Unit 3 - Time Series Analysis

Class notes: Vibha Masti

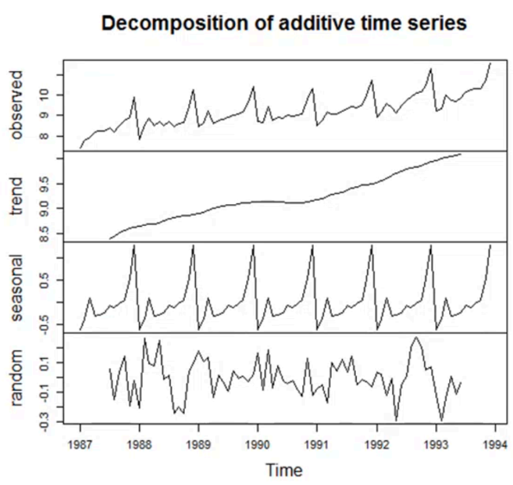
Feedback/corrections: vibha@pesu.pes.edu

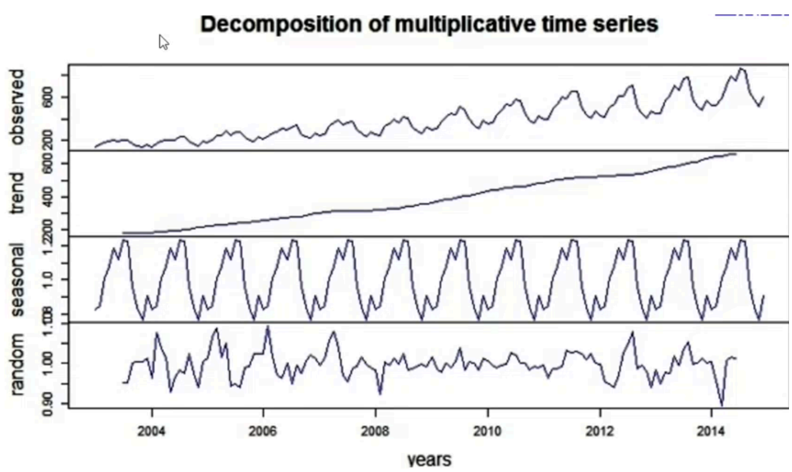
Table of Contents

More

Additive & Multiplicative Time Series

- Composed of trend (overall trend), seasonal (repeat seasonally - eg Diwali season), cyclic (similar to seasonal but slower-moving - 10-20 years, recession) and irregulars (random)
- Additive: $Y_t = T_t + S_t + C_t + I_t$
- Multiplicative: $Y_t = T_t \times S_t \times C_t \times I_t$





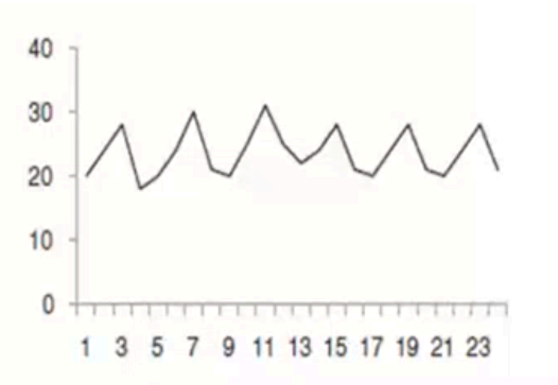
Components

1. Trend Component

- Consistent long-term upward/downward movement of data

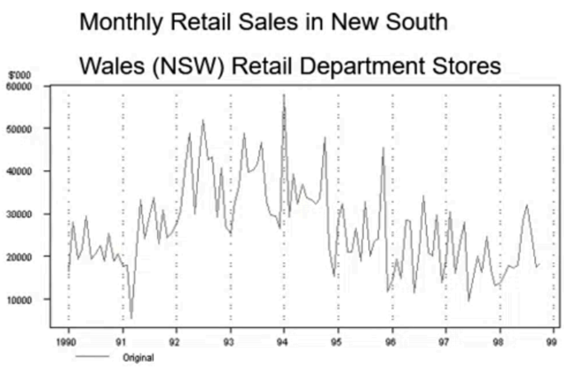
2. Seasonal Components

- Repetitive upward/downward fluctuations from the trend
- Within calendar year
- Festivals, seasons, customs, business practices, market
- India: Oct-Dec demand
- Conditions
  - Natural (weather)
  - Business and administrative procedures (school term)
  - Social/cultural (festivals)
  - Trading day effects (number of weekends in a month)
  - Moving holiday effects (Ramadan, Diwali, Easter)
- Identify seasonal components: regularly spaced peaks and troughs



(b) Seasonality (fixed periodicity)

- Notice trend and seasonality



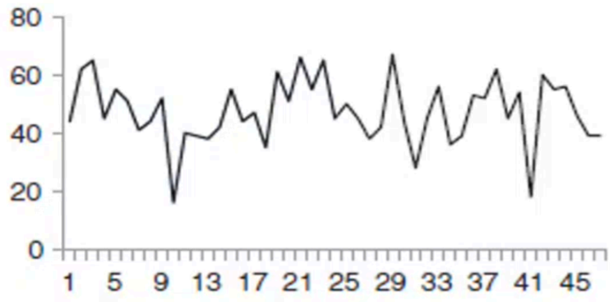
Obvious large seasonal increase in December retail sales in New South Wales due to Christmas shopping

3. Cyclical Components

- Fluctuations due to macroeconomic changes
- Recession, unemployment

4. Irregular Components

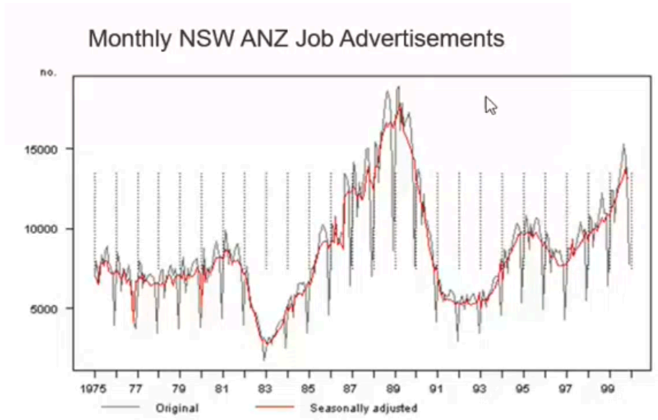
- White noise/random noise
- Uncorrelated changes
- Normal distribution with mean value of 0 and constant variance
- After seasonal and trend components have been estimated and removed
- Short term fluctuations in series (not systematic or predictable)



(d) Irregular

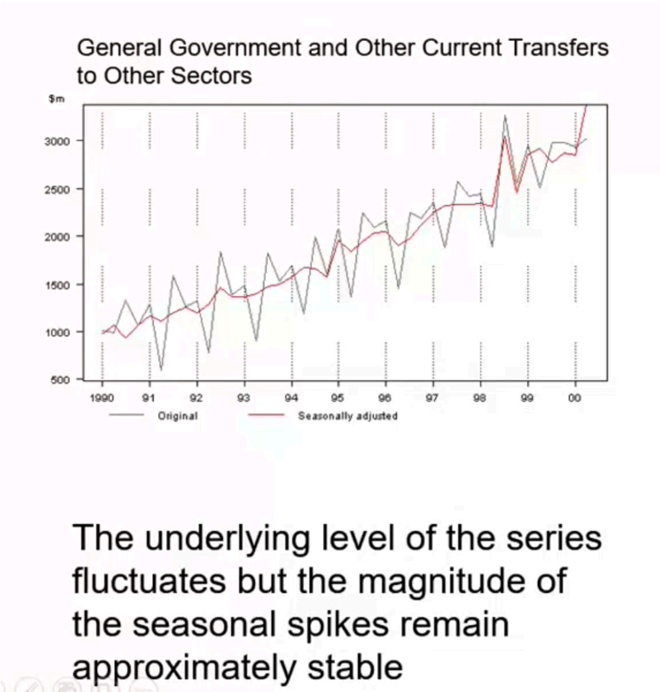
Additive vs Multiplicative

- Multiplicative more often used, better fit
  - $Y_t = T_t \times S_t$ used
  - Cyclic estimation - large dataset required
  - Appropriate if seasonal correlated with level/local mean



The trend has the same units as the original series, but the seasonal and irregular components are unitless factors, distributed around 1

- Additive only appropriate if
  - Seasonal and cyclical independent of trend
  - Seasonal remains constant about level/mean



Decomposition of TS Model

1. Additive decomposition

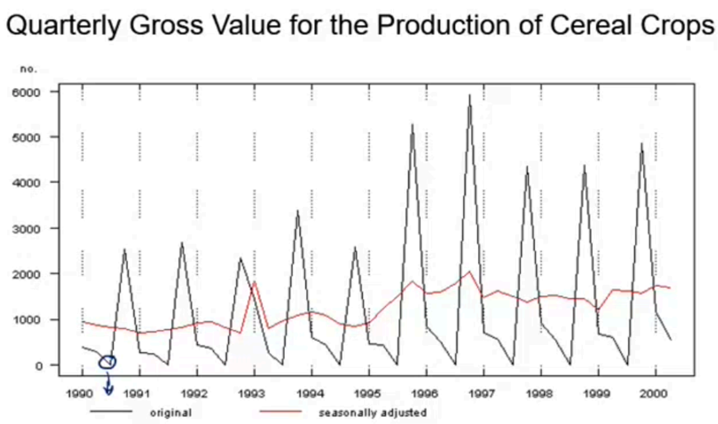
- When amplitude of seasonal and irregular do not change with level
- Observed time series O_t
  - $O_t = T_t + S_t + I_t$
- Seasonally adjusted series SA_t
  - Observed $-$ approximation for Seasonal $\hat{S}_t$
  - $SA_t = O_t - \hat{S}_t$
  - $SA_t = T_t + I_t$

2. Multiplicative decomposition

- Amplitude of seasonal and irregular increase as the level of the trend rises
- Observed time series O_t
  - $O_t = T_t \times S_t \times I_t$
- Seasonally adjusted series
- SA_t
  - Observed $\div$ approximation for Seasonal $\hat{S}_t$
  - $SA_t = \frac{O_t}{\hat{S}_t}$
  - $SA_t = T_t \times I_t$

3. Pseudo-Additive decomposition

- Multiplicative model not used when data contains small or zero values
  - Cannot divide by 0
- Pseudo-additive combines additive and multiplicative
- Assumption: seasonal and irregular independent of each other but dependent on trend
  - $O_t = T_t + T_t \times (S_t - 1) + T_t \times (I_t - 1)$
  - $O_t = T_t \times (S_t + I_t - 1)$
  - Both S_t and I_t centered around 1
  - Subtract 1 from both to center around 0
- Example that requires pseudo-additive



This model is used as cereal crops are only produced during certain months, with crop production being virtually zero for one quarter each year.

Extraction of Time Series Components

Error Metrics

- Y_t is actual value
- F_t is forecasted value

1. Mean absolute error (MAE)

- $MAE = \frac{1}{n}\sum_{t=1}^{n} |Y_t - F_t|$

2. Mean absolute percentage error (MAPE)

- $MAPE = \frac{1}{n}\sum_{t=1}^{n} (\frac{|Y_t - F_t|}{Y_t} \times 100\%)$

3. Mean squared error (MSE)

- $MSE = \frac{1}{n}\sum_{t=1}^{n}(Y_t - F_t)^2$

4. Root mean square error (RMSE)

- $RMSE = \sqrt{MSE}$

Forecasing Methods

1. Total Mean

- Mean of total data
- Good when there is no overall trend
- Does not account for trend or seasonality
- $F_{t+1} = \frac{1}{N}\sum_{i=1}^{N} Y_i$

2. Simple Moving Average (SMA)

- Mean of k most recent observations
- Moving average value F_{t+1}
- $F_{t+1} = \frac{1}{k}\sum_{i=t-k+1}^{t} Y_i$
- Large k: Infrequient fluctuations in series
- To forecast F_{t+1}
- Use error metrics to find optimal value of k
- Each observation given equal importance

3. Weighted Moving Average

- Each observation given a weight w_i
- Eg: Most recent observation given most weightage
- $F_{t+1} = \dfrac{1}{k}\sum_{i=t-k+1}^{t} w_i \times Y_i$

4. Single Exponential Smoothing

- α is smoothing constant
- No trend, no seasonal
- Weigh the most recent forecast F_t and most recent observed value y_t
- $F_{t+1} = \alpha\, Y_t + (1-\alpha)\, F_t$

  $F_{t+1} = \alpha\, Y_t + (1-\alpha)\, (\alpha\, Y_{t-1} + (1-\alpha)\, F_{t-1})$

  $F_{t+1} = \alpha\, Y_t + \alpha\, (1-\alpha)\, Y_{t-1} + (1-\alpha)^2\, F_{t-1}$

  $\ldots$

  $F_{t+1} = \sum_{i=0}^{t-1} \alpha\, (1-\alpha)^i\, Y_{t-i}$
- Influence of α

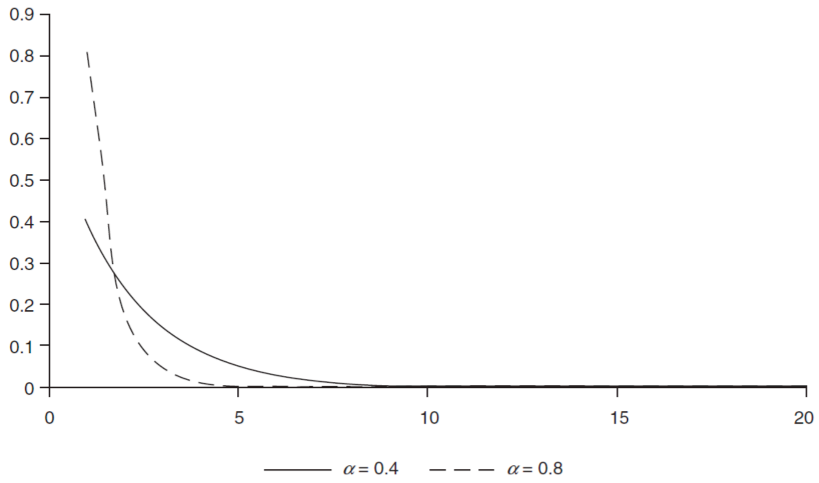


FIGURE 13.3 Exponential decay of weights to older observations.

Pros

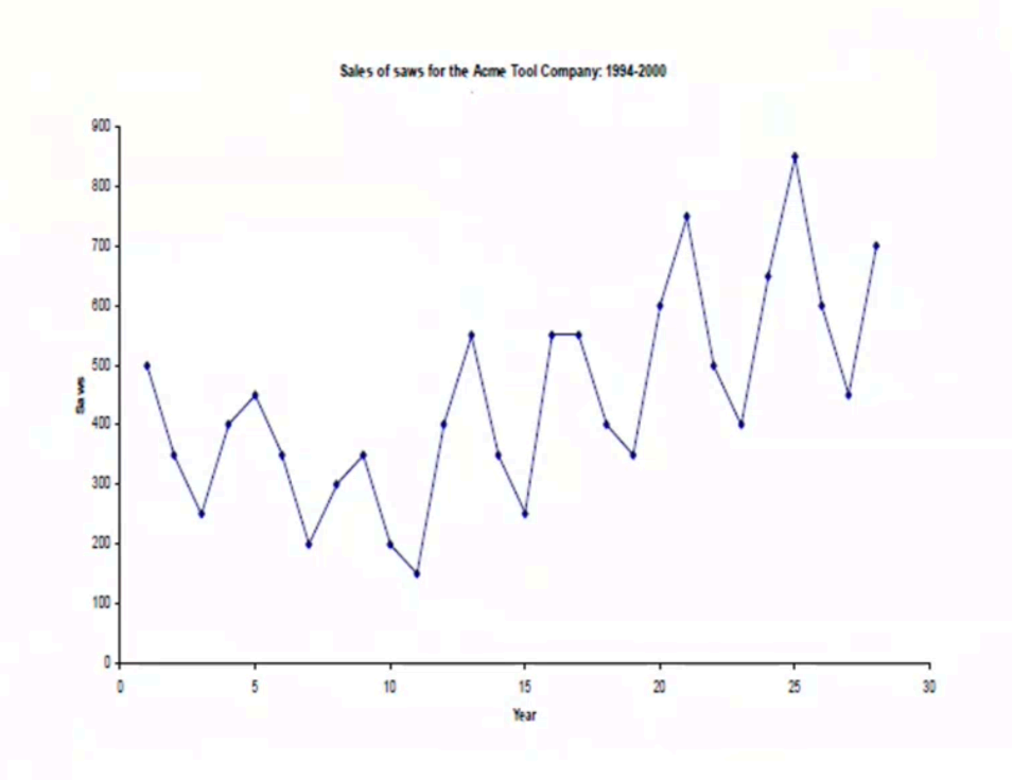
- Uses all historic data unlike SMA
- Assigns progressively decreasing weights to older data

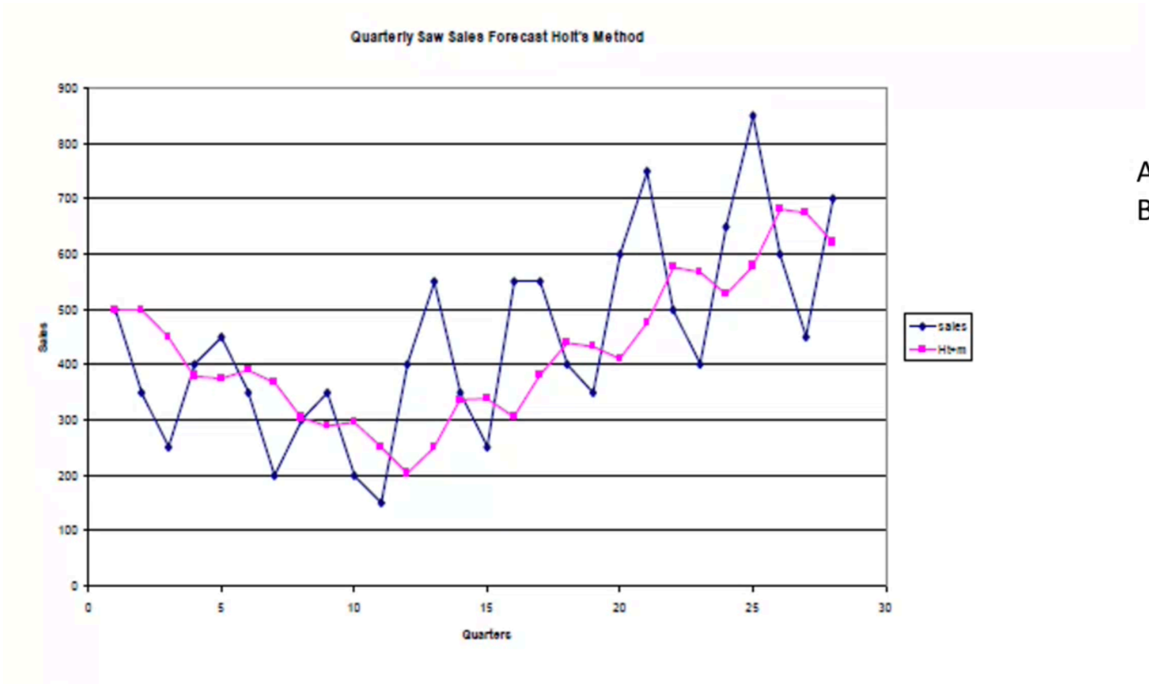
Cons

- Increasing n makes forecast less sensitive to changes
- Lags behind trend
- Larger n, larger lag
- Forecase bias and systematic error when strong trend/seasonality

5. Double Explonential Smoothing (Holt's Two Parameter ES)

- Builds trend into model
- Adds growth factor
- Level (intercept) equation
    - $L_t = \alpha \times Y_t + (1 - \alpha) \times F_t$
- Trend equation
    - $T_t = \beta \times (L_t - L_{t-1}) + (1 - \beta) \times T_{t-1}$
- The forecast for time $t + 1$ is
    - $F_{t+1} = L_t + T_t$
- The forecast for time $t + n$
    - $F_{t+n} = L_t + n\, T_t$





Alpha = 0.3
Beta = 0.1

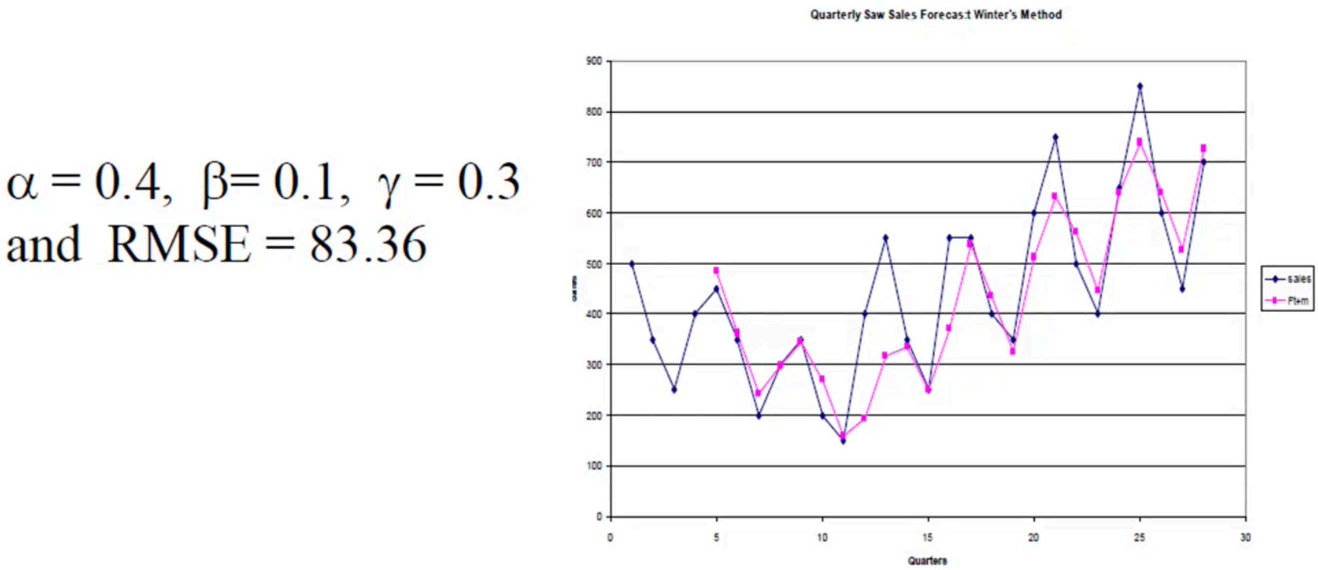
6. Triple Explonential Smoothing (Holt Winter's Method)

- c is the duration/length of seasonality

6.1 Multiplicative

- Level (intercept) equation
  - $L_t = \alpha \frac{Y_t}{S_{t-c}} + (1-\alpha)\left[L_{t-1} + T_{t-1}\right]$
- Trend equation
  - $T_t = \beta \times (L_t - L_{t-1}) + (1-\beta) \times T_{t-1}$
- Seasonal equation
  - $S_t = \gamma \frac{Y_t}{L_t} + (1-\gamma)\, S_{t-c}$
- Forecast F_{t+1}
  - $F_{t+1} = [L_t + T_t] \times S_{t+1-c}$
- Initialisations for HW Method
  - Example of annual seasonal index (12 months)
  - $L_t = Y_t$
  - $L_t = \frac{1}{c}\,(Y_1 + Y_2 + \ \ldots \ + Y_c)$
  - $T_t = \frac{1}{c}\,[\frac{Y_t - Y_{t-c}}{12} + \frac{Y_{t-1} - Y_{t-c-1}}{12} + \frac{Y_{t-2} - Y_{t-c-2}}{12} + \ldots + \frac{Y_{t-c+1} - Y_{t-2c+1}}{12}]$

- Example



6.2 Additive

- Level (intercept) equation
  - $L_t = \alpha\,(Y_t - S_{t-c}) + (1-\alpha)\,(L_{t-1} + T_{t-1})$
- Trend equation
  - $T_t = \beta \times (L_t - L_{t-1}) + (1-\beta) \times T_{t-1}$
- Seasonal equation
  - $S_t = \gamma\,(Y_t - L_t) + (1-\gamma)\,S_{t-c}$
- Forecast F_{t+m}
  - $F_{t+m} = L + m\,T_{t-1} + S_{t+m-c}$
- Initial values
  - $L_t = Y_t$
  - $T_t = \frac{1}{c}\,[\frac{Y_t - Y_{t-c}}{12} + \frac{Y_{t-1} - Y_{t-c-1}}{12} + \frac{Y_{t-2} - Y_{t-c-2}}{12} + \ldots + \frac{Y_{t-c+1} - Y_{t-2c+1}}{12}]$
  - $S_1 = Y_1 - L_c$
  - $S_2 = Y_2 - L_c$
  - $S_c = Y_c - L_c$

Predicting Seasonality Index Using Method of Averages

The seasonality index based on first 3 years of data using method of averages is shown in Table 1.

Month	Sale Quantity (2012)	Sale Quantity (2013)	Sale Quantity (2014)	Monthly Average $\bar{Y}_K$	Seasonality Index $\bar{Y}_k / \bar{\bar{Y}}$
January	3002666	4447581	4634047	4028098.00	1.087932
February	4401553	3675305	3772879	3949912.33	1.066815
March	3205279	3477156	3187110	3289848.33	0.888541
April	4245349	3720794	3093683	3686608.67	0.9957
May	3001940	3834086	4557363	3797796.33	1.02573
June	4377766	3888913	3816956	4027878.33	1.087872
July	2798343	3871342	4410887	3693524.00	0.997568
August	4303668	3679862	3694713	3892747.67	1.051375
September	2958185	3358242	3822669	3379698.67	0.912808
October	3623386	3361486	3689286	3558053.33	0.960979
November	3279115	3670362	3728654	3559377.00	0.961337
December	2843766	3123966	4732677	3566803.00	0.963342
Average of monthly averages				3702528.22	

Seasonality index can be interpreted as percentage change from the trend line.

For example, the seasonality index for January is approximately 1.088 or 108.8%.

This implies that in January, the demand will be approximately 8.8% more from the trend line.

The seasonality index for March is 0.8885 or 88.85%.

- Steps
    - **STEP 1**
    - Calculate the average of value of *Y* for each season that is, if the data is monthly data, then we need to calculate the average for each month based on the training data.
    - Let these averages be $\bar{Y}_1, \bar{Y}_2, \bar{Y}_3 \ldots\ldots \bar{Y}_c$
    - **STEP 2**
    - Calculate the average of the seasons' averages calculated in step 1 (say $\bar{\bar{Y}}$).
    - **STEP 3**
    - The seasonality index for season *k* is given by the ratio $\bar{Y}_k / \bar{\bar{Y}}$.
    - to the procedure explained above is first divide the value of Y_t with its yearly average and calculate the seasonal average
    - We will use first 3 years of data to calculate the seasonality index for various months.
- Parameter tuning for α, β and γ

7. Croston's Forecasting Method for Intermittent Demand

- Intermittent demand: spare parts, diyas, Christmas trees
- Exponential smoothing will produce bias
- Two components
  - Predicting time between demands
  - Magnitude of demand
- Forecast: mean demand per period
- Symbols
  - Y_t = demand at time t (maybe 0)
  - F_t = forecasted demand (predicted)
  - TD_t = time between latest and previous non-zero demand in period t
  - FTD_t = forecased time between demand in period t
- Steps:
  - If $Y_t = 0$ then $F_{t+1} = F_t$ and $FTD_{t+1} = FTD_t$
  - If $Y_t \neq 0$ then $F_{t+1} = \alpha\, Y_t + (1-\alpha)\, F_t$ and $FTD_{t+1} = \beta\, TD_t + (1-\beta)\, FTD_t$
- Mean demand per period D_{t+1}
  - $D_{t+1} = \dfrac{F_{t+1}}{FTD_{t+1}}$

Question

- Quarterly demand for spare parts of avionics system of an aircraft
- Use demand during Q1 to Q4 to forecast for quarters 5 to 16 using Croston's method

Quarter	1	2	3	4	5	6	7	8
Demand	20	12	0	18	16	0	20	22
Quarter	9	10	11	12	13	14	15	16
Demand	0	28	0	0	30	26	0	34

TABLE 13.8 Quarterly demand for avionic system spares

- Zero demand: 3, 6, 9, 11, 12, 15

- Procedure used for starting values of F_t and FTD_t is shown in the table here:
- TD_4 = 2 since the elapsed time from the previous demand and current demand period is 2 (4 – 2).
- The forecasted time between demand is the average TD_t values up to t = 4.
- So, FTD_4 = (1+2)/2 = 1.5.
- The forecasted demand F_4 for t = 4 is (20 + 12 + 18)/3 = 16.67.
- Note that the total value is divided by 3 (not 4) since only 3 quarters had non-zero demand.
- So, the starting values for Croston's method are.

Quarter	Demand	TD_t	FTD_t	F_t
1	20			
2	12	1		
3	0			
4	18	2	1.5	16.67

$TD_4 = 2$, $FTD_4 = 1.5$, and $F_4 = 16.67$

Let $\alpha = \beta = 0.2$. Then

$F_5 = 0.2 \times 18 + (1-0.2) * 16.67 = 16.936$

$FTD_5 = 0.2 \times 2 + (1 - 0.2) \times 1.5 = 1.6$

Forecasted demand for periods 5 to 16 using Croston's method.

Quarter	Demand	TD_t	FTD_t	F_t	$D_t = (F_t/FTD_t)$
1	20				
2	12	1			
3	0				
4	18	2	1.5000	16.67	11.11333
5	16	1	1.6000	16.936	10.585
6	0		1.4800	16.7488	11.31676
7	20	2	1.4800	16.7488	11.31676
8	22	1	1.5840	17.39904	10.98424
9	0		1.4672	18.31923	12.48585
10	28	2	1.4672	18.31923	12.48585
11	0		1.5738	20.25539	12.8707
12	0		1.5738	20.25539	12.8707
13	30	3	1.5738	20.25539	12.8707
14	26	1	1.8590	22.20431	11.94417
15	0		1.6872	22.96345	13.61034
16	34	2	1.6872	22.96345	13.61034

Question

- **Example: lubricant sales**
- Several years ago, an oil company requested forecasts of monthly lubricant sales
- One of the time series is shown in the table below.
- The data contain small counts, with many months registering no sales at all, and only small numbers of items sold in other months.

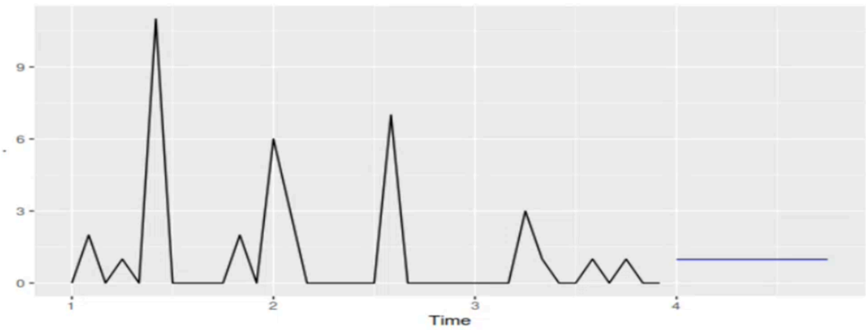
Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	0	2	0	1	0	11	0	0	0	0	2	0
2	6	3	0	0	0	0	0	7	0	0	0	0
3	0	0	0	3	1	0	0	1	0	1	0	0

- There are 11 non-zero demand values in the series, denoted by q.
- The corresponding arrival series a is also shown in the following table.

i	1	2	3	4	5	6	7	8	9	10	11
q	2	1	11	2	6	3	7	3	1	1	1
a	2	2	2	5	2	1	6	8	1	3	2

- Applying Croston's method gives the demand forecast 2.750 and the arrival forecast 2.793.
- So the forecast of the original series is $\hat{y}_{T+h|T} = 2.750/2.793 = 0.985$

In R



- `tsinterminent` package R

Case studies - Power of seasonality index

1. Forecasting study 1

- Following are the weekly attainment figures: Week 1: 75% Week 2: 77% Week 3: 79% Week 4: 81%.
- What was amiss? What is the real-life forecasting story?

CASE 1 : Forecast over-indexed in April

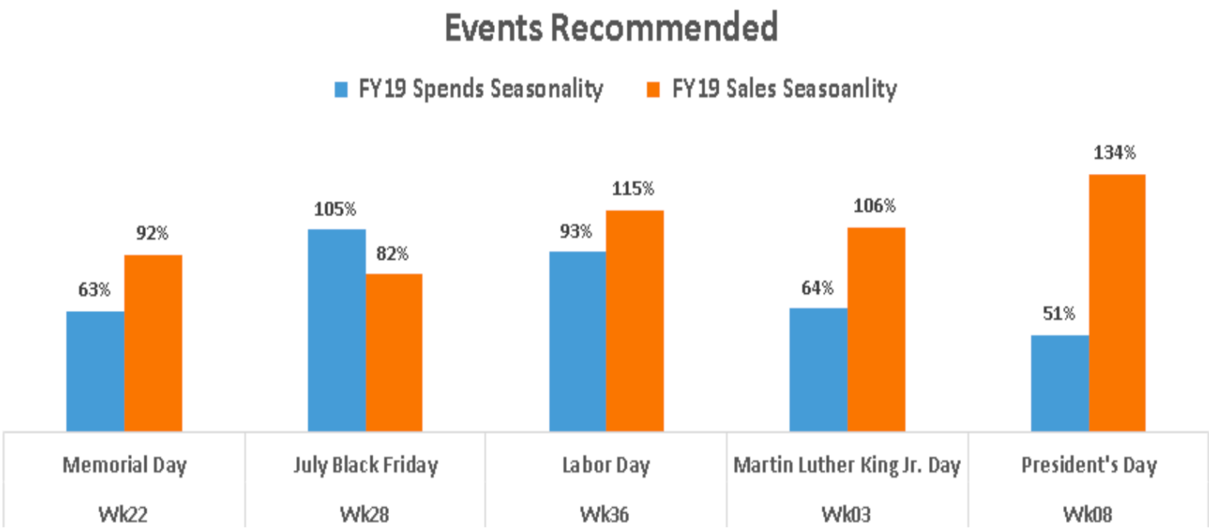
April Forecast and Actuals Comparison				
	FY19 Forecast	FY19 Actuals	FY18 Actuals	FY17 Actuals
April Average(Mn)	16.5	14.3	15.5	15.6
Year Average(Mn)	18.6	18.4	19.1	19.5
April Seasonality	89%	78%	81%	80%

- Here is what happened: As we can see from the image,
- **April forecast** seasonality was **over indexed by 11%**, i.e. at 89% of yearly average while actuals were trending towards 78%.
- What does a seasonality index mean?

Seasonality Index Calculation				
	FY19 Forecast	FY19 Actuals	FY18 Actuals	FY17 Actuals
Week 1(Mn)	18.2	17.1	18.5	17.9
Week 2(Mn)	18.4	17.0	17.6	18.7
Week 3(Mn)	19.6	16.5	17.5	18.6
Year Average(Mn)	18.6	18.4	19.1	19.5
Week 1	98%	93%	97%	92%
Week 2	99%	=F73/F$75	92%	96%
Week 3	105%	90%	91%	95%

- SI: Normalised
- Divide each number by their yearly average to calculate the index

2. Forecasting study 2



- Conclusions
    - High proportion of marketing on black friday
    - **Reallocate the spends** from July Black Friday day to President's week, for higher ROI

Regression for Forecasting

- Forecast at time t
    - $F_t = \beta_0 + \beta_1 \, X_{1t} + \beta_2 \, X_{2t} + \ \ldots \ + \beta_n \, X_{nt}$
- Example
    - $F_t = \beta_0 + \beta_1 \times (promotion\_expenses_t) + \beta_2 \times (competition\_promotion_t)$
    -

Model	*R*	*R*-Square	Adjusted *R*-Square	Std. Error of the Estimate	Durbin–Watson
1	0.928	0.862	0.853	207017.359	1.608

- Two factors not entirely independent
- Need high R^2 for forecasting (this is high enough)
- If DW = 2, no autocorrelation exists
- D = 1.608 => no autocorrelation between errors
- Autocorrelation leads to inclusion of nonsignificant variables in the equation
- Standard error of regression coefficient is underestimated

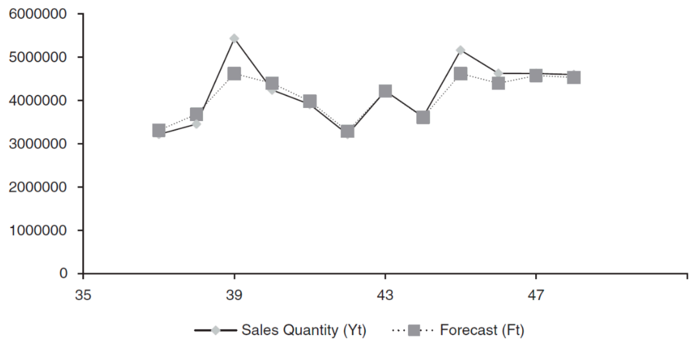
Model		Unstandardized Coefficients		Standardized Coefficients	*t*	Sig.
		B	Std. Error	Beta		
1	(Constant)	808471.843	278944.970		2.898	0.007
	Promotion Expenses	22432.941	1953.674	0.825	11.482	0.000
	Competition Promotion	−212646.036	77012.289	−0.198	−2.761	0.009

  - If competition spends more, sales reduce

$$F_t = 808471.843 + 22432.941 X_{1t} - 212646.036 X_{2t}$$

$$X_{1t} = \text{Promotion expenses at time } t$$

$$X_{2t} = \begin{cases} 1 & \text{Competition is on promotion} \\ 0 & \text{Otherwise} \end{cases}$$



- Sales increases when promotions expenses increase and the sales decrease when the competition is on the promotion.

Comparing methods

- For given example, regression performs better

Method	MAPE	RMSE
Moving Average	734725.84	14.03%
Exponential Smoothing	742339.22	13.94%
Regression	302969	**4.19%**

Forecasting with Regression - Seasonality

Steps:

1. Estimate the seasonality index (using method of averages or ratio to moving average)
2. De-seasonalise the data using additive or multiplicative model (eg: de-seasonalised data in multiplicative: $Y_{d,t} = \frac{Y_t}{S_t}$ where S_t is the seasonality index for period t)
3. Develop a forecasting model on de-seasonalised data $F_{d,t}$
4. The forecast for period $t+1$ is $F_{t+1} = F_{d,t+1} \times S_{t+1}$ (re-seasonalise)

Autoregressive Models

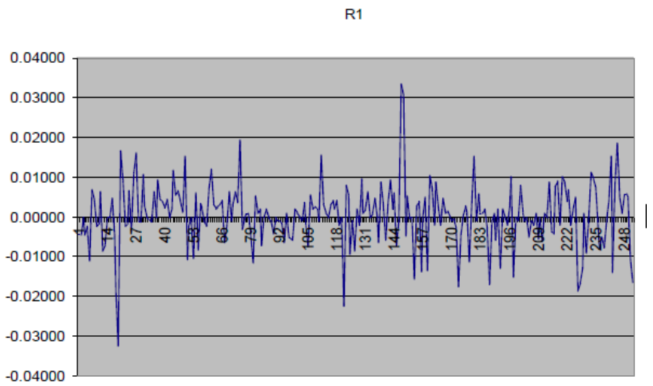
- Regression of variable on itself (AR)
- Assumption: time-series is assumed to be a stationary process
  - If TS data not stationary, must be converted to stationary before applying AR models
- Assumption: errors follow white noise (normal distribution: $\epsilon \sim N(0,\ \sigma_\epsilon^2)$)

Conditions for Stationary Time-Series data Y_t

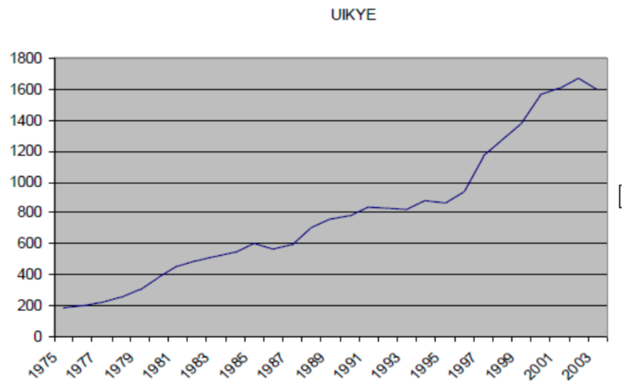
1. Means of Y_t at different values of t are constant
2. Variances of Y_t at different time periods are constant (homoscedasticity)
3. Covariance of Y_t and Y_{t-k} for different lags depend only on k and not on time t (only the interval and not the time)

Concept of Stationarity

- **Strictly stationary:** distribution of values remains same as time proceeds
- **Weakly stationary:**
    1. Constant mean: $E(y_t) = \mu$
    2. Constant variance: $E(y_t - \mu)^2 = \sigma^2$
    3. Constant auto covariance structure: $E(y_{t1} - \mu)(y_{t2} - \mu) = \gamma_{t2-t1}, \forall t_1, t_2$
- Point 3: covariance between y_{t-1} and y_{t-2} being the same as y_{t-5} and y_{t-6}



Stationary Series

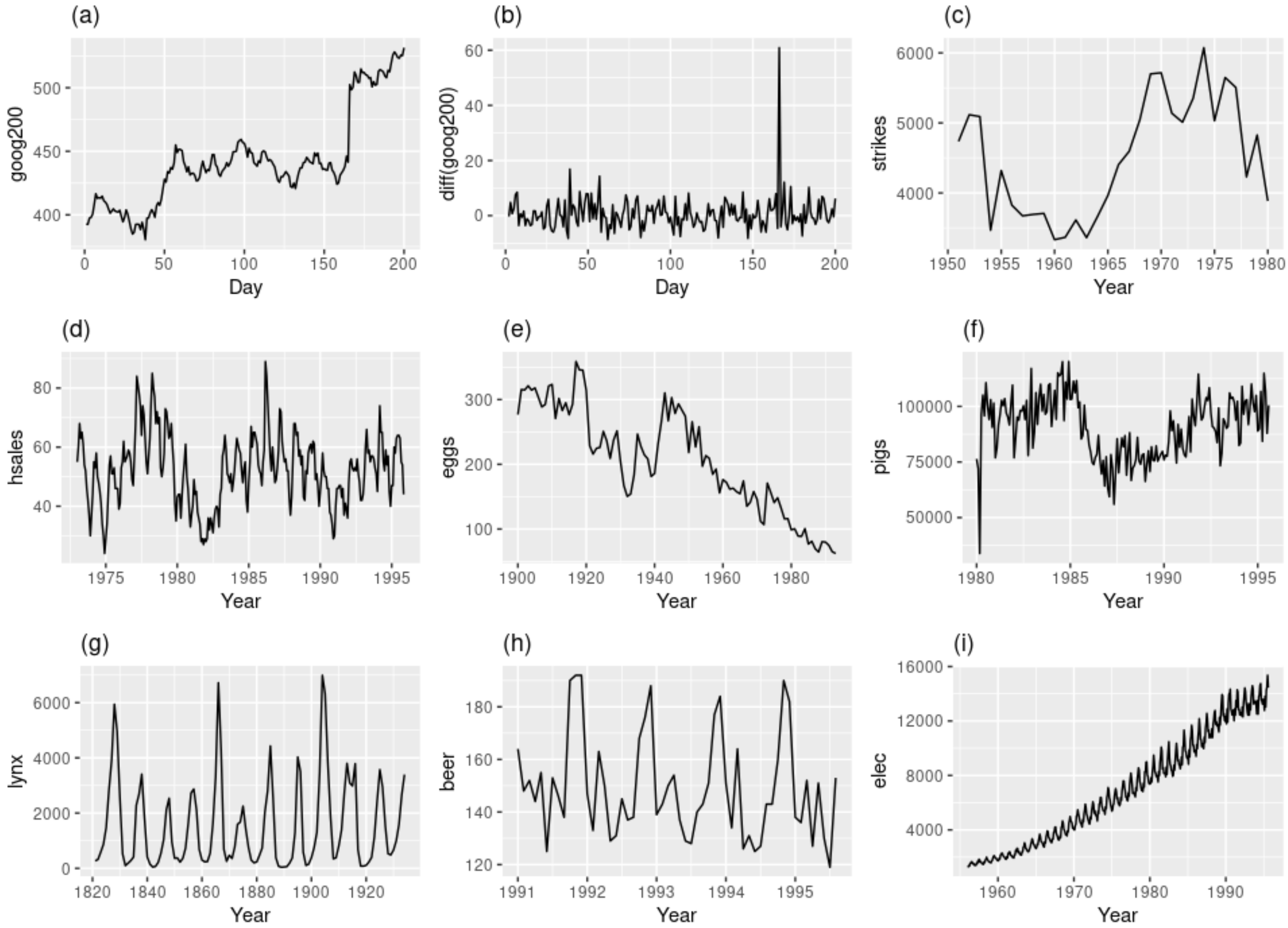


Non-stationary Series

- Graph 2 has an upward trend
- If variables in OLS are not stationary, high R^2 and low DW statistic indicate high autocorrelation
    - Caused by drift in variables
- Determine if signal is stationary
    - Plotting
    - Assesing autocorrelation function
    - Use DF, ADF tests on significance of autocorrelation coefficients

Example: which of the following are stationary?

- Source: https://otexts.com/fpp2/stationarity.html



(a) Google stock price for 200 consecutive days

- Upward trend

(b) Daily change in the Google stock price for 200 consecutive days

- **Stationary** ([first order differenced])

(c) Annual number of strikes in the US

- Seasonality and trend

(d) Monthly sales of new one-family houses sold in the US

- Seasonality

(e) Annual price of a dozen eggs in the US (constant dollars)

- Downward trend

(f) Monthly total of pigs slaughtered in Victoria, Australia

- Seasonality, trends, levels

(g) Annual total of lynx trapped in the McKenzie River district of north-west Canada

- Non-stationary upon first glance
- Cycles are aperiodic — they are caused when the lynx population becomes too large for the available feed, so that they stop breeding and the population falls to low numbers, then the regeneration of their food sources allows the population to grow again, and so on
- In the long-term, the timing of these cycles is not predictable
- Hence the series is **stationary**

(h) Monthly Australian beer production

- Trend, seasonality

(i) Monthly Australian electricity production

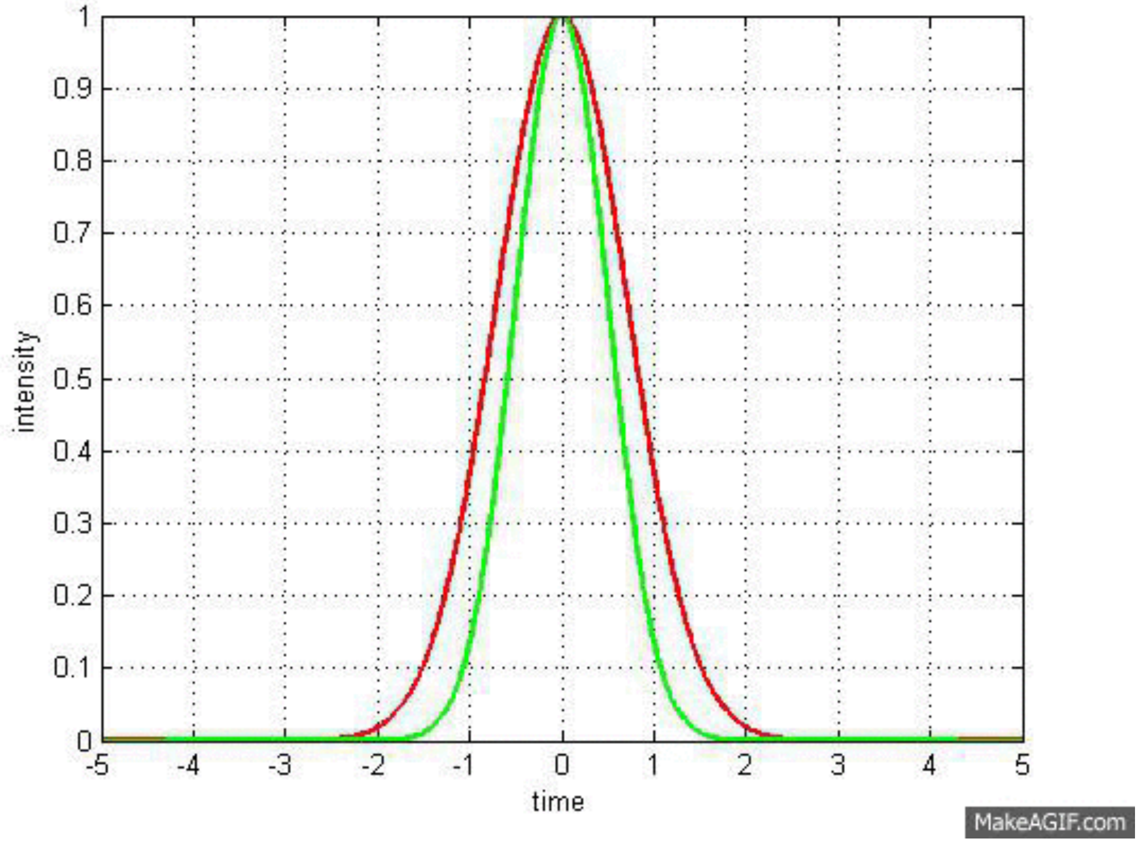
- Trend, seasonality

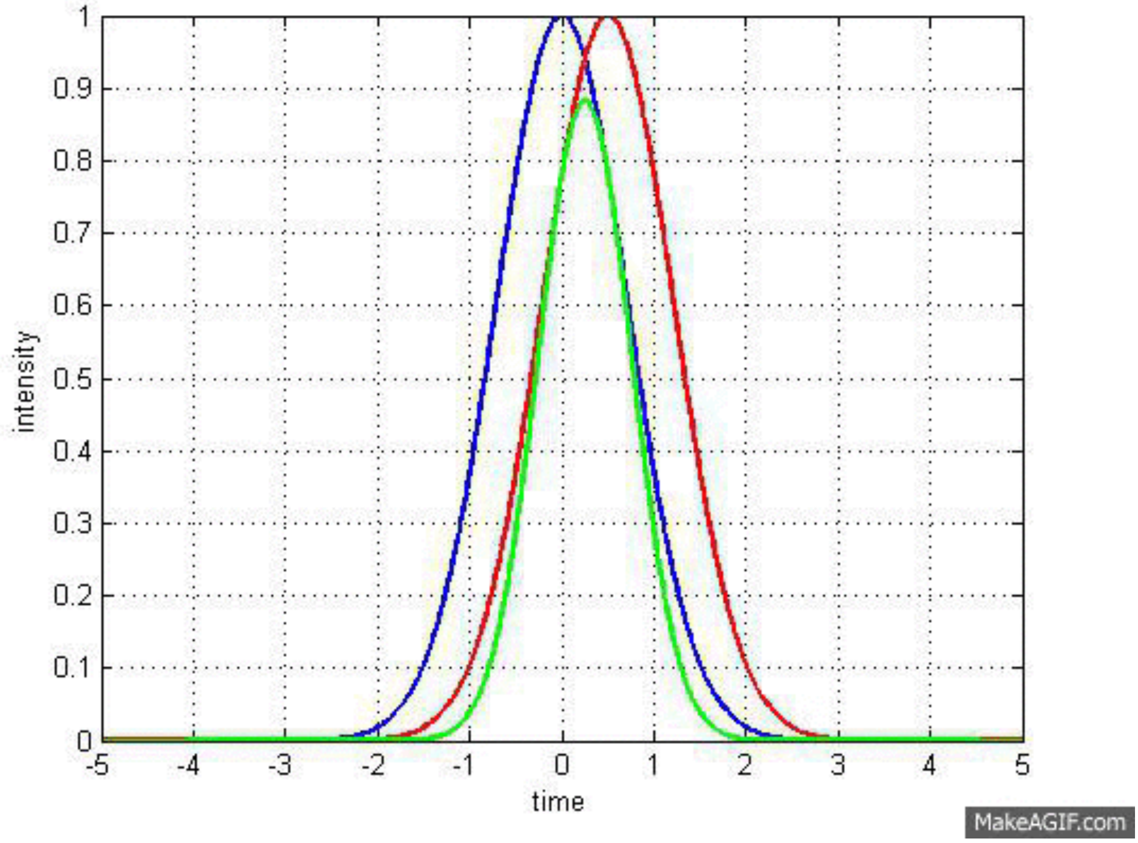
ACF and PACF

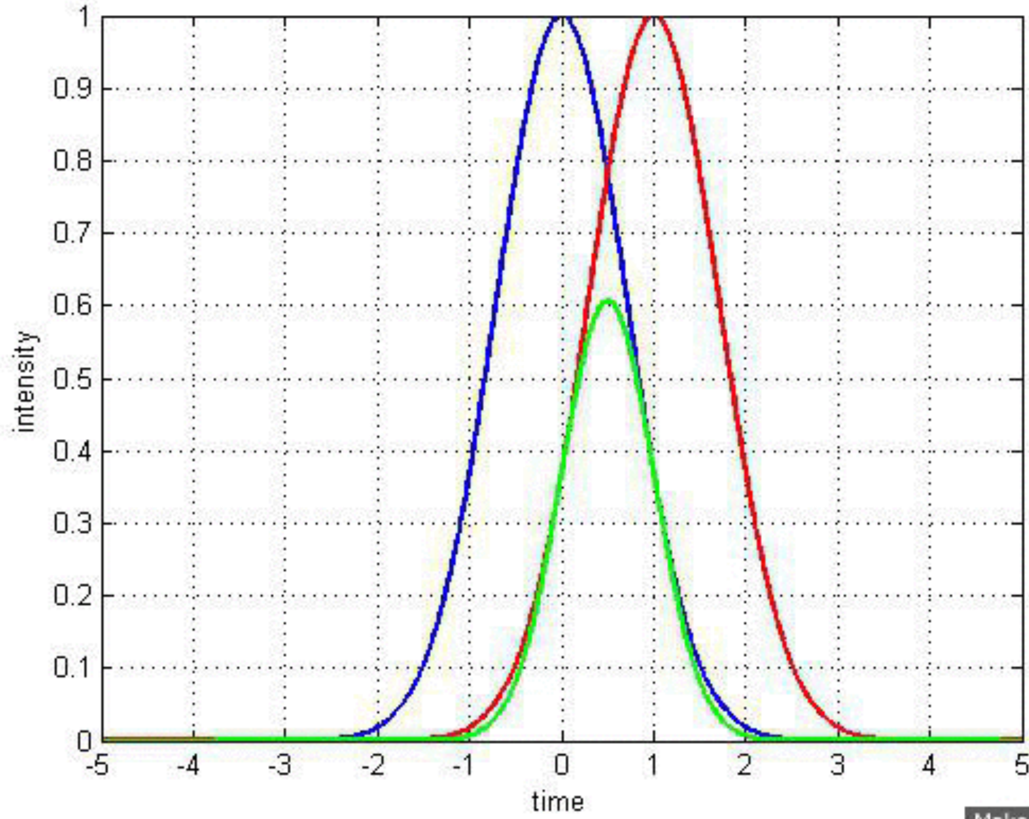
- ACF: autocorrelation function
- PACF: partial autocorrelation function

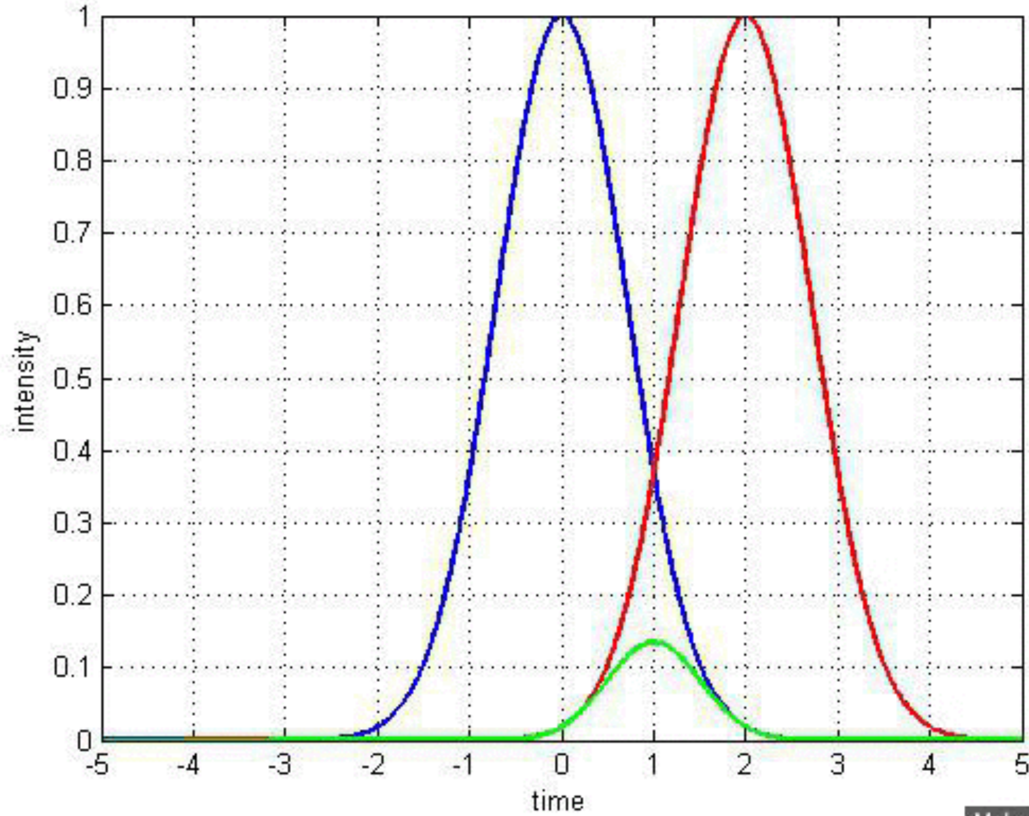
ACF at lag k

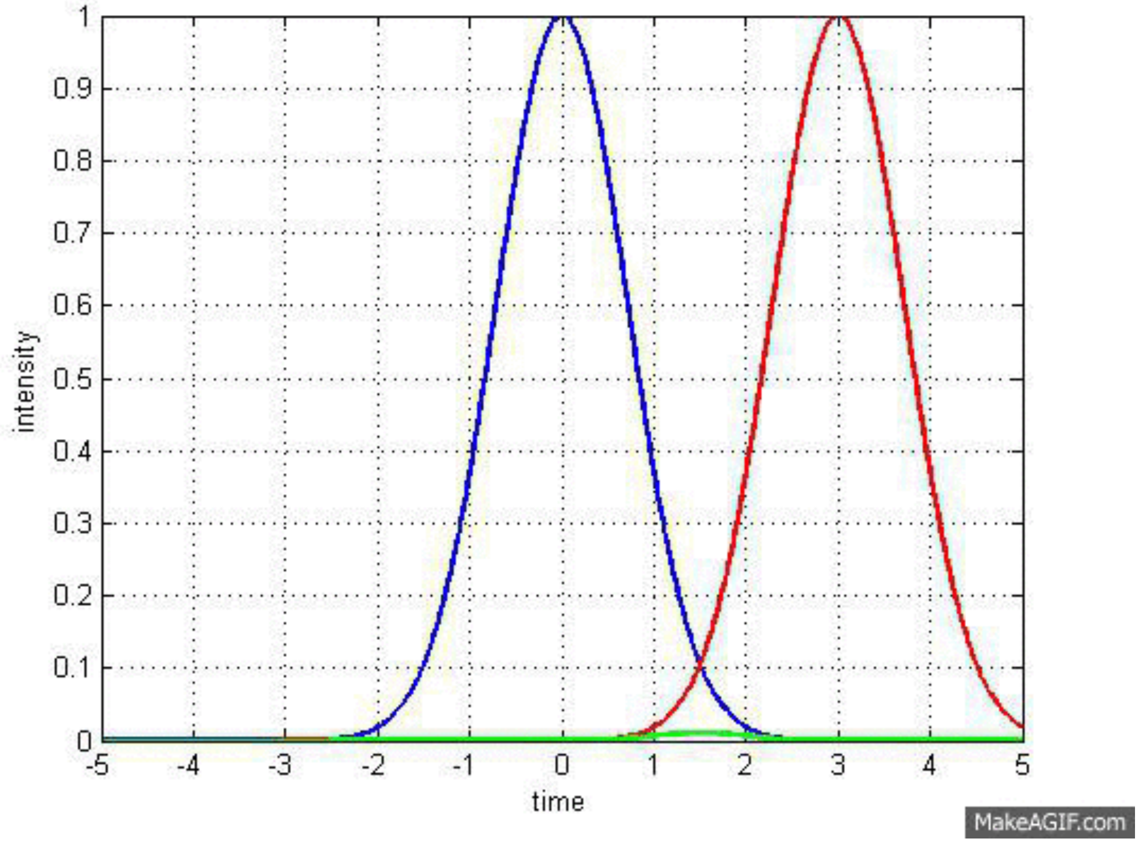
- Stationary TS: ACF function of lag and not time
- $\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{covariance\ at\ lag\ k}{variance}$
- $\rho_k = \frac{\sum_{t=k+1}^{n} (Y_{t-k} - \bar{Y})(Y_t - \bar{Y})}{\sum_{t=1}^{n} (Y_t - \bar{Y})^2}$
- ACF between -1 and 1

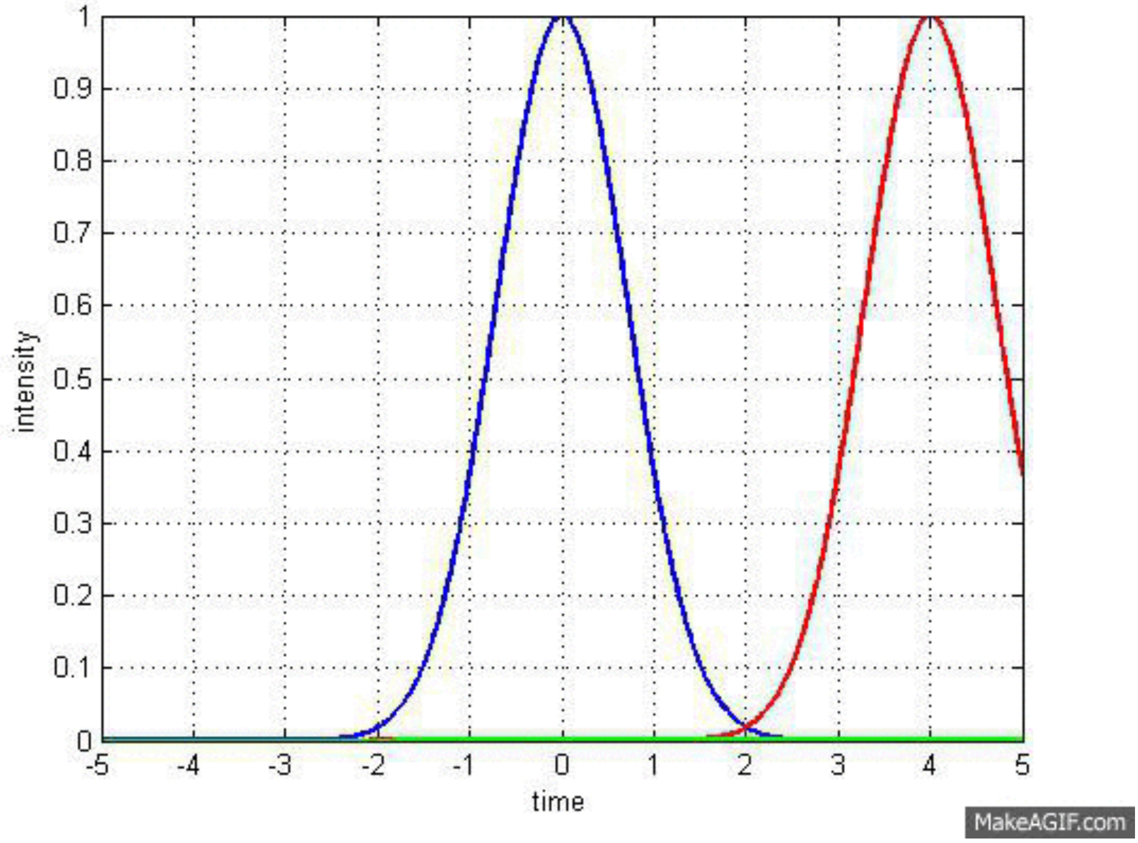

intensity
time
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-5
-4
-3
-2
-1
0
1
2
3
4
5
MakeAGIF.com

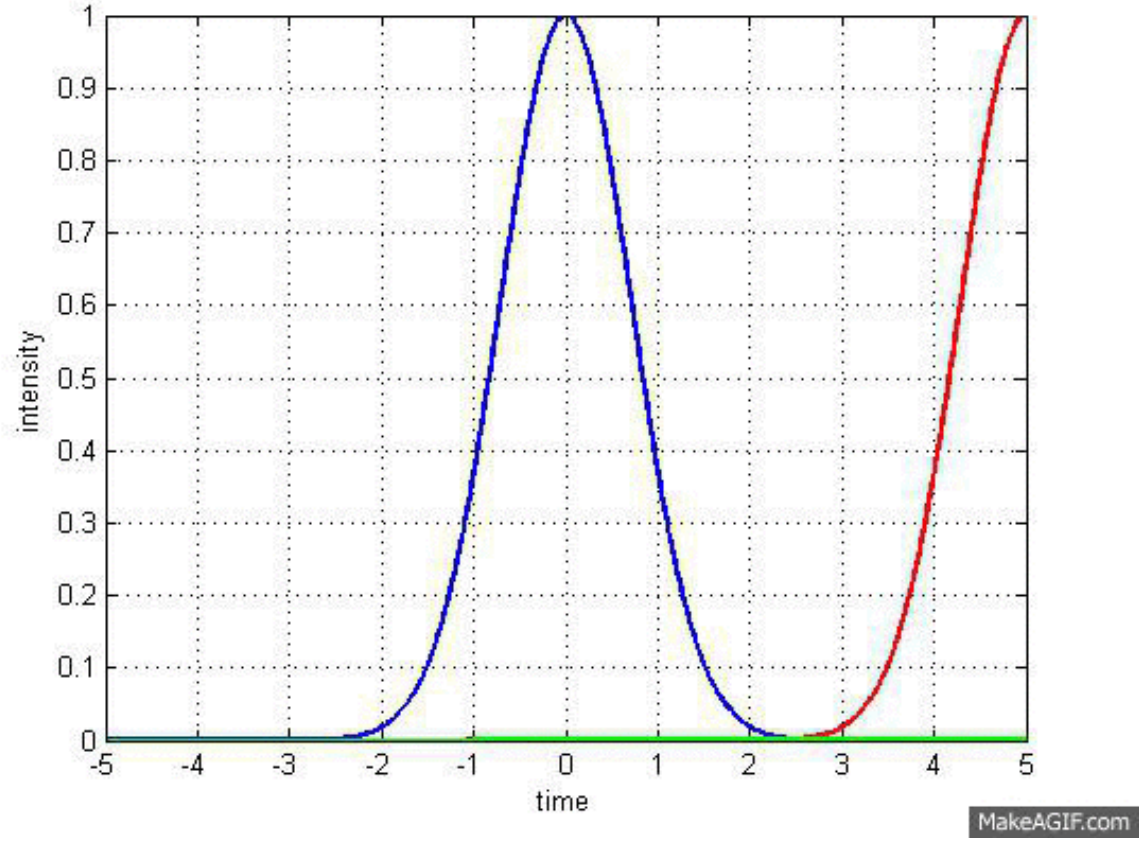

intensity
time
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-5
-4
-3
-2
-1
0
1
2
3
4
5
MakeAGIF.com


intensity
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-5
-4
-3
-2
-1
0
1
2
3
4
5
time

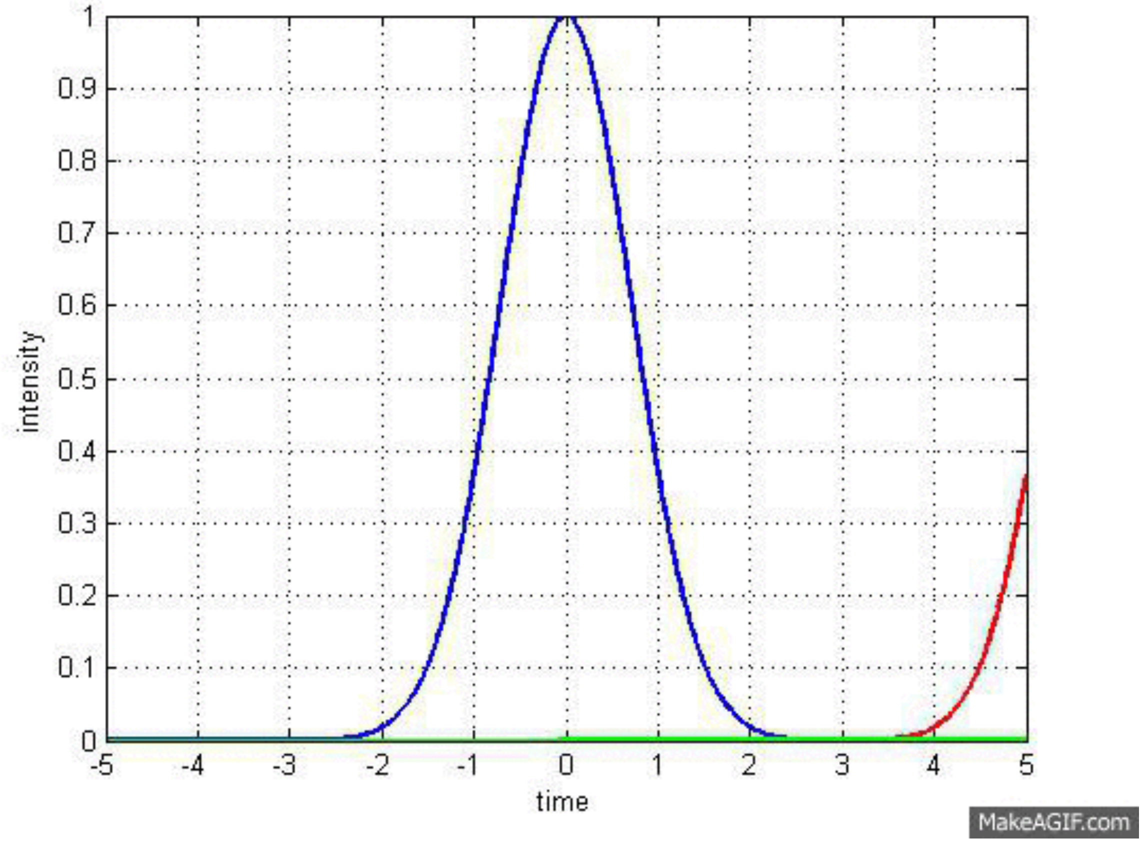

intensity
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-5
-4
-3
-2
-1
0
1
2
3
4
5
time


intensity
time
MakeAGIF.com


intensity
time
MakeAGIF.com



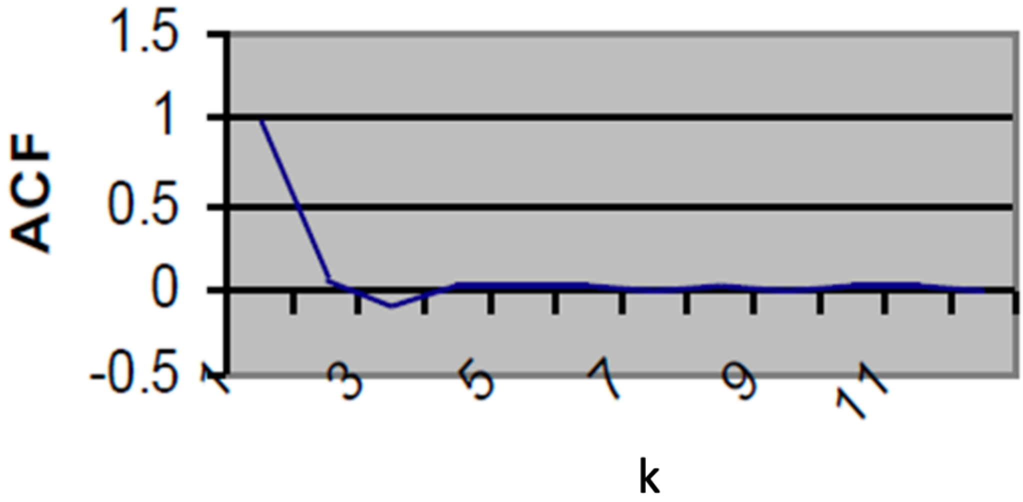
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
intensity
-5
-4
-3
-2
-1
0
1
2
3
4
5
time
MakeAGIF.com



1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
intensity
-5
-4
-3
-2
-1
0
1
2
3
4
5
time
MakeAGIF.com

Correlogram

- ACF against k - sample correlogram



- Determine stationarity: if ACF falls immediately from 1 to 0 and then equals about 0 thereafter, series is stationary
- If ACF declines gradually from 1 to 0 over a long period of time, it is not stationary
- Plot shown above: stationary

Statistical Significance of ACF

- Q-statistic: if sample ACFs are jointly equal to 0
- $Q = n \sum_{k=1}^{m} \hat{\rho}_k{}^2$
  - n: sample size
  - m: lag length
- If jointly equal to 0, TS is stationary
- Null hypothesis: sample ACFs jointly equal 0
- Follows $\chi^2(m)$: m degrees of freedom

PACF

- Partial Autocorrelation Function
- Correlations between observations k time periods apart after controlling for correlations at intermediate lags
- First order (k=1) ACF and PACF are same
- Second order (k=2) PACF
  - $\frac{cov(y_t,\ y_{t-2}|y_{t-1})}{\sqrt{var(y_t|y_{t-1})\ var(y_{t-2}|y_{t-1})}}$

- Partial correlogram (Box-Jenkins methodology)

AR, MA and ARMA Model

1. AR (Autoregression)

- Assumption: stationary
- Model: (no level or trend $\implies \beta_0 = 0$)
  - $Y_{t+1} = \beta\, Y_t + \epsilon_{t+1}$
- If level/mean present
  - $Y_{t+1} - \mu = \beta \times (Y_t - \mu) + \epsilon_{t+1}$
- Expanding Y_t
  - $Y_{t+1} - \mu = \beta \times (\beta \times (Y_{t-1} - \mu) + \epsilon_t) + \epsilon_{t+1}$
- Expanding fully
  - $Y_{t+1} - \mu = \beta^t (Y_0 - \mu) + \beta^{t-1}\epsilon_1 + \beta^{t-2}\epsilon_2 + \ \ldots \ + \beta\epsilon_t + \epsilon_{t+1}$
  - $Y_{t+1} - \mu = \beta^t (Y_0 - \mu) + \sum_{k=1}^{t-1} \beta^{t-k} \times \epsilon_k + \epsilon_{t+1}$
- Practical purposes: $|\beta| < 1$
- The second part of the equation can also become infinitely large if the errors do not follow a white noise

Estimating β

- $\sum\limits_{t=2}^{n} \epsilon_t^2 = \sum\limits_{t=2}^{n} [(Y_t - \mu) - \beta \times (Y_{t-1} - \mu)]^2$
- Take derivative and equate to 0, solve for β
- $\hat{\beta} = \dfrac{\sum\limits_{t=2}^{n} (Y_t - \mu)(Y_{t-1} - \mu)}{\sum\limits_{t=2}^{n} (Y_{t-1} - \mu)^2}$

AR Model Identification

ACF:

- $\rho_k = \dfrac{\sum\limits_{t=k+1}^{n} (Y_{t-k} - \bar{Y})(Y_t - \bar{Y})}{\sum\limits_{t=1}^{n} (Y_t - \bar{Y})^2}$: autocorrelation coefficient for order k
- $H_0 : \rho_k = 0$
- $H_a : \rho_k \neq 0$

- Null hypothesis rejected when $|\rho_k| > \dfrac{1.96}{\sqrt{n}}$

PACF: partial autocorrelation of order k

- $H_0 : \rho_{pk} = 0$
- $H_a : \rho_{pk} \neq 0$
- Null hypothesis rejected when $|\rho_{pk}| > \dfrac{1.96}{\sqrt{n}}$

Order of AR(p)

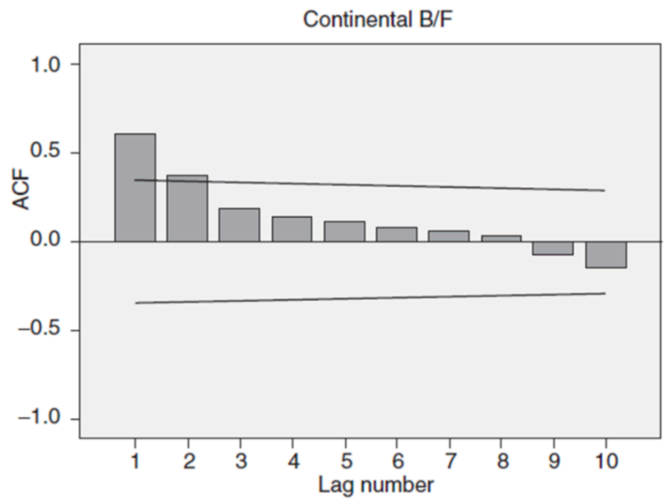
- ACF: spikes decay towards zero, coefficients may oscillate
- PACF: spikes decay to zero after lag p

Example

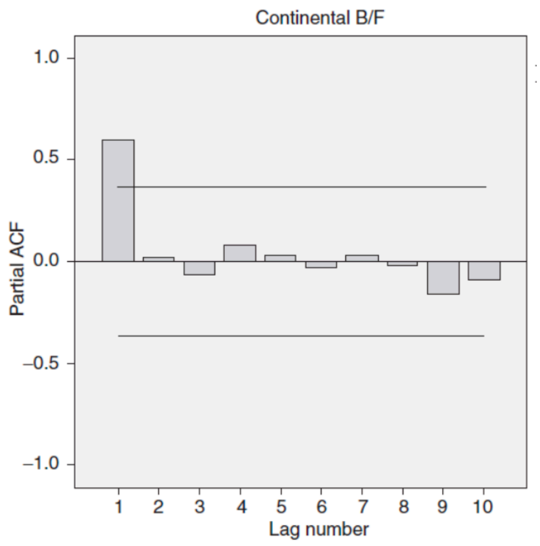
Build an auto-regressive model based on the first 30 days of data and forecast the demand for continental breakfast on days 31 to 37. Comment on the accuracy of the forecast.

Day	Demand CB	Day	Demand CB
1	25	20	43
2	25	21	41
3	25	22	46
4	35	23	41
5	41	24	40
6	30	25	32
7	40	26	41
8	40	27	41
9	40	28	40
10	40	29	43
11	40	30	46
12	40	31	45
13	44	32	45
14	49	33	46
15	50	34	43
16	45	35	40
17	40	36	41
18	42	37	41
19	40		

- Finding p using ACF and PACF plots (first 30 observations)



ACF



PACF

- Critical values: horizontal lines
- Reject null hypothesis where vertical bar beyond critical values
- ACF: spikes decay towards zero, coefficients may oscillate
- PACF: spikes decay to zero after lag p
- PACF hits lag 0 at 2 $\implies$ p = 1
- $\therefore$ model is $AR(1)$

Results of AR(1)

Model	Model Fit Statistics			
	R-Square	RMSE	MAPE	Normalized BIC
Continental B/F-Model_1	0.373	5.133	10.518	3.498

- Left: using actual value Y_t, right: using forecasted value F_t (here, k = 1)

$$(F_{t+1} - 38.890) = 0.731(Y_t - 38.890)$$

Day	Y_t	F_t	$(Y_t - F_t)^2$	$\|Y_t - F_t\|/Y_t$
31	45	44.08741	0.832821	0.02028
32	45	43.35641	2.701388	0.036524
33	46	43.35641	6.988568	0.057469
34	43	44.08741	1.182461	0.025289
35	40	41.89441	3.588789	0.04736
36	41	39.70141	1.686336	0.031673
37	41	40.43241	0.322158	0.013844

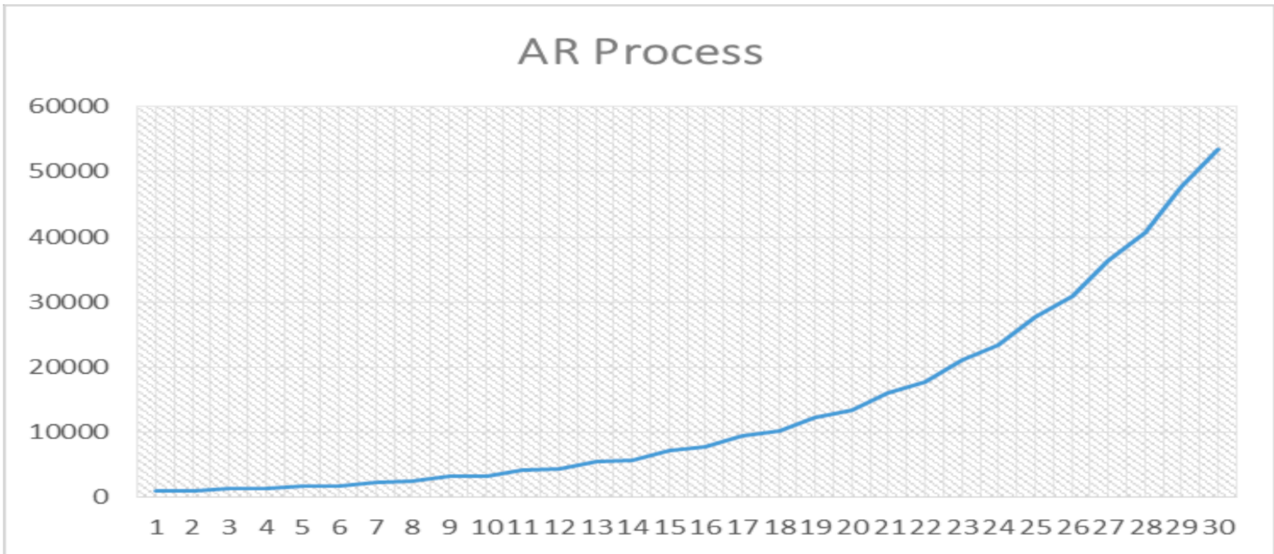
MAPE 1.5721
RMSE 0.0332 (3.32%)

$$(F_{t+k} - 38.890) = 0.731(F_{t+k-1} - 38.890)$$

Day	Y_t	F_t	$(Y_t - F_t)^2$	$\|Y_t - F_t\|/Y_t$
31	45	44.0874	0.8328	0.0203
32	45	42.6893	5.3393	0.0513
33	46	41.6673	18.7723	0.0942
34	43	40.9202	4.3256	0.0484
35	40	40.3741	0.1399	0.0094
36	41	39.9749	1.0509	0.0250
37	41	39.6830	1.7344	0.0321

MAPE 2.1446
RMSE 0.04009 (4.009%)

Example of AR Model



AR(1) $y_t = a1^* \ y_{t-1}$
AR(2) $y_t = a1^* \ y_{t-1} + a2^* \ y_{t-2}$
AR(3) $y_t = a1^* \ y_{t-1} + a2^* \ y_{t-2} + a3^* \ y_{t-2}$

2. MA (Moving Average)

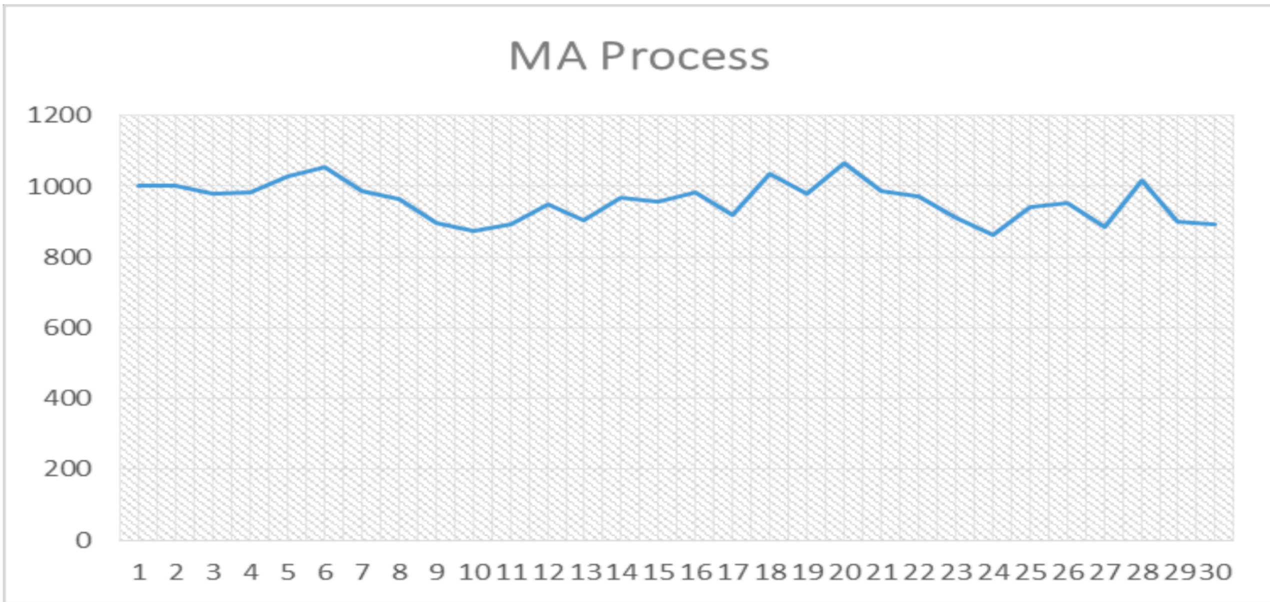
- Dependent model regressed against lagged values of past terms or error terms
- MA(q)
- Modelling the errors and not the terms themselves
- $Y_{t+1} = \mu + \alpha_1 \epsilon_t + \epsilon_{t+1}$
- MA(q) given by

- $Y_{t+1} = \mu + \alpha_1 \epsilon_t + \alpha_2 \epsilon_{t-1} + \ \ldots \ + \alpha_q \epsilon_{t-q+1} + \epsilon_{t+1}$

Order of MA(q)

- ACF: spikes decay to zero after lag of q
- PACF: spikes decay towards zero, coefficients may oscillate

Example of MA Model



MA(1) $\varepsilon_t = b1 * \varepsilon_{t-1}$

MA(2) $\varepsilon_t = b1 * \varepsilon_{t-1} + b2 * \varepsilon_{t-2}$

MA(3) $\varepsilon_t = b1 * \varepsilon_{t-1} + b2 * \varepsilon_{t-2} + b3 * \varepsilon_{t-3}$

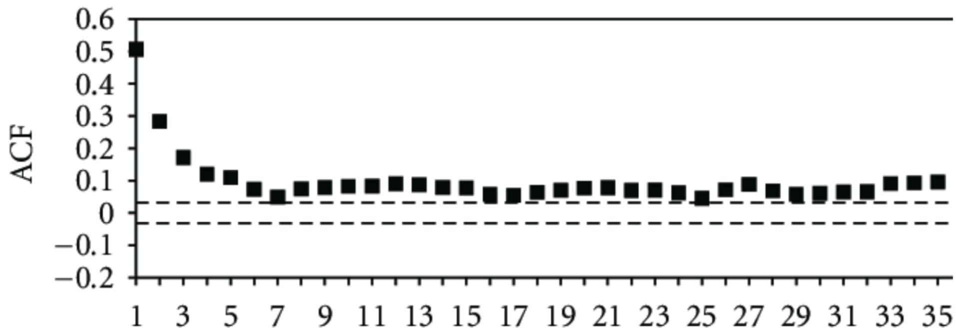
3. AR(p) and MA(q) - ARMA(p,q)

- Stationary time series - ACF and correlogram, Q-statistic
- AR(p): p lags of the dependent variable
- MA(q): q lags of the error term
- ARMA(p, q): autoregressive, moving average
- $Y_{t+1} = \overbrace{\beta_1 Y_t + \beta_2 Y_{t-1} + \ldots + \beta_p Y_{t-p+1}}^{\text{Auto Regressive Part}} + \overbrace{\alpha_1 \varepsilon_t + \alpha_2 \varepsilon_{t-1} + \ldots + \alpha_q \varepsilon_{t-q+1}}^{\text{Moving Average Part}} + \varepsilon_{t+1}$

1. Auto-correlation value, $|\rho_p| > 1.96/\sqrt{n}$ for first q values (first q lags) and cuts off to zero.
2. Partial auto-correlation function, $|\rho_{pk}| > 1.96/\sqrt{n}$ for first p values and cuts off to zero.

Summary of Parameter Selections for AR, MA and ARMA

Model	ACF	PACF
AR (p)	Spikes decay towards zero. Coefficients may oscillate.	Spikes decay to zero after lag p
MA (q)	Spikes decay to zero after lag q	Spikes decay towards zero. Coefficients may oscillate.
ARMA (p, q)	Spikes decay (either direct or oscillatory) to zero beginning after lag q	Spikes decay (either direct or oscillatory) to zero beginning after lag p



Summary of AR, MA and ARMA Models

- Autoregressive AR process:
  - Series current values depend on its own previous values
  - AR(p) - Current values depend on its own p-previous values
  - P is the order of AR process
- Moving average MA process:
  - The current deviation from mean depends on previous deviations
  - MA(q) - The current deviation from mean depends on q- previous deviations
  - q is the order of MA process
- Autoregressive Moving average ARMA process

Example of ARMA

Monthly demand for avionic system spares used in Vimana 007 aircraft is provided. Build an ARMA model based on the first 30 months of data and forecast the demand for spares for months 31 to 37. Comment on the accuracy of the forecast.

Month	Demand for Spares	Month	Demand for Spares
1	457	20	516
2	439	21	656
3	404	22	558
4	392	23	647
5	403	24	864
6	371	25	610
7	382	26	677
8	358	27	609
9	594	28	673
10	482	29	400
11	574	30	443
12	704	31	503
13	486	32	688
14	509	33	602
15	537	34	629
16	407	35	823
17	523	36	671
18	363	37	487
19	479		

1. Plot ACF and PACF (use confidence limits)

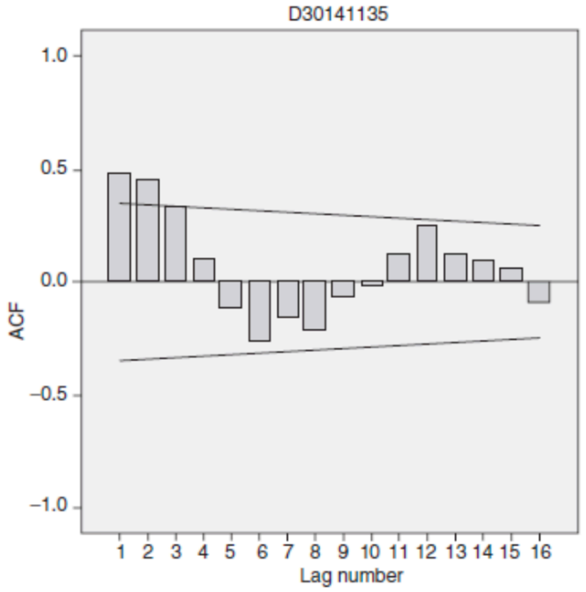


FIGURE 13.9 ACF plot for avionic system spares deman

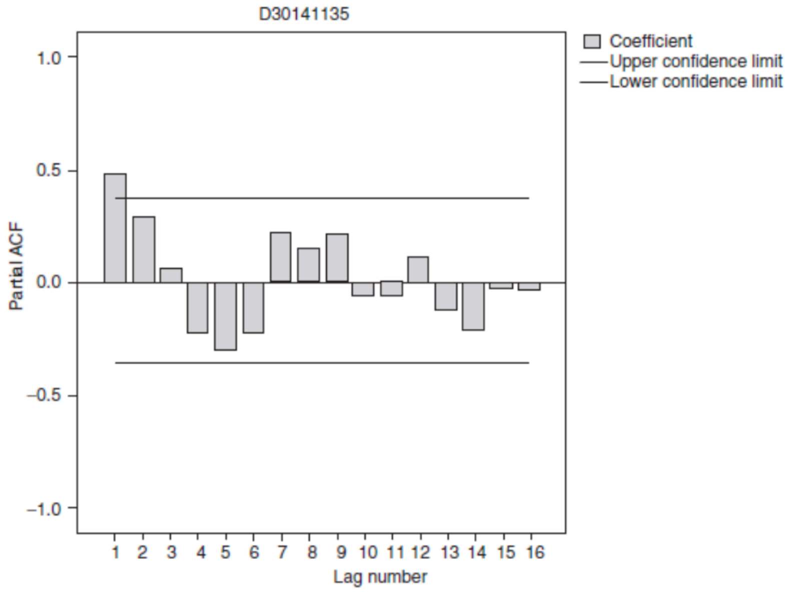


FIGURE 13.10 PACF plot for avionic system spares demand.

2. Forecast ARMA (1, 2)

Model	Model Fit Statistics		
	Stationary *R*-Squared	RMSE	MAPE
Avionic Spares	0.429	98.824	14.231

TABLE 13.26 model parameters

			Estimate	*SE*	*T*	Sig.
Avionic Spares		Constant	496.699	57.735	8.603	0.000
	AR	Lag 1	0.706	0.170	4.153	0.000
	MA	Lag 1	0.694	0.173	4.006	0.000
		Lag 2	−0.727	0.170	−4.281	0.000

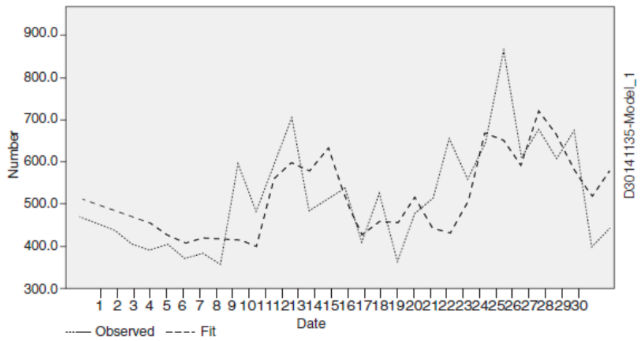


FIGURE 13.11 Observed versus forecasted demand.

All the three components in the ARMA model (AR lag 1 and MA lags 1 and 2) are statistically significant (Table 13.26). The model equation using SPSS is given by

$$Y_{t+1} - 496.669 = 0.706 \times (Y_t - 496.699) - 0.694 \times \varepsilon_t + 0.727 \times \varepsilon_{t-1} \qquad (13.45)$$

3. Compute MAPE, RMSE

TABLE 13.27 ARMA(1, 2) model forecast

Month	Y_t	F_t	$(Y_t - F_t)^2$	$\lvert Y_t - F_t \rvert / Y_t$
31	503	464.8107	1458.423	0.075923
32	688	378.5341	95769.15	0.449805
33	602	444.6372	24763.04	0.2614
34	629	685.8851	3235.909	0.090437
35	823	743.5124	6318.281	0.096583
36	671	630.7183	1622.614	0.060032
37	487	649.3491	26357.22	0.333366

The RMSE and MAPE for the validation data (months 31 and 37) are 150.961 0.1953 (19.53%), respectively (Table 13.27).

The forecasted values using F_t instead of Y_t when forecasting for more than one period ahead in time are shown in Table 13.28.

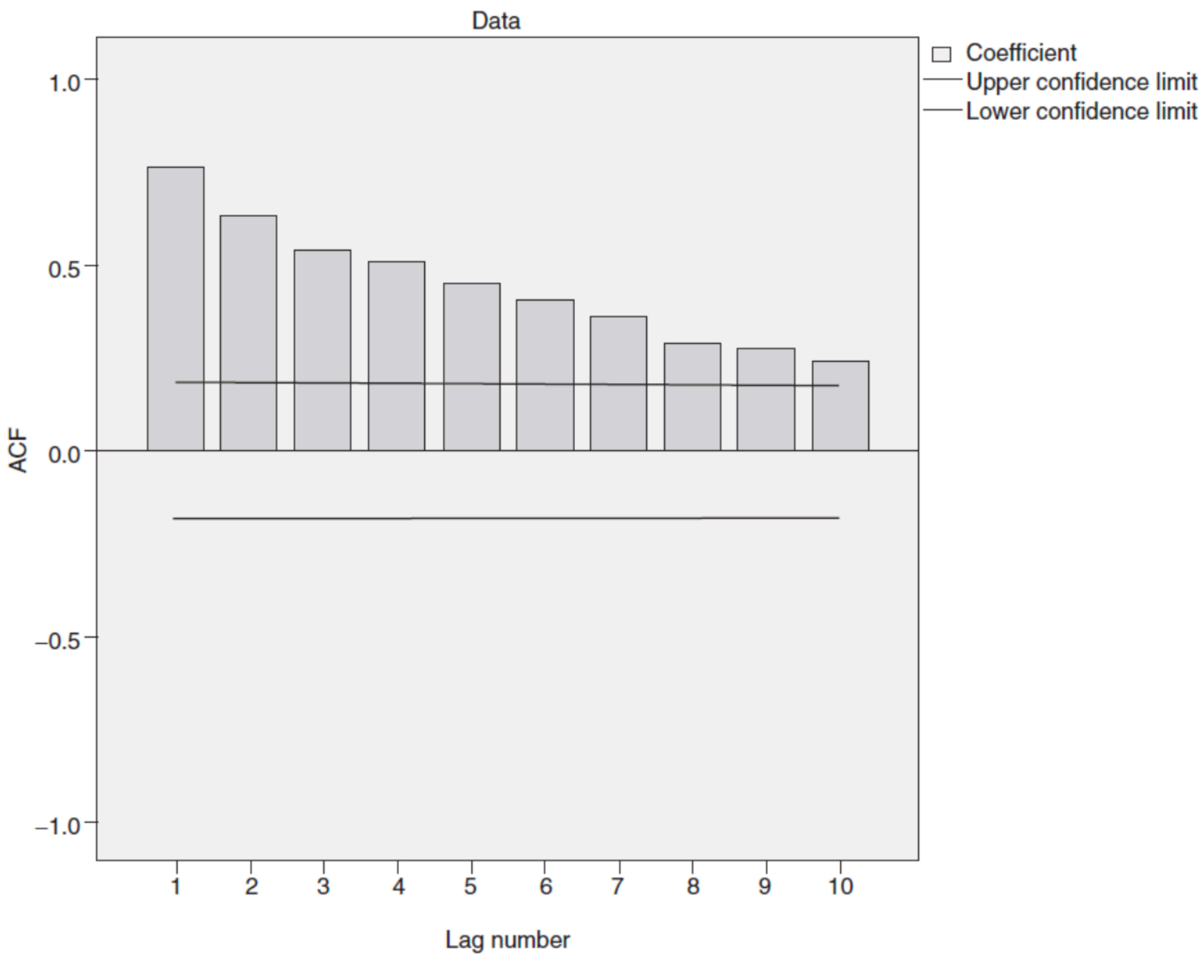
TABLE 13.28 ARMA (1, 2) forecast

Month	Y_t	F_t	$(Y_t - F_t)^2$	$\lvert Y_t - F_t \rvert / Y_t$
31	503	464.4239	1488.1147	0.0767
32	688	377.8374	96200.8258	0.4508
33	602	444.5195	24800.1101	0.2616
34	629	687.2082	3388.1980	0.0925
35	823	744.9583	6090.4998	0.0948
36	671	630.5592	1635.4571	0.0603
37	487	648.3959	26048.6313	0.3314

The RMSE and MAPE for the validation data (months 31 and 37) are 151.02 and 0.1954 (19.54%), respectively.

Concept of Stationarity, DF, ADF

1. Identifying Stationarity using ACF



- Slow decline and no cut-off to 0 $\Longrightarrow$ non-stationarity

2. Quantitative Test - Dickey-Fuller (DF) Test

- AR(1) defined as
  - $Y_{t+1} = \beta Y_t + \epsilon_{t+1}$
- If $|\beta| > 1$, AR(1) process can become very large
- If $|\beta| = 1$, non-stationary
- DF test is hypothesis test
  - $H_0 : \beta = 1$ (time-series is non-stationary)
  - $H_a : \beta < 1$ (time series is stationary)
- AR(1) written as
  - $Y_{t+1} - Y_t = \Delta Y_t = (\beta - 1)Y_t + \epsilon_{t+1} = \psi Y_t + \epsilon_{t+1}$
  - $\psi = \beta - 1$
- DF test in terms of ψ
  - $H_0 : \psi = 0$ (time-series is non-stationary)
  - $H_a : \psi < 0$ (time series is stationary)
- DF test statistic = $\frac{\psi}{S_e(\psi)}$
- $S_e(\psi)$ is the standard error of ψ

3. Augmented DF (ADF) Test

- DF test only valid if residual ϵ_{t+1} follows a white noise
- When ϵ_{t+1} is not white noise, actual series may not be AR(1)
  - May have more significant lags
- Solution: augment p-lags of the dependent variable Y
  - $\Delta Y_t = \psi Y_t + \sum_{i=0}^{p} \alpha_i \Delta Y_{t-i} + \epsilon_{t+1}$
- Augmented DF test hypotheses
  - $H_0 : \psi = 0$ (time-series is non-stationary)
  - $H_a : \psi < 0$ (time series is stationary)

Differencing - Transforming Non-Stationary Signal to Stationary

- Order of differencing d to convert a non-stationary signal to a stationary signal
- Left and right difference (usually left difference)
- Due to trend
  - De-trending: fit a trend line and subtract it from the time-series

- Otherwise
    - Differencing TS process into difference stationary

1. First Difference

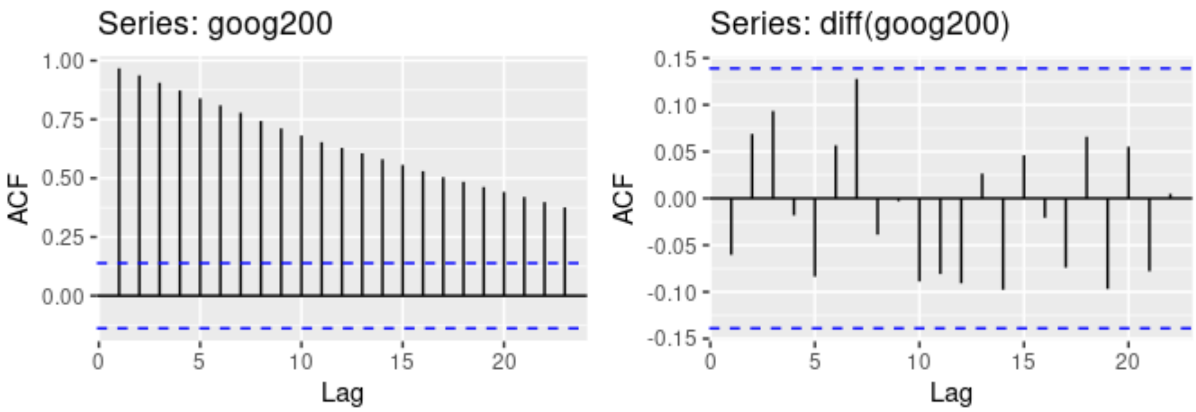
- $d = 1$
- Difference between consecutive values of the TS
- $\nabla Y_t = Y_t - Y_{t-1}$

2. Second Difference

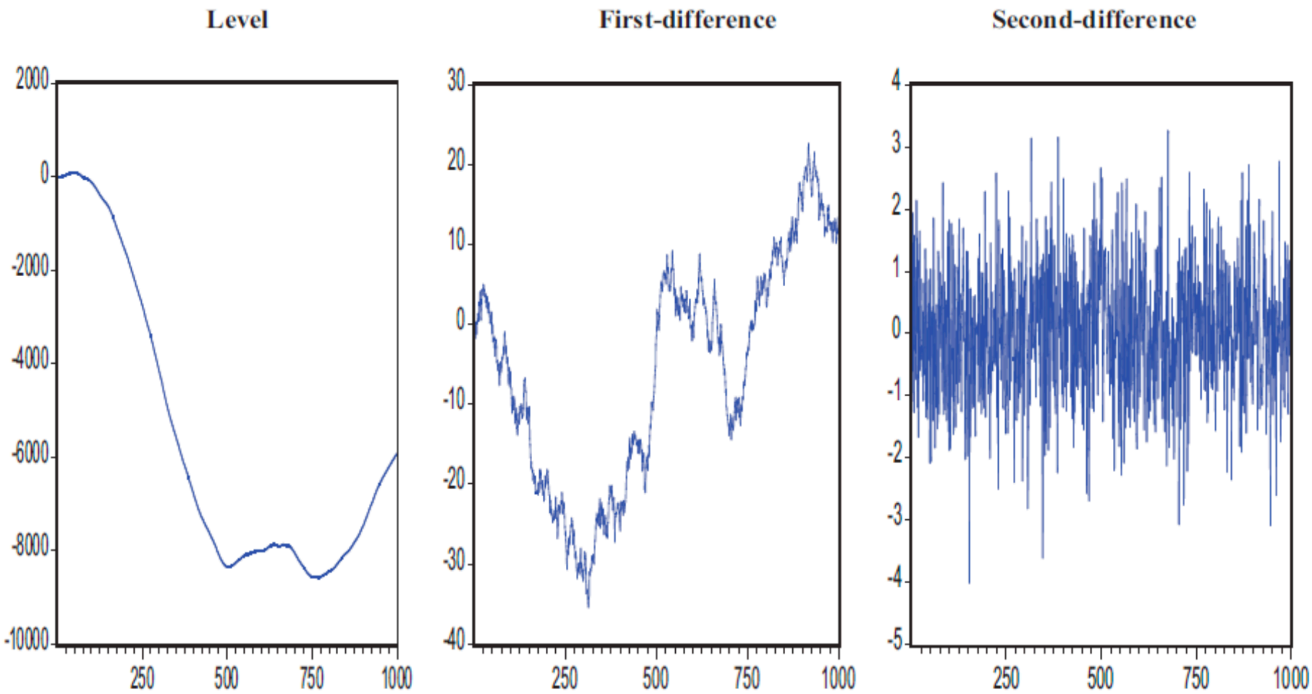
- $d = 2$
- Difference of first differences
- $\nabla^2 Y_t = \nabla(\nabla Y_t) = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})$
- $\nabla^2 Y_t = Y_t - 2Y_{t-1} + Y_{t-2}$

Differencing Example

- Source: https://otexts.com/fpp2/stationarity.html



- The ACF of the Google stock price (left) and of the daily changes in Google stock price (right)
- The ACF of the differenced Google stock price looks just like that of a white noise series
- No autocorrelations outside 95% confidence interval



Random Walk Model

- Differenced series $\nabla Y_t = Y'_t = Y_t - Y_{t-1}$
- First order differenced has only $T - 1$ values (from second observation)
- If differenced series is white noise, model for original series
  - $Y_t - Y_{t-1} = \epsilon_t$
- Rearranging: random walk model
  - $Y_t = Y_{t-1} + \epsilon_t$
- Used for non-stationary data (financial, economic)
- Random walks
  - Long periods of apparent trends
  - Sudden unpredictable changes in direction
- Forecast of RW model: previous observation
- Future movements unpredictable
- Used as a benchmark to compare other models' performance

Random Walk Model with Non-Zero Mean

- Mean c
- $Y_t - Y_{t-1} = c + \epsilon_t$
- $Y_t = Y_{t-1} + c + \epsilon_t$
- Value of c is average of changes between consecutive observations

- If $c > 0$, average change is an increase in the value of Y_t
    - Y_t tends to drift upwards
- Else, drifts downwards

ARIMA Model

- Auto Regressive Integrated Moving Average Model
- $ARIMA(p, d, q)$
- Integrated (I) series: series which needs to be differenced to be made stationary
- Lags of the stationarised series are called AR terms
- Lags of the forecast errors are called MA terms

Step 1: Model Identification

- Identify right values of p, d, q

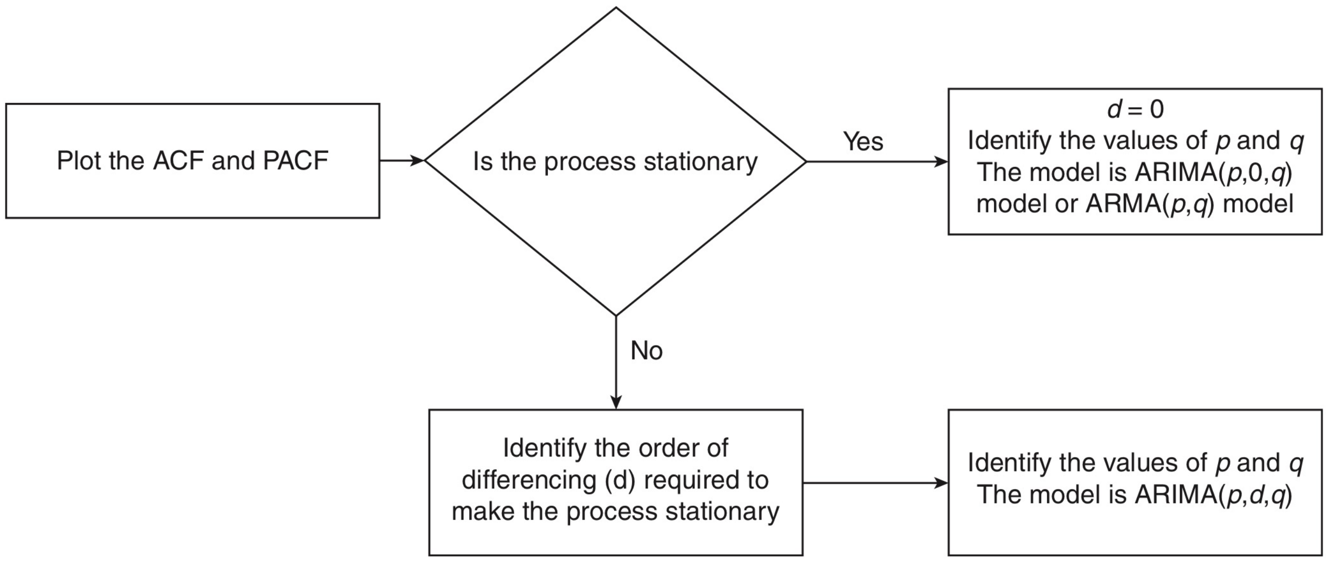


FIGURE 13.14 Model identification in ARIMA model.

Step 2: Parameter Estimation and Model Selection

- Estimation of coefficients in AR and MA using OLS
- Criteria: RMSE, MAPE, Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC)
    - AIC, BIC: measures of distances from actual values to forecasted values
    - $AIC = -2LL + 2K$ where
        - LL is the log likelihood function
        - K is number of parameters (orders) estimated $(p + q)$
    - $BIC = -2LL + K \ln{(n)}$
        - n is number of observations

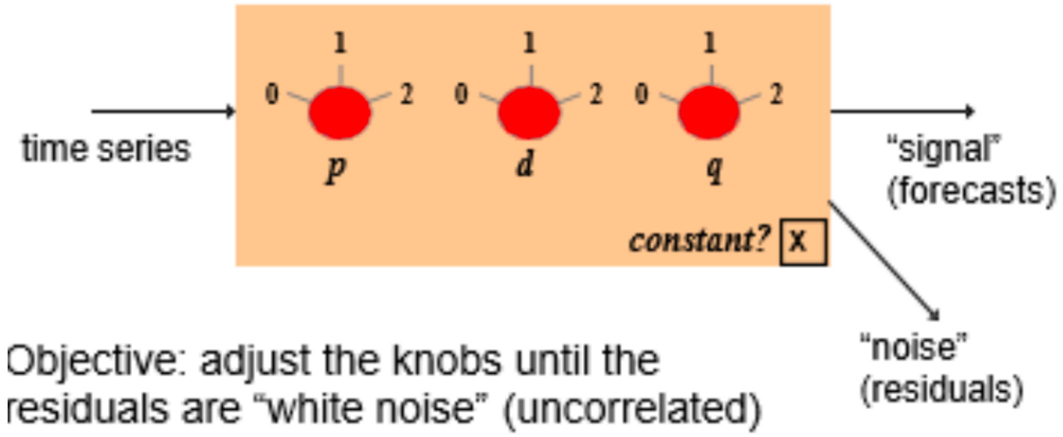
        - Higher penalty than AIC for every additional variable added to model
    - Low values of AIC, BIC preferred

Step 3: Model Validation

- Satisfy all regression conditions
- Goodness of fit test: Ljung-Box test

ARIMA Filtering Box

- Think of p, d and q as knobs
- Adjust knobs until residuals are white noise



ARIMA Models

1. $ARIMA(0,0,0) + c$ = constant model
    - $Y_{t+1} = c + \epsilon_{t+1}$
2. $ARIMA(0,1,0)$ = random walk model
    - $Y_{t+1} - Y_t = \epsilon_{t+1}$
3. $ARIMA(0,1,0) + c$ = random walk with drift
    - $Y_{t+1} - Y_t = c + \epsilon_{t+1}$
4. $ARIMA(1,0,0) + c$ = regress Y on Y_{lag1}
    - $Y_{t+1} = \beta_1 \, Y_t + c + \epsilon_{t+1}$
5. $ARIMA(1,1,0) + c$ = regress Y_{diff1} on Y_{diff1_lag1}
    - $Y_{t+1} - Y_t = \beta_1 \times (Y_t - Y_{t-1}) + c + \epsilon_{t+1}$
6. $ARIMA(2,1,0) + c$ = regress Y_{diff1} on Y_{diff1_lag1} and Y_{diff1_lag2}
    - $\Delta Y_{t+1} = \beta_1 \Delta Y_t + \beta_2 \Delta Y_{t-1} + c + \epsilon_{t+1}$
7. $ARIMA(0,1,1)$ = SES model
    - $Y_{t+1} - Y_t = -\alpha_1 \epsilon_t + \epsilon_{t+1}$
    - $Y_{t+1} - \epsilon_{t+1} = F_{t+1} = Y_t - \alpha_1 (Y_t - F_t)$

- $F_{t+1} = (1 - \alpha_1)Y_t + \alpha_1 F_t$

8. $ARIMA(0,1,1) + c$ = SES model with constant linear trend
9. $ARIMA(1,1,2)$ = LES with damped trend
10. $ARIMA(0,2,2)$ = generalised LES

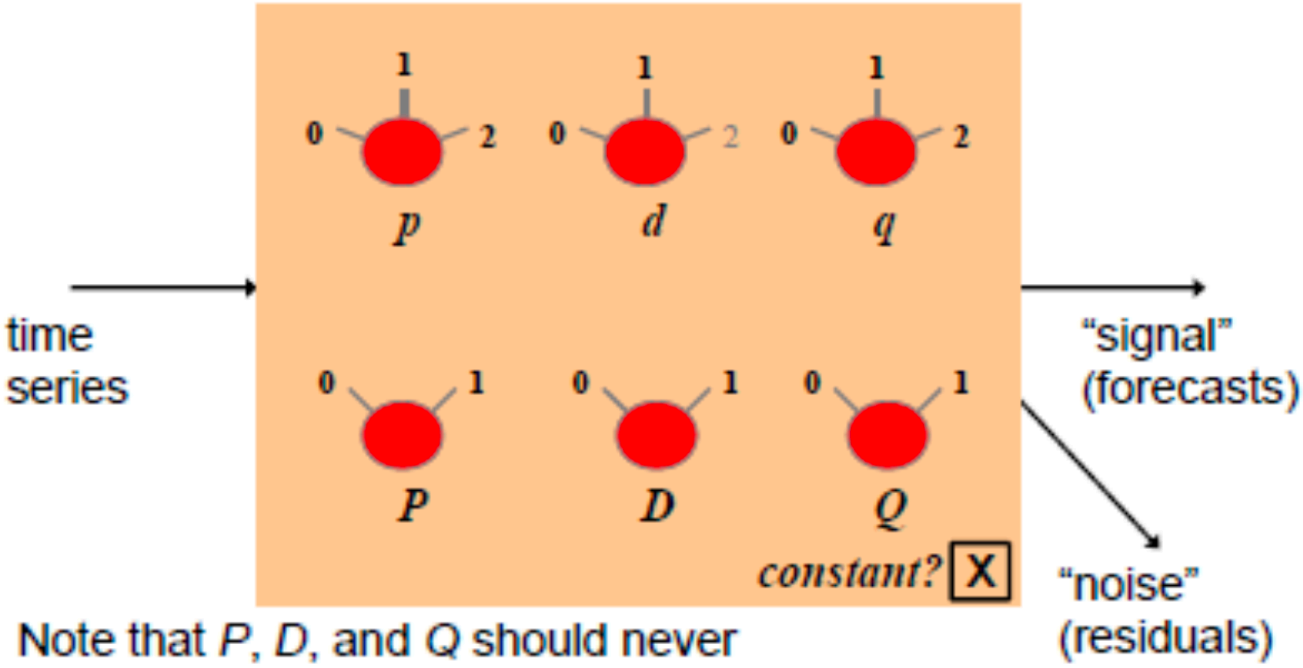
- Usually, $p + q \leq 2$
- If differenced, it must be un-differenced

Rules of Thumb

- If stationarised series has positive autocorrelation at lag 1, AR terms oftern work best
  - Compensate for the lack of nonseasonal difference
- If stationarised series has negative autocorrelation at lag 1, MA terms oftern work best
  - Fine-tune the effect of nonseasonal difference
- Look at 05.1 slides

SARIMA Model

- Seasonal ARIMA
- Seasonality
  - P: number of seasonal autoregressive terms
  - D: number of seasonal differences
  - Q: number of seasonal moving average terms
- Complete model: $SARIMA(p,d,q)(P,D,Q)$
- Filtering box (tune like knobs)



Note that *P*, *D*, and *Q* should never be larger than 1 !!

Seasonal Differences

- Combine non-seasonal and seasonal differences

If d=0, D=1: $y_t = Y_t - Y_{t-s}$

s is the seasonal period, e.g., *s*=12 for monthly data

If d=1, D=1: $y_t = (Y_t - Y_{t-1}) - (Y_{t-s} - Y_{t-s-1})$

$$= Y_t - Y_{t-1} - Y_{t-s} + Y_{t-s-1}$$

D should never be more than 1, and *d*+*D* should never be more than 2. Also, if *d*+*D* =2, the constant term should be suppressed.

SAR and SMA terms

- Setting $P = 1$ (SAR) adds multiple of y_{t-s} to the forecast for y_t
- Setting $Q = 1$ (SMA) adds a multiple of ϵ_{t-s} to the forecast for y_t
- SAR + SMA should never exceed 1

Ljung-Box Test for Autocorrelations

- Checks if auto-correlations are non-zero
- Null and alt hypotheses
    - H_0: model does not show lack of fit (model is a good fit)
    - H_a: model shows lack of fit
- Test statistic: Q-statistic
    - $Q(m) = n(n+2) \sum_{k=1}^{m} \frac{\rho_k^2}{n-k}$
    - m is total number of lags
    - n is number of observations
    - k is number of lags
    - ρ_k is the autocorrelation of lag k
- Q-statistic is chi-square distribution with m-p-q degrees of freedom

- The *Q*-statistic for ARIMA(1, 1, 1) is 10.216 (Table 1) and the corresponding *p*-value is 0.855 and thus we fail to reject the null hypothesis.
- Table 1: ARIMA (1, 1, 1) model summary for Omelette demand

Model	Model Fit Statistics			Ljung–Box *Q*(18)		
	R-Squared	RMSE	MAPE	Statistics	*Df*	Sig.
Omellette-Model_1	0.584	3.439	20.830	10.216	16	0.855

- *Q*(*m*) measures accumulated auto-correlation up to lag *m*.

Thiel's Coefficient

- Comparision of forecasting model to naive forecast
- $F_{t+1} = Y_t$

Day	Y_t	ARMA (1,2) Forecast	$(Y_t - F_t)^2$	Naïve Forecast ($F_{t+1} = Y_t$)	$(Y_t - F_t)^2$
31	503	464.8107	1458.423	443	3600
32	688	378.5341	95769.15	503	34225
33	602	444.6372	24763.04	688	7396
34	629	685.8851	3235.909	602	729
35	823	743.5124	6318.281	629	37636
36	671	630.7183	1622.614	823	23104
37	487	649.3491	26357.22	671	33856
		Total	159524.6	Total	140546

- U-statistic
  - $U = \frac{\sum_{t=1}^{n}(Y_{t+1} - F_{t+1})^2}{\sum_{t=1}^{n}(Y_{t+1} - Y_t)^2}$
  - Ratio of SSE of forecasting to SSE of naive model
  - If U < 1, forecasting doing better than naive
  - If U > 1, forecasting doing worse than naive

The X Factor (ARX, ARIMAX)

- X: exogenous variables
- Other factors that influence the forecast (domain knowledge)
- How to integrate it into the model?

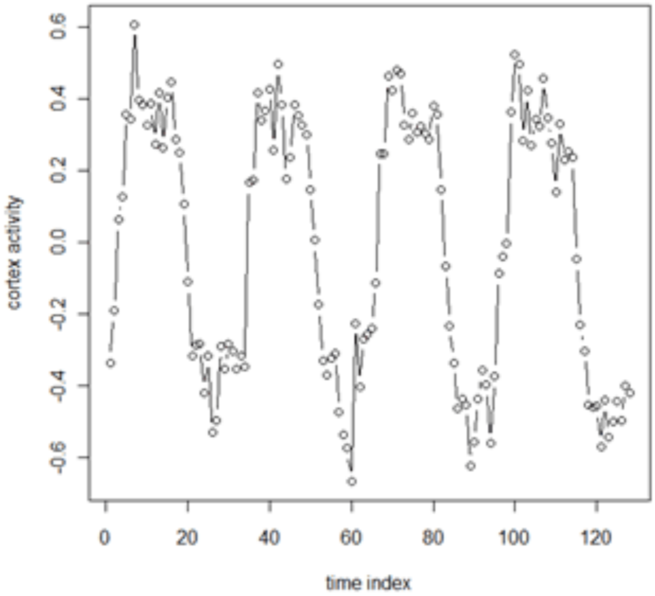
Spectral Analysis of TS Data

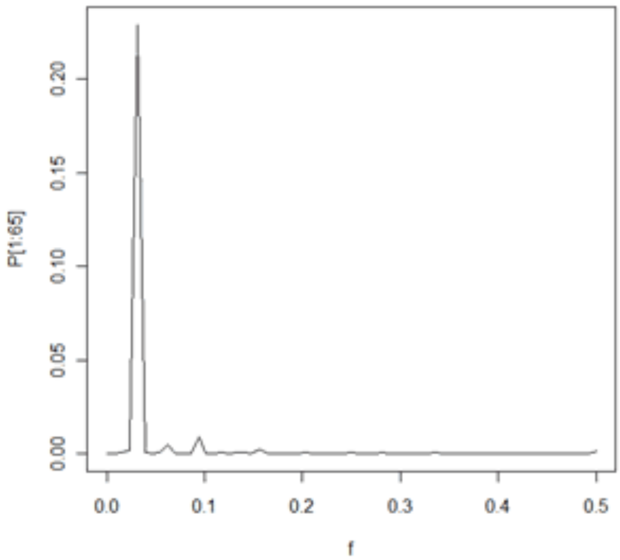
Discrete Fourier Transform of the Time Series

- Function of time to function of frequencies
- Discrete FT of time series $x_1, \ \ldots \ , x_n$
  - $d(\omega_j) = \sqrt{n}\sum_{t=1}^{n} x_t e^{j \times -2\pi t\omega_j}$
  - $d(\omega_j) = \sqrt{n}\sum_{t=1}^{n} x_t \cos{(j\ 2\pi\omega_j t)} - j\sqrt{n}\sum_{t=1}^{n} x_t \sin{(j\ 2\pi\omega_j t)}$
- Periodogram: $I(\omega_j)$
  - $|d(\omega_j)|^2 = d_c^2(\omega_j) + d_s^2(\omega_j)$
  - Cosine and sine components, re and im components
  - If no periodic trend in data, $E[d(\omega_j)] = 0$ and periodogram expresses variance of x_t at frequency ω_j

- If periodic trend exists, $E[d(\omega_j)]$ is the contribution to the preiodic trend at the frequency ω_j

- Eg: The series is n = 128 values of brain cortex activity, measured every 2 seconds for 256 seconds. A stimulus, brushing of the back of the hand, was applied for 32 seconds and then was stopped for 32 seconds. This pattern was repeated for a total of 256 seconds. The series is actually the average of this process for five different subjects.





Wavelet Transformation

- Finite portion of a signal
- Similar to FT
- Read slides for more

More

- DL
- Classifier chains
- MDP, RL
- Read slides