Time Series Modeling

Manish Gupta

http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc4.htm

Definitions, Applications

- *Time Series: An ordered sequence of values of a variable at equally spaced time intervals*.
- Application: Fit a model and proceed to forecasting and monitoring.
- Domains
  - Economic Forecasting
  - Sales Forecasting
  - Budgetary Analysis
  - Stock Market Analysis
  - Yield Projections
  - Process and Quality Control
  - Inventory Studies
  - Workload Projections
  - Utility Studies
  - Census Analysis
  - …

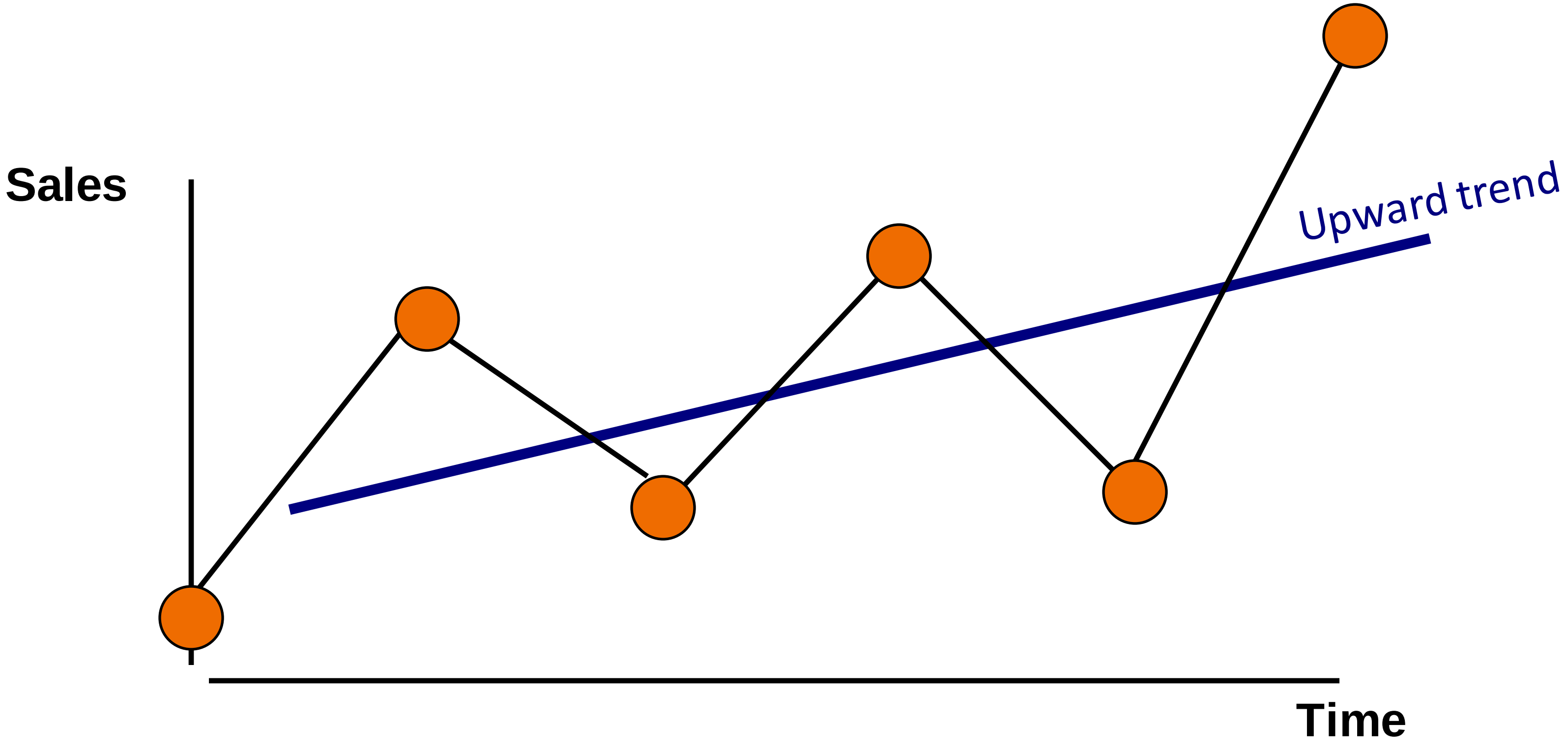
NAÏVE MODELS

- They are based solely on the most recent information available.
- Is sometimes called the "no change" forecast.
- Suitable for very small data sets.
- The simplest model is:

$$\hat{Y}_{t+1} = Y_t$$

Trend Component

- Overall Upward or Downward Movement
- Data Taken Over a Period of Years



- The technique can be adjusted to take trend into consideration:

$$\hat{Y}_{t+1} = Y_t + (Y_t - Y_{t-1})$$

- The rate of change might be more appropriate than the absolute amount of change:

$$\hat{Y}_{t+1} = Y_t \frac{Y_t}{Y_{t-1}}$$

Simple Averages

- Uses the mean of all the relevant historical observations as the forecast of the next period.

$$\hat{Y}_{t+1} = \frac{1}{t}\sum_{i=1}^{t} Y_i$$

- New observation is added:

$$\hat{Y}_{t+1} = \frac{t\hat{Y}_{t+1} + Y_{t+1}}{t+1}$$

Moving Averages

- A moving average of order k is the mean value of the k most recent observations

$$\hat{Y}_{t+1} = \frac{(Y_t + Y_{t-1} + Y_{t-2} + \cdots . + Y_{t-k+1})}{k}$$

$$K \ = \ number\ of\ terms\ in\ the\ moving\ average$$

- The method does not handle trend or seasonality very well, although it does work better than the simple average method.

Use 5 week moving average (last data)

Week t	Purchases
1	275
2	291
3	307
4	281
5	295
6	268
7	252
8	279
9	264
10	288

Week t	Purchases
11	302
12	287
13	290
14	311
15	277
16	245
17	282
18	277
19	298
20	303

Week t	Purchases
21	310
22	299
23	285
24	250
25	260
26	245
27	271
28	282
29	302
30	285

Results

Week	Purchases	AVER1	RESI1
1	275	*	*
2	291	*	*
3	307	*	*
4	281	*	*
5	295	289.8	*
6	268	288.4	-21.8
7	252	280.6	-36.4
8	279	275.0	-1.6
9	264	271.6	-11.0
10	288	270.2	16.4
11	302	277.0	31.8
12	287	284.0	10.0
13	290	286.2	6.0
14	311	295.6	24.8
15	277	293.4	-18.6
16	245	282.0	-48.4
17	282	281.0	0.0
18	277	278.4	-4.0
19	298	275.8	19.6
20	303	281.0	27.2
21	310	294.0	29.0
22	299	297.4	5.0
23	285	299.0	-12.4
24	250	289.4	-49.0
25	260	280.8	-29.4
26	245	267.8	-35.8
27	271	262.2	3.2
28	282	261.6	19.8
29	302	272.0	40.4
30	285	277.0	13.0

$$\frac{275 + 291 + 307 + 281 + 295}{5} = 289.8$$

Exponential smoothing

- Assigns *exponentially decreasing weights* as the observation get older
- *Recent observations are given relatively more weight in forecasting than the older observations*.
- Let S denote a smoothed value
  - Set $S_2 = y_1$ and then $S_t = \alpha y_{t-1} + (1-\alpha)S_{t-1} \quad 0 \leq \alpha \leq 1 \quad t \geq 3$
- The speed at which the older responses are dampened (smoothed) is a function of the value of α. When α is close to 1, dampening is quick and when α is close to 0, dampening is slow.
- We choose the best value for α as the value which results in the smallest MSE.

Exponential smoothing

Time	y_t	$S(\alpha = 0.1)$	Error	Error squared
1	71			
2	70	71	-1.00	1.00
3	69	70.9	-1.90	3.61
4	68	70.71	-2.71	7.34
5	64	70.44	-6.44	41.47
6	65	69.80	-4.80	23.04
7	72	69.32	2.68	7.18
8	78	69.58	8.42	70.90
9	75	70.43	4.57	20.88
10	75	70.88	4.12	16.97
11	75	71.29	3.71	13.76
12	70	71.67	-1.67	2.79

- For α=0.1, the sum of the squared errors (SSE) = 208.94. The mean of the squared errors (MSE) is the SSE /11 = 19.0.
- The MSE was again calculated for α=0.5 and turned out to be 16.29, so in this case we would prefer an α of 0.5.
- We could do a complete search for α from 0 to 1.
- Forecasting formula:
  - $S_{t+1} = \alpha y_t + (1 - \alpha) S_t, \quad 0 < \alpha \leq 1, t > 0$
- Not good when there is trend in the data.

Exponential smoothing

- $\alpha = 0.5$ or $\alpha = 0.9$

- $S_{(t+1)} = \alpha y_t + (1-\alpha)S_t$, where $S_2 = y_1$

- $S_4 = \alpha y_3 + (1-\alpha)S_3$
- $S_4 = \alpha y_3 + (1-\alpha)[\alpha y_2 + (1-\alpha)S_2]$
- $S_4 = \alpha y_3 + (1-\alpha)[\alpha y_2 + (1-\alpha)y_1]$

- When $\alpha = 0.5$, $S_4 = 0.5y_3 + 0.25y_2 + 0.25y_1$
- When $\alpha = 0.9$, $S_4 = 0.9y_3 + 0.09y_2 + 0.01y_1$

Double exponential smoothing (Holt's 2 parameter method)

- Formulas

$$S_t = \alpha y_{t-1} + (1-\alpha)(S_{t-1} + b_{t-1}) \quad 0 \le \alpha \le 1$$

$$b_t = \gamma(S_{t-1} - S_{t-2}) + (1-\gamma) b_{t-1} \quad 0 \le \gamma \le 1$$

- Initialization

S_1 is in general set to y_1.

$$b_3 = y_2 - y_1$$

$$b_5 = \frac{1}{3}[(y_2 - y_1) + (y_3 - y_2) + (y_4 - y_3)]$$

$$b_n + 1 = \frac{y_n - y_1}{n-1}$$

- The first smoothing equation adjusts S_t directly for the trend of the previous period, b_{t-1} by adding it to the last smoothed value,. This helps to eliminate the lag and brings S_{t-1} to the appropriate base of the current value. S_{t-1}

- The second smoothing equation then updates the trend, which is expressed as the difference between the last two values. The equation is similar to the basic form of single smoothing, but here applied to the updating of the trend.

Forecast using Double smoothing

- Consider this dataset:
  6.4, 5.6, 7.8, 8.8, 11, 11.6, 16.7, 15.3, 21.6, 22.4
- Fit a double smoothing model with α=0.3623 and γ=1.0. These are the estimates that result in the lowest possible MSE. For comparison's sake we also fit a single smoothing model with α=0.977 (this results in the lowest MSE for single exponential smoothing).
- The MSE for double smoothing is 3.7024.
- The MSE for single smoothing is 8.8867.

The one-period-ahead forecast is given by:

$$F_{t+1} = S_t + b_t \, .$$

The m-periods-ahead forecast is given by:

$$F_{t+m} = S_t + mb_t \, .$$

Seasonality

- By seasonality, we mean periodic fluctuations.
  - For example, retail sales tend to peak for the Christmas season and then decline after the holidays. So, time series of retail sales will typically show increasing sales from September through December and declining sales in January and February.

- If seasonality is present, it must be incorporated into the time series model.

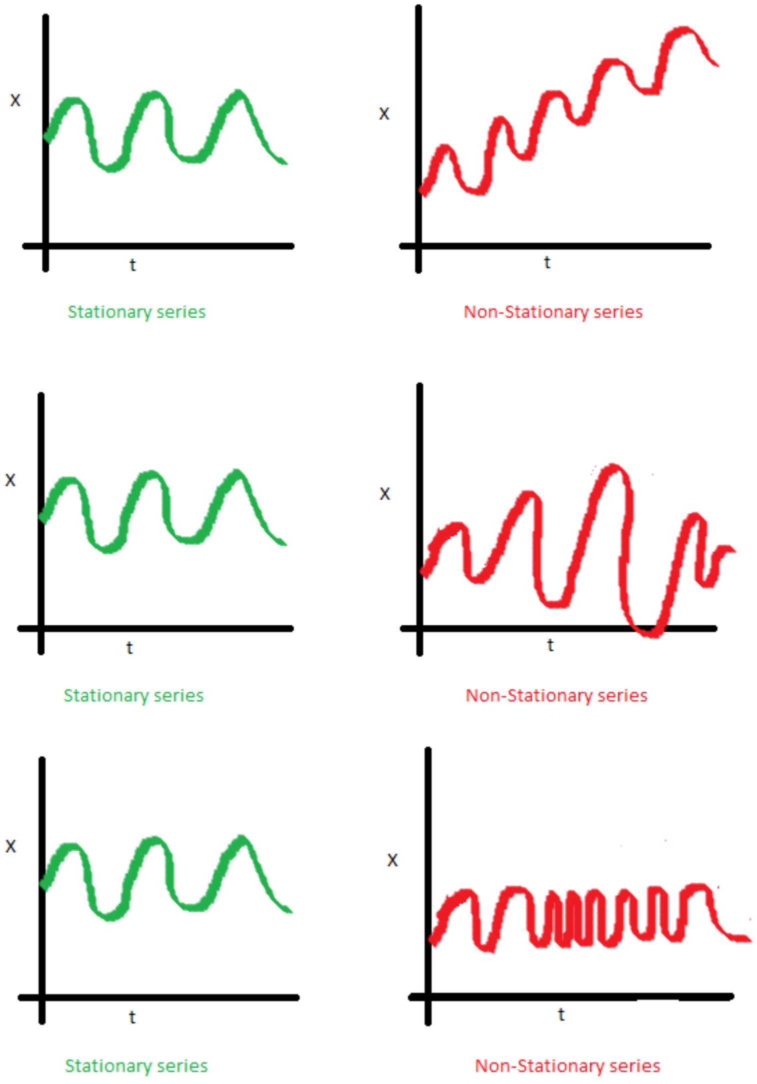
- Techniques for detecting seasonality.
  - A run sequence plot will often show seasonality.
  - The autocorrelation plot can help identify seasonality: If there is significant seasonality, the autocorrelation plot should show spikes at lags equal to the period.

Triple exponential smoothing

- What happens if the data show trend and seasonality?
- Double smoothing will not work.
- So, we bring in a third equation to take care of seasonality (sometimes called periodicity). The resulting set of equations is called the "Holt-Winters" (HW) method after the names of the inventors.

Stationary series

- The mean of the series should not be a function of time rather should be a constant.
- The variance of the series should not be a function of time. This property is known as homoscedasticity.
- The covariance of ith term and the (i + m)th term should not be a function of time.
- A stationary process has the property that the mean, variance, and autocorrelation structure do not change over time.



Why care about stationarity of a time series?

- Until unless your time series is stationary, you cannot build a good time series model.
- In cases where the stationary criterion is violated, the first requisite becomes to stationarize the time series and then try models to predict this time series.
- There are multiple ways of bringing this stationarity. Some of them are Detrending, Differencing, etc.

Making a series stationary

- We can difference the data. That is, given the series Z_t, we create the new series
  - $Y_i = Z_i - Z_{i-1}$

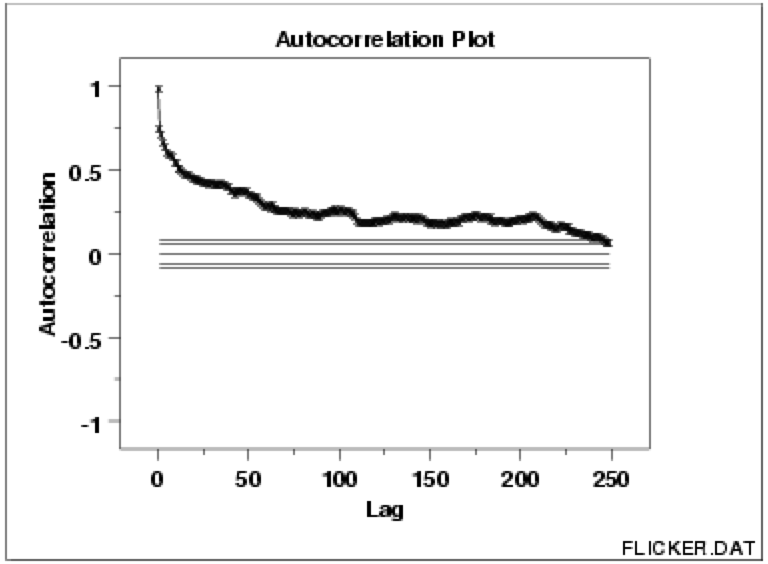
- If the data contains a trend, we can fit some type of curve to the data and then model the residuals from that fit. Since the purpose of the fit is to simply remove long term trend, a simple fit, such as a straight line, is typically used.

- For non-constant variance, taking the logarithm or square root of the series may stabilize the variance.

- A simple way to correct for a seasonal component is to use differencing. If there is a seasonal component at the level of one week, then we can remove it on an observation today by subtracting the value from last week.

Autocorrelation Plot and PACF

- **Autocorrelation**, also known as **serial correlation**, is the correlation of a signal with a delayed copy of itself as a function of delay. Informally, it is the similarity between observations as a function of the time lag between them.
- The partial autocorrelation at lag k is the autocorrelation between X_t and X_{t-k} that is not accounted for by lags 1 through k−1.



$$\rho(k) = \frac{\frac{1}{n-k}\sum_{t=k+1}^{n}(y_t-\bar{y})(y_{t-k}-\bar{y})}{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t-\bar{y})}\sqrt{\frac{1}{n-k}\sum_{t=k+1}^{n}(y_{t-k}-\bar{y})}}$$

Each partial autocorrelation could be obtained as a series of regressions of the form:

$$\tilde{y}_t = \phi_{21}\tilde{y}_{t-1} + \phi_{22}\tilde{y}_{t-2} + e_t \,,$$

where $\tilde{y}_t$ is the original series minus the sample mean, $y_t - \bar{y}$. The estimate of ϕ_{22} will give the value of the partial autocorrelation of order 2. Extending the regression with k additional lags, the estimate of the last term will give the partial autocorrelation of order k.

Framework of ARIMA Time series modeling

1. Visualize the time series
2. Stationarize the series
3. Plot ACF/PACF charts and find optimal parameters
4. Build the ARIMA model
5. Make Predictions

ARIMA model description

- Decompose into trend, seasonal and residual components.
- AR (auto-regressive) models
- MA (moving average) models
- **I**: *Integrated*. The use of differencing of raw observations (i.e. subtracting an observation from an observation at the previous time step) in order to make the time series stationary.
- A moving average model is conceptually a [linear regression]() of the current value of the series against the white noise or random shocks of one or more prior values of the series.
- Box and Jenkins popularized an approach that combines the moving average and the autoregressive approaches.

$$X_t = \delta + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{p-1} + A_t$$

where X_t is the time series, A_t is white noise, and

$$\delta = \left(1 - \sum_{i=1}^{p} \phi_i\right) \mu \,,$$

with μ denoting the process mean.

An autoregressive model is simply a [linear regression]() of the current value of the series against one or more prior values of the series. The value of p is called the order of the AR model.

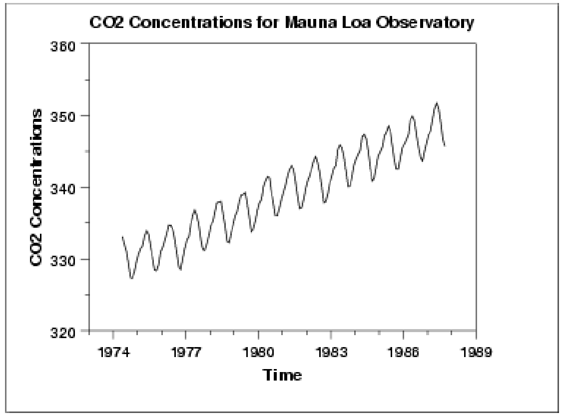
$$X_t = \mu + A_t - \theta_1 A_{t-1} - \theta_2 A_{t-2} - \cdots - \theta_q A_{t-q}$$

where X_t is the time series, μ is the mean of the series, A_{t-i} are white noise terms, and $\theta_1, \ldots, \theta_q$ are the parameters of the model. The value of q is called the order of the MA model.

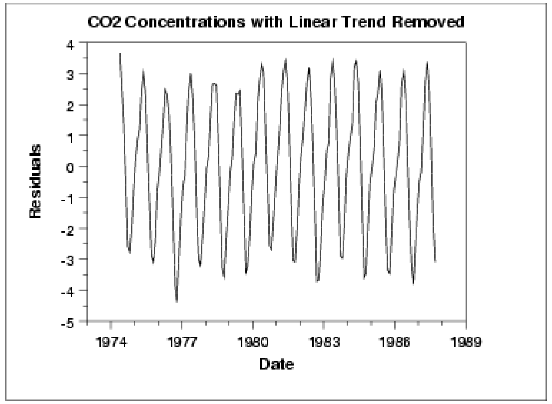
Shape of Autocorrelation Function

SHAPE	INDICATED MODEL
Exponential, decaying to zero	Autoregressive model. Use the partial autocorrelation plot to identify the order of the autoregressive model.
Alternating positive and negative, decaying to zero	Autoregressive model. Use the partial autocorrelation plot to help identify the order.
One or more spikes, rest are essentially zero	Moving average model, order identified by where plot becomes zero.
Decay, starting after a few lags	Mixed autoregressive and moving average model.
All zero or close to zero	Data is essentially random.
High values at fixed intervals	Include seasonal autoregressive term.
No decay to zero	Series is not stationary.

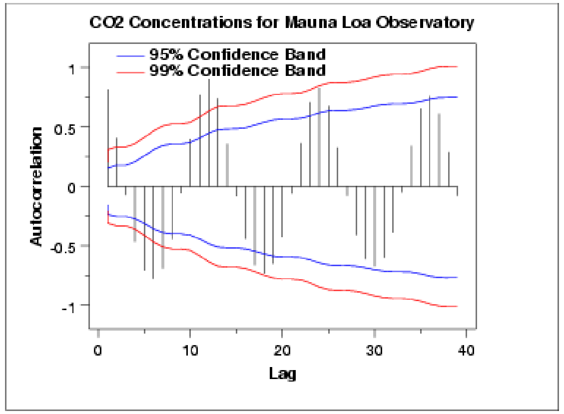
Model Identification example



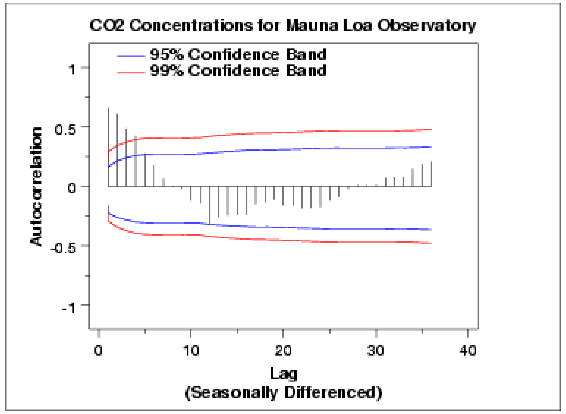
run sequence plot of the data indicates a rising trend. simple linear fit should be sufficient to remove this upward trend.



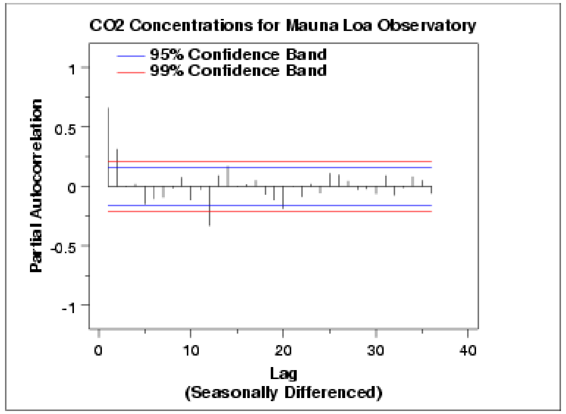
This plot contains the residuals from a linear fit to the original data. But it has seasonality. To know the period draw autocorrelation plot.



The autocorrelation plot shows an alternating pattern of positive and negative spikes. It also shows a repeating pattern every 12 lags, which indicates a seasonality effect.

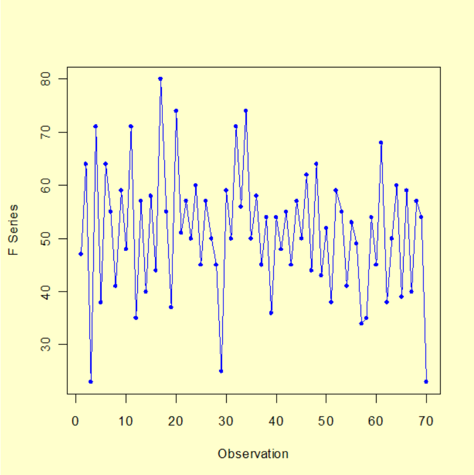


autocorrelation plot on seasonally differenced data shows a mixture of exponential decay and a damped sinusoidal pattern.

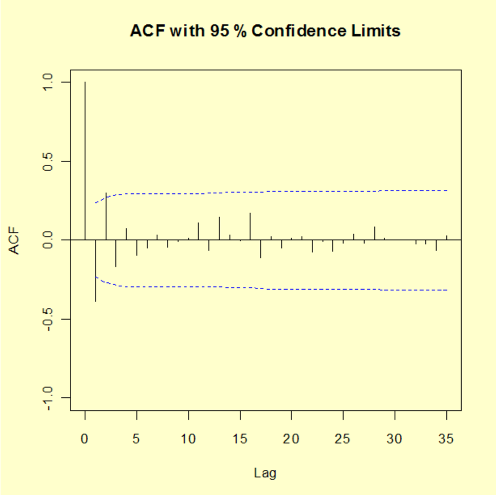


The partial autocorrelation plot suggests that an AR(2) model might be appropriate since the partial autocorrelation becomes zero after the second lag.

Univariate Box-Jenkins Analysis example



Run sequence plot shows no seasonal component or a noticeable trend. We compute the autocorrelation function (ACF) of the data for the first 35 lags to determine the type of model to fit to the data.



The ACF values alternate in sign and decay quickly after lag 2, indicating that an AR(2) model is appropriate for the data.

We fit an AR(2) model to the data.

$$X_t = \delta + \phi_1 X_{t-1} + \phi_2 X_{t-2} + A_t$$

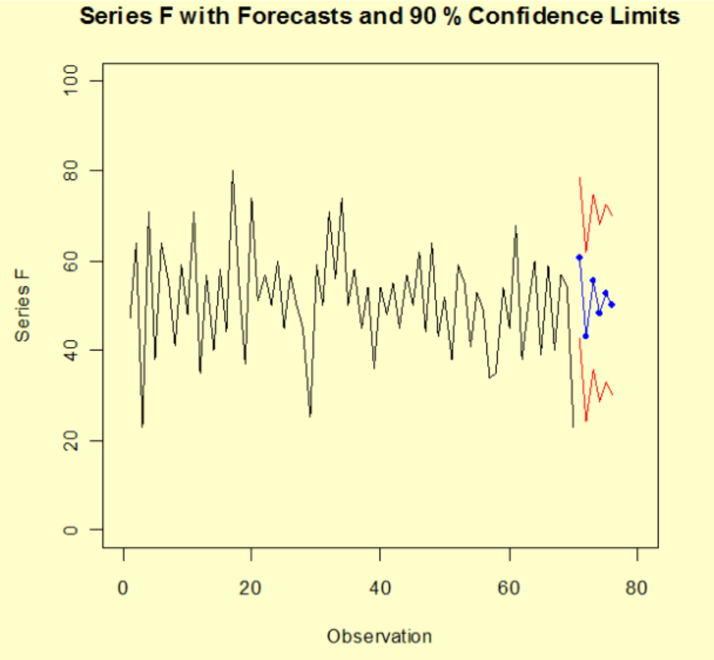
The model fitting results are shown below.

```
Source    Estimate    Standard Error
------    --------    --------------
φ1         -0.3198        0.1202
φ2          0.1797        0.1202

δ = 51.1286
Residual standard deviation = 10.9599

Test randomness of residuals:
Standardized Runs Statistic Z = 0.4887, p-value = 0.625
```

Using our AR(2) model, we forcast values six time periods into the future.

Period	Prediction	Standard Error
71	60.6405	10.9479
72	43.0317	11.4941
73	55.4274	11.9015
74	48.2987	12.0108
75	52.8061	12.0585
76	50.0835	12.0751

The "historical" data and forecasted values (with 90 % confidence limits) are shown in the graph below.



Thanks!!

Questions?