Data Analysis Course
Time Series Analysis & Forecasting

Venkat Reddy

Contents

Drawbacks of the use of traditional models

- There is no systematic approach for the identification and selection of an appropriate model, and therefore, the identification process is mainly trial-and-error
- There is difficulty in verifying the validity of the model
  - Most traditional methods were developed from intuitive and practical considerations rather than from a statistical foundation

ARIMA

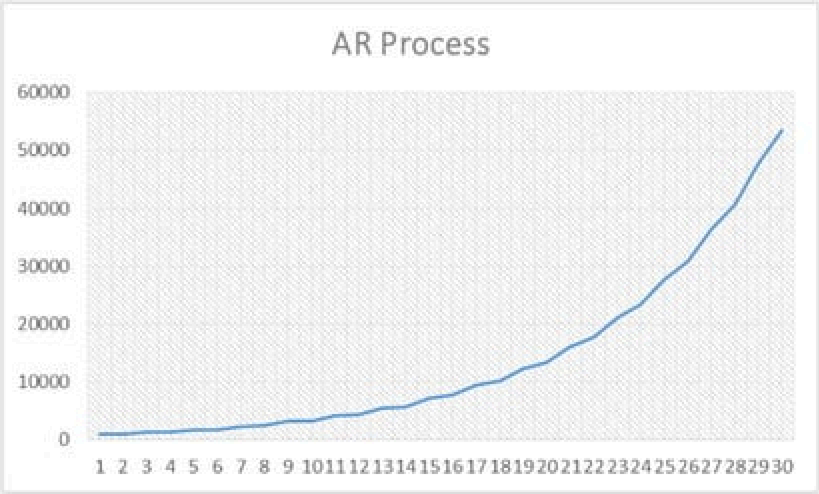
ARIMA Models

- Autoregressive Integrated Moving-average
- A "stochastic" modeling approach that can be used to calculate the probability of a future value lying between two specified limits

AR & MA 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

AR Process → p



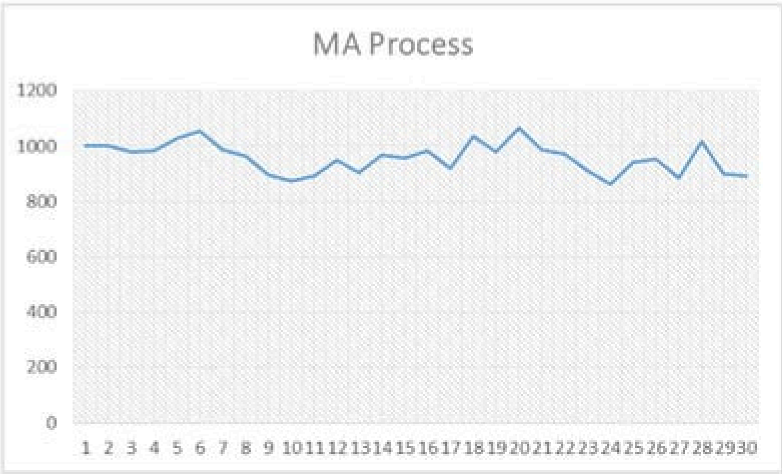
AR(1) $y_t = a1 * y_{t-1}$

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

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

Data Analysis Course
Venkat Reddy

MA Process $\rightarrow q$



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

ARIMA Models

- Autoregressive (AR) process:
  - Series current values depend on its own previous values
- Moving average (MA) process:
  - The current deviation from mean depends on previous deviations
- Autoregressive Moving average (ARMA) process
- Autoregressive Integrated Moving average (ARIMA)process.

- ARIMA is also known as Box-Jenkins approach. It is popular because of its generality;
- It can handle any series, with or without seasonal elements, and it has well-documented computer programs

ARIMA Model

Y_t → AR filter → Integration filter → MA filter → ε_t

(long term) (stochastic trend) (short term) (white noise error)

ARIMA (2,0,1) $y_t = a_1 y_{t-1} + a_2 y_{t-2} + b_1 \varepsilon_{t-1}$

ARIMA (3,0,1) $y_t = a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + b_1 \varepsilon_{t-1}$

ARIMA (1,1,0) $\Delta y_t = a_1 \Delta y_{t-1} + \varepsilon_t$, where $\Delta y_t = y_t - y_{t-1}$

ARIMA (2,1,0) $\Delta yt = a_1 \Delta y_{t-1} + a_2 \Delta y_{t-2} + \varepsilon_t$ where $\Delta yt = yt - yt\text{-}1$

To build a time series model issuing ARIMA, we need to study the time series and identify p,d,q

ARIMA equations

- ARIMA(1,0,0)
  - $y_t = a_1 y_{t-1} + \varepsilon_t$
- ARIMA(2,0,0)
  - $y_t = a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t$
- ARIMA (2,1,1)
  - $\Delta yt = a_1 \Delta y_{t-1} + a_2 \Delta y_{t-2} + b_1 \varepsilon_{t-1}$ where $\Delta yt = yt - yt-1$

Overall Time series Analysis & Forecasting Process

- Prepare the data for model building- Make it stationary
- Identify the model type
- Estimate the parameters
- Forecast the future values

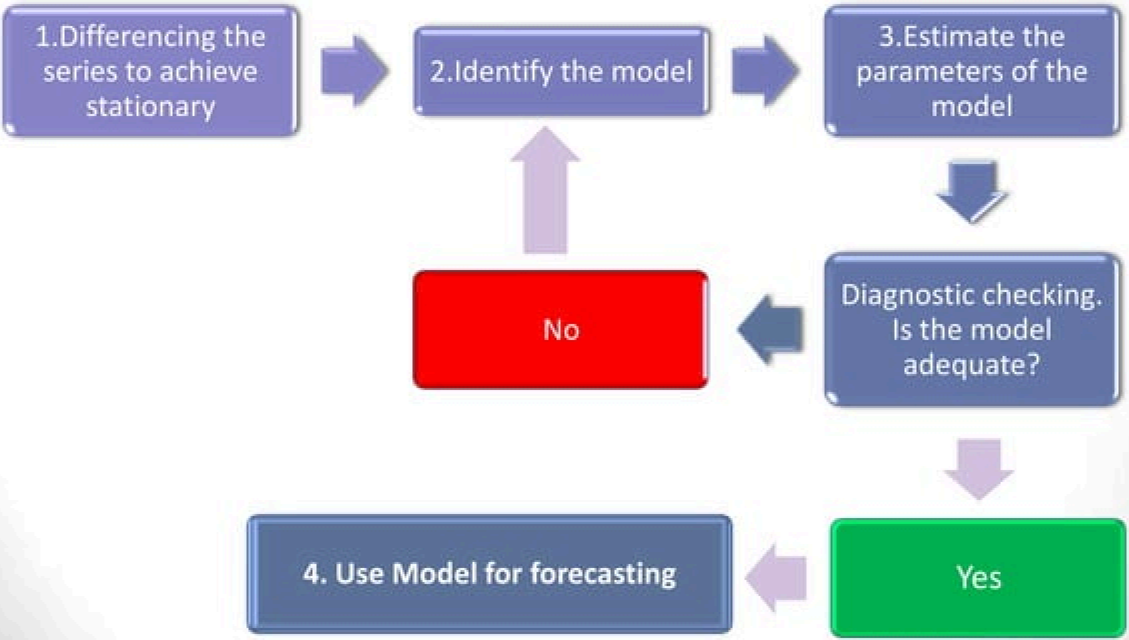
ARIMA (p,d,q) modeling

To build a time series model issuing ARIMA, we need to study the time series and identify p,d,q

- **Ensuring Stationarity**
  - Determine the appropriate values of d
- **Identification:**
  - Determine the appropriate values of p & q using the ACF, PACF, and unit root tests
  - p is the AR order, d is the integration order, q is the MA order
- **Estimation :**
  - Estimate an ARIMA model using values of p, d, & q you think are appropriate.
- **Diagnostic checking:**
  - Check residuals of estimated ARIMA model(s) to see if they are white noise; pick best model with well behaved residuals.
- **Forecasting:**
  - Produce out of sample forecasts or set aside last few data points for in-sample forecasting.

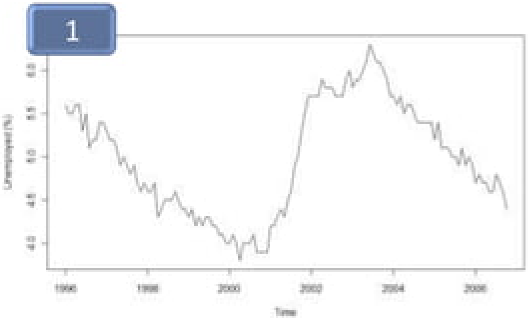
Data Analysis Course
Venkat Reddy

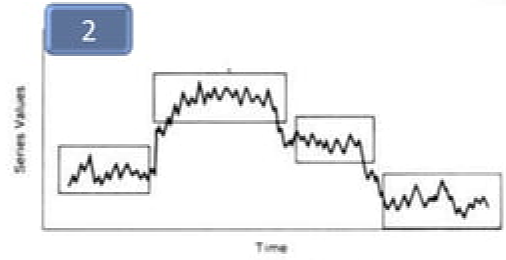
The Box-Jenkins Approach



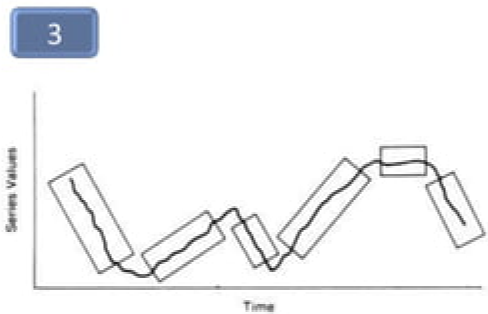
Step-1 : Stationarity

Some non stationary series

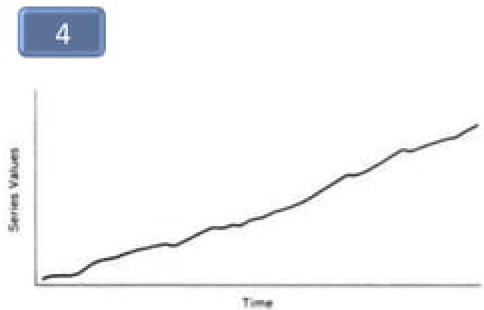




A Nonstationary Series: Random Changes in Level



A Nonstationary Series: Random Changes in Both Level and Slope



A Nonstationary Series: Overall Trend

Stationarity

- In order to model a time series with the Box-Jenkins approach, the series has to be stationary
- In practical terms, the series is stationary if tends to wonder more or less uniformly about some fixed level
- In statistical terms, a stationary process is assumed to be in a particular state of statistical equilibrium, i.e., $p(x_t)$ is the same for all t
- In particular, if z_t is a stationary process, then the first difference $\nabla z_t = z_t - z_{t-1}$ and higher differences $\nabla^d z_t$ are stationary

Testing Stationarity

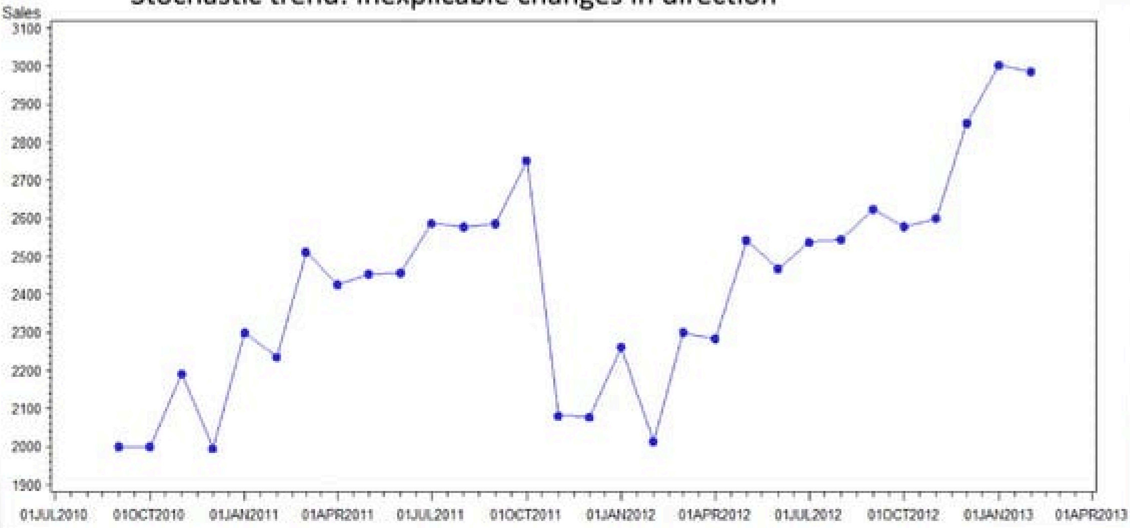
- Dickey-Fuller test
  - P value has to be less than 0.05 or 5%
  - If p value is greater than 0.05 or 5%, you accept the null hypothesis, you conclude that the time series has a unit root.
  - In that case, you should first difference the series before proceeding with analysis.

- What DF test ?
  - Imagine a series where a fraction of the current value is depending on a fraction of previous value of the series.
  - DF builds a regression line between fraction of the current value Δy_t and fraction of previous value δy_{t-1}
  - The usual t-statistic is not valid, thus D-F developed appropriate critical values. **If P value of DF test is <5% then the series is stationary**

Demo: Testing Stationarity

- Sales_1 data

Stochastic trend: Inexplicable changes in direction



Data Analysis Course
Venkat Reddy

Demo: Testing Stationarity

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	0.3251	0.7547	0.74	0.8695		
	1	0.3768	0.7678	1.26	0.9435		
	2	0.3262	0.7539	1.05	0.9180		
Single Mean	0	-6.9175	0.2432	-1.77	0.3858	2.05	0.5618
	1	-3.5970	0.5662	-1.06	0.7163	1.52	0.6913
	2	-3.7030	0.5522	-0.88	0.7783	1.02	0.8116
Trend	0	-11.8936	0.2428	-2.50	0.3250	3.16	0.5624
	1	-7.1620	0.6017	-1.60	0.7658	1.34	0.9063
	2	-9.0903	0.4290	-1.53	0.7920	1.35	0.9041

Achieving Stationarity

- Differencing : Transformation of the series to a new time series where the values are the differences between consecutive values
- Procedure may be applied consecutively more than once, giving rise to the "first differences", "second differences", etc.
- Regular differencing (RD)

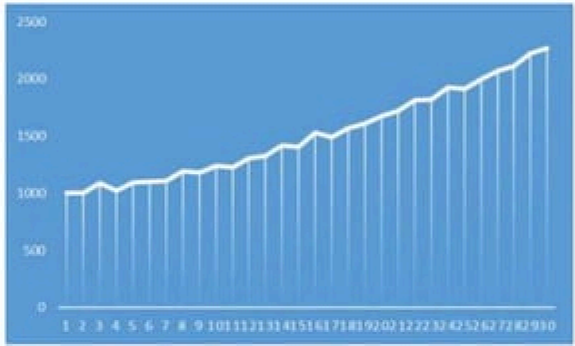
(1st order) $\nabla x_t = x_t - x_{t-1}$

(2nd order) $\nabla^2 x_t = (\nabla x_t - \nabla x_{t-1}) = x_t - 2x_{t-1} + x_{t-2}$

- It is unlikely that more than two regular differencing would ever be needed
- Sometimes regular differencing by itself **is not** sufficient and prior transformation is also needed

Differentiation

Actual Series



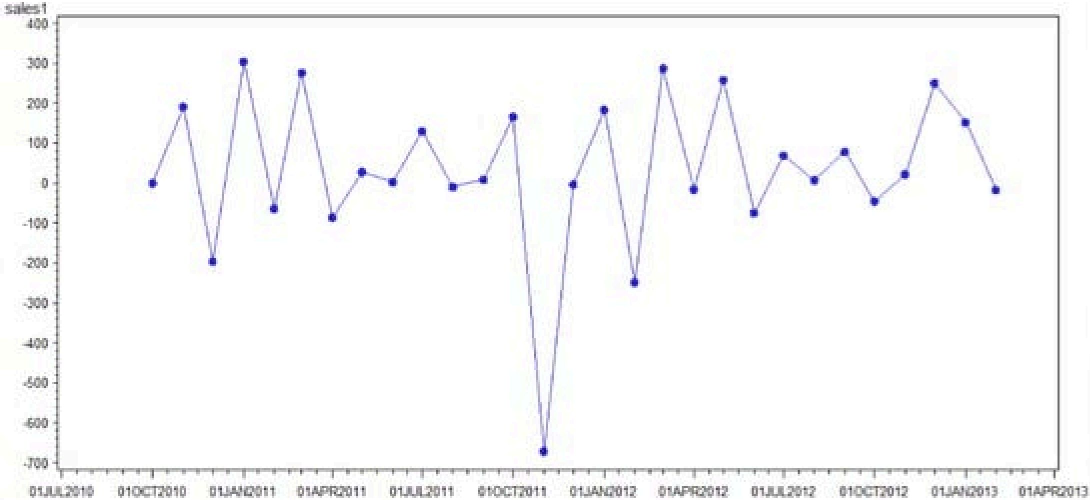
Series After Differentiation

Demo: Achieving Stationarity

```
data lagsales_1;
set sales_1;
sales1=sales-lag1(sales);
run;
```

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-37.7155	<.0001	-7.46	<.0001		
	1	-32.4406	<.0001	-3.93	0.0003		
	2	-19.3900	0.0006	-2.38	0.0191		
Single Mean	0	-38.9718	<.0001	-7.71	0.0002	29.70	0.0010
	1	-37.3049	<.0001	-4.10	0.0036	8.43	0.0010
	2	-25.6253	0.0002	-2.63	0.0992	3.50	0.2081
Trend	0	-39.0703	<.0001	-7.58	0.0001	28.72	0.0010
	1	-37.9046	<.0001	-4.08	0.0180	8.35	0.0163
	2	-25.7179	0.0023	-2.59	0.2875	3.37	0.5234

Demo: Achieving Stationarity



Achieving Stationarity-Other methods

- Is the trend stochastic or deterministic?
  - If stochastic (inexplicable changes in direction): use differencing
  - If deterministic(plausible physical explanation for a trend or seasonal cycle) : use regression
- Check if there is variance that changes with time
  - YES : make variance constant with log or square root transformation
- Remove the trend in mean with:
  - 1st/2nd order differencing
  - Smoothing and differencing (seasonality)
- If there is seasonality in the data:
  - Moving average and differencing
  - Smoothing

Step2 : Identification

Identification of orders p and q

- Identification starts with d
- ARIMA(p,d,q)
- What is Integration here?
- First we need to make the time series stationary
- We need to learn about ACF & PACF to identify p,q

- Once we are working with a stationary time series, we can examine the ACF and PACF to help identify the proper number of lagged y (AR) terms and ε (MA) terms.

Autocorrelation Function (ACF)

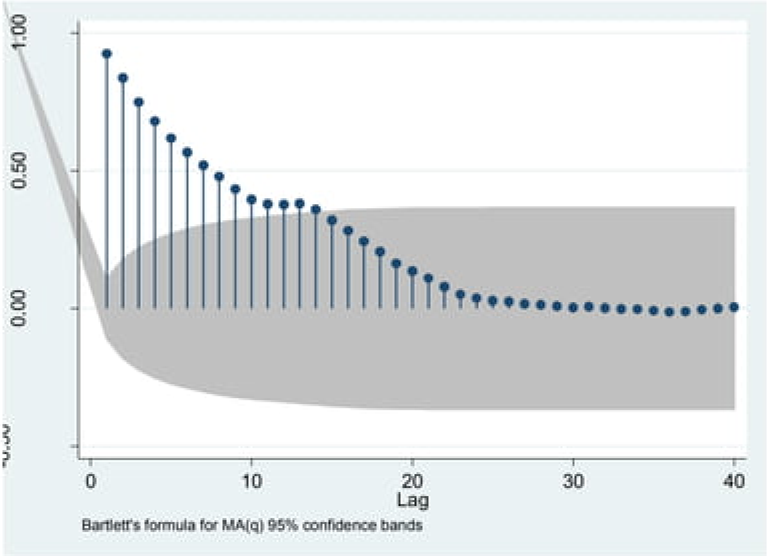
- Autocorrelation is a correlation coefficient. However, instead of correlation between two different variables, the correlation is between two values of the same variable at times $\boldsymbol{X_i}$ and $\boldsymbol{X_{i+k}}$.
- Correlation with lag-1, lag2, lag3 etc.,
- The ACF represents the degree of persistence over respective lags of a variable.

$\rho_k = \gamma_k / \gamma_0$ = covariance at lag k/ variance

$$\rho_k = \frac{E[(y_t - \mu)(y_{t-k} - \mu)]^2}{E[(y_t - \mu)^2]}$$

ACF (0) = 1, ACF (k) = ACF (-k)

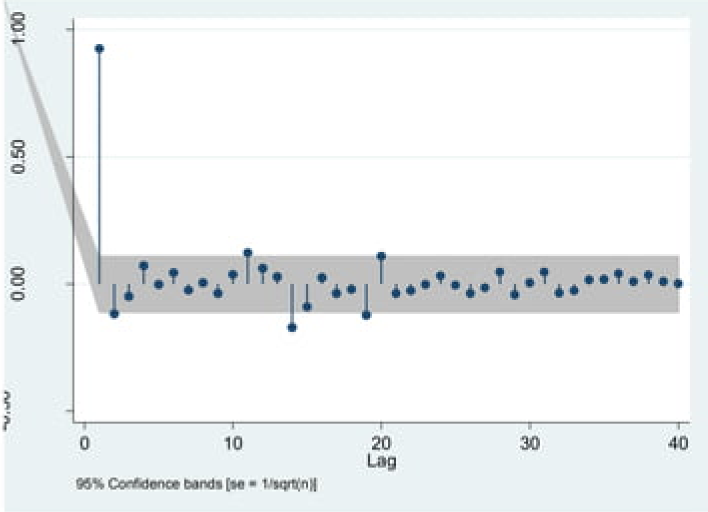
ACF Graph


1.00
0.50
0.00
0
10
20
30
40
Lag
Bartlett's formula for MA(q) 95% confidence bands

Partial Autocorrelation Function (PACF)

- The exclusive correlation coefficient
- Partial regression coefficient - The lag k partial autocorrelation is the partial regression coefficient, θ_{kk} in the k^{th} order auto regression
- In general, the "partial" correlation between two variables is the amount of correlation between them which is not explained by their mutual correlations with a specified set of other variables.
- For example, if we are regressing a variable Y on other variables X1, X2, and X3, the partial correlation between Y and X3 is the amount of correlation between Y and X3 that is not explained by their common correlations with X1 and X2.
- $y_t = \theta_{k1} y_{t-1} + \theta_{k2} y_{t-2} + \ldots + \theta_{kk} y_{t-k} + \varepsilon_t$
- **Partial correlation** measures the degree of association between two random variables, with the effect of a set of controlling random variables removed.

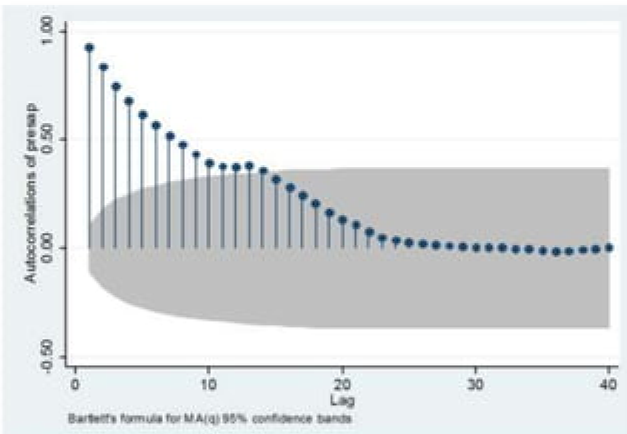
PACF Graph



Data Analysis Course
Venkat Reddy

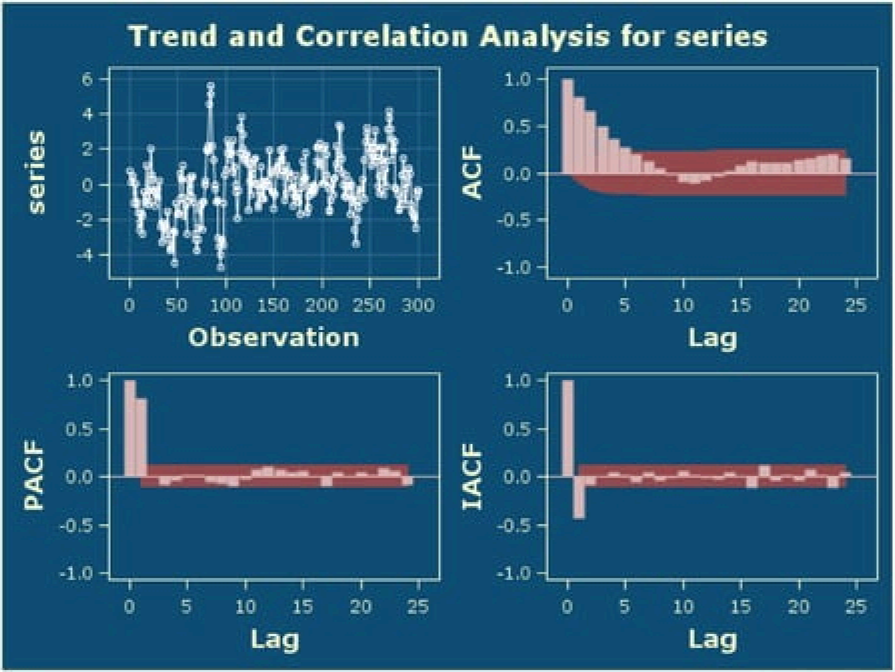
Identification of AR Processes & its order -p

- For AR models, the ACF will dampen exponentially
- The PACF will identify the order of the AR model:
  - The AR(1) model ($y_t = a_1 y_{t-1} + \varepsilon_t$) would have one significant spike at lag 1 on the PACF.
  - The AR(3) model ($y_t = a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + \varepsilon_t$) would have significant spikes on the PACF at lags 1, 2, & 3.



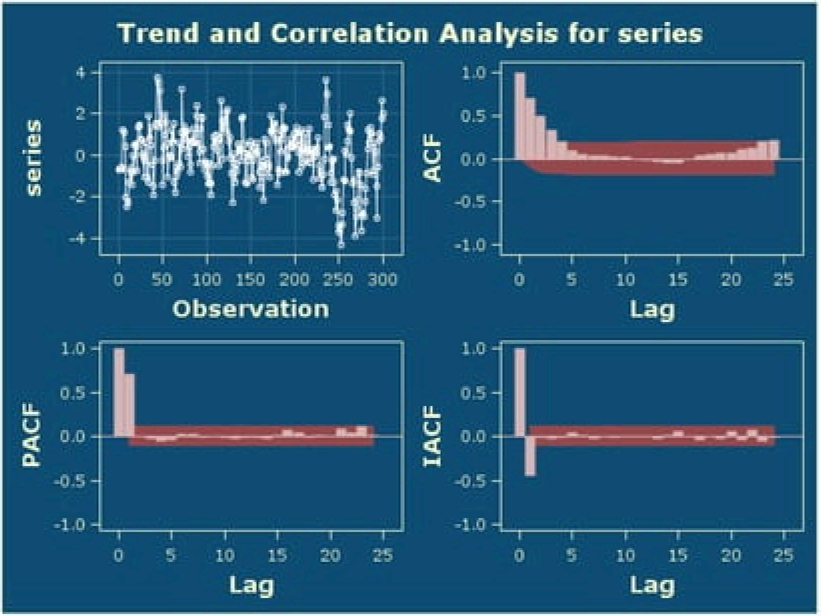
AR(1) model

$$y_t = 0.8y_{t-1} + \varepsilon_t$$



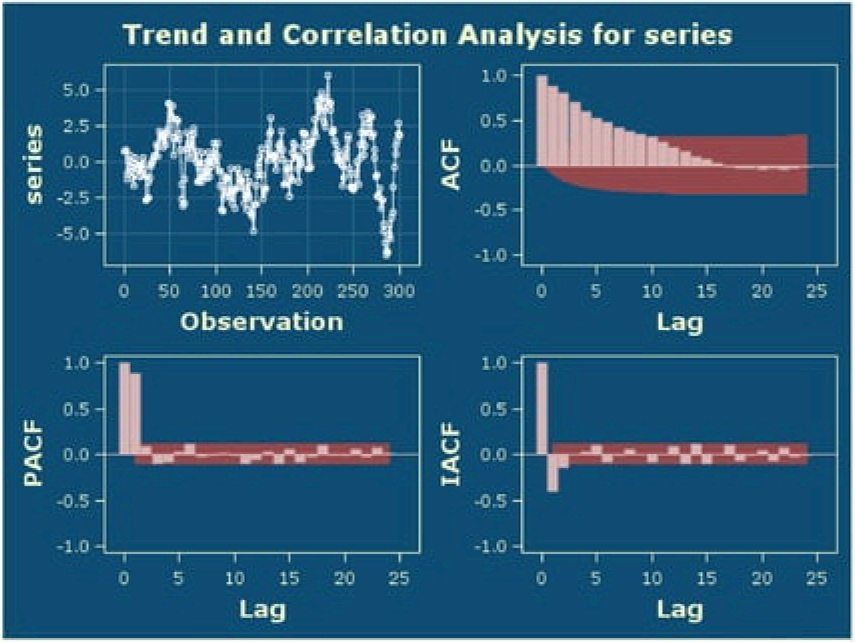
AR(1) model

$$y_t = 0.77y_{t-1} + \varepsilon_t$$



AR(1) model

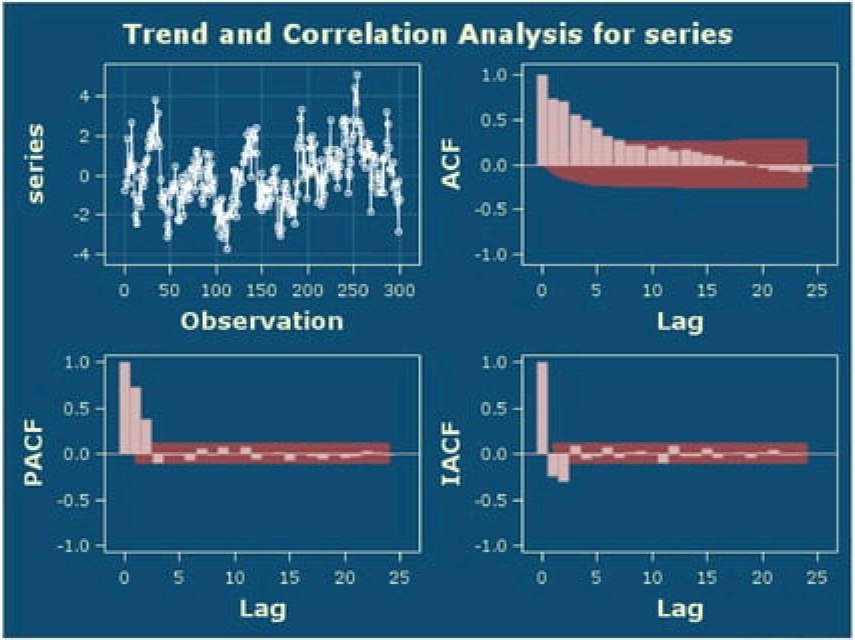
$$y_t = 0.95y_{t-1} + \varepsilon_t$$



Data Analysis Course
Venkat Reddy

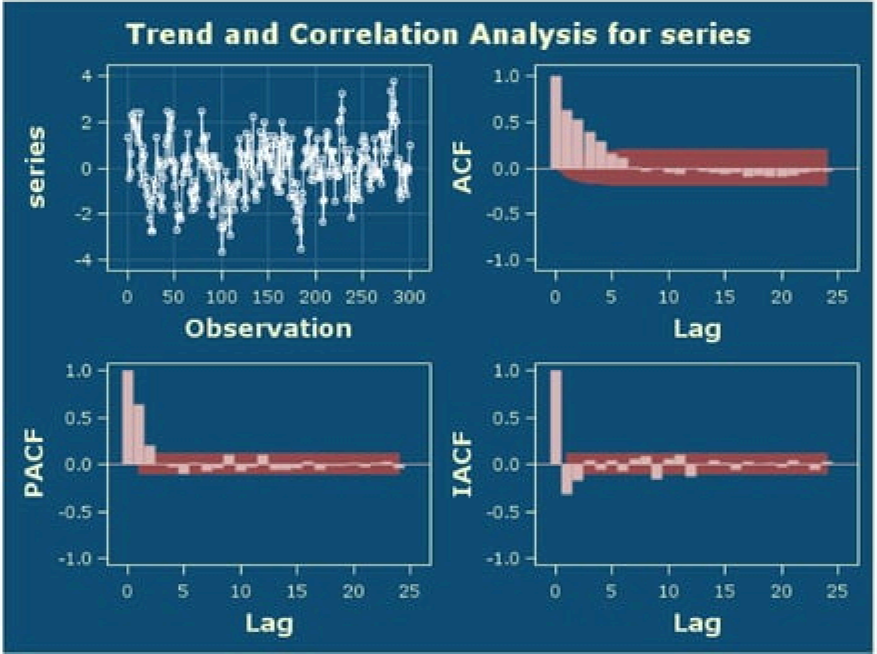
AR(2) model

$$y_t = 0.44y_{t-1} + 0.4y_{t-2} + \varepsilon_t$$



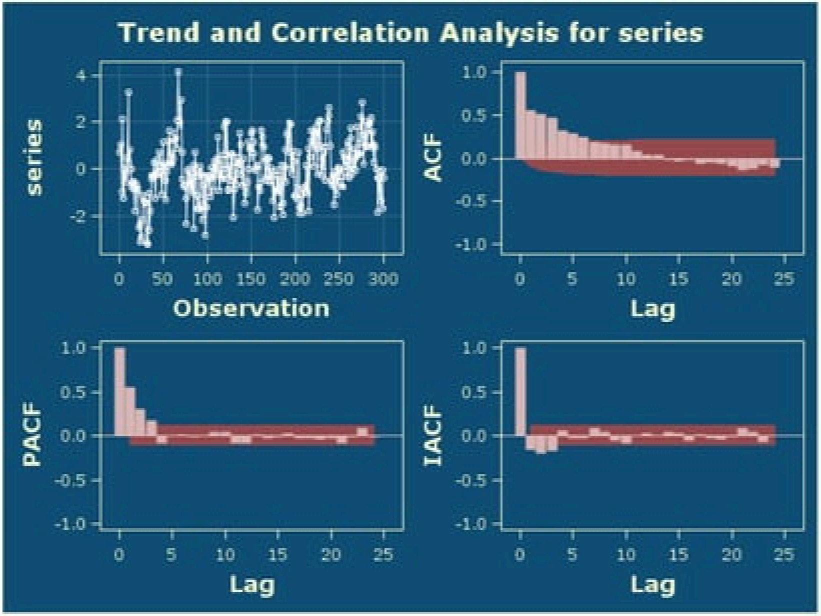
AR(2) model

$$y_t = 0.5y_{t-1} + 0.2y_{t-2} + \varepsilon_t$$



AR(3) model

$$y_t = 0.3y_{t-1} + 0.3y_{t-2} + 0.1y_{t-3} + \varepsilon_t$$



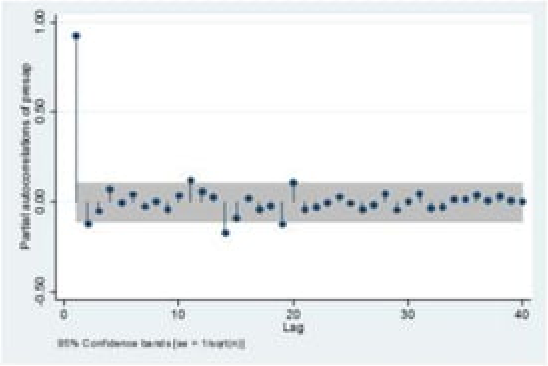
Once again

Properties of the ACF and PACF of MA, AR and ARMA Series

Process	MA(q)	AR(p)	ARMA(p,q)
Auto-correlation function	Cuts off	Infinite. Tails off. Damped Exponentials and/or Cosine waves	Infinite. Tails off. Damped Exponentials and/or Cosine waves after q-p.
Partial Autocorrelation function	Infinite. Tails off. Dominated by damped Exponentials & Cosine waves.	Cuts off	Infinite. Tails off. Dominated by damped Exponentials & Cosine waves after p-q.

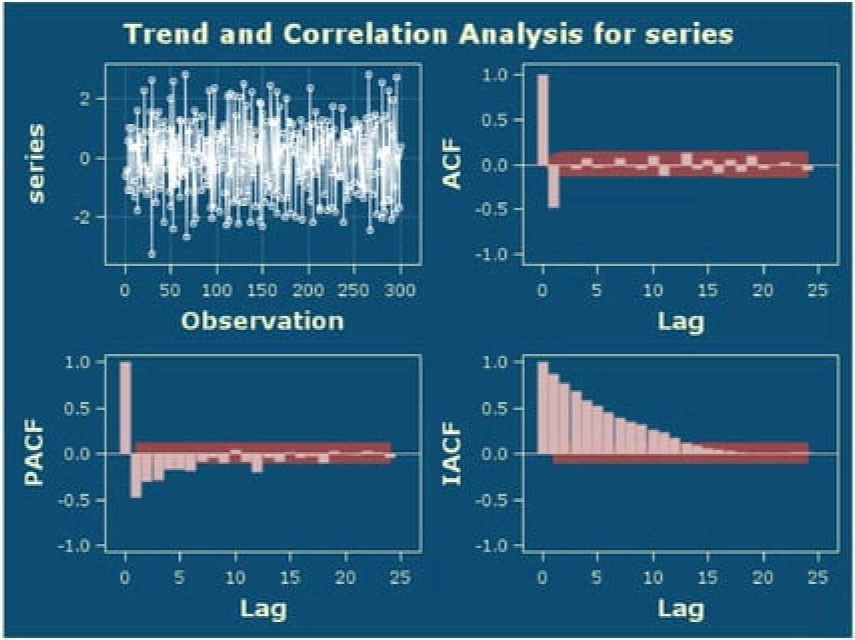
Identification of MA Processes & its order - q

- Recall that a MA(q) can be represented as an AR(∞), thus we expect the opposite patterns for MA processes.
- The PACF will dampen exponentially.
- The ACF will be used to identify the order of the MA process.
- MA(1) ($y_t = \varepsilon_t + b_1 \varepsilon_{t-1}$) has one significant spike in the ACF at lag 1.
- MA (3) ($y_t = \varepsilon_t + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2} + b_3 \varepsilon_{t-3}$) has three significant spikes in the ACF at lags 1, 2, & 3.



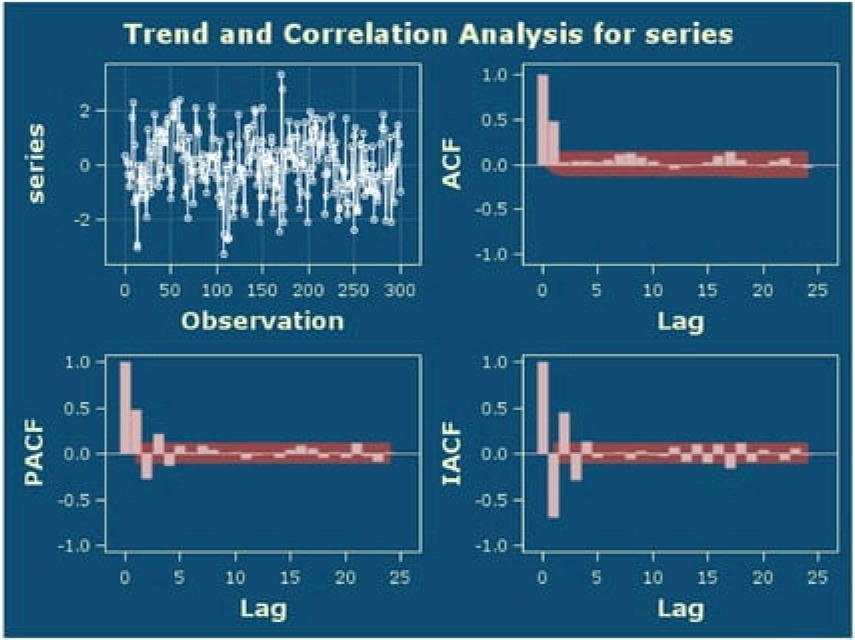
MA(1)

$y_t = -0.9\varepsilon_{t-1}$



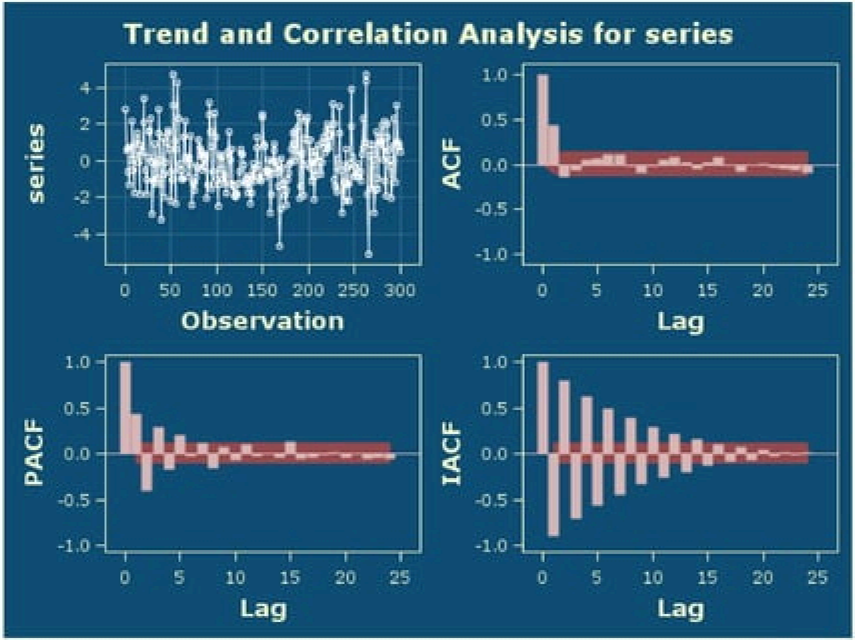
MA(1)

$y_t = 0.7\varepsilon_{t-1}$



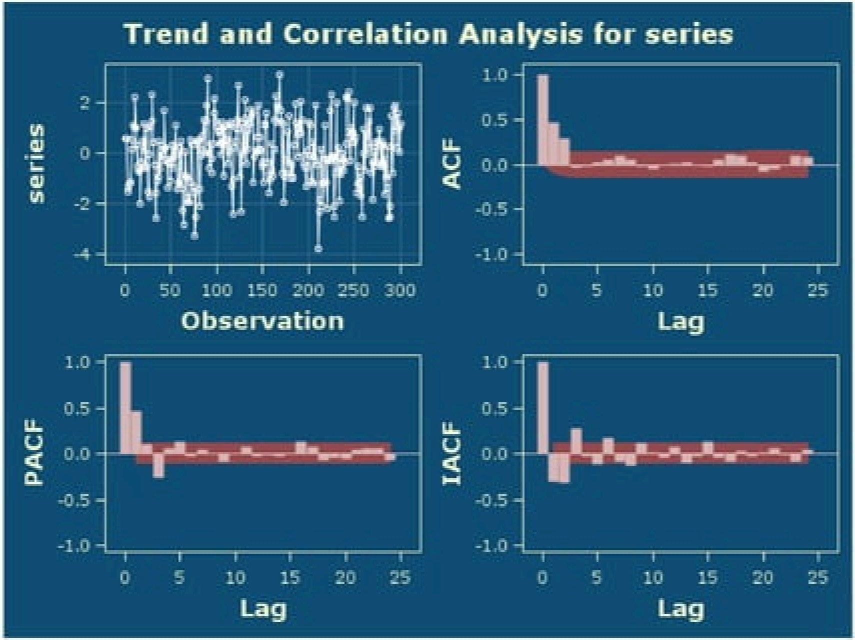
MA(1)

$y_t = 0.99\varepsilon_{t-1}$



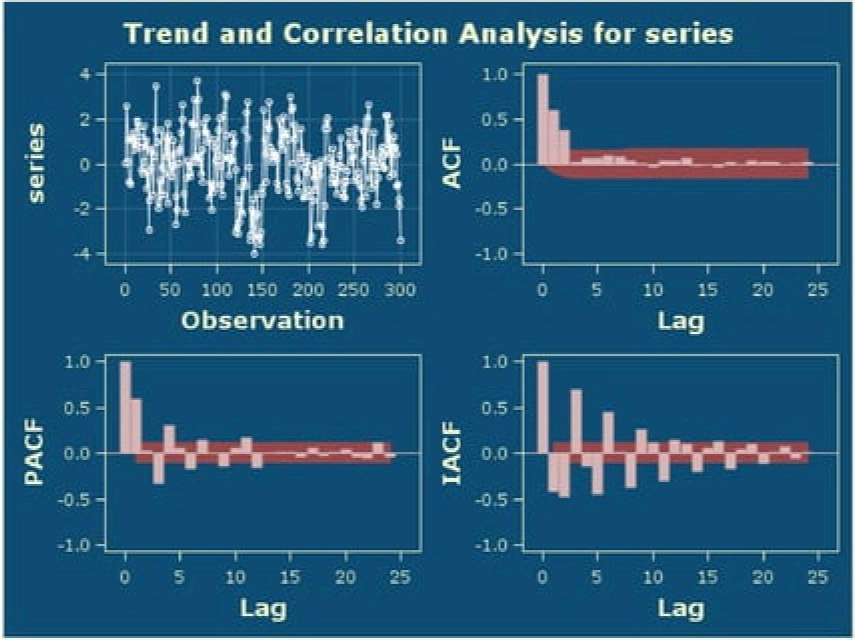
MA(2)

$$y_t = 0.5\varepsilon_{t-1} + 0.5\varepsilon_{t-2}$$



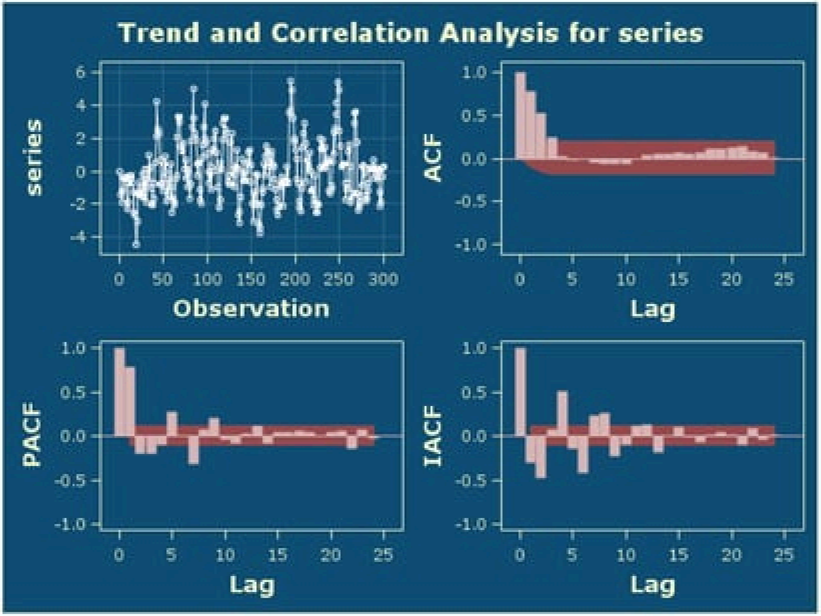
MA(2)

$y_t = 0.8\varepsilon_{t-1} + 0.9\varepsilon_{t-2}$



MA(3)

$y_t = 0.8\varepsilon_{t-1} + 0.9\varepsilon_{t-2} + 0.6\varepsilon_{t-3}$



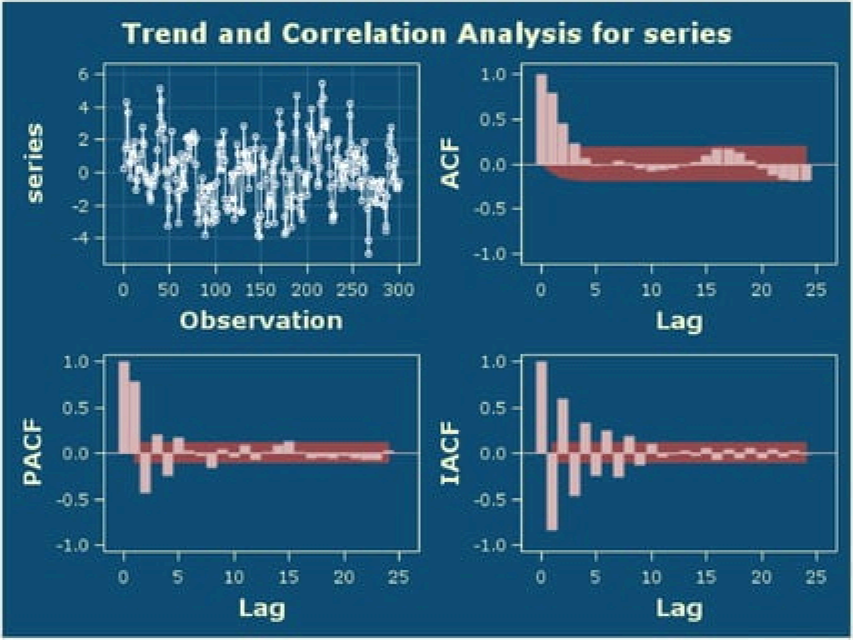
Once again

Properties of the ACF and PACF of MA, AR and ARMA Series

Process	MA(q)	AR(p)	ARMA(p,q)
Auto-correlation function	Cuts off	Infinite. Tails off. Damped Exponentials and/or Cosine waves	Infinite. Tails off. Damped Exponentials and/or Cosine waves after q-p.
Partial Autocorrelation function	Infinite. Tails off. Dominated by damped Exponentials & Cosine waves.	Cuts off	Infinite. Tails off. Dominated by damped Exponentials & Cosine waves after p-q.

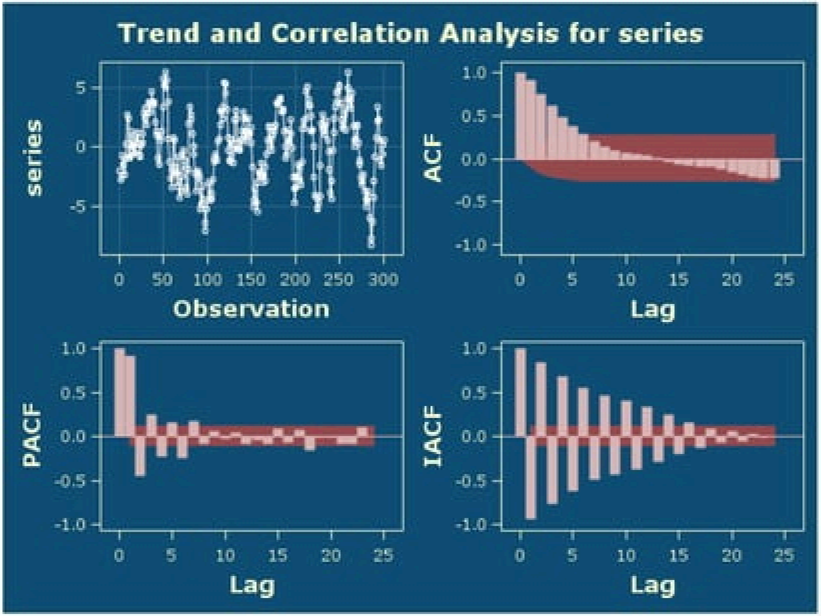
ARMA(1,1)

$$y_t = 0.6y_{t-1} + 0.8\varepsilon_{t-1}$$



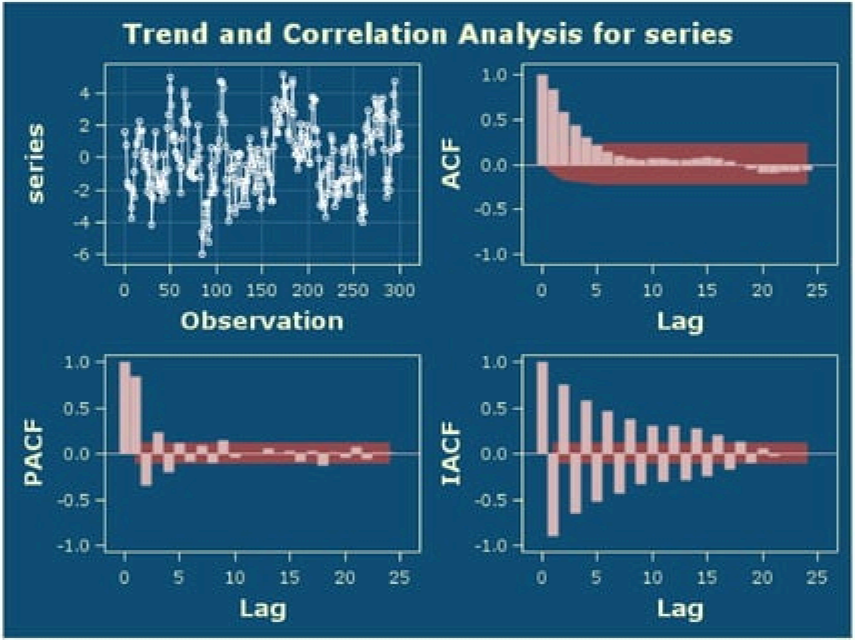
ARMA(1,1)

$$y_t = 0.78y_{t-1} + 0.9\varepsilon_{t-1}$$



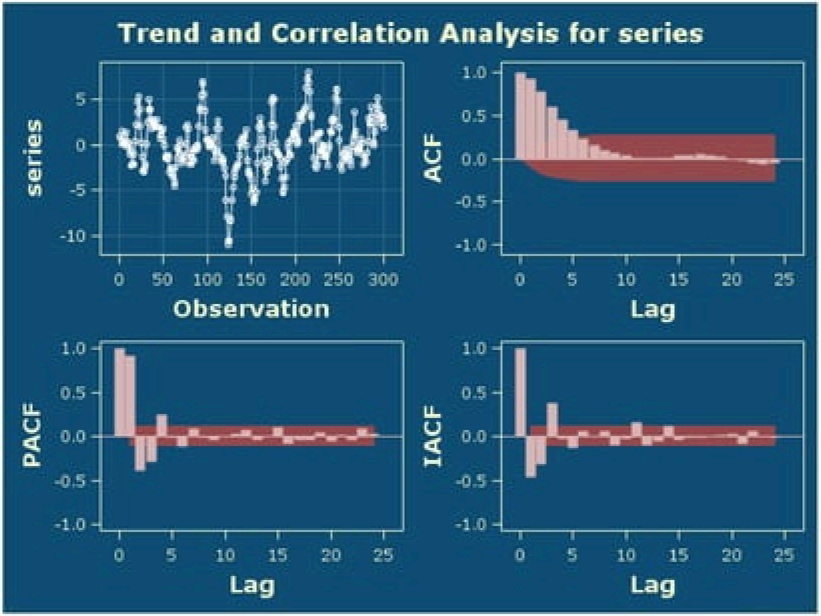
ARIMA(2,1)

$$y_t = 0.4y_{t-1} + 0.3y_{t-2} + 0.9\varepsilon_{t-1}$$



ARMA(1,2)

$$y_t = 0.8y_{t-1} + 0.4\varepsilon_{t-1} + 0.55\varepsilon_{t-2}$$



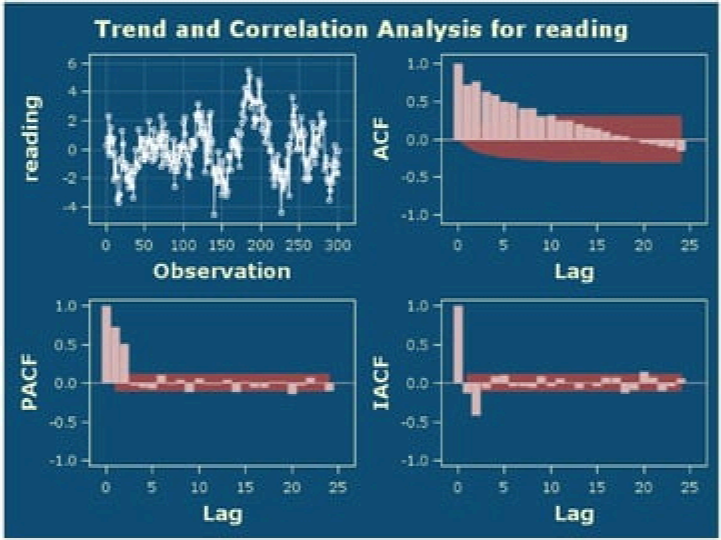
ARMA Model Identification

Properties of the ACF and PACF of MA, AR and ARMA Series

Process	MA(q)	AR(p)	ARMA(p,q)
Auto-correlation function	Cuts off	Infinite. Tails off. Damped Exponentials and/or Cosine waves	Infinite. Tails off. Damped Exponentials and/or Cosine waves after q-p.
Partial Autocorrelation function	Infinite. Tails off. Dominated by damped Exponentials & Cosine waves.	Cuts off	Infinite. Tails off. Dominated by damped Exponentials & Cosine waves after p-q.

Demo1: Identification of the model

```
proc arima data= chem_readings plots=all;
identify var=reading  scan esacf center  ;
run;
```



- ACF is dampening, PCF graph cuts off. - Perfect example of an AR process

Demo: Identification of the model

PACF cuts off after lag 2

1. $d = 0, p = 2, q = 0$

SAS ARMA(p+d,q) Tentative Order Selection Tests			
SCAN		ESACF	
p+d	q	p+d	q
2	0	2	3
1	5	4	4
		5	3

$$y_t = a1y_{t-1} + a2y_{t-2} + \varepsilon_t$$

LAB: Identification of model

- Download web views data
- Use sgplot to create a trend chart
- What does ACF & PACF graphs say?
- Identify the model using below table
- Write the model equation

Properties of the ACF and PACF of MA, AR and ARMA Series

Process	MA(q)	AR(p)	ARMA(p,q)
Auto-correlation function	Cuts off	Infinite. Tails off. Damped Exponentials and/or Cosine waves	Infinite. Tails off. Damped Exponentials and/or Cosine waves after q-p.
Partial Autocorrelation function	Infinite. Tails off. Dominated by damped Exponentials & Cosine waves.	Cuts off	Infinite. Tails off. Dominated by damped Exponentials & Cosine waves after p-q.

Step3 : Estimation

Parameter Estimate

- We already know the model equation. AR(1,0,0) or AR(2,1,0) or ARIMA(2,1,1)
- We need to estimate the coefficients using Least squares. Minimizing the sum of squares of deviations

$$min \quad \sum_t \varepsilon_t^2$$

$$min \quad \sum_{t=2}^{T} (y_t - \phi y_{t-1})^2$$

Demo1: Parameter Estimation

- Chemical reading data

```
proc arima data=chem_readings;
identify var=reading scan esacf  center;
estimate p=2 q=0 noint method=ml;
run;
```

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
AR1,1	0.42444	0.06928	6.13	<.0001	1
AR1,2	0.25315	0.06928	3.65	0.0003	2

$$y_t = 0.\ 424y_{t-1} + 0.2532y_{t-2} + \varepsilon_t$$

Lab: Parameter Estimation

- Estimate the parameters for webview data

Step4 : Forecasting

Forecasting

- Now the model is ready
- We simply need to use this model for forecasting

```
proc arima data=chem_readings;
identify var=reading scan esacf  center;
estimate p=2 q=0 noint method=ml;
forecast lead=4 ;
run;
```

Forecasts for variable Reading				
Obs	Forecast	Std Error	95% Confidence Limits	
198	17.2405	0.3178	16.6178	17.8633
199	17.2235	0.3452	16.5469	17.9000
200	17.1759	0.3716	16.4475	17.9043
201	17.1514	0.3830	16.4007	17.9020

LAB: Forecasting using ARIMA

- Forecast the number of sunspots for next three hours

Validation: How good is my model?

- Does our model really give an adequate description of the data
- Two criteria to check the goodness of fit
  - Akaike information criterion (AIC)
  - Schwartz Bayesiancriterion (SBC)/Bayesian information criterion (BIC).
- These two measures are useful in comparing two models.
- The smaller the AIC & SBC the better the model

Goodness of fit

- Remember… **Residual analysis** and Mean deviation, Mean Absolute Deviation and Root Mean Square errors?
- Four common techniques are the:
  - Mean absolute deviation, $$\text{MAD} = \sum_{i=1}^{n} \frac{\left|Y_i - \hat{Y}_i\right|}{n}$$
  - Mean absolute percent error $$\text{MAPE} = \frac{100}{n} \sum_{i=1}^{n} \frac{\left|Y_i - \hat{Y}_i\right|}{Y_i}$$
  - Mean square error, $$\text{MSE} = \sum_{i=1}^{n} \frac{\left(Y_i - \hat{Y}_i\right)^2}{n}$$
  - Root mean square error. $$\text{RMSE} = \sqrt{\text{MSE}}$$

Lab: Overall Steps on sunspot example

- Import the time series data
- Prepare the data for model building- Make it stationary
- Identify the model type
- Estimate the parameters
- Forecast the future values

Thank you