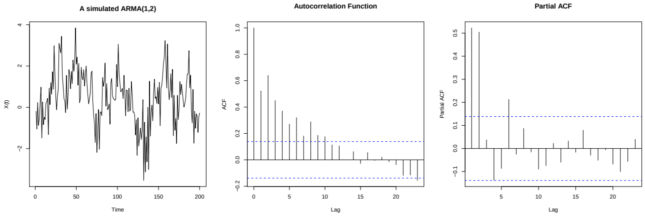


Figure : A simulated process: $X_t = 0.8X_{t-1} + \epsilon_t - 0.6\epsilon_{t-1} + 0.4\epsilon_{t-2}$

Some Examples of ARMA (p, q) Process

- AR (1) ≡ARMA (1,0) Process may be written as:

$$X_t = \phi X_{t-1} + \epsilon_t \text{ or } (1 - \phi \mathbf{B}) X_t = \epsilon_t$$

- AR (2) ≡ARMA (2,0) Process may be written as:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t \text{ or } (1 - \phi_1 \mathbf{B} - \phi_2 \mathbf{B}^2) X_t = \epsilon_t$$

- MA (1) ≡ARMA (0,1) Process may be written as:

$$X_t = \epsilon_t + \theta \epsilon_{t-1} \text{ or } X_t = (1 + \theta \mathbf{B}) \epsilon_t$$

- MA (2) ≡ARMA (0,2) Process may be written as:

$$X_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} \text{ or } X_t = (1 + \theta_1 \mathbf{B} + \theta_2 \mathbf{B}^2) \epsilon_t$$

Some Examples of ARMA (p, q) Process

- ARMA $(1,1)$ Process may be written as:

$$X_t = \phi X_{t-1} + \epsilon_t + \theta\epsilon_{t-1} \text{ or } (1-\phi\mathbf{B})X_t = (1+\theta\mathbf{B})\epsilon_t$$

- ARMA $(2,1)$ Process may be written as:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t + \theta\epsilon_{t-1} \text{ or } (1-\phi_1\mathbf{B}-\phi_2\mathbf{B}^2)X_t = (1+\theta\mathbf{B})\epsilon_t$$

- ARMA $(1,2)$ Process may be written as:

$$X_t = \phi X_{t-1} + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} \text{ or } (1-\phi\mathbf{B})X_t = (1+\theta_1\mathbf{B}+\theta_2\mathbf{B}^2)\epsilon_t$$

- ARMA $(2,2)$ Process may be written as:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} \text{ or } \Phi_2(\mathbf{B})X_t = \Theta_2(\mathbf{B})\epsilon_t,$$

$$\text{where } \Phi_2(\mathbf{B}) = 1 - \phi_1\mathbf{B} - \phi_2\mathbf{B}^2 \ \& \ \Theta_2(\mathbf{B}) = 1 + \theta_1\mathbf{B} + \theta_2\mathbf{B}^2$$

Autoregressive-Integrated-Moving Average (ARIMA) Models

Definition

A time series $\{X_t\}$, with zero mean, is called Autoregressive-Integrated-Moving Average of order (p, d, q), denoted by ARIMA (p, d, q) if the dth differences of the $\{X_t\}$ series is an ARMA (p, q) process.

The process in the backshift operator may be written as:

$$\Phi_p(\mathbf{B})(1-\mathbf{B})^d X_t = \Theta_q(\mathbf{B})\epsilon_t,$$

where $\Phi_p(\mathbf{B}) = 1 - \phi_1\mathbf{B} - \phi_2\mathbf{B}^2 - \ldots - \phi_p\mathbf{B}^p$ polynomial of degree p, $\Theta_q(\mathbf{B}) = 1 + \theta_1\mathbf{B} + \theta_2\mathbf{B}^2 + \ldots + \theta_q\mathbf{B}^q$ polynomial of degree q

Remarks

- ARIMA models are applied in some cases where data show evidence of non stationarity, where an initial differencing step can be applied one or more times to eliminate the non stationarity.
- AR (p) $\equiv$ARIMA $(p, 0, 0)$,
- MA (q) $\equiv$ARIMA $(0, 0, q)$,
- ARI (p, d) $\equiv$ARIMA $(p, d, 0)$,
- IMA (d, q) $\equiv$ARIMA $(0, d, q)$,
- ARMA (p, q) $\equiv$ARIMA $(p, 0, q)$,
- WN $\equiv$ARIMA (0,0,0),
- $I(d)$ $\equiv$ARIMA $(0, d, 0)$. Note that $I(1)$ is the well-known random walk, $X_t = X_{t-1} + \epsilon_t$, that is widely used to describe the behavior of the series of a stock price (as we will discuss later).

Some Examples of ARIMA (p, d, q) Process

- ARI (1,1)≡ARIMA (1,1,0) Process may be written as:

$$X_t = \phi X_{t-1} + X_{t-1} - \phi X_{t-2} + \epsilon_t \equiv (1 - \phi \mathbf{B})(1 - \mathbf{B})X_t = \epsilon_t$$

- ARIMA (1,1,1) Process may be written as:

$$(1 - \phi \mathbf{B})(1 - \mathbf{B})X_t = (1 + \theta \mathbf{B})\epsilon_t$$

- ARIMA (2,1,1) Process may be written as:

$$(1 - \phi_1 \mathbf{B} - \phi_2 \mathbf{B}^2)(1 - \mathbf{B})X_t = (1 + \theta \mathbf{B})\epsilon_t$$

- ARIMA (1,2,2) Process may be written as:

$$(1 - \phi \mathbf{B})(1 - \mathbf{B})^2 X_t = (1 + \theta_1 \mathbf{B} + \theta_2 \mathbf{B}^2)\epsilon_t$$

A Simulated ARIMA (2,1,1) Series

```
R> set.seed(6436);par(mfrow=c(1,3))
R> arima.sim<-arima.sim(n=200,list(order=c(2,1,1),
+ ar=c(0.8,-0.4),ma=-0.6))
R> plot(arima.sim,ylab="X(t)",main="A simulated ARIMA(2,1,1
R> acf(arima.sim, main="Autocorrelation Function")
R> acf(arima.sim, type="partial", main="Partial ACF")
```

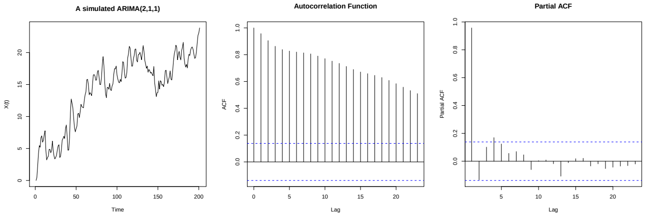


Figure : $(1-0.8\mathbf{B}+0.4\mathbf{B}^2)(X_t-X_{t-1})=\epsilon_t-0.6\epsilon_{t-1}$

Seasonal ARIMA (SARIMA) Models

A non-stationary process is often possess a seasonal component that repeats itself after a regular period of time, where the smallest time period denoted by s is called the seasonal period.

- Monthly observations: $s = 12$ (12 observations per a year).
- Quarterly observations: $s = 4$ (4 observations per a year).
- Daily observations: $s = 365$ (365 observations per a year).
- Weekly observations: $s = 52$ (52 observations per a year).
- Daily observations by week: $s = 5$ (5 working days).

Seasonal ARIMA (SARIMA) Models

Definition

The multiplicative Seasonal ARIMA (SARIMA) model denoted by ARIMA $(p, d, q) \times (P, D, Q)_s$, where s is the number of seasons:

$$\Phi_p(\mathbf{B})\boldsymbol{\Phi}_P(\mathbf{B}^s)(1-\mathbf{B})^d(1-\mathbf{B}^s)^D X_t = \Theta_q(\mathbf{B})\boldsymbol{\Theta}_Q(\mathbf{B}^s)\epsilon_t,$$

$\Phi_p(\mathbf{B}) = 1 - \sum_{i=1}^{p} \phi_i \mathbf{B}^i$ polynomial in $\mathbf{B}$ of degree p,
$\Theta_q(\mathbf{B}) = 1 + \sum_{i=1}^{q} \theta_i \mathbf{B}^i$ polynomial in $\mathbf{B}$ of degree q,
$\boldsymbol{\Phi}_P(\mathbf{B}^s) = 1 - \sum_{i=1}^{P} \Phi_i \mathbf{B}^{is}$ polynomial in $\mathbf{B}^s$ of degree P,
$\boldsymbol{\Theta}_Q(\mathbf{B}^s) = 1 + \sum_{i=1}^{Q} \Theta_i \mathbf{B}^{is}$ polynomial in $\mathbf{B}^s$ of degree Q,
with no common roots between $\boldsymbol{\Phi}_P(\mathbf{B}^s)$ and $\boldsymbol{\Theta}_Q(\mathbf{B}^s)$, p, d, and q are the order of non-seasonal AR model, MA model and ordinary differencing respectively, whereas P, D, and Q are the order of Seasonal Autoregressive (SAR) model, Seasonal Moving Average (SMA) model, and seasonal differencing respectively.

Remarks

- The idea is that SARIMA are ARIMA (p, d, q) models whose residuals ϵ_t are ARIMA (P, D, Q), whose operators are defined on $\mathbf{B}^s$ and successive powers, where p, q, and d are the non-seasonal AR order, non-seasonal MA order, and non-seasonal differencing respectively, while P, Q and D are the seasonal AR (SAR) order, seasonal MA (SMA) order, and seasonal differencing at lag s respectively.
- Seasonal differencing $\Delta_s X_t = (1 - \mathbf{B}^\mathbf{s})X_t = X_t - X_{t-s}$ will remove seasonality in the same way that ordinary differencing $\Delta X_t = X_t - X_{t-1}$ will remove a polynomial trend.

Some Examples of SARIMA $(p, d, q) \times (P, D, Q)_s$ Process

- ARIMA $(0, 1, 0) \times (0, 1, 0)_5$ model can be written as:

$$(1 - \mathbf{B})(1 - \mathbf{B}^5)X_t = \epsilon_t$$

- ARIMA $(0, 1, 0) \times (0, 1, 1)_4$ model can be written as:

$$(1 - \mathbf{B})(1 - \mathbf{B}^4)X_t = (1 + \Theta\mathbf{B}^4)\epsilon_t$$

- ARIMA $(1, 0, 0) \times (0, 1, 1)_{12}$ model can be written as:

$$(1 - \phi\mathbf{B})(1 - \mathbf{B}^{12})X_t = (1 + \Theta\mathbf{B}^{12})\epsilon_t$$

- ARIMA $(0, 1, 1) \times (0, 1, 1)_{12}$ model can be written as:

$$(1 - \mathbf{B})(1 - \mathbf{B}^{12})X_t = (1 + \theta\mathbf{B})(1 + \Theta\mathbf{B}^{12})\epsilon_t$$

Seasonal Patterns and Correlogram

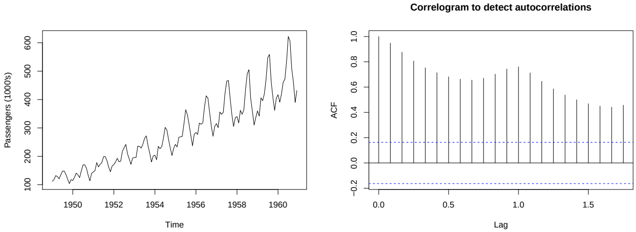
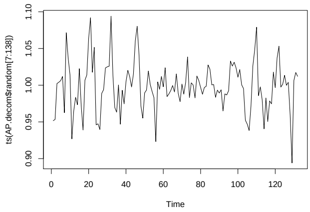


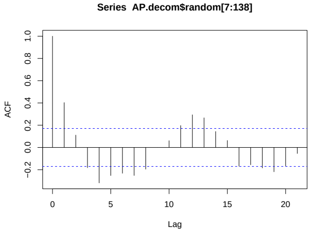
Figure : International air passenger bookings in US for years 1949 – 1960.

There is a clear seasonal variation. At the time, bookings were highest during the summer months June-August and lowest during the autumn month of November and winter month of February.

Remove Trend and Seasonality Using decompose()

```
R> par(mfrow=c(1,2))
R> AP.decom <- decompose(AirPassengers, "multiplicative")
R> plot(ts(AP.decom$random[7:138]))
R> acf(AP.decom$random[7:138]); par(mfrow=c(1,1))
```

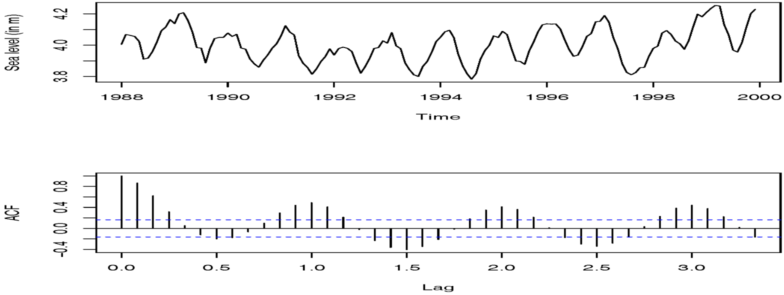




The correlogram suggests either MA (2) or that the seasonal adjustment has not been entirely effective. (Try SARIMA models!).

Example of SARIMA $(p, d, q) \times (P, D, Q)_s$ Process

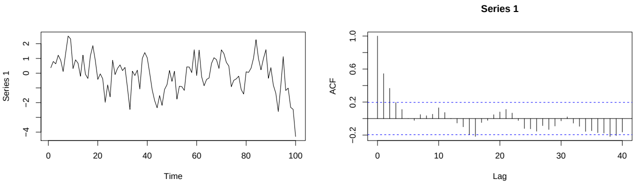
The plots of the average monthly sea level at Darwin, Australia from January 1988 to December 1999 (in metres), available from `http://www.sci.usq.edu.au/courses/STA3303/resources/climatology/darwinsl.txt`, show consistent seasonal pattern.



A Simulated SARIMA $(2,0,0)\times(0,1,1)_{12}$ Series

portes source: `http://site.iugaza.edu.ps/emahdi/courses/time-series-analysis-stat3305/lectures/`

```
R> library("portes")#Try sim_sarima in sarima package!
R> set.seed(13); phi<-c(1.3,-0.35); theta.season<-0.8
R> Z1<-varima.sim(list(ar=phi,d=0,ma.season=theta.season,
+     d.season=1,period=12),n=100,trunc.lag=50)
R> par(mfrow=c(1,2)); plot(Z1); acf(Z1,lag.max =40)
```



$$(1-1.3\mathbf{B}+0.5\mathbf{B}^2)(1-\mathbf{B}^{12})X_t=\epsilon_t+0.8\epsilon_{t-12}$$

Second order properties of MA (1) Models

For the MA (1) process: $X_t = \epsilon_t + \theta\epsilon_{t-1}$ where $\epsilon_t \sim WN(0, \sigma^2)$.
The mean of X_t: $\mathbb{E}(X_t) = 0$ (which is independent of t).
The variance: $\gamma_X(0) = \mathbb{E}(\epsilon_t^2 + 2\theta\epsilon_t\epsilon_{t-1} + \theta^2\epsilon_{t-1}^2) = (1+\theta^2)\sigma^2$.
The autocovariance and the autocorrelation are independent of t without any restrictions on the parameter θ as shown below:

$$\begin{aligned}\gamma_X(h) &= \mathbb{E}(\epsilon_t\epsilon_{t+h}) + \theta\mathbb{E}(\epsilon_{t-1}\epsilon_{t+h}) + \theta\mathbb{E}(\epsilon_t\epsilon_{t+h-1}) + \theta^2\mathbb{E}(\epsilon_{t-1}\epsilon_{t+h-1}) \\ &= \begin{cases} (1+\theta^2)\sigma^2, & h = 0 \\ \theta\sigma^2, & h = \pm 1, \\ 0, & h = \pm 2, \pm 3, \ldots \end{cases}\end{aligned}$$

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \begin{cases} 1, & h = 0 \\ \dfrac{\theta}{1+\theta^2}, & h = \pm 1 \\ 0, & h = \pm 2, \pm 3, \ldots \end{cases}$$

Second Order Properties of MA (q) Models

For MA (q) process: $X_t = \sum_{i=0}^{q} \theta_i \epsilon_{t-i}, \theta_0 = 1, \epsilon_t \sim WN(0, \sigma^2)$.

The mean of X_t: $\mathbb{E}(X_t) = 0$ (which is independent of t).

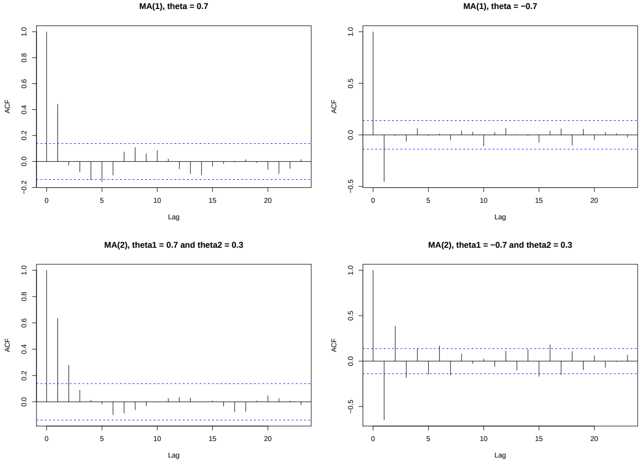
The variance: $\gamma_X(0) = (1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2)\sigma^2 = \sigma^2 \sum_{i=0}^{q} \theta_i^2$.

The autocovariance and the autocorrelation are independent of t without any restrictions on the parameter θ as shown below:

$$\gamma_X(h) = \begin{cases} \sigma^2 \sum_{i=h}^{q} \theta_i \theta_{i-h}, & h = 0, \pm 1, \pm 2, \ldots, \pm q \\ 0, & h > q \end{cases}$$

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \begin{cases} \frac{\sum_{i=h}^{q} \theta_i \theta_{i-h}}{\sum_{i=0}^{q} \theta_i^2}, & h = 0, \pm 1, \pm 2, \ldots, \pm q \\ 0, & h > q \end{cases}$$

ACF for Simulated MA (1) and MA (2) process



Example: The ACF of MA (q) Models

Example: Find the first and second moments of the process $X_t = \epsilon_t + 0.6\epsilon_{t-1} - 0.3\epsilon_{t-2}$, where $\epsilon \sim WN(0, \sigma^2)$.

Solution:

The process is MA (2) with $\theta_0 = 1, \theta_1 = 0.6$, and $\theta_2 = -0.3$

$$\begin{aligned}
\mathbb{E}(X_t) &= 0 \\
\mathbb{V}ar(X_t) &= \gamma_X(0) = (1 + \theta_1^2 + \theta_2^2)\sigma^2 = 1.45\sigma^2 \\
\gamma_X(1) &= \sigma^2 \sum_{i=1}^{2} \theta_i\theta_{i-1} = (\theta_1\theta_0 + \theta_2\theta_1)\sigma^2 = 0.42\sigma^2 \\
\gamma_X(2) &= \sigma^2 \sum_{i=2}^{2} \theta_i\theta_{i-2} = \theta_2\theta_0\sigma^2 = -0.30\sigma^2 \\
\rho_X(1) &= \frac{\gamma_X(1)}{\gamma_X(0)} = 0.27, \rho_X(2) = \frac{\gamma_X(2)}{\gamma_X(0)} = 0.19, \rho_X(h) = 0 \forall h > 2.
\end{aligned}$$

Invertibility of MA (q) Models

- If we replace θ by $1/\theta$, the autocorrelation function for MA (1), $\rho = \frac{\theta}{1+\theta^2}$, will not be changed. There are thus two processes, $X_t = \epsilon_t + \theta\epsilon_{t-1}$ and $X_t = \epsilon_t + \frac{1}{\theta}\epsilon_{t-1}$, show identical autocorrelation pattern, and hence the MA (1) coefficient is not uniquely identified.
- For the general moving average process, MA (q), there is a similar identifiability problem.
- The problem can be resolved by assuming that the operator $1 + \theta_1\mathbf{B} + \theta_2\mathbf{B}^2 + \ldots + \theta_q\mathbf{B}^q$ be invertible, i.e., all absolute *roots* of $\Theta_q(\mathbf{B})$: $1 + \theta_1\mathbf{B} + \theta_2\mathbf{B}^2 + \ldots + \theta_q\mathbf{B}^q = 0$ lie outside the unit circle.

Invertibility of MA (q) Models

Definition

The MA (q) process is invertible if it can be represented as a convergent infinite AR form AR (∞).

The MA (q), $X_t = \Theta_q(\mathbf{B})\epsilon_t$, where $\Theta_q(\mathbf{B}) = 1 + \sum_{j=1}^{q} \theta_j \mathbf{B}^q$ and $\epsilon_t \sim WN(0, \sigma^2)$, in an infinite AR representation AR (∞):

$$\Theta_q(\mathbf{B})^{-1} X_t = \Theta_q(\mathbf{B})^{-1} \Theta_q(\mathbf{B})\epsilon_t = \epsilon_t$$

$$\Rightarrow \Pi_\infty(\mathbf{B}) X_t = \epsilon_t,$$

where $\Pi_\infty(\mathbf{B}) = \Theta_q(\mathbf{B})^{-1} = 1 - \sum_{i=1}^{\infty} \pi_i \mathbf{B}^i = -\sum_{i=0}^{\infty} \pi_i \mathbf{B}^i$ with $\pi_0 = -1$, and π_i satisfy $\sum_{i=0}^{\infty} |\pi_i| < \infty$.
Note that the condition of finite sum $(\sum_{i=0}^{\infty} |\pi_i| < \infty)$ ensures that the AR (∞) series is convergent (i.e., MA (q) is invertible).

MA (1) in an Infinite AR Representation AR (∞)

Consider the MA (1) process $X_t = (1 + \theta \mathbf{B})\epsilon_t$. where we multiply both sides by $(1 + \theta \mathbf{B})^{-1}$ to get

$$\epsilon_t = \frac{1}{1 + \theta \mathbf{B}} X_t = \Pi_\infty(\mathbf{B}) X_t,$$

where $\Pi_\infty(\mathbf{B}) = (1 + \theta \mathbf{B})^{-1} = 1 - \sum_{i=1}^{\infty} \pi_i \mathbf{B}^i = -\sum_{i=0}^{\infty} \pi_i \mathbf{B}^i$. Recall that the Taylor series of $\dfrac{1}{1-x}$, $|x| < 1$ is the geometric series $1 + x + x^2 + x^3 + \ldots$, which implies to

$$\frac{1}{1 + \theta \mathbf{B}} = \frac{1}{1 - (-\theta \mathbf{B})} = 1 - \theta \mathbf{B} + \theta^2 \mathbf{B}^2 - \theta^3 \mathbf{B}^3 + \ldots = \sum_{i=0}^{\infty} (-\theta)^i \mathbf{B}^i$$

So that, (see the next slide)

MA (1) in an Infinite AR Representation AR (∞)

$$\sum_{i=0}^{\infty}(-\theta)^i\mathbf{B}^i = \sum_{i=0}^{\infty} -\pi_i\mathbf{B}^i.$$

By equating the coefficients of the corresponding powers $\mathbf{B}^i$ we get

$$\pi_i = (-1)^{i+1}(\theta)^i \ \forall i \geq 0.$$

$\Rightarrow$ MA (1) process $X_t = (1+\theta\mathbf{B})\epsilon_t$ in an AR (∞) representation is

$$X_t = \sum_{i=1}^{\infty}\pi_i\mathbf{B}^i X_t + \epsilon_t, \text{ where } \pi_i = (-1)^{i+1}\theta^i, i = 1, 2, \ldots.$$

Note that the absolute root of $1+\theta\mathbf{B} = 0$ implies that $|\frac{1}{\theta}| > 1$ or equivalently $|\theta| < 1$ ensures that the MA (1) is an invertible series.

Examples - Invertibility of MA (1) Models

1.) The MA (1) process $X_t = \epsilon_t + 0.4\epsilon_{t-1}$ is an invertible process because $|\theta| = |0.4| < 1$ (or equivalently the root of $(1 + 0.4\mathbf{B}) = 0$ in absolute value is $\mathbf{B} = 1/0.4 = 2.5 > 1$). This process can be written in the AR (∞) representation

$$X_t = \sum_{i=1}^{\infty} \pi_i X_{t-i} + \epsilon_t, \text{ where } \pi_i = (-1)^{i+1}(0.4)^i, i = 1, 2, \ldots,$$

$$X_t = \epsilon_t + 0.4X_{t-1} - (0.4)^2 X_{t-2} + (0.4)^3 X_{t-3} - \ldots$$

2.) The MA (1) process $X_t = \epsilon_1 + 1.8\epsilon_{t-1}$, is not invertible process because the root of $(1 + 1.8\mathbf{B}) = 0$ in absolute value is $\mathbf{B} = 1/1.8 = 0.56 < 1$. Thus, we can not write this process in an AR (∞) representation.

MA (q) in an Infinite AR Representation AR (∞)

Because $\Pi_\infty(\mathbf{B}) = \Theta_q(\mathbf{B})^{-1}$, the coefficients π_i can be obtained by equating the coefficients in the relation $\Theta_q(\mathbf{B})\Pi_\infty(\mathbf{B}) = 1$. Thus,

$$\begin{aligned}
\Theta_q(\mathbf{B})\Pi_\infty(\mathbf{B}) &= (1+\theta_1\mathbf{B}+\ldots+\theta_q\mathbf{B}^q)(1-\pi_1\mathbf{B}-\pi_2\mathbf{B}^2-\ldots) \\
&= 1-(\pi_1-\theta_1)\mathbf{B}-(\pi_2+\theta_1\pi_1-\theta_2)\mathbf{B}^2-\ldots \\
&- (\pi_j+\theta_1\pi_{j-1}+\ldots+\theta_{q-1}\pi_{j-q+1}+\theta_q\pi_{j-q})\mathbf{B}^j-\ldots
\end{aligned}$$

by equating coefficients of various powers $\mathbf{B}^j$ in the relation $\Theta_q(\mathbf{B})\Pi_\infty(\mathbf{B}) = 1$ for $j = 1, 2, \ldots$, we have

$$\pi_j = -\theta_1\pi_{j-1} - \theta_2\pi_{j-2} - \ldots - \theta_{q-1}\pi_{j-q+1} - \theta_q\pi_{j-q},$$

where $\pi_0 = -1$ and $\pi_j = 0$, for $j < 0$.

Example: MA (2) in an AR (∞) Representation

Example: An example of an invertible MA (2) process is

$$X_t = \epsilon_t - 0.1\epsilon_{t-1} + 0.42\epsilon_{t-2}.$$

The |roots| of $(1 - 0.1\mathbf{B} + 0.42\mathbf{B}^2) = (1 - 0.7\mathbf{B})(1 + 0.6\mathbf{B}) = 0$ are $\mathbf{B} = 1/0.7 = 1.43 > 1$ and $|\mathbf{B}| = |\frac{1}{-0.6}| = 1.67 > 1$. Thus, the process is invertible.

Now the process in an AR (∞) representation is written as follows:

$$X_t = \sum_{i=1}^{\infty} \pi_i X_{t-i} + \epsilon_t,$$

where $\pi_j = -\theta_1\pi_{j-1} - \theta_2\pi_{j-2} - \ldots - \theta_q\pi_{j-q}$, $q = 2$, $\theta_1 = -0.1, \theta_2 = 0.42, \theta_j = 0\ \forall j > q, \pi_0 = -1$, and $\pi_j = 0\ \forall j < 0$. (see the next slide).

Continue with the Previous Example

$\pi_j = -\theta_1 \pi_{j-1} - \theta_2 \pi_{j-2} - \ldots - \theta_q \pi_{j-q}$, where $q = 2$,
$\theta_1 = -0.1, \theta_2 = 0.42, \theta_j = 0 \ \forall j > q, \pi_0 = -1$, and $\pi_j = 0 \ \forall j < 0$.

$$\begin{aligned}
\pi_1 &= -\theta_1 \pi_0 - 0 - \ldots = -(-0.1)(-1) = -0.1 \\
\pi_2 &= -\theta_1 \pi_1 - \theta_2 \pi_0 = -(-0.1)(-0.1) - (0.42)(-1) = 0.41 \\
\pi_3 &= -\theta_1 \pi_2 - \theta_2 \pi_1 - 0 - \ldots \\
&= -(-0.1)(0.41) - (0.42)(-0.1) - 0 = 0.083 \\
\pi_4 &= -\theta_1 \pi_3 - \theta_2 \pi_2 - 0 - \ldots \\
&= -(-0.1)(0.083) - (0.42)(0.41) - 0 = -0.1639 \\
\vdots &= \vdots
\end{aligned}$$

Thus, the MA (2) process in the AR (∞) representation is

$$X_t = \epsilon_t - 0.1X_{t-1} + 0.41X_{t-2} + 0.083X_{t-3} - 0.1639X_{t-4} + \ldots$$

Example: MA (2) in an AR (∞) Representation

Example: Consider the MA (2) process

$$X_t = \epsilon_t - \epsilon_{t-1} + 0.5\epsilon_{t-2}.$$

Recall that the *roots* of the quadratic equation $ax^2 + bx + c = 0$ are $x = \dfrac{-b \pm \sqrt{b^2 - 4ac}}{2a}$. Thus |roots| of $1 - \mathbf{B} + 0.5\mathbf{B}^2 = 0$ are $|\mathbf{B}| = \dfrac{|-(-1) \pm \sqrt{(-1)^2 - 4(0.5)(1)}|}{2(0.5)} = |1 \pm i|$, where $i = \sqrt{-1}$.
Recall that the absolute value of a complex number $|a + ib| = \sqrt{a^2 + b^2}$, so that the absolute values of the two roots are $\sqrt{2} \approx 1.41 > 1$; Hence, the process is invertible. (see the next slide)

Continue with the Previous Example

Now the process in an AR (∞) representation is written as follows:

$$X_t = \sum_{i=1}^{\infty} \pi_i X_{t-i} + \epsilon_t,$$

$\pi_j = -\theta_1 \pi_{j-1} - \theta_2 \pi_{j-2}$, $\theta_1 = -1, \theta_2 = 0.5, \pi_0 = -1$ & $\pi_{-1} = 0$.

$$\begin{aligned}
\pi_1 &= -\theta_1\pi_0 = -(-1)(-1) = -1 \\
\pi_2 &= -\theta_1\pi_1 - \theta_2\pi_0 = -(-1)(-1) - (0.5)(-1) = -0.5 \\
\pi_3 &= -\theta_1\pi_2 - \theta_2\pi_1 = -(-1)(-0.5) - (0.5)(-1) = 0 \\
\pi_4 &= -\theta_1\pi_3 - \theta_2\pi_2 = -(-1)(0) - (0.5)(-0.5) = 0.25 \\
\vdots &= \vdots
\end{aligned}$$

Thus, the MA (2) process in the AR (∞) representation is

$$X_t = \epsilon_t - X_{t-1} - 0.5X_{t-2} + 0.25X_{t-4} + \dots$$

Finding the Roots in R

- In R, the function `polyroot(a)`, where a is the vector of polynomial coefficients in increasing order can be used to find the zeros of a real or complex polynomial of degree $n-1$
$$p(x) = a_1 + a_2 x + \ldots + a_n x^{n-1}.$$
- The function `Mod()` can be used to check whether the roots of $p(x)$ have modulus > 1 or not.

Example

The MA (4) process
$X_t = \epsilon_t - 0.3\epsilon_{t-1} + 0.7\epsilon_{t-2} - 1.2\epsilon_{t-3} + 0.1\epsilon_{t-4}$ is not invertible as

```
R> roots <- polyroot(c(1,-0.3, 0.7,-1.2,0.1))
R> Mod(roots)

[1]  0.8829242  0.8829242  1.1250090 11.4024243
```

AR (1) in MA (∞) Representation

Let X_t is an AR (1) process, where $\epsilon_t \sim WN(0, \sigma^2)$.

$$\begin{aligned} X_t &= \phi X_{t-1} + \epsilon_t \\ X_t &= \phi(\phi X_{t-2} + \epsilon_{t-1}) + \epsilon_t \\ &= \epsilon_t + \phi\epsilon_{t-1} + \phi^2 X_{t-2} \\ &= \epsilon_t + \phi\epsilon_{t-1} + \phi^2(\phi X_{t-3} + \epsilon_{t-2}) \\ &= \epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \phi^3 X_{t-3} \\ \vdots &= \vdots \\ X_t &= \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j} \end{aligned}$$

which is in the MA (∞) representation. Note that (apart of ϵ_{t-j}) this representation may be seen as a geometric series, which is converge if and only if $|\phi| < 1$, (see the next slide).

AR (1) in MA (∞) Representation - Second Order Properties of AR (1) Models

The AR (1) process with backshift operator may be written as $(1-\phi\mathbf{B})X_t = \epsilon_t$, where $\epsilon_t \sim WN(0,\sigma^2) \Rightarrow X_t = \dfrac{\epsilon_t}{1-\phi\mathbf{B}}$.
If $|\phi\mathbf{B}| < 1$, Taylor series of $\dfrac{1}{1-\phi\mathbf{B}} = 1 + \phi\mathbf{B} + \phi^2\mathbf{B}^2 + \dots$

$$\Rightarrow X_t = (1 + \phi\mathbf{B} + \phi^2\mathbf{B}^2 + \phi^3\mathbf{B}^3 + \dots)\epsilon_t = \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j}$$

Which is in the MA (∞) Wold representation. (See the next slide.)

The AR(1) process in MA(∞) representation: $X_t = \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}$

The mean of X_t: $\mathbb{E}(X_t) = \sum_{i=0}^{\infty} \phi^i \mathbb{E}(\epsilon_{t-i}) = 0$ (independent of t).

The variance: $\gamma_X(0) = \mathbb{E}(\sum_{i=0}^{\infty} \phi^{2i} \epsilon_{t-i}^2) = \sigma^2 \sum_{i=0}^{\infty} \phi^{2i} = \frac{\sigma^2}{1-\phi^2}$.

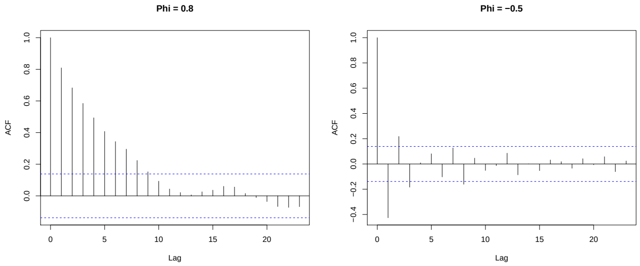
The autocovariance is independent of t as follows:

$$\begin{aligned}
\gamma_X(h) &= \mathbb{E}(X_t X_{t+h}) = \sum_{i=0}^{\infty} \sum_{j=h}^{\infty} \phi^i \phi^j \mathbb{E}(\epsilon_{t-i} \epsilon_{t+h-j}) \\
&= \sigma^2 \sum_{i=0}^{\infty} \phi^i \phi^{i+h} = \phi^h \sigma^2 \sum_{i=0}^{\infty} \phi^{2i} = \sigma^2 \frac{\phi^h}{1-\phi^2}
\end{aligned}$$

The autocorrelation: $\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \phi^h$ (independent of t).

Autocorrelation Function of AR(1) Model- Example

```
R> set.seed(125)
R> ar.1<-arima.sim(n=200,list(order=c(1,0,0),ar=0.8))
R> ar.2<-arima.sim(n=200,list(order=c(1,0,0),ar=-0.5))
R> par(mfrow=c(1,2))
R> acf(ar.1,main="Phi = 0.8"); acf(ar.2,main="Phi = -0.5")
R> par(mfrow=c(1,1))
```



Example: The Autocorrelation of AR (1)

Example: The autocorrelation of the stationary AR (1) process:

$$X_t = 0.6X_{t-1} + \epsilon_t, \epsilon_t \sim WN(0, 4)$$

are

$$\rho_X(h) = \phi^h = 0.6^h \; \forall h \geq 0,$$

Thus $\rho_X(0) = 1, \rho_X(1) = 0.6, \rho_X(2) = 0.6^2 = 0.36, \ldots$.

```
R> Lag <- 0:8
R> rho <- 0.6^Lag
R> round(rho,3)
```

```
[1] 1.000 0.600 0.360 0.216 0.130 0.078 0.047 0.028 0.017
```

Second Order Properties of AR (p) Models

Consider the AR (p) model:

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \epsilon_t, \text{ where } \epsilon_t \sim WN(0, \sigma^2),$$

- The expectation of X_t is $\mathbb{E}(X_t) = 0$ (Why?).
- The variance of X_t is $\mathbb{V}ar(X_t) = \gamma_X(0) = \sum_{i=1}^{p} \phi_i \gamma_X(i) + \sigma^2$, where $\gamma_X(i) = \mathbb{C}ov(X_t, X_{t-i})$. Note that the variance of X_t is obtained by multiplying both sides of the above equation by X_t and taking expectation, where $\mathbb{E}(X_t \epsilon_t) = \sigma^2$ (Why?).

(see the next slide)!

Second Order Properties of AR(p) Models (Cont.)

Multiplying both sides of the equation $X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \epsilon_t$ by X_{t-h} and taking expectations gives

$$\gamma_X(h) = \sum_{i=1}^{p} \phi_i \gamma_X(h-i) \; \forall h > 0$$

Now divide both sides by $\gamma_X(0)$ to get the autocorrelations

$$\rho_X(h) = \sum_{i=1}^{p} \phi_i \rho_X(h-i) \; \forall h > 0$$

These are the Yule-Walker equations that we will discuss later (see Partial Autocorrelation Function (PACF) Section).

The population autocorrelations $\rho_X(h)$ can be found recursively by solving the Yule-Walker equations given the values of ϕ_i. In practice, as we will see later, we use the sample autocorrelation coefficients to estimate the values of ϕ_i.

Stationarity of AR (1) Models

Definition

The condition for the Autoregressive process of order 1, AR (1): $X_t = \phi X_{t-1} + \epsilon_t$, to be stationary is that the absolute value of the root of $(1 - \phi \mathbf{B}) = 0$ must lie outside the unit circle. That is, the AR (1) is stationary if $|\mathbf{B}| = |\frac{1}{\phi}| > 1$, or equivalently, if $|\phi| < 1$,

- $(1 - 0.4\mathbf{B})X_t = \epsilon_t$ is a stationary process because the absolute value of the root of $(1 - 0.4\mathbf{B}) = 0$ is $|\mathbf{B}| = |1/0.4| = 2.5 > 1$.
- $(1 + 1.8\mathbf{B})X_t = \epsilon_t$ is not stationary process because $|\text{root}|$ of $(1 + 1.8\mathbf{B}) = 0$ is $|\mathbf{B}| = |\frac{1}{-1.8}| = 0.56 < 1$.
- $X_t = 0.5X_{t-1} + \epsilon_t$ is a stationary process because $|0.5| < 1$.

Stationarity of AR (2) Models

Theorem

For the AR (2) model, $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t$, the necessary and sufficient conditions for stationarity are:

$$\begin{aligned} |\phi_2| &< 1 \\ \phi_1 + \phi_2 &< 1 \\ \phi_2 - \phi_1 &< 1 \end{aligned}$$

Examples

- $X_t = 1.1X_{t-1} - 0.4X_{t-2} + \epsilon_t$ is stationary.
- $X_t = 0.6X_{t-1} - 1.3X_{t-2} + \epsilon_t$ is not stationary $(|\phi_2| \not< 1)$.
- $X_t = 0.6X_{t-1} + 0.8X_{t-2} + \epsilon_t$ is not stationary $(\phi_1 + \phi_2 \not< 1)$.
- $X_t = -0.4X_{t-1} + 0.7X_{t-2} + \epsilon_t$ is not stationary $(\phi_2 - \phi_1 \not< 1)$.

Stationarity of AR(p) Models

Definition

The condition for the Autoregressive process of order p, AR(p): $X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \epsilon_t$, to be stationary is that the absolute values of the roots of $(1 - \phi_1 \mathbf{B} - \phi_2 \mathbf{B}^2 - \ldots - \phi_p \mathbf{B}^p) = 0$ must lie outside the unit circle.

- $X_t = -0.4X_{t-1} + 0.21X_{t-2} + \epsilon_t$ is a stationary process as $|\text{roots}|$ of $(1 + 0.4\mathbf{B} - 0.21\mathbf{B}^2) = (1 - 0.3\mathbf{B})(1 + 0.7\mathbf{B}) = 0$ are $|\frac{1}{0.3}| = 3.3 > 1$ and $|\frac{1}{-0.7}| = 1.4 > 1$.
- $X_t = 1.7X_{t-1} - 0.42X_{t-2} + \epsilon_t$ is not a stationary process because the absolute value of one of the two roots of $(1 - 1.7\mathbf{B} + 0.42\mathbf{B}^2) = (1 - 0.3\mathbf{B})(1 - 1.4\mathbf{B}) = 0$ lies inside the unit circle, which is $|\frac{1}{1.4}| = 0.71 < 1$.
- $X_t = -0.25X_{t-2} + \epsilon_t$ is a stationary process because $|\text{roots}|$ of $1 + 0.25\mathbf{B}^2 = 0$ are $|\pm 2i| = 2 > 1$, where $i = \sqrt{-1}$.

Causality of AR (p) in an Infinite MA (∞) Representation

The AR (p), $\Phi_p(\mathbf{B})X_t = \epsilon_t$, where $\Phi_p(\mathbf{B}) = 1 - \sum_{j=1}^{p} \phi_j \mathbf{B}^p$ and $\epsilon_t \sim WN(0, \sigma^2)$, is causal process if it can be written in an infinite MA representation MA (∞):

$$\Phi_p(\mathbf{B})^{-1}\Phi_p(\mathbf{B})X_t = \Phi_p(\mathbf{B})^{-1}\epsilon_t$$

$$\Rightarrow X_t = \Phi_p(\mathbf{B})^{-1}\epsilon_t = \Psi_\infty(\mathbf{B})\epsilon_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$$

where $\Psi_\infty(\mathbf{B}) = \Phi_p(\mathbf{B})^{-1}$, $\Psi_\infty(\mathbf{B}) = 1 + \sum_{i=1}^{\infty} \psi_i \mathbf{B}^i$, ψ_i satisfy $\sum_{i=0}^{\infty} |\psi_i| < \infty$ with $\psi_0 = 1$. (see the next slide).

AR (p) in an Infinite MA (∞) Representation

Because $\Psi_\infty(\mathbf{B}) = \Phi_p(\mathbf{B})^{-1}$, the coefficients ψ_i can be obtained by equating the coefficients in the relation $\Phi_p(\mathbf{B})\Psi_\infty(\mathbf{B}) = 1$. Thus,

$$\begin{aligned} \Phi_p(\mathbf{B})\Psi_\infty(\mathbf{B}) &= (1 - \phi_1\mathbf{B} - \ldots - \phi_p\mathbf{B}^p)(1 + \psi_1\mathbf{B} + \psi_2\mathbf{B}^2 + \ldots) \\ &= 1 + (\psi_1 - \phi_1)\mathbf{B} + (\psi_2 - \phi_1\psi_1 - \phi_2)\mathbf{B}^2 + \ldots \\ &+ (\psi_j - \phi_1\psi_{j-1} - \ldots - \phi_p\psi_{j-p})\mathbf{B}^j + \ldots \end{aligned}$$

by equating coefficients of various powers $\mathbf{B}^j$ in the relation $\Phi_p(\mathbf{B})\Psi_\infty(\mathbf{B}) = 1$ for $j = 1, 2, \ldots$, we have

$$\psi_j = \phi_1\psi_{j-1} + \phi_2\psi_{j-2} + \ldots + \phi_{p-1}\psi_{j-p+1} + \phi_p\psi_{j-p},$$

where $\psi_0 = 1$ and $\psi_j = 0$, for $j < 0$.

Example: AR (p) in an Infinite MA (∞) Representation

Example: An example of a stationary AR (2) process:

$$X_t = 0.1X_{t-1} - 0.42X_{t-2} + \epsilon_t$$

where $p = 2, \phi_1 = 0.1, \phi_2 = -0.42, \phi_j = 0 \; \forall j > p$.
The |roots| of $(1 - 0.1\mathbf{B} + 0.42\mathbf{B}^2) = (1 - 0.7\mathbf{B})(1 + 0.6\mathbf{B}) = 0$ are $\mathbf{B} = 1/0.7 = 1.43 > 1$ and $|\mathbf{B}| = |\frac{1}{-0.6}| = 1.67 > 1$. Thus, the process is stationary-causal.
Now the process in a MA (∞) representation is written as follows:

$$X_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-1},$$

where $\psi_j = \phi_1\psi_{j-1} + \phi_2\psi_{j-2} + \ldots + \phi_{p-1}\psi_{j-p+1} + \phi_p\psi_{j-p}$, with $\psi_0 = 1$ and $\psi_j = 0 \; \forall j < 0$. (see the next slide).

Continue with the Previous Example

Use $\psi_j = \phi_1\psi_{j-1} + \phi_2\psi_{j-2} + \ldots + \phi_p\psi_{j-p}$, where $p = 2$, $\phi_1 = 0.1, \phi_2 = -0.42, \phi_j = 0\ \forall j > p, \psi_0 = 1$ and $\psi_j = 0\ \forall j < 0$,

$$\begin{aligned}
\psi_1 &= \phi_1\psi_0 = (0.1)(1) = 0.1 \\
\psi_2 &= \phi_1\psi_1 + \phi_2\psi_0 = (0.1)(0.1) + (-0.42)(1) = -0.41 \\
\psi_3 &= \phi_1\psi_2 + \phi_2\psi_1 \\
&= (0.1)(-0.41) + (-0.42)(0.1) + 0 = -0.083 \\
\psi_4 &= \phi_1\psi_3 + \phi_2\psi_2 \\
&= (0.1)(-0.083) + (-0.42)(-0.41) + 0 = 0.1639 \\
\vdots &= \vdots
\end{aligned}$$

Thus, the AR(2) process in the MA(∞) representation:

$$X_t = \epsilon_t + 0.1\epsilon_{t-1} - 0.41\epsilon_{t-2} - 0.083\epsilon_{t-3} + 0.1639\epsilon_{t-4} + \ldots$$

Impulse Response Sequence of ARMA (p, q) Models

Recall that the ARMA (p, q) models may be represented as

$$\Phi_p(\mathbf{B})X_t = \Theta_q(\mathbf{B})\epsilon_t,$$

$$\Phi_p(\mathbf{B}) = 1 - \sum_{i=1}^{p} \phi_i \mathbf{B}^i \text{ and } \Theta_q(\mathbf{B}) = 1 + \sum_{j=1}^{q} \theta_j \mathbf{B}^j.$$

Under the stationarity conditions, we have the MA (∞) form

$$X_t = \Psi_\infty(\mathbf{B})\epsilon_t,$$

where

$$\Psi_\infty(\mathbf{B}) = \Phi_p(\mathbf{B})^{-1}\Theta_q(\mathbf{B}) = \sum_{i=0}^{\infty} \psi_i \mathbf{B}^i,$$

Ψ is known as the impulse response sequence.

(see next page)

Impulse Response Sequence of ARMA (p, q) Models

Similar to the pure AR model situation, the coefficients ψ_i can be derived by equating the coefficients of the relation

$$\Phi_p(\mathbf{B})\Psi_\infty(\mathbf{B}) = \Theta_q(\mathbf{B}).$$

The ψ_j satisfy

$$\psi_j = \phi_1\psi_{j-1} + \phi_2\psi_{j-2} + \ldots + \phi_p\psi_{j-p} + \theta_j,\ j = 1, 2, \ldots.$$

where $\psi_0 = 1, \psi_j = 0 \forall j < 0$, and $\theta_j = 0 \forall j > q$ (see next page)

Impulse Response Sequence of ARMA (p, q) Models

Under the invertibility conditions, we have the convergent casual infinite AR representation as

$$\epsilon_t = \Pi_\infty(\mathbf{B})X_t,$$

where

$$\Pi_\infty(\mathbf{B}) = \Theta_q(\mathbf{B})^{-1}\Phi_p(\mathbf{B}) = 1 - \sum_{i=1}^{\infty} \pi_i \mathbf{B}^i.$$

Similar to the pure MA model situation, the coefficients π_i can be determined from the equating coefficients of relation

$$\Theta_q(\mathbf{B})\Pi_\infty(\mathbf{B}) = \Phi_p(\mathbf{B}).$$

The π_j satisfy

$$\pi_j = -\theta_1\pi_{j-1} - \theta_2\pi_{j-2} - \ldots - \theta_q\pi_{j-q} + \phi_j, \; j = 1, 2, \ldots$$

where $\pi_0 = -1, \pi_j = 0 \forall j < 0$, and $\phi_j = 0 \forall j > p$.

ACF of ARMA(1,1) Models

Consider the casual ARMA(1,1) model:

$$X_t = \phi X_{t-1} + \epsilon_t + \theta\epsilon_{t-1}, \text{ where } |\phi| < 1 \Rightarrow$$

$$\begin{aligned}
\gamma(h) &= \mathbb{C}ov(X_{t+h}, X_t) = \mathbb{E}(X_{t+h}X_t) \\
&= \mathbb{E}[(\phi X_{t+h-1} + \epsilon_{t+h} + \theta\epsilon_{t+h-1})X_t] \\
&= \phi\mathbb{E}(X_{t+h-1}X_t) + \mathbb{E}(\epsilon_{t+h}X_t) + \theta\mathbb{E}(\epsilon_{t+h-1}X_t) \qquad (1)
\end{aligned}$$

Recall that $X_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$, where the coefficients ψ_j can be calculated from the relation $\Psi_\infty(\mathbf{B}) = \Phi_p(\mathbf{B})^{-1}\Theta_q(\mathbf{B}) = \dfrac{1+\theta\mathbf{B}}{1-\phi\mathbf{B}}$ (see impulse response sequence of ARMA(p,q) models), where $\psi_0 = 1$ and $\psi_1 = \phi + \theta$ for the ARMA(1,1) case.

ACF of ARMA(1, 1) Models

$$\begin{aligned}
\mathbb{E}(\epsilon_{t+h}X_t) &= \mathbb{E}(\epsilon_{t+h}\sum_{j=0}^{\infty}\psi_j\epsilon_{t-j}) = \sum_{j=0}^{\infty}\psi_j\mathbb{E}(\epsilon_{t+h}\epsilon_{t-j}) \\
&= \begin{cases} \psi_0\sigma^2, & \text{for } h = 0; \\ 0, & \text{for } h \geq 1. \end{cases} \qquad (2)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}(\epsilon_{t+h-1}X_t) &= \mathbb{E}(\epsilon_{t+h-1}\sum_{j=0}^{\infty}\psi_j\epsilon_{t-j}) = \sum_{j=0}^{\infty}\psi_j\mathbb{E}(\epsilon_{t+h-1}\epsilon_{t-j}) \\
&= \begin{cases} \psi_1\sigma^2, & \text{for } h = 0; \\ \psi_0\sigma^2, & \text{for } h = 1; \\ 0, & \text{for } h \geq 2. \end{cases} \qquad (3)
\end{aligned}$$

Furthermore, $\psi_0 = 1$ and $\psi_1 = \phi + \theta$.

ACF of ARMA(1, 1) Models

Putting all Equations (1-3) together we obtain

$$\begin{aligned}
\gamma(h) &= \phi\mathbb{E}(X_{t+h-1}X_t) + \mathbb{E}(\epsilon_{t+h}X_t) + \theta\mathbb{E}(\epsilon_{t+h-1}X_t) \\
&= \begin{cases} \phi\gamma(1) + \sigma^2(1+\phi\theta+\theta^2), & \text{for } h = 0; \\ \phi\gamma(0) + \sigma^2\theta, & \text{for } h = 1; \\ \phi\gamma(h-1), & \text{for } h \geq 2. \end{cases}
\end{aligned}$$

Note that $\gamma(h) = \phi\gamma(h-1), h \geq 2$ has an iterative form

$$\begin{aligned}
\gamma(2) &= \phi\gamma(1) \\
\gamma(3) &= \phi\gamma(2) = \phi^2\gamma(1) \\
\dots &= \dots \\
\gamma(h) &= \phi^{h-1}\gamma(1),
\end{aligned}$$

with initial conditions:

$$\begin{cases} \gamma(0) = \phi\gamma(1) + \sigma^2(1+\phi\theta+\theta^2) \\ \gamma(1) = \phi\gamma(0) + \sigma^2\theta \end{cases}$$

ACF of ARMA (1, 1) Models

Solving for $\gamma(0)$ and $\gamma(1)$ we obtain

$$\begin{aligned}\gamma(0) &= \sigma^2 \frac{(1 + 2\phi\theta + \theta^2)}{1 - \phi^2} \\ \gamma(1) &= \sigma^2 \frac{(1 + \phi\theta)(\phi + \theta)}{1 - \phi^2}\end{aligned}$$

This gives us

$$\gamma(h) = \sigma^2 \frac{(1 + \phi\theta)(\phi + \theta)}{1 - \phi^2} \phi^{h-1}, \text{ for } h \geq 1$$

Finally dividing by $\gamma(0)$ to get the ACF of ARMA (1,1)

$$\rho(h) = \frac{(1 + \phi\theta)(\phi + \theta)}{1 + 2\phi\theta + \theta^2} \phi^{h-1}, \text{ for } h \geq 1$$

An ARMA (1,1) Simulated Process - Example

```
R> set.seed(13675)
R> sim<-arima.sim(n=200,list(order=c(1,0,1),ar=0.9,ma=0.5))
R> par(mfrow=c(1,2))
R> plot(sim,ylab=""); acf(sim,lag.max=85)
R> par(mfrow=c(1,1))
```

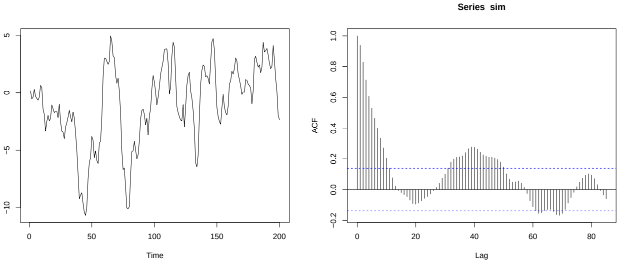


Figure : A simulated ARMA (1,1) process: $X_t = 0.9X_{t-1} + \epsilon_t + 0.5\epsilon_{t-1}$

ACF for ARMA (p, q) Models

Assume that the ARMA (p, q) model

$$\Phi_p(\mathbf{B})X_t = \Theta_q(\mathbf{B})\epsilon_t,$$

is causal-stationary, that is the roots of $\Phi_p(\mathbf{B})$ are outside the unit circle. Then we can write

$$X_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j},$$

where the coefficients ψ_j can be calculated from the relation $\Psi_\infty(\mathbf{B}) = \Phi_p(\mathbf{B})^{-1}\Theta_q(\mathbf{B})$. It follows that $\mathbb{E}(X_t) = 0$.

- Note that X_t and ϵ_{t+j} are uncorrelated for $j > 0$ and $\psi_j = 0$ for $j < 0$.
- As in the case for ARMA (1,1), We can obtain a homogeneous differential equation in terms of $\gamma(h)$ with some initial conditions as follows

ACF for ARMA (p, q) Models

For an ARMA model $X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \sum_{j=0}^{q} \theta_j \epsilon_{t-j}$, with $\theta_0 = 1$,

$$\begin{aligned}
\gamma(h) &= \mathbb{C}ov(X_{t+h}, X_t) = \mathbb{E}(X_{t+h}X_t) \\
&= \mathbb{E}[(\sum_{j=1}^{p} \phi_j X_{t+h-j} + \sum_{j=0}^{q} \theta_j \epsilon_{t+h-j})X_t] \\
&= \sum_{j=1}^{p} \phi_j \mathbb{E}[X_{t+h-j}X_t] + \sum_{j=0}^{q} \theta_j \mathbb{E}\left[\epsilon_{t+h-j} \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}\right] \\
&= \sum_{j=1}^{p} \phi_j \gamma(h-j) + \sigma^2 \sum_{j=h}^{q} \theta_j \psi_{j-h}, \text{ for } h \geq 0. \qquad (4)
\end{aligned}$$

This gives the general homogeneous difference equation for $\gamma(h)$.
(see the next slide!)

ACF for ARMA (p, q) Models

$$\gamma(h) - \phi_1\gamma(h-1) - \cdots - \phi_p\gamma(h-p) = 0, \text{ for } h \geq \max(p, q+1)$$

with initial conditions

$$\gamma(h) - \sum_{j=1}^{p} \phi_j\gamma(h-j) = \sigma^2 \sum_{j=h}^{q} \theta_j\psi_{j-h}, \text{ for } 0 \leq h \leq \max(p, q+1).$$

Dividing these two equations through by $\gamma(0)$ will allow us to solve for the ACF of ARMA (p, q) models, $\rho(h) = \gamma(h)/\gamma(0)$.

$$\rho(h) - \phi_1\rho(h-1) - \cdots - \phi_p\rho(h-p) = 0, \text{ for } h \geq \max(p, q+1)$$

with initial conditions

$$\rho(h) - \sum_{j=1}^{p} \phi_j\rho(h-j) = \frac{\sigma^2}{\gamma(0)} \sum_{j=h}^{q} \theta_j\psi_{j-h}, \text{ for } 0 \leq h \leq \max(p, q+1).$$

ACF of an AR (p) Model

For a causal AR (p), it follows from the previous slide that

$$\rho(h) - \phi_1 \rho(h-1) - \cdots - \phi_p \rho(h-p) = 0, \ \text{for } h \geq p.$$

with initial conditions

$$\begin{cases} \rho(0) - \phi_1\rho(-1) - \phi_2\rho(-2) - \cdots - \phi_p\rho(-p) & = \dfrac{\sigma^2}{\gamma(0)} \\ \rho(1) - \phi_1\rho(0) - \phi_2\rho(-1) - \cdots - \phi_p\rho(1-p) & = 0 \\ \rho(2) - \phi_1\rho(1) - \phi_2\rho(0) - \cdots - \phi_p\rho(2-p) & = 0 \\ \vdots & \vdots \\ \rho(p-1) - \phi_1\rho(p-2) - \phi_2\rho(p-3) - \cdots - \phi_p\rho(-1) & = 0 \\ \rho(p) - \phi_1\rho(p-1) - \phi_2\rho(p-2) - \cdots - \phi_p\rho(0) & = 0 \end{cases}$$

where $\rho(-h) = \rho(h)$, for $h = 1, 2, \ldots, p$.

Example - ACF of an AR (2) Model

Consider the case AR (2). Then

$$\rho(h) - \phi_1\rho(h-1) - \phi_2\rho(h-2) = 0, \text{ for } h \geq 2.$$

with initial conditions

$$\begin{cases} \rho(0) - \phi_1\rho(-1) - \phi_2\rho(-2) = \sigma^2/\gamma(0) \text{ (what is the value of } \gamma(0)\text{?)} \\ \rho(1) - \phi_1\rho(0) - \phi_2\rho(-1) = 0 \\ \rho(2) - \phi_1\rho(1) - \phi_2\rho(0) = 0 \text{ (you may neglect this condition!)} \end{cases}$$

where $\rho(-h) = \rho(h)$, $\forall h$. Hence,

$$\begin{cases} \rho(0) = 1 \\ \rho(1) = \frac{\phi_1}{1-\phi_2} \\ \rho(h) = \phi_1\rho(h-1) + \phi_2\rho(h-2), \text{ for } h \geq 2. \end{cases}$$

Cramer's Rule

Definition

Consider the system of n linear equations for n unknowns, represented in matrix multiplication form: $AX = \mathbf{b}$, where the $n \times n$ matrix A has a nonzero determinant, and the vector $X = (x_1, x_2, \ldots, x_n)^T$ is the column vector of the unknowns variables. Then the system has a unique solution, whose individual values for the unknowns are given by

$$x_i = \frac{det(A_i)}{det(A)}, i = 1, 2, \ldots, n,$$

where A_i is the matrix formed by replacing the i-th column of A by the column vector $\mathbf{b}$.

Partial Autocorrelation Function (PACF)

- The partial autocorrelation function (PACF), $\phi_{h,h}$, is a useful tool used in identifying the order of the AR(p) process.
- The partial autocorrelation measures the correlation between two random variables X_t and X_{t+h} at different lags h after removing linear dependence of X_{t+1} through X_{t+h-1}: The PACF thus represents the sequence of conditional correlations:

$$\phi_{h,h} = Corr(X_t, X_{t+h}|X_{t+1}, \ldots, X_{t+h-1}), h = 1, 2, \ldots$$

(see the next slide for the definition of the conditional correlation!)

- The autocorrelation function (ACF) between two random variables X_t and X_{t+h} at different lags h does not adjust for the influence of the intervening lags: The ACF thus represents the sequence of unconditional correlations.

Partial Autocorrelation Function (PACF)

$$\begin{aligned}
\phi_{h,h} &= Corr(X_t, X_{t+h}|X_{t+1}, \ldots, X_{t+h-1}) \\
&= \frac{\mathbb{C}ov[(X_t|X_{t+1}, \ldots, X_{t+h-1}), (X_{t+h}|X_{t+1}, \ldots, X_{t+h-1})]}{\sqrt{\mathbb{V}ar(X_t|X_{t+1}, \ldots, X_{t+h-1})}\sqrt{\mathbb{V}ar(X_{t+h}|X_{t+1}, \ldots, X_{t+h-1})}} \\
&= \frac{\mathbb{C}ov[(X_t - \hat{X}_t), (X_{t+h} - \hat{X}_{t+h})]}{\sqrt{\mathbb{V}ar(X_t - \hat{X}_t)}\sqrt{\mathbb{V}ar(X_{t+h} - \hat{X}_{t+h})}},
\end{aligned}$$

where

$$\hat{X}_t = \alpha_1 X_{t+1} + \alpha_2 X_{t+2} + \ldots + \alpha_{h-1} X_{t+h-1},$$
$$\hat{X}_{t+h} = \beta_1 X_{t+1} + \beta_2 X_{t+2} + \ldots + \beta_{h-1} X_{t+h-1},$$

and $\alpha_i,\ \beta_i$ $(1 \leq i \leq h-1)$ are the mean squared linear regression coefficients obtained from minimizing the $\mathbb{E}(X_t - \hat{X}_t)^2$ and $\mathbb{E}(X_{t+h} - \hat{X}_{t+h})^2$ respectively.

Yule-Walker Equations and PACF for AR (p) Process

- The Yule-Walker equations can be used to derive the partial autocorrelation coefficients at lags $1, 2, \ldots, h$ as follows:
  1. Fit the regression model, where the dependent variable X_t from a zero mean stationary process is regressed on the h lagged variables $X_{t-1}, X_{t-2}, \ldots, X_{t-h}$. i.e.,
  $$X_t = \phi_{h,1}X_{t-1} + \phi_{h,2}X_{t-2} + \ldots + \phi_{h,h}X_{t-h} + \epsilon_t,$$
  where $\phi_{h,h}$ denotes the h-th regression parameter and ϵ_t is an error term with mean 0 and uncorrelated with X_{t-h}, for $h \neq 0$.
  2. Multiply this equation by X_{t-1}, take expectations and divide the results by the variance of X_t. Do the same operation with $X_{t-2}, X_{t-3}, \ldots, X_{t-h}$ successively to get the following set of h-Yule-Walker equations. (see the next slide)!
- Yule-Walker equations is a technique that can be used to estimate the autoregression parameters of the AR (h) model $X_t = \sum_{i=1}^{h} \phi_i X_{t-i} + \epsilon_t$ from data.

Yule-Walker Equations and PACF for AR(p) Process

The $h \times h$ Yule-Walker Equations associated with AR(h) model:

$$\begin{aligned}
\rho_1 &= \phi_{h,1} + \phi_{h,2}\rho_1 + \phi_{h,3}\rho_2 + \ldots + \phi_{h,h}\rho_{h-1} \\
\rho_2 &= \phi_{h,1}\rho_1 + \phi_{h,2} + \phi_{h,3}\rho_1 \ldots + \phi_{h,h}\rho_{h-2} \\
\ldots &= \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
\rho_h &= \phi_{h,1}\rho_{h-1} + \phi_{h,2}\rho_{h-2} + \phi_{h,h-1}\rho_1 + \ldots + \phi_{h,h}
\end{aligned}$$

Which may represented in a matrix form as: $AX = \mathbf{b}$,

$$\begin{pmatrix} 1 & \rho_1 & \rho_2 & \ldots & \rho_{h-1} \\ \rho_1 & 1 & \rho_1 & \ldots & \rho_{h-2} \\ \rho_2 & \rho_1 & 1 & \ldots & \rho_{h-3} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \rho_{h-1} & \rho_{h-2} & \ldots & \rho_1 & 1 \end{pmatrix} \begin{pmatrix} \phi_{h,1} \\ \phi_{h,2} \\ \phi_{h,3} \\ \ldots \\ \phi_{h,h} \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \ldots \\ \rho_h \end{pmatrix}$$

X is the column vector variables $(\phi_{h,1}, \phi_{h,2}, \ldots, \phi_{h,h})^T$ and $\mathbf{b}$ is the column vector $(\rho_1, \rho_2, \ldots, \rho_h)^T$ (see the next slide)!

Yule-Walker Equations and PACF for AR (p) Process

Use Cramer's rule successively for $j = 1, 2, \ldots$ to get:

$$
\begin{aligned}
\phi_{1,1} &= \rho_1 \\
\phi_{2,2} &= \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \\
\phi_{3,3} &= \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} = \frac{\rho_3(1 - \rho_1^2) + \rho_1\rho_2(\rho_2 - 2) + \rho_1^3}{1 - 2\rho_1^2 + 2\rho_1^2\rho_2 - \rho_2^2} \\
\vdots &= \vdots
\end{aligned}
$$

(see next page)!

Yule-Walker Equations and PACF for AR (p) Process

$$\phi_{h,h} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{h-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{h-3} & \rho_2 \\ \vdots & \cdots & \cdots & \cdots & \cdots & \vdots \\ \rho_{h-1} & \rho_{h-2} & \rho_{h-3} & \cdots & \rho_1 & \rho_h \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{h-2} & \rho_{h-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{h-3} & \rho_{h-2} \\ \vdots & \cdots & \cdots & \cdots & \cdots & \vdots \\ \rho_{h-1} & \rho_{h-2} & \rho_{h-3} & \cdots & \rho_1 & 1 \end{vmatrix}}$$

A Numerical Example: How to Calculate PACF

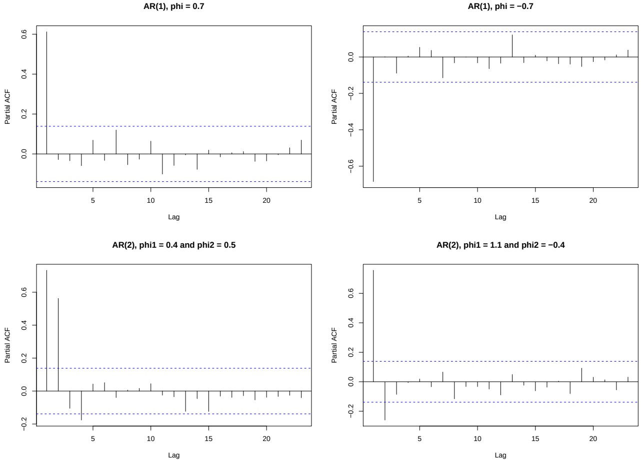
Suppose that $\rho_1 = 0.7, \rho_2 = 0.5$, and $\rho_3 = 0.2$, then

$$
\begin{aligned}
\phi_{1,1} &= \rho_1 = 0.7 \\
\phi_{2,2} &= \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\begin{vmatrix} 1 & 0.7 \\ 0.7 & 0.5 \end{vmatrix}}{\begin{vmatrix} 1 & 0.7 \\ 0.7 & 1 \end{vmatrix}} = 0.0196 \\
\phi_{3,3} &= \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} = \frac{\begin{vmatrix} 1 & 0.7 & 0.7 \\ 0.7 & 1 & 0.5 \\ 0.5 & 0.7 & 0.2 \end{vmatrix}}{\begin{vmatrix} 1 & 0.7 & 0.5 \\ 0.7 & 1 & 0.7 \\ 0.5 & 0.7 & 1 \end{vmatrix}} = -0.392
\end{aligned}
$$

Remarks

- The partial autocorrelation function $\phi_{h,h}$ is a function of the autocorrelations $\rho_1, \ldots, \rho_h$. Thus, $-1 \leq \phi_{h,h} \leq 1 \; \forall h > 0$.
- If $\{\epsilon_t\}$ is a white noise process, then the partial autocorrelation function $\phi_{h,h} = 0$ for all $h \neq 0$, whereas $\phi_{0,0} = \rho_0 = 1$.
- If the underlying process is AR(p), $\phi_{h,h} = 0 \; \forall h > p$, so the plot of the PACF should show a cutoff after lag p (see the next slide!).
- Replacing ρ_h (population autocorrelations) by $\hat{\rho}_h$ (sample autocorrelations) $\forall h$, will give the sample PACF $\hat{\phi}_{h,h}$ (see Levinson-Durbin recursion method).

PACF for Simulated AR (1) and AR (2) Process



Sample PACF : Levinson-Durbin Recursive Method

In practice the sample PACF is obtained by Levinson-Durbin recursion method starting with $\hat{\phi}_{1,1} = \hat{\rho}_1$ as follows:

$$\hat{\phi}_{h+1,h+1} = \frac{\hat{\rho}_{h+1} - \sum_{j=1}^{h} \hat{\phi}_{h,j}\hat{\rho}_{h+1-j}}{1 - \sum_{j=1}^{h} \hat{\phi}_{h,j}\hat{\rho}_j},$$

and $\hat{\phi}_{h+1,j} = \hat{\phi}_{h,j} - \hat{\phi}_{h+1,h+1}\hat{\phi}_{h,h+1-j}$, for $j = 1, 2, \ldots, h$.

$$\Longrightarrow \hat{\phi}_{1,1} = \hat{\rho}_1, \quad \hat{\phi}_{2,2} = \frac{\hat{\rho}_2 - \hat{\rho}_1^2}{1 - \hat{\rho}_1^2}, \quad \hat{\phi}_{2,1} = \hat{\phi}_{1,1} - \hat{\phi}_{2,2}\hat{\phi}_{1,1},$$

$$\hat{\phi}_{3,3} = \frac{\hat{\rho}_3 - \hat{\phi}_{2,1}\hat{\rho}_2 - \hat{\phi}_{2,2}\hat{\rho}_1}{1 - \hat{\phi}_{2,1}\hat{\rho}_1 - \hat{\phi}_{2,2}\hat{\rho}_2}, \hat{\phi}_{3,2} = \hat{\phi}_{2,2} - \hat{\phi}_{3,3}\hat{\phi}_{2,1},$$

$\hat{\phi}_{3,1} = \hat{\phi}_{2,1} - \hat{\phi}_{3,3}\hat{\phi}_{2,2}$, other $\hat{\phi}_{h,h}, \hat{\phi}_{h,j}$ can be calculated similarly

Levinson-Durbin Algorithm for PACF in AR Models

- For AR (1),

$$\hat{\phi}_{1,1} = \hat{\rho}_1, \text{ and } \hat{\phi}_{h,h} = 0 \; \forall h > 1.$$

- For AR (2),

$$\hat{\phi}_{1,1} = \hat{\rho}_1, \hat{\phi}_{2,2} = \frac{\hat{\rho}_2 - \hat{\rho}_1^2}{1 - \hat{\rho}_1^2}, \hat{\phi}_{2,1} = \hat{\phi}_{1,1} - \hat{\phi}_{2,2}\hat{\phi}_{1,1}, \hat{\phi}_{h,h} = 0 \; \forall h > 2.$$

- For AR (3),

$$\hat{\phi}_{1,1} = \hat{\rho}_1, \hat{\phi}_{2,2} = \frac{\hat{\rho}_2 - \hat{\rho}_1^2}{1 - \hat{\rho}_1^2}, \hat{\phi}_{2,1} = \hat{\phi}_{1,1} - \hat{\phi}_{2,2}\hat{\phi}_{1,1},$$

$$\hat{\phi}_{3,3} = \frac{\hat{\rho}_3 - \hat{\phi}_{2,1}\hat{\rho}_2 - \hat{\phi}_{2,2}\hat{\rho}_1}{1 - \hat{\phi}_{2,1}\hat{\rho}_1 - \hat{\phi}_{2,2}\hat{\rho}_2}, \hat{\phi}_{3,2} = \hat{\phi}_{2,2} - \hat{\phi}_{3,3}\hat{\phi}_{2,1},$$

$$\hat{\phi}_{3,1} = \hat{\phi}_{2,1} - \hat{\phi}_{3,3}\hat{\phi}_{2,2}, \text{ and } \hat{\phi}_{h,h} = 0 \; \forall h > 3.$$

A Numerical Example: How to Calculate PACF

Suppose that $\hat{\rho}_1 = -0.188, \hat{\rho}_2 = -0.201$, and $\hat{\rho}_3 = 0.181$, then

$$
\begin{aligned}
\hat{\phi}_{1,1} &= \hat{\rho}_1 = -0.188, \\
\hat{\phi}_{2,2} &= \frac{\hat{\rho}_2 - \hat{\rho}_1^2}{1 - \hat{\rho}_1^2} = \frac{-0.201 - (-0.188)^2}{1 - (-0.188)^2} = -0.245, \\
\hat{\phi}_{2,1} &= \hat{\phi}_{1,1} - \hat{\phi}_{2,2}\hat{\phi}_{1,1} = -0.188 - (-0.245)(-0.188) = -0.234, \\
\hat{\phi}_{3,3} &= \frac{\hat{\rho}_3 - \hat{\phi}_{2,1}\hat{\rho}_2 - \hat{\phi}_{2,2}\hat{\rho}_1}{1 - \hat{\phi}_{2,1}\hat{\rho}_1 - \hat{\phi}_{2,2}\hat{\rho}_2} \\
&= \frac{0.181 - (-0.234)(-0.201) - (-0.245)(-0.188)}{1 - (-0.234)(-0.188) - (-0.245)(-0.201)} = 0.097, \\
\hat{\phi}_{3,2} &= \hat{\phi}_{2,2} - \hat{\phi}_{3,3}\hat{\phi}_{2,1} = -0.245 - (0.097)(-0.234) = -0.222, \\
\hat{\phi}_{3,1} &= \hat{\phi}_{2,1} - \hat{\phi}_{3,3}\hat{\phi}_{2,2} = (-0.234) - (0.097)(-0.245) = -0.210.
\end{aligned}
$$

The PACF of Invertible MA and ARMA models

- For an MA (1), $X_t = \epsilon_t + \theta\epsilon_{t-1}, |\theta| < 1$, one can show that
$$\phi_{h,h} = \frac{(-\theta)^h(1-\theta^2)}{1-\theta^{2(h+1)}}, h \geq 1.$$
- For an invertible MA (q), no finite representation exists; Hence, the PACF will never cut off, as in the case of an AR (p).
- Because an invertible ARMA model has an infinite AR representation, the PACF will tails off (not cut off).
- The PACF for MA models behaves much like the ACF for AR models (tails off).
- Also, the PACF for AR models behaves much like the ACF for MA models (cuts off after the lags p and q for the AR (p) and MA (q) models respectively.

Testing Randomness Based on Individual Autocorrelations

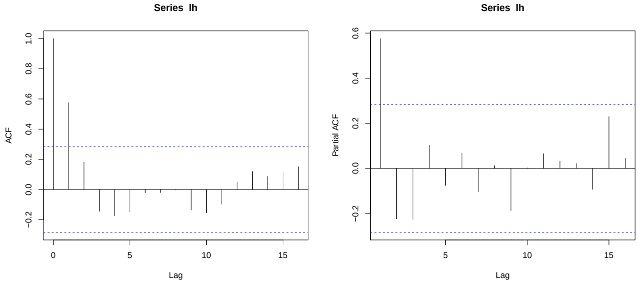
- To determine whether the values of the ACF , or the PACF , are negligible, we can use the approximation that they both have standard deviation $\approx 1/\sqrt{n}$.
- At 5% significant level, the approximate confidence bounds are $\pm 1.96/\sqrt{n} \approx \pm 2/\sqrt{n}$, where the values outside this range can be regarded as significant.
- In R, the acf() and pacf() functions can compute and plot the sample ACF and the sample PACF respectively, where the confidence bounds are shown as blue dotted lines.
- Testing randomness based on individual ACF or PACF is misleading because $\rho(h), h = 1, 2, ...$ are not independent. This may produce a large number of $\rho(h)$ exceeds $\pm 2/\sqrt{n}$ even if the underlying time series is a White Noise series.

We prefer the portmanteau test statistic for testing randomness (as we will see later).

Plot of Autocorrelation & Partial Autocorrelation Functions

Consider the regular time series data available from R with the name lh. The data is 48 samples giving the luteinizing hormone in blood samples at 10 minute intervals from a human female.

```
R> par(mfrow=c(1,2))
R> acf(lh)  ##Try this code: acf(lh,plot=FALSE)
R> pacf(lh) ##Try this code: pacf(lh,plot=FALSE)
```



Random Walk Process

Definition

The process $X_t = X_{t-1} + \epsilon_t$, where ϵ_t is a White Noise $WN(0, \sigma^2)$ for all $t \geq 1$, (i.e., an AR(1) process with $\phi = 1$) is called a random walk (unit-root) processes.

In words, the value of X in period t is equal to its value in $t-1$, plus a random step due to the white-noise shock ϵ_t.

Random Walk Process is Not Stationary

The process $X_t = X_{t-1} + \epsilon_t$ is not stationary as seen below: Consider $X_0 = a$, where a is a constant $\Rightarrow$

$$\begin{aligned}
X_1 &= x_0 + \epsilon_1 = a + \epsilon_1 \\
X_2 &= x_1 + \epsilon_2 = a + \epsilon_1 + \epsilon_2 \\
X_3 &= x_2 + \epsilon_3 = a + \epsilon_1 + \epsilon_2 + \epsilon_3 \\
\vdots &= \vdots \\
X_t &= x_{t-1} + \epsilon_t = a + \sum_{i=1}^{t} \epsilon_i
\end{aligned}$$

Although, the expected value of $X_t = \mathbb{E}(X_t) = a$, which independent of time, the autocovariance and autocorrelation are functions of time (see the next slide!)

Random Walk Process is Not Stationary

The autocovariance function is:

$$\gamma_X(h) = \mathbb{C}ov(X_t, X_{t+h}) = \mathbb{C}ov(a + \sum_{i=1}^{t} \epsilon_i, a + \sum_{j=1}^{t+h} \epsilon_j) = t\sigma^2.$$

The autocorrelation function is:

$$\rho_X(h) = \frac{\mathbb{C}ov(X_t, X_{t+h})}{\sqrt{\mathbb{V}ar(X_t)\mathbb{V}ar(X_{t+h})}} = \frac{t\sigma^2}{\sqrt{t\sigma^2(t+h)\sigma^2}} = \frac{1}{\sqrt{1+h/t}},$$

Note that: For large t with h considerably less than t, $\rho_X(h)$ is nearly 1, Hence, the correlogram for a random walk is characterised by positive autocorrelations with very slow decay down from unity.

Test for Non-stationarity: Dickey-Fuller Test Statistics

For a simple AR(1) model $X_t = \phi X_{t-1} + \epsilon_t$, the unit root presents if $\phi = 1$. The model would be non-stationary in this case.

- The regression model can be written as:

$$\Delta X_t = (\phi - 1) X_{t-1} + \epsilon_t = \delta X_{t-1} + \epsilon_t,$$

  where Δ is the first difference operator.

- This model can be estimated and testing for a unit root which is equivalent to testing:

$$H_0 : \delta = 0 \text{ (or } \phi = 1) \quad \text{vs.} \quad H_1 : \delta < 0 \text{ (or } \phi < 1).$$

- The test is a one-sided left tail test.

Dickey-Fuller Unit Root Test Statistics

- Dickey-Fuller test statistics derived (modified t-distribution) to test whether a unit root is present in an AR model.
- There are 3 versions of the Dickey-Fuller unit root test:
  1. Test for a unit root without drift (constant) and without trend

$$\Delta X_t = \delta X_{t-1} + \epsilon_t.$$

  2. Test for a unit root with drift only (no trend)

$$\Delta X_t = a_0 + \delta X_{t-1} + \epsilon_t.$$

  3. Test for a unit root with drift and deterministic time trend

$$\Delta X_t = a_0 + a_1 t + \delta X_{t-1} + \epsilon_t.$$

Test for Non-stationarity: Dickey-Fuller Test Statistic

- If X_t is stationary (i.e., $|\phi| < 1$) then it can be shown
$$\hat{\phi} \sim \mathcal{N}(\phi, \frac{1}{n}(1-\phi^2)).$$
Under the null hypothesis the above result gives
$$\hat{\phi} \sim \mathcal{N}(1, 0),$$
which clearly does not make any sense.
- Phillips (1987) showed that the sample moments of X_t converge to random functions of Brownian motion so that
$$n(\hat{\phi} - 1) \xrightarrow{d} \frac{\int_0^1 W(r)dW(r)}{\int_0^1 W(r)^2 dr}.$$
- $\hat{\phi}$ is not asymptotically normally distributed. Consequently, the critical values of the test were computed by simulation. We reject H_0 if $n(\hat{\phi} - 1)$ is $\leq$ the tabulated critical value.

Critical Values for the Dickey-Fuller Unit Root Statistics

Model	N	1%	2.5%	5%	10%	90%	95%	97.5%	99%
Model I (no constant, no trend)									
	25	-2.66	-2.26	-1.95	-1.60	0.92	1.33	1.70	2.16
	50	-2.62	-2.25	-1.95	-1.61	0.91	1.31	1.66	2.08
	100	-2.60	-2.24	-1.95	-1.61	0.90	1.29	1.64	2.03
	250	-2.58	-2.23	-1.95	-1.61	0.89	1.29	1.63	2.01
	500	-2.58	-2.23	-1.95	-1.61	0.89	1.28	1.62	2.00
	>500	-2.58	-2.23	-1.95	-1.61	0.89	1.28	1.62	2.00
Model II (constant, no trend)									
	25	-3.75	-3.33	-3.00	-2.62	-0.37	0.00	0.34	0.72
	50	-3.58	-3.22	-2.93	-2.60	-0.40	-0.03	0.29	0.66
	100	-3.51	-3.17	-2.89	-2.58	-0.42	-0.05	0.26	0.63
	250	-3.46	-3.14	-2.88	-2.57	-0.42	-0.06	0.24	0.62
	500	-3.44	-3.13	-2.87	-2.57	-0.43	-0.07	0.24	0.61
	>500	-3.43	-3.12	-2.86	-2.57	-0.44	-0.07	0.23	0.60
Model III (constant, trend)									
	25	-4.38	-3.95	-3.60	-3.24	-1.14	-0.80	-0.50	-0.15
	50	-4.15	-3.80	-3.50	-3.18	-1.19	-0.87	-0.58	-0.24
	100	-4.04	-3.73	-3.45	-3.15	-1.22	-0.90	-0.62	-0.28
	250	-3.99	-3.69	-3.43	-3.13	-1.23	-0.92	-0.64	-0.31
	500	-3.98	-3.68	-3.42	-3.13	-1.24	-0.93	-0.65	-0.32
	>500	-3.96	-3.66	-3.41	-3.12	-1.25	-0.94	-0.66	-0.33

Test for Unit Roots in ARMA models: Augmented Dickey-Fuller Test

- The augmented Dickey-Fuller (ADF) test tests the null hypothesis that a time series X_t is not stationary against the alternative that it is stationary, assuming that the dynamics in the data have an ARMA structure.
- In R, the ADF tests are implemented in the functions:
  1. The function `adf.test()` in the package **tseries**.
  2. The function `nsdiffs()` in the package **forecast**.
  3. The function `ur.df()` in the package **urca**.
  4. `adfTest()` and `urdfTest()` in the package **fUnitRoots**.
- In each case, we reject the null hypothesis if p-value is less than or equal to the significance level (usually $\alpha = 0.05$).

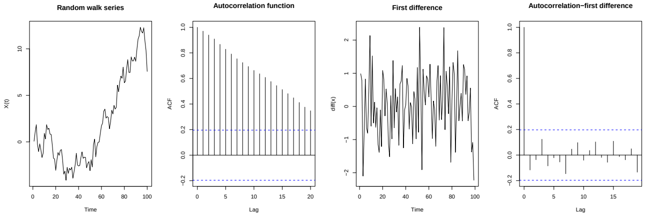
Transform a Non-stationary Series to a Stationary Series

- We can eliminate the non-stationarity in a random walk process by taking the first difference of X_t. Thus, $Y_t = \Delta X_t = X_t - X_{t-1} = \epsilon_t$ is WN stationary process.
- To achieve stationarity in the presence of d unit roots, we must apply d differences to X_t; $Y_t = \Delta^d X_t = (1 - \mathbf{B})^d X_t$.

Note that: A random walk often provides a good fit to data with stochastic trends, such as stock price changes, although even better fits are usually obtained from more general model formulations, such as the ARIMA models.

Simulated Random Walk Series

```
R> set.seed(348)
R> n<-100; epsilon<-rnorm(n); x<-epsilon[1]
R> for (i in 1:(n-1)){x[i+1] <- x[i]+epsilon[i+1]}
R> par(mfrow=c(1,4))  ## Try x <- cumsum(epsilon)
R> plot(ts(x),ylab="X(t)",main="Random walk series")
R> acf(ts(x), main="Autocorrelation function")
R> ts.plot(diff(x),main="First difference")
R> acf(ts(diff(x)),main="Autocorrelation-first difference")
```



Example: Test for Unit root in an AR(1) Model

Example

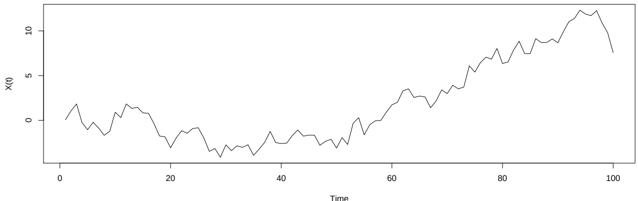
Consider the AR(1) model $\hat{X}_t = 0.946X_{t-1}$, where $n = 34$.
The Dickey-Fuller test statistic is

$$n(\hat{\phi} - 1) = 34(0.946 - 1) = -1.836.$$

At the significant level $\alpha = 0.05$, the tabulated critical value associated with $n = 25$ is -1.95.
Since the value of the test statistic is not less than or equal to the critical value , we will not reject H_0; Hence, there exist a unit root. We need to take a difference to transform it to stationary in order to be able to estimate a model for the series.

Test for a unit root with drift and deterministic time trend

```
R> set.seed(348);n<-100;epsilon<-rnorm(n);x<-epsilon[1]
R> x <-cumsum(epsilon);plot(ts(x),ylab="X(t)")
R> library("tseries");adf.test(x)
Augmented Dickey-Fuller Test

data:  x
Dickey-Fuller = -1.849, Lag order = 4, p-value = 0.6392
alternative hypothesis: stationary
```



List of Some Useful R Functions

R-Function	Description
`read.table()`	reads data into a data frame.
`scan()`	read data into a vector or list.
`attach()`	makes names of column variables available.
`ts()`	produces a time series object.
`ts.plot()`	produces a time plot for one or more series.
`points()`	add a sequence of points centered to a specified coordinates.
`window()`	extracts a subset of a time series.
`aggregate()`	creates an aggregated series.
`time()`	extracts the time from a time series object.
`decompose()`	decomposes a series into the components trend, seasonal effect, and residual.
`summary()`	summarises an R object.
`mean()`	returns the mean (average).
`var()`	returns the variance with denominator $n-1$.
`sd()`	returns the standard deviation.
`polyroot()`	find zeros of a real or complex polynomial.
`Mod()`	check whether the roots of the polynomial have modulus > 1 or not.
`adf.test()`	Computes the Augmented Dickey-Fuller test for the null that x has a unit root. (**t**series package).
`nsdiffs()`	Number of differences required for a stationary series. If `test="adf"`, the Augmented Dickey-Fuller test is used. (**f**orecast package).
`ur.df()`	Performs the augmented Dickey-Fuller unit root test. (**u**rca package).
`adfTest()`	Augmented Dickey-Fuller test for unit roots.(**f**UnitRoots package).
`urdfTest()`	Augmented Dickey-Fuller test for unit roots.(**f**UnitRoots package).

List of Some Useful R Functions

R-Function	Description
`cov()`	returns the covariance with denominator $n-1$.
`cor()`	returns the correlation.
`acf()`	returns the correlogram (or sets the argument to obtain autocovariance function).
`pacf()`	returns the partial autocorrelation function.
`lm()`	linear models (least squares fit).
`predict()`	forecasts future values.
`HoltWinters()`	estimates the parameters of the Holt-Winters or exponential smoothing model.
`coef()`	extracts the coefficients of a fitted model.
`resid()`	extracts the residuals from a fitted model.
`diff()`	returns suitably lagged and iterated differences.
`round()`	round the values in its first argument to the specified number of decimal places.
`set.seed()`	provide a seed for simulations ensuring that the simulations can be reproduced.
`par(mfrow=`$c(i,j)$`)`	set a graphical device to insert $i \times j$ pictures on one plot.
`ar()`	fit an autoregressive time series model to the data, by default selecting the complexity by AIC .
`arima()`	fit an ARIMA model to a univariate time series.
`arima.sim()`	simulate from an ARIMA model.
`varima.sim()`	Simulate data from seasonal/nonseasonal ARIMA models (portes package).

List of Some Useful R Functions

R-Function	Description
`AIC()`	Akaike information criterion for selection model.
`BIC()`	Bayesian information criterion for selection model.
`Box.test()`	compute the Box-Pierce or Ljung-Box portmanteau tests.
`BoxPierce()`	compute the univariate or multivariate Box-Pierce portmanteau test (portes package).
`LjungBox()`	compute the univariate or multivariate Ljung-Box portmanteau test (portes package).
`gvtest()`	compute the univariate Peňa-Rodríguez or multivariate Mahdi-McLeod portmanteau test (portes).
`tsdiag()`	diagnostic plots for time series fits.
`qqplot()`	produces a Q-Q plot of two datasets.
`qqnorm()`	produces a normal Q-Q plot of points.
`qqline()`	draw the diagonal line for normal Q-Q plots produced by `qqnorm()`.
`layout()`	divides the device up into as many rows and columns as there are in matrix `mat`, with the column-widths and the row-heights specified in the respective arguments.
`boxplot()`	produce box-and-whisker plot(s) of the given (grouped) values.
`start()`	extract and encode the times the first observation were taken.
`end()`	extract and encode the times the last observation were taken.
`frequency()`	returns the number of samples per unit time.
`polyroot()`	finds zeros of polynomials and roots of the characteristic equation to check for stationarity.
`det()`	calculate the Determinant of a Matrix.
`matrix()`	create a matrix from a given set of values.

Homework

- Install R on your own PC (laptop).
- Install the following R packages: **t**series, **f**orecast,**T**SA,**z**oo **a**stsa, **t**imeSeries, **p**ortes, **s**arima, **M**ASS, **l**attice, **n**lme, **M**TS, **v**ars, **m**vtnorm, **f**racdiff, **s**spir, **a**kima, **f**Garch, **F**itAR, **F**GN.
- Download (as txt files) all data sets needed for this course, create a "ts" object, plot and explain each one: http://www.stat.pitt.edu/stoffer/tsa4/, http://staff.elena.aut.ac.nz/Paul-Cowpertwait/ts/, and http://astro.temple.edu/~wwei/data.html
- Read the summarized time series analysis on CRAN task view: https://cran.r-project.org/web/views/TimeSeries.html.

Homework

For the macroeconomic Longley's Economic Regression Data with the name longley available from the R package **datasets**:

1. Import the data into R.
2. Use the `summary()` function with this data. Explain it?
3. Calculate the correlation coefficient between the number of unemployed (Unemployed) and the population size (Population) for those who are ≥ 14 years of age.
4. Plot `Unemployed` ($Y-$axis) versus `Population` ($X-$axis) using the function `plot()`.
5. Plot `Unemployed` ($Y-$axis) as time series data versus `Year` ($X-$axis).
6. Fit the linear regression using R with the `lm()` function, where `Unemployed` is the dependent variable.
7. Interpret the results you get from 6.

Homework

The data jj (Quarterly earnings per share for 1960Q1 to 1980Q4 of the U.S. company, Johnson & Johnson, Inc.) available from the **R** package **a**stsa.

1. Compute the base trend using centered moving averages.
2. Use 1 to estimate the normalized seasonal indices.
3. Deseasonalize the series
4. Fit a regression line to the deseasonalized observations.
5. Use trend to make deseasonalized predictions.
6. Reseasonalize predictions and plot the forecasts.

Homework

The data with the name `w6` available from the link `http://astro.temple.edu/~wwei/data.html` represents the realization yearly US Tobacco production from the 1871 to 1984 (in millions of pounds).

1. Import the data into R using `scan()` R function.
2. Plot the data as time series object.
3. Estimate the trend line using the least squares method.
4. Compute the mean squares of the residuals.
5. Plot the forecast 4 years ahead.
6. Smooth the series using a three-year moving average.
7. Smooth the series using a four-year moving average.
8. Use the `decompose()` R function to estimate trends and seasonal effects. Explain!

Homework

For the data with the name `w1` available from the link `http://astro.temple.edu/~wwei/data.html`.

1. Use the exponential smoothing method with four different smoothing parameters $\alpha = 0.2, 0.4, 0.5$ and 0.8 to forecast the 1-day ahead.
2. What is the optimal value of the smoothing parameter α?.
3. Use the R function `HoltWinters()` to find the exponential smoothing model for the above values of α.
4. Plot the original series together with the smoothing series.

Homework

Compute and plot the values of $\hat{\gamma}_X(h)$, $\hat{\rho}_X(h)$, and $\hat{\phi}_{h,h}$ for lag $h = 0, 1, 2, 3$ and 4. Check your results using R.

t	X_t	t	X_t
1	2	11	1
2	1	12	1
3	4	13	4
4	3	14	3
5	3	15	5
6	5	16	6
7	2	17	4
8	1	18	2
9	0	19	5
10	3	20	3

Homework

Which of the following time sequence is stationary (Weak Stationary)?

- $X_t = a + bt + \epsilon_t$, where a and b are constants, $\epsilon_t \sim i.i.d.(0, 1)$ for all $t \in \mathbb{R}$.
- $X_t = \epsilon_t + t\epsilon_{t-1}$, $\epsilon_t \sim i.i.d.(0, 1)$ for all $t \in \mathbb{R}$.
- $X_t = (-1)^t A$, where A is a random variable with a zero mean and a unit variance.
- $X_t = cos(t + Y)$, where Y is a random variable with a Uniform distribution $(0, 2\pi)$ and $t \in \mathbb{R}$.
- $X_t = Usin(2\pi t) + Vcos(2\pi t)$, where U and V are independent random variables, each with mean 0 and variance 1.

Homework

1. Classify the following AR models (that is, state if they are AR (1), or AR (2), etc.), determine the mean of each model, and rewrite each one using the backshift operator.
   - $Y_t - 0.17Y_{t-1} + 0.19Y_{t-2} = \epsilon_t$
   - $Z_{n+1} = 77 - 0.55Z_n - 0.24Z_{n-1} + 0.19Z_{n-2} + \epsilon_{n+1}$
   - $A_t + 0.2A_{t-1} + 0.035A_{t-2} - 0.13A_{t-3} = 5 + \epsilon_t$
2. Classify the following MA models (that is, state if they are MA (1), or MA (2), etc.), determine the mean of each model, and rewrite each one using the backshift operator.
   - $B_{t+1} = 10 + \epsilon_{t+1} - 0.06\epsilon_t + 0.35\epsilon_{t-1}$
   - $Q_t - \epsilon_t - 0.29\epsilon_{t-1} + 0.19\epsilon_{t-2} + 0.61\epsilon_{t-3} - 0.26\epsilon_{t-4} = 0$
   - $W_t - \epsilon_t + 0.4\epsilon_{t-1} + 0.25\epsilon_{t-2} - 0.2\epsilon_{t-3} = 9$

Homework

1. Find the autocovariance and autocorrelation functions of the time series models:

$$X_t = 0.45x_{t-1} - 0.2X_{t-2} + \epsilon_t, \text{ where } \epsilon_t \sim (0, 9)$$

$$X_t = \epsilon_t + 0.2\epsilon_{t-1} - 0.1\epsilon_{t-2}, \text{ where } \epsilon_t \sim (0, 9)$$

2. Derive $\gamma_X(h)$ and $\rho_X(h)$ for the ARMA (1,1).

Homework

1. Classify the following ARMA models (that is, state if they are ARMA (1,1), or ARMA (1,2), etc.), determine the mean of each model, rewrite each one using the backshift operator, and finally convert each model into a pure AR model, and into a pure MA model.
   - $Y_t - 0.17Y_{t-1} + 0.19Y_{t-2} = 12 + \epsilon_t + 0.2\epsilon_{t-1}$
   - $Z_{n+1} = 0.5Z_n - 0.2Z_{n-1} + \epsilon_{n+1} - 0.6\epsilon_n - 0.18\epsilon_{n-1} + 0.1\epsilon_{n-2}$
   - $A_t + 0.2A_{t-1} + 0.04A_{t-2} = \epsilon_t + 0.4\epsilon_{t-1} + 0.3\epsilon_{t-2} - 0.2\epsilon_{t-3}$
2. Find the autocovariance and autocorrelation functions of the time series models:

$$X_t = 0.45x_{t-1} - 0.2X_{t-2} + \epsilon_t, \text{ where } \epsilon_t \sim (0, 9)$$

$$X_t = \epsilon_t + 0.2\epsilon_{t-1} - 0.1\epsilon_{t-2}, \text{ where } \epsilon_t \sim (0, 9)$$

Homework

1. Which one of the following MA (q) process is invertible?. Represent the invertible process in an AR (∞) form.
   - $X_t = (1 + \mathbf{B})\epsilon_t$
   - $X_t = \epsilon_t - 1.3\epsilon_{t-1} + 0.36\epsilon_{t-2}$
   - $X_t = \epsilon_t + 1.6\epsilon_{t-1} - 0.36\epsilon_{t-2}$
   - $X_t = (1 + 0.8\mathbf{B} - 0.6\mathbf{B}^2)\epsilon_t$
2. Which one of the following AR (p) process is stationary?. Represent the stationary process in a MA (∞) form.
   - $X_t = 1.3X_{t-1} + \epsilon_t$.
   - $(1 + 0.5\mathbf{B})X_t = \epsilon_t$.
   - $X_t = 1.25X_{t-1} - 0.25X_{t-2} + \epsilon_t$.
   - $(1 - 1.5\mathbf{B} + 0.56\mathbf{B}^2)X_t = \epsilon_t$.
   - $(1 - 0.2\mathbf{B} + 0.8\mathbf{B}^2)X_t = \epsilon_t$.

Homework

1. Which one of the following ARMA (p, q) process is invertible and stationary?. Represent the invertible process in an AR (∞) form and represent the stationary process in a MA (∞) form.
   - $X_t = -0.5X_{t-1} + \epsilon_t + 0.7\epsilon_{t-1}$
   - $X_t = 0.7X_{t-1} + 0.2X_{t-2} + \epsilon_t - 0.5\epsilon_{t-1}$
   - $X_t = 0.9X_{t-1} - 0.4X_{t-2} + \epsilon_t + 1.2\epsilon_{t-1} - 0.3\epsilon_{t-2}$
2. Determine the ACF of the following processes, $\epsilon \sim WN(0, \sigma^2)$ and represent each of them in an infinite form.
   - $X_t = \epsilon_t - 1.1\epsilon_{t-1} + 0.28\epsilon_{t-2},$
   - $X_t = 0.7X_{t-1} + \epsilon_t.$
   - $X_t = 0.1X_{t-1} + 0.3X_{t-2} + \epsilon_t.$
   - $X_t = 0.3X_{t-1} + 0.54X_{t-2} - 0.112X_{t-3} + \epsilon_t.$

Homework

1. Write the following equations in operator form:
   - $\dot{X}_t = \phi_1 \dot{X}_{t-1} + \ldots + \phi_p \dot{X}_{t-p} + \epsilon_t$, where $\dot{X} = X - \mu$.
   - $\dot{X} = \sum_{i=1}^{p} \phi_i \dot{X}_{t-i} + \sum_{j=0}^{q} \theta_j \epsilon_{t-j}$, where $\dot{X} = X - \mu$ and $\theta_0 = 1$
2. Consider the ARMA (1,1): $X_t = \phi X_{t-1} + \epsilon_t + \theta \epsilon_{t-1}$, where $\phi = -\theta$. Show that this mode is not really an ARMA (1,1), but it is a White Noise model ARMA (0,0).
3. Consider the ARMA (2,1): $X_t = -0.3X_{t-1} + 0.18X_{t-2} + \epsilon_t + 0.6\epsilon_{t-1}$. Show that this mode is not really an ARMA (2,1), but it is an AR (1)$\equiv$ARMA (1,0).

Homework

For the following models written using backshift operators, expand the model and write down the model in standard form. In addition, write down the model using ARIMA notation.

- $X_t = (1 - \phi\mathbf{B} - \theta\mathbf{B}^2)\epsilon_t$
- $(1 - \phi\mathbf{B})X_t = \epsilon_t$
- $(1 - \phi_1\mathbf{B} - \Phi_2\mathbf{B}^2)X_t = (1 + \theta_1\mathbf{B} + \theta_2\mathbf{B}^2)\epsilon_t$
- $(1 - \Phi\mathbf{B})(1 - \mathbf{B})^2 X_t = (1 + \theta\mathbf{B})\epsilon_t$
- $(1 - \Phi\mathbf{B}^4)X_t = \epsilon_t$
- $\Phi_p(\mathbf{B})(1 - \mathbf{B})^3(Z_t - \mu) = \Theta_q(\mathbf{B})\epsilon_t$, where $p = 3, q = 2$, and $\Phi_p(\mathbf{B}) = 1 - \sum_{i=1}^{p} \phi_i\mathbf{B}^i$ and $\Theta_q(\mathbf{B}) = 1 + \sum_{i=1}^{q} \theta_i\mathbf{B}^i$

Homework

For the following models written using backshift operators, expand the model and write down the model in standard form. In addition, write down the model using ARIMA notation.

- $(1+0.3\mathbf{B})(1-\mathbf{B})^2X_t = \epsilon_t$.
- $(1+0.3\mathbf{B})(1-\mathbf{B})(1-\mathbf{B}^{12})X_t = \epsilon_t$.
- $X_t = (1-0.4\mathbf{B})(1+0.3\mathbf{B}^7)\epsilon_t$.
- $(1-\mathbf{B}^7)X_t = (1-0.5\mathbf{B}-0.2\mathbf{B}^2)\epsilon_t$.
- $(1-0.27\mathbf{B})(1-0.51\mathbf{B}^{365})(1-\mathbf{B})(1-\mathbf{B}^{365})^2X_t = (1-0.46\mathbf{B})\epsilon_t$.
- $(1-0.4\mathbf{B})(1-\mathbf{B})X_t = (1+0.5\mathbf{B}^{12}+0.6\mathbf{B}^{24})\epsilon_t$.
- $X_t = (1+0.7\mathbf{B})(1+0.7\mathbf{B}^{12})\epsilon_t$
- $(1-0.7\mathbf{B})(1-0.1\mathbf{B}^{12})X_t = (1+0.6\mathbf{B}^4)\epsilon_t$

Homework

Use the build in R function `arima.sim()` to simulate the following models (assume that $n = 300$ and $\epsilon_t \overset{i.i.d.}{\sim} N(0,1)$). Plot each one and compute the autocovariance, autocorrelation, and partial autocorrelation functions using the `acf()` and `pacf()` R functions.

$$
\begin{aligned}
X_t &= 0.4X_{t-1} + \epsilon_t \\
X_t &= -0.7X_{t-1} + \epsilon_t \\
X_t &= \epsilon_t + 0.2\epsilon_{t-1} \\
X_t &= 0.1X_{t-1} + 0.3X_{t-2} + \epsilon_t \\
X_t &= \epsilon_t - 0.6\epsilon_{t-1} + 0.3\epsilon_{t-2} \\
X_t &= 0.2X_{t-1} + \epsilon_t + 0.8\epsilon_{t-1} \\
X_t &= 0.7X_{t-1} + 0.2X_{t-2} + \epsilon_t + 0.5\epsilon_{t-1} \\
X_t &= 0.9X_{t-1} - 0.4X_{t-2} + \epsilon_t + 1.2\epsilon_{t-1} - 0.3\epsilon_{t-2}
\end{aligned}
$$

Homework

- Write a short piece of R-code (use set.seed(345)) to simulate the AR (1) model $X_t = \phi X_t + \epsilon_t$ of length 500, where $\epsilon_t \sim N(2, 8)$, for each $\phi : \phi = 0.9, -0.9, 0.6, -0.6, 0.3, -0.3, 1.5, -1.5, 0.1, -0.1$. Plot each simulated series and plot each ACF and PACF . Comment on your findings: What effect does the value of ϕ have on the stationarity of the series?
- Write a short piece of R-code (use set.seed(345)) to simulate the MA (1) model $X_t = \epsilon_t + \theta\epsilon_{t-1}$ of length 500, where $\epsilon_t \sim N(0, 1)$, for each $\theta : \theta = 0.9, -0.9, 0.6, -0.6, 0.3, -0.3, 1.5, -1.5, 0.1, -0.1$. Plot each simulated series and plot each ACF and PACF . Comment on your findings: What effect does the value of θ have on the stationarity of the series?