Time series analysis

Lecture 4. Forecasting in the framework of Box-Jenkins model

Dr. Khamidov Obidjon

Removing non-stationarity in time series

- The non-stationary pattern in a time series data needs to be removed in order that other correlation structure present in the series can be seen before proceeding with model building.
- One way of removing non-stationarity is through the method of differencing.

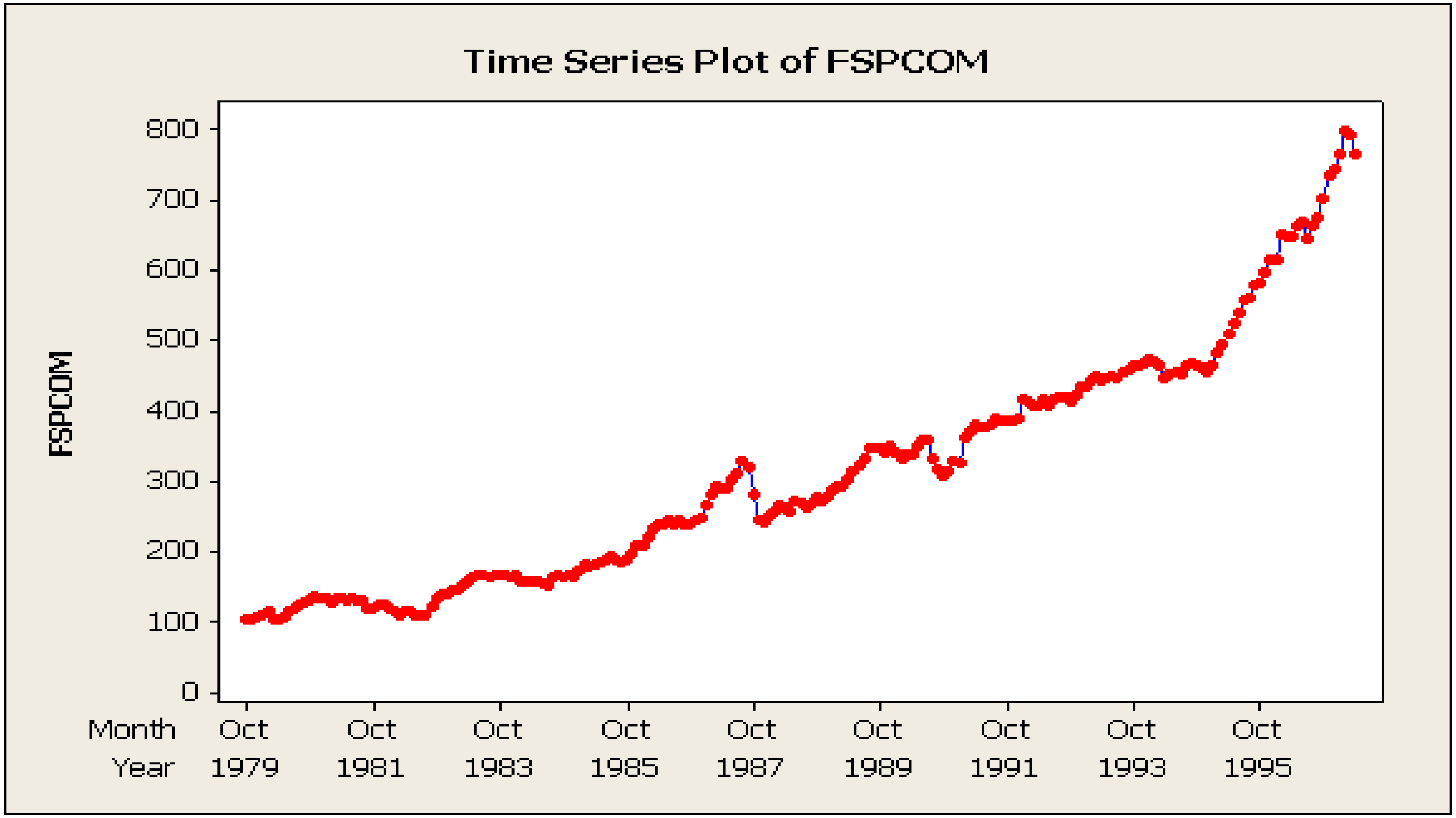
Removing non-stationarity in time series

- The differenced series is defined as:

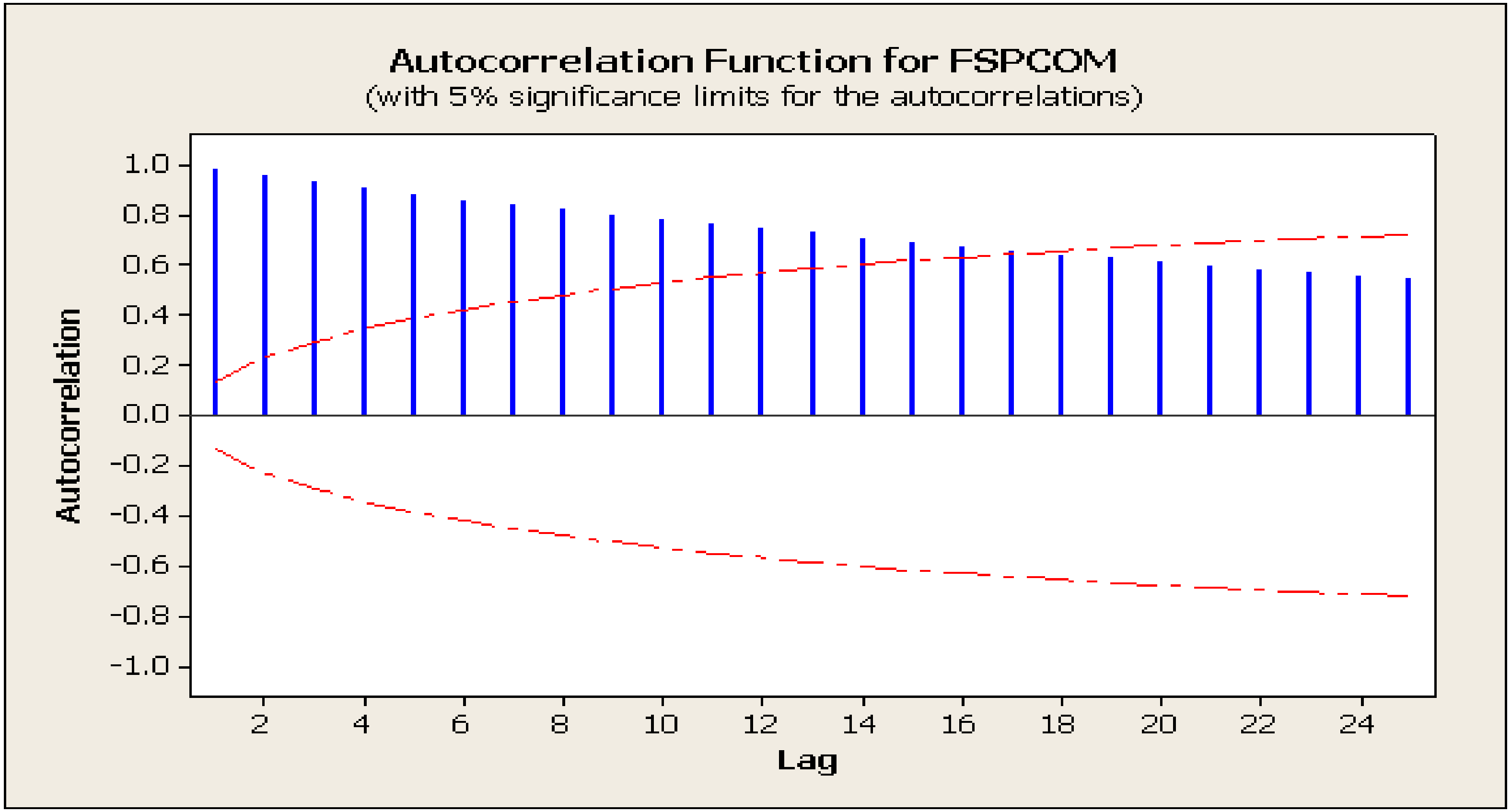
$$y'_t = y_t - y_{t-1}$$

- The following two slides shows the time series plot and the ACF plot of the monthly S&P 500 composite index from 1979 to 1997.

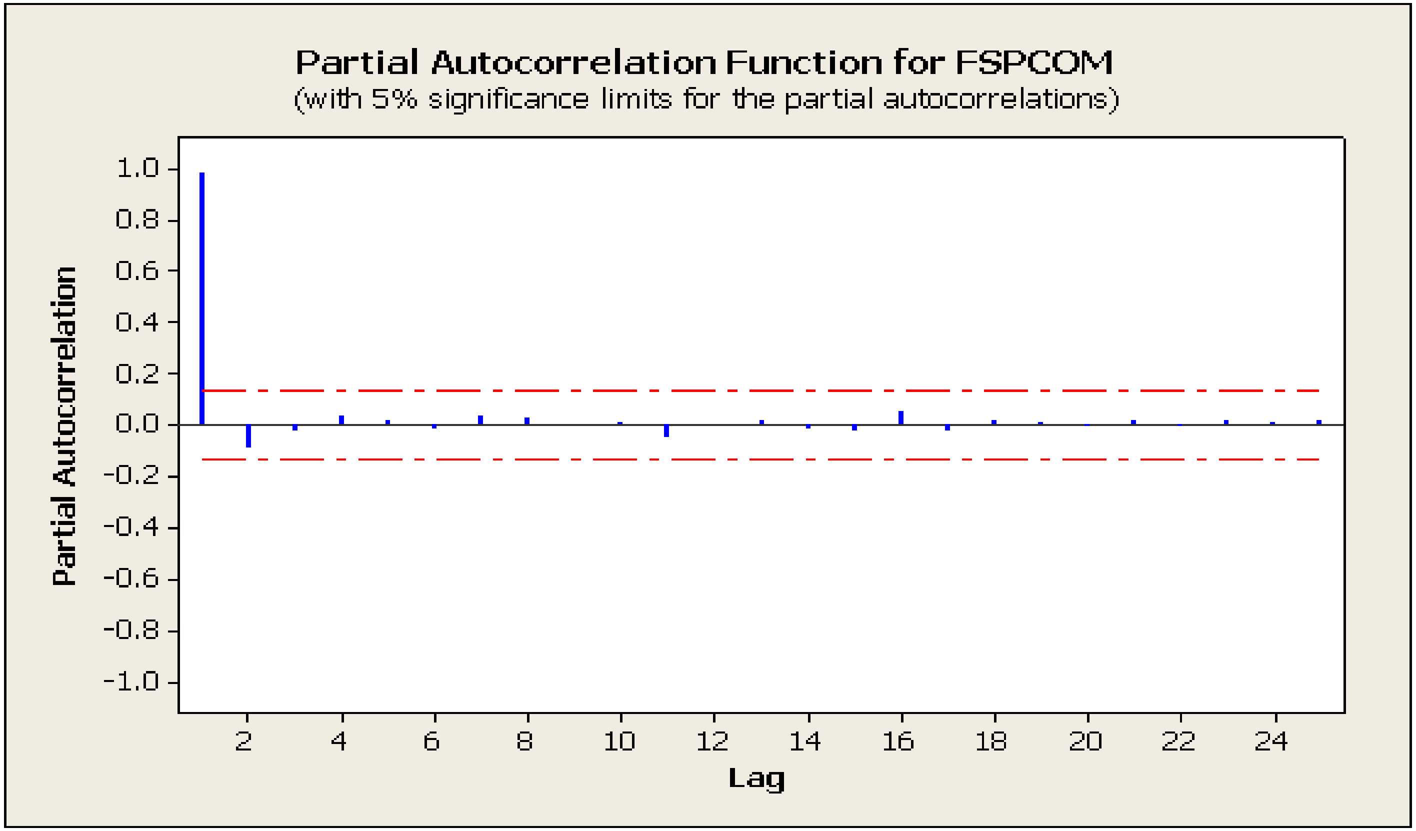
Removing non-stationarity in time series



Removing non-stationarity in time series



Removing non-stationarity in time series



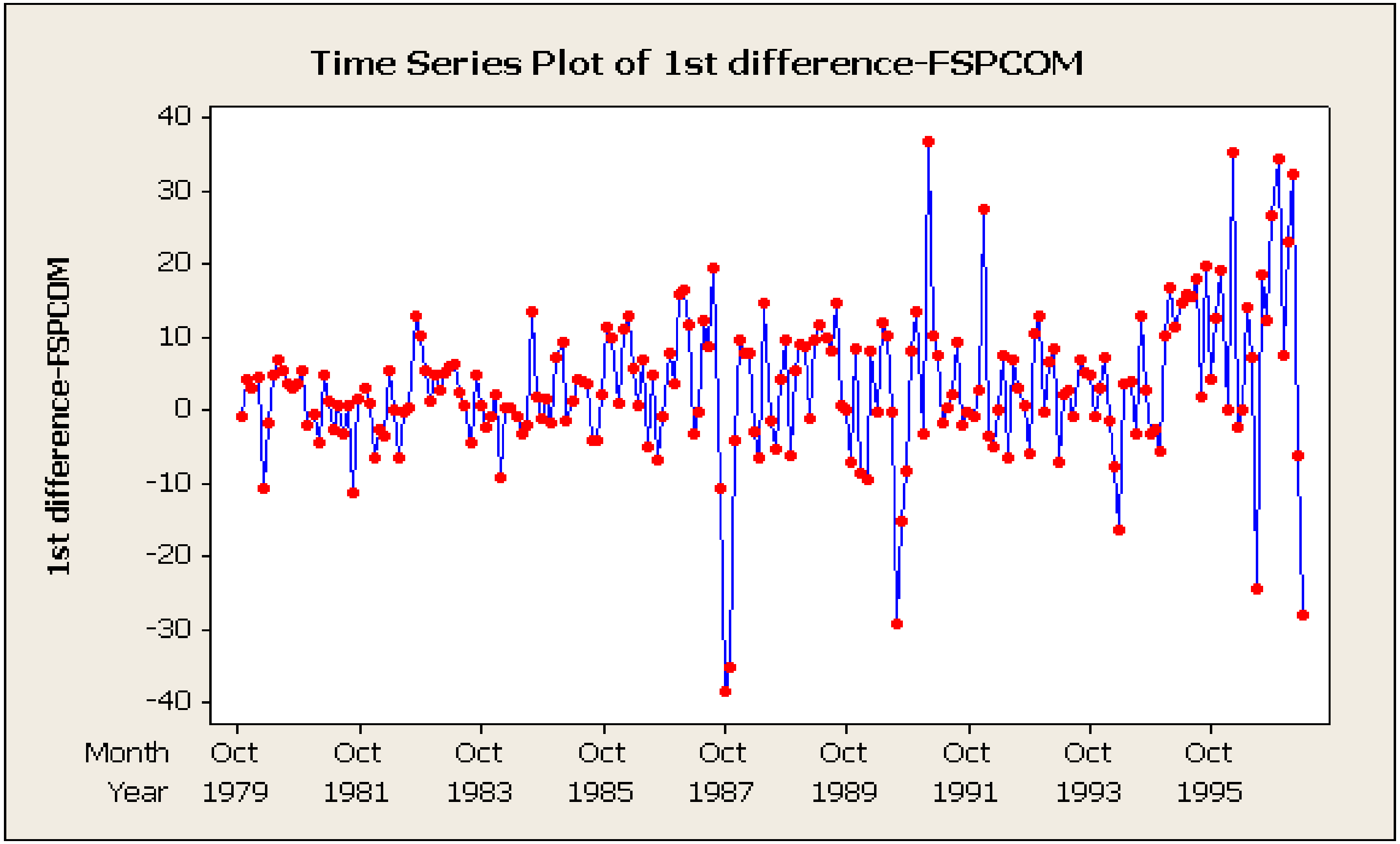
Removing non-stationarity in time series

- The time plot shows that it is not stationary in the mean.
- The ACF and PACF plot also display a pattern typical for non-stationary pattern.
- Taking the first difference of the S& P 500 composite index data represents the monthly changes in the S&P 500 composite index.

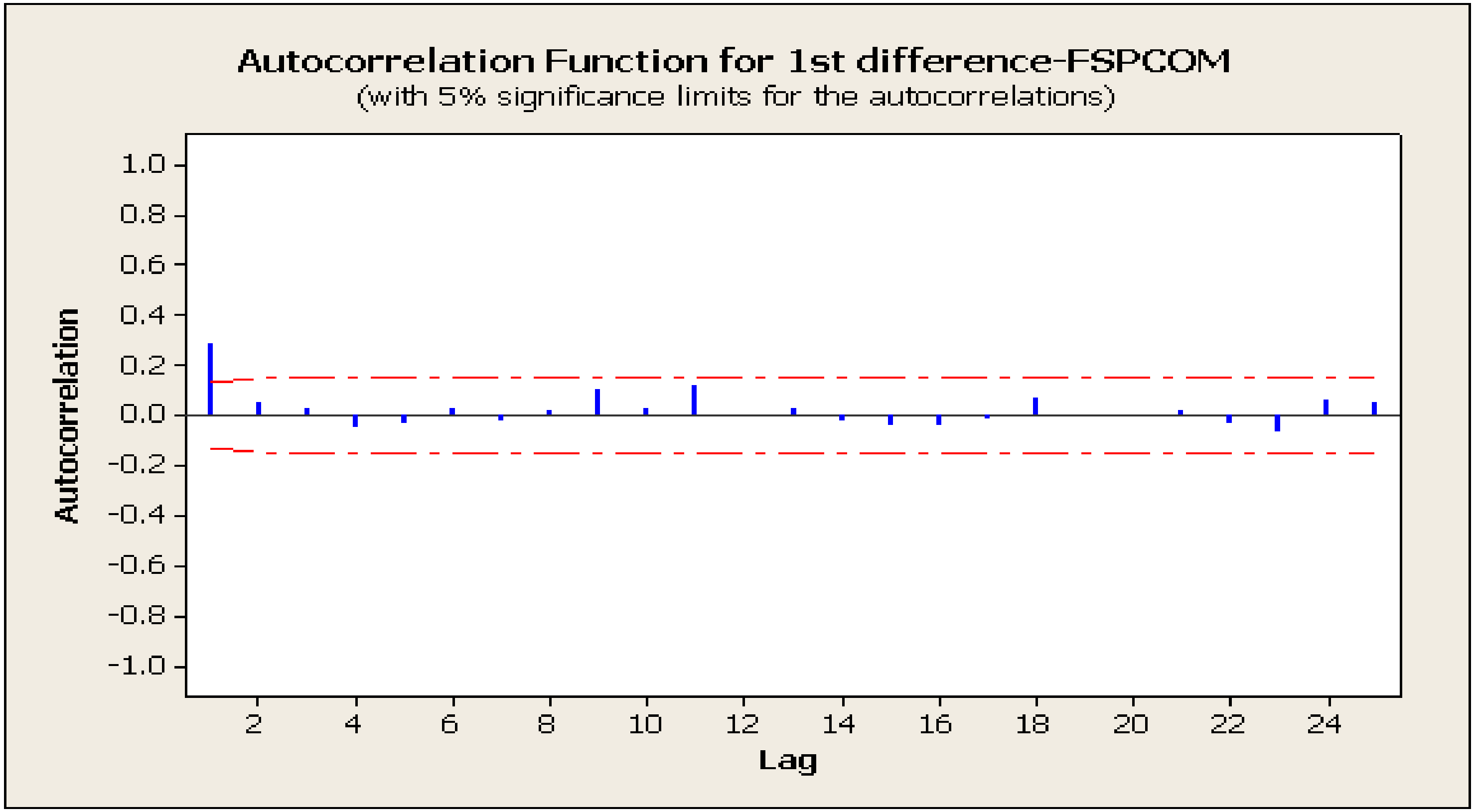
Removing non-stationarity in time series

- The time series plot and the ACF and PACF plots indicate that the first difference has removed the growth in the time series data.
- The series looks just like a white noise with almost no autocorrelation or partial autocorrelation outside the 95% limits.

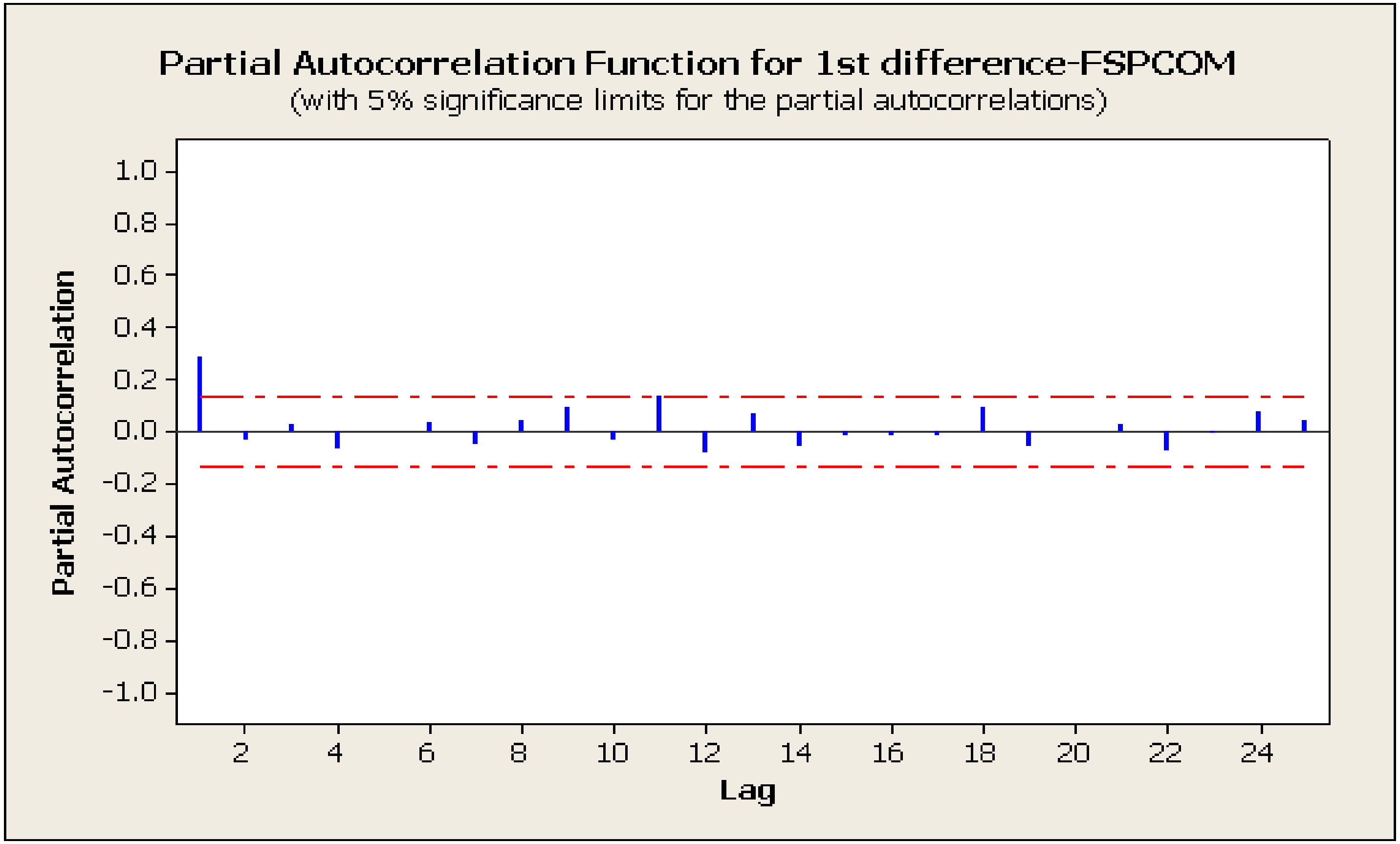
Removing non-stationarity in time series



Removing non-stationarity in time series



Removing non-stationarity in time series



Removing non-stationarity in time series

- Note that the ACF and PACF at lag 1 is outside the limits, but it is acceptable to have about 5% of spikes fall a short distance beyond the limit due to chance.

Random Walk

- Let y_t denote the S&P 500 composite index, then the time series plot of differenced S&P 500 composite index suggests that a suitable model for the data might be

$$y_t - y_{t-1} = e_t$$

- Where e_t is white noise.

Random Walk

- The equation in the previous slide can be rewritten as

$$y_t = y_{t-1} + e_t$$

- This model is known as "random walk" model and it is widely used for non-stationary data.

Random Walk

- Random walks typically have long periods of apparent trends up or down which can suddenly change direction unpredictably
- They are commonly used in analyzing economic and stock price series.

Removing non-stationarity in time series

- Taking first differencing is a very useful tool for removing non-statioanarity, but sometimes the differenced data will not appear stationary and it may be necessary to difference the data a second time.

Removing non-stationarity in time series

- The series of second order difference is defined:

$$y''_t = y'_t - y'_{t-1} = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2}$$

- In practice, it is almost never necessary to go beyond second order differences.

Seasonal differencing

- With seasonal data which is not stationary, it is appropriate to take seasonal differences.
- A seasonal difference is the difference between an observation and the corresponding observation from the previous year.

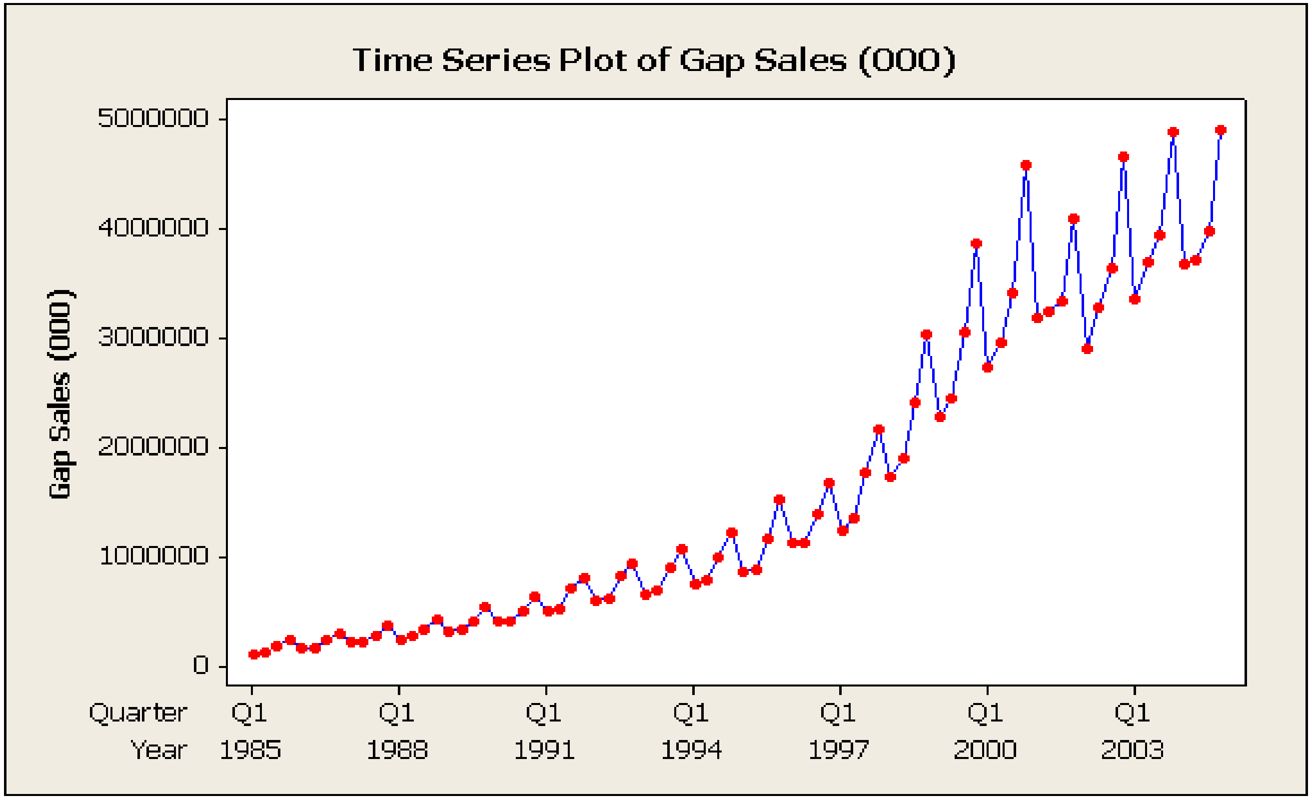
$$y'_t = y_t - y_{t-s}$$

  – Where s is the length of the season

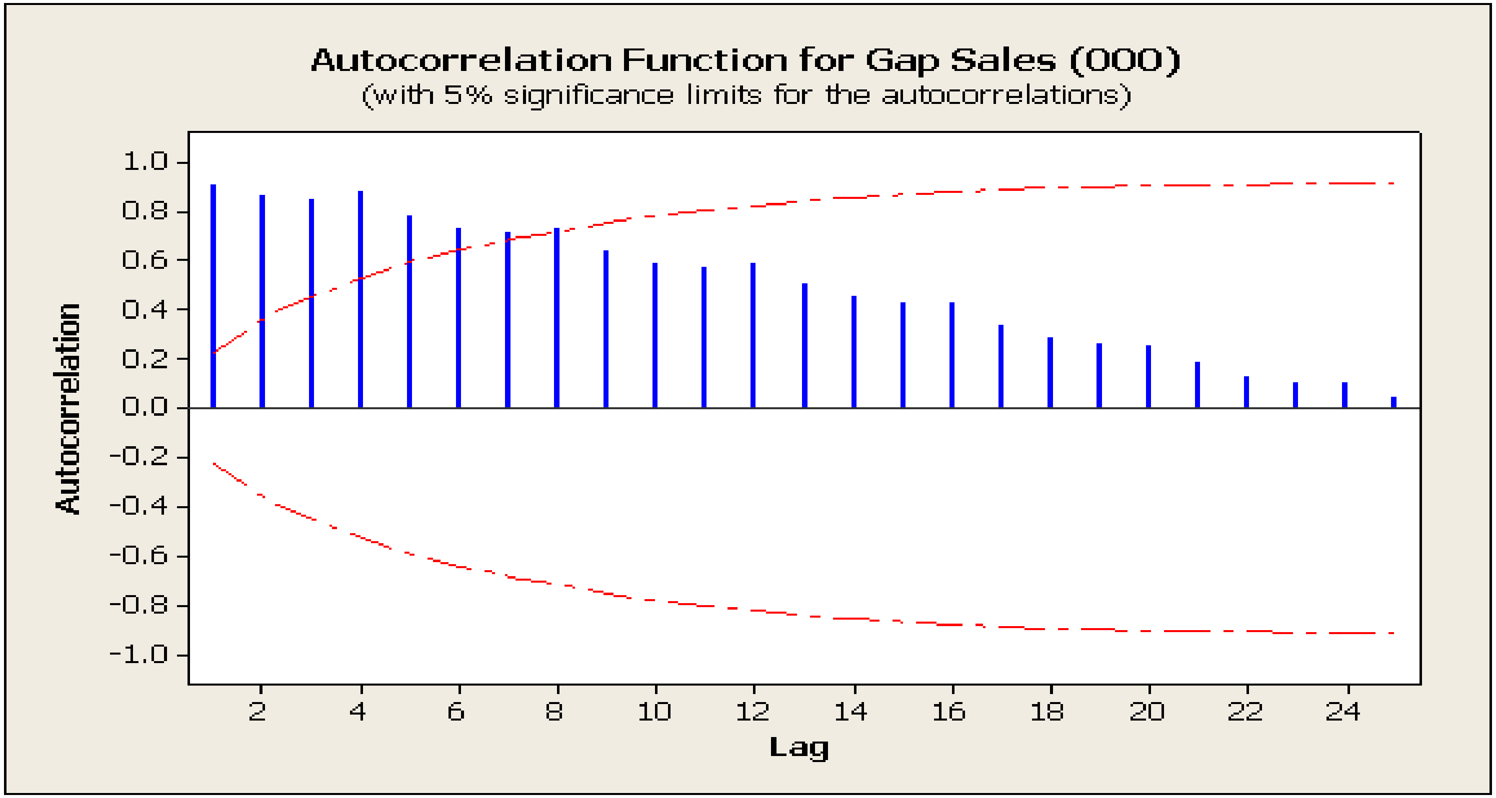
Seasonal differencing

- The Gap quarterly sales is an example of a non-stationary seasonal data.
- The following time series plot show a trend with a pronounced seasonal component
- The auto correlations show that
  - The series is non-stationary.
  - The series is seasonal.

Seasonal differencing



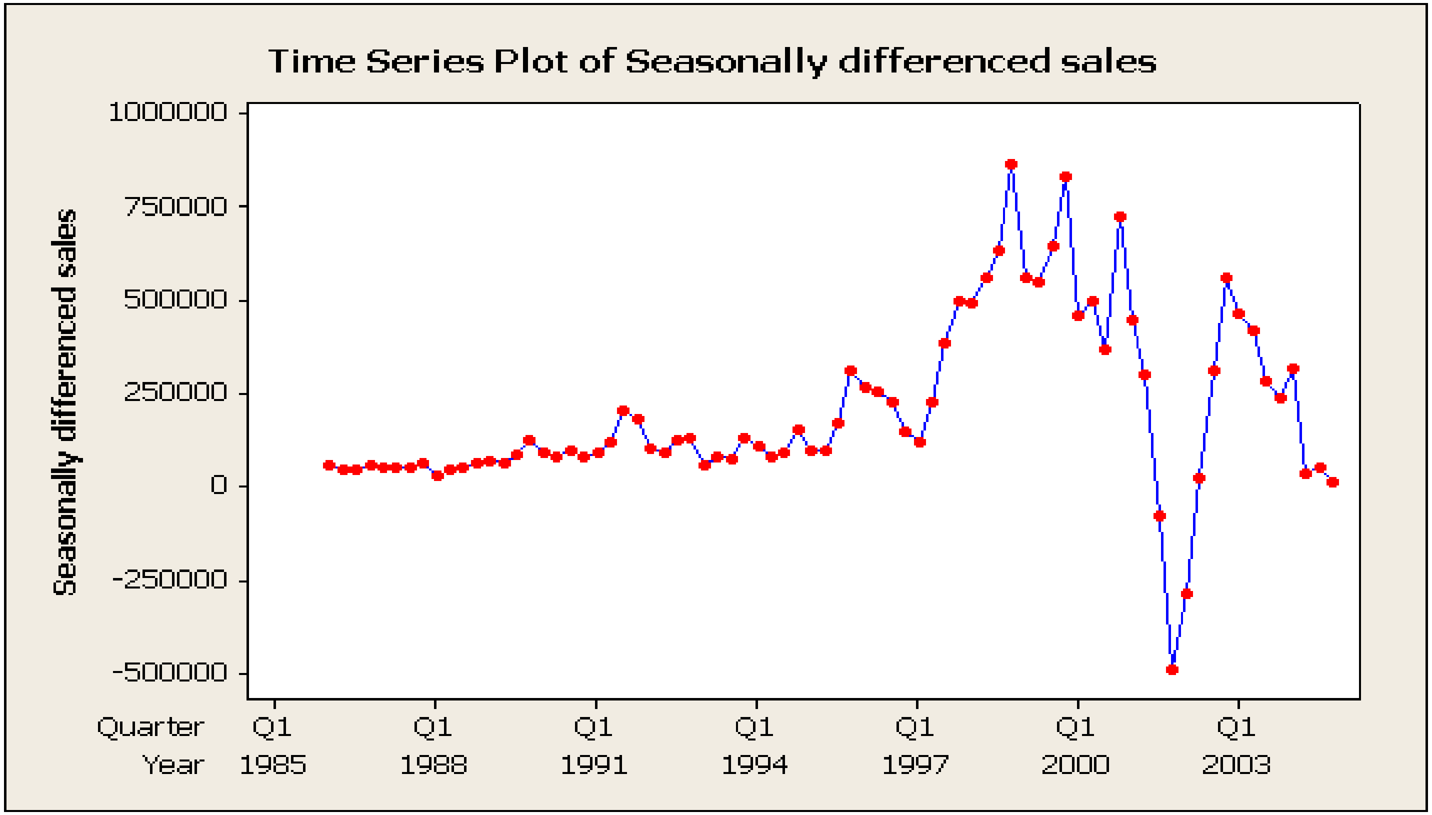
Seasonal differencing



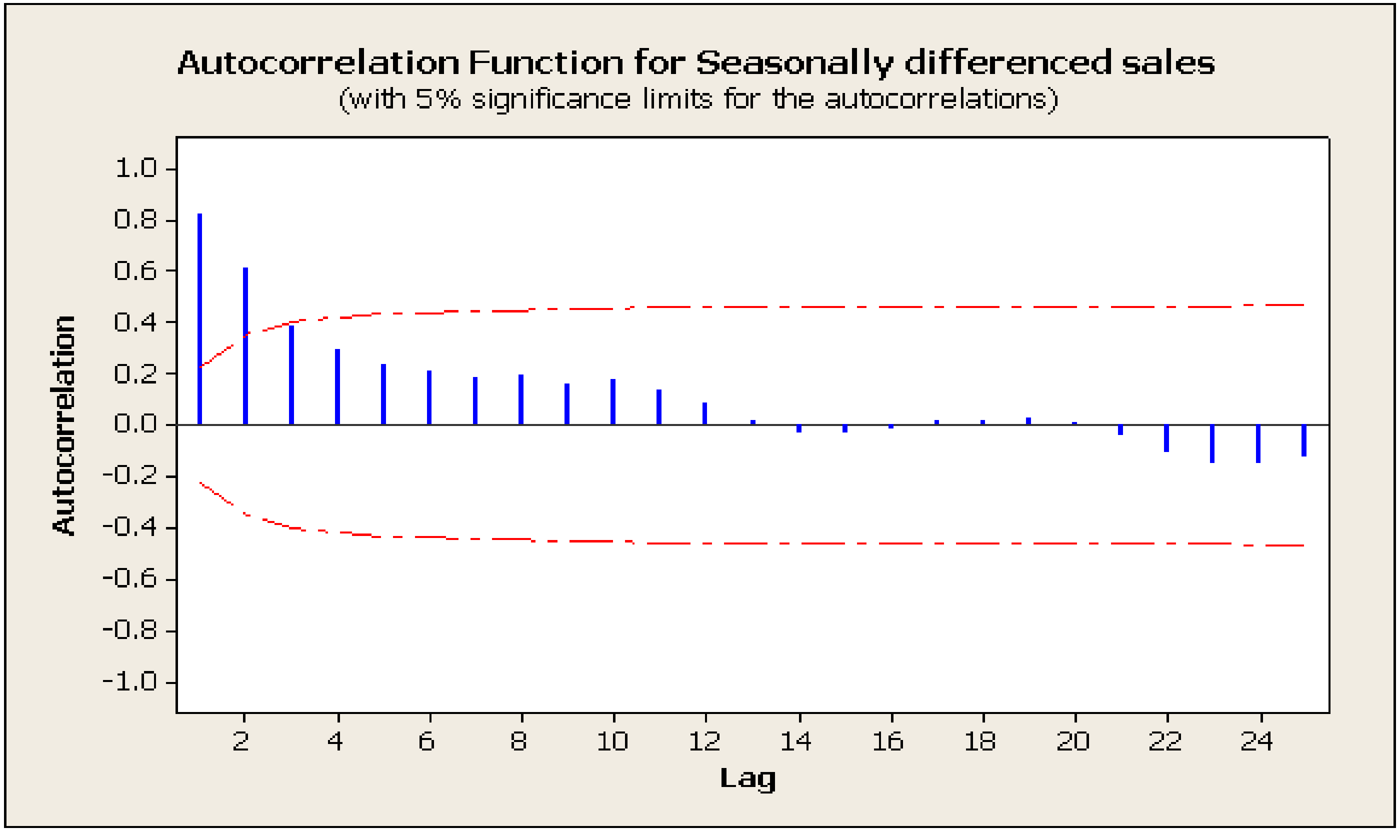
Seasonal differencing

- The seasonally differenced series represents the change in sales between quarters of consecutive years.
- The time series plot, ACF and PACF of the seasonally differenced Gap's quarterly sales are in the following three slides.

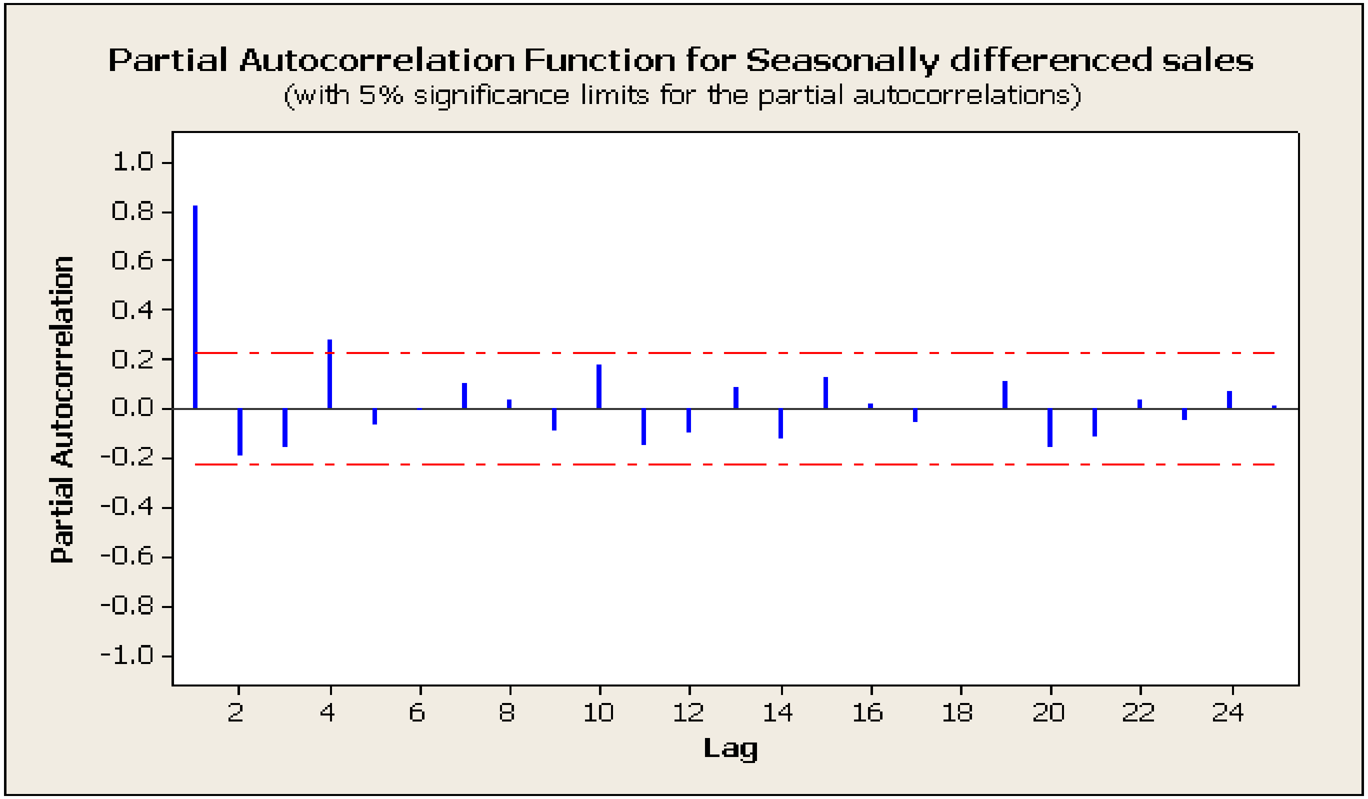
Seasonal differencing



Seasonal differencing



Seasonal differencing



Seasonal differencing

- The series is now much closer to being stationary, but more than 5% of the spikes are beyond 95% critical limits and autocorrelation show gradual decline in values.
- The seasonality is still present as shown by spike at time lag 4 in the PACF.

Seasonal differencing

- The remaining non-stationarity in the mean can be removed with a further first difference.
- When both seasonal and first differences are applied, it does not make no difference which is done first.

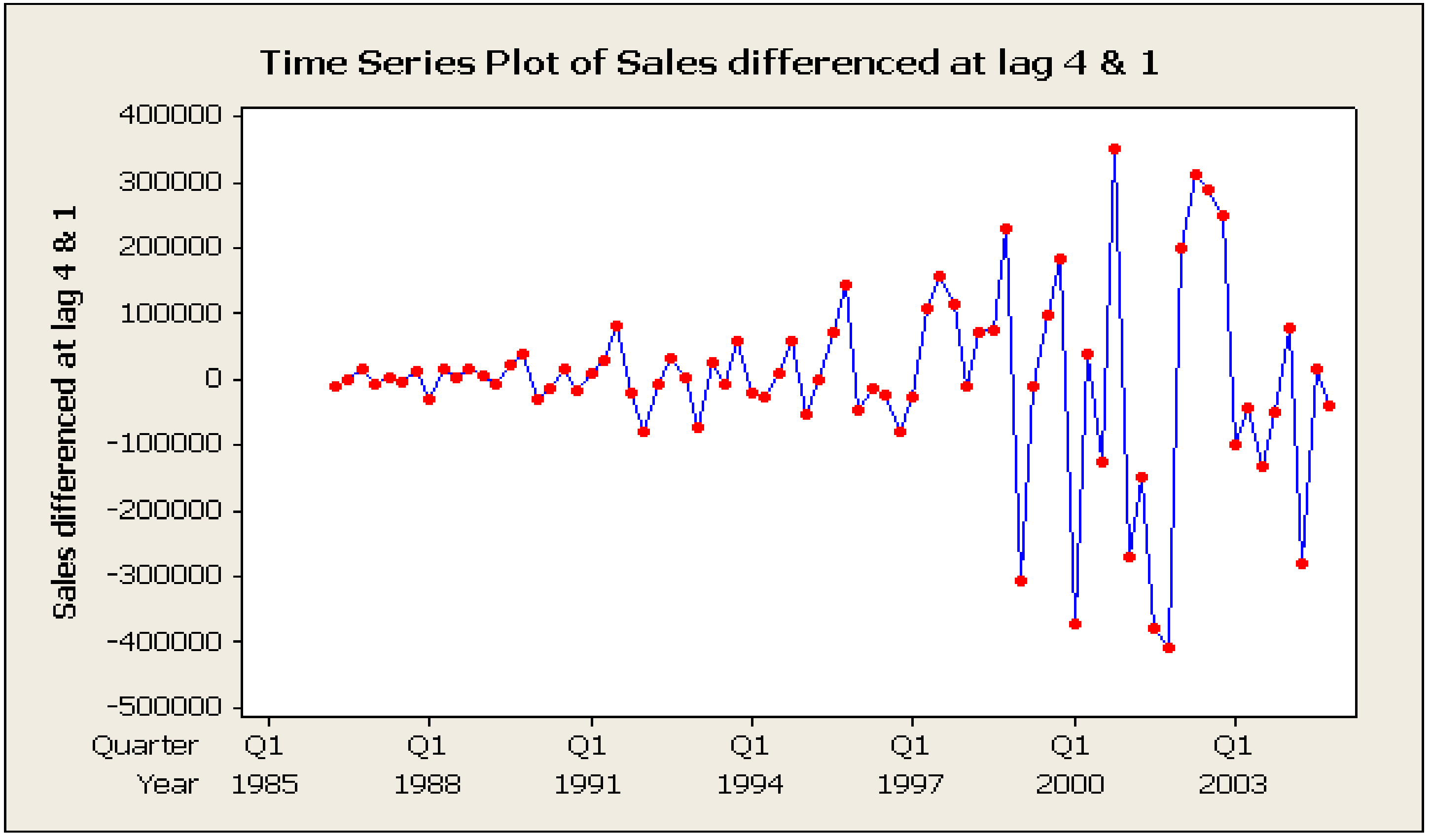
Seasonal differencing

- It is recommended to do the seasonal differencing first since sometimes the resulting series will be stationary and hence no need for a further first difference.
- When differencing is used, it is important that the differences be interpretable.

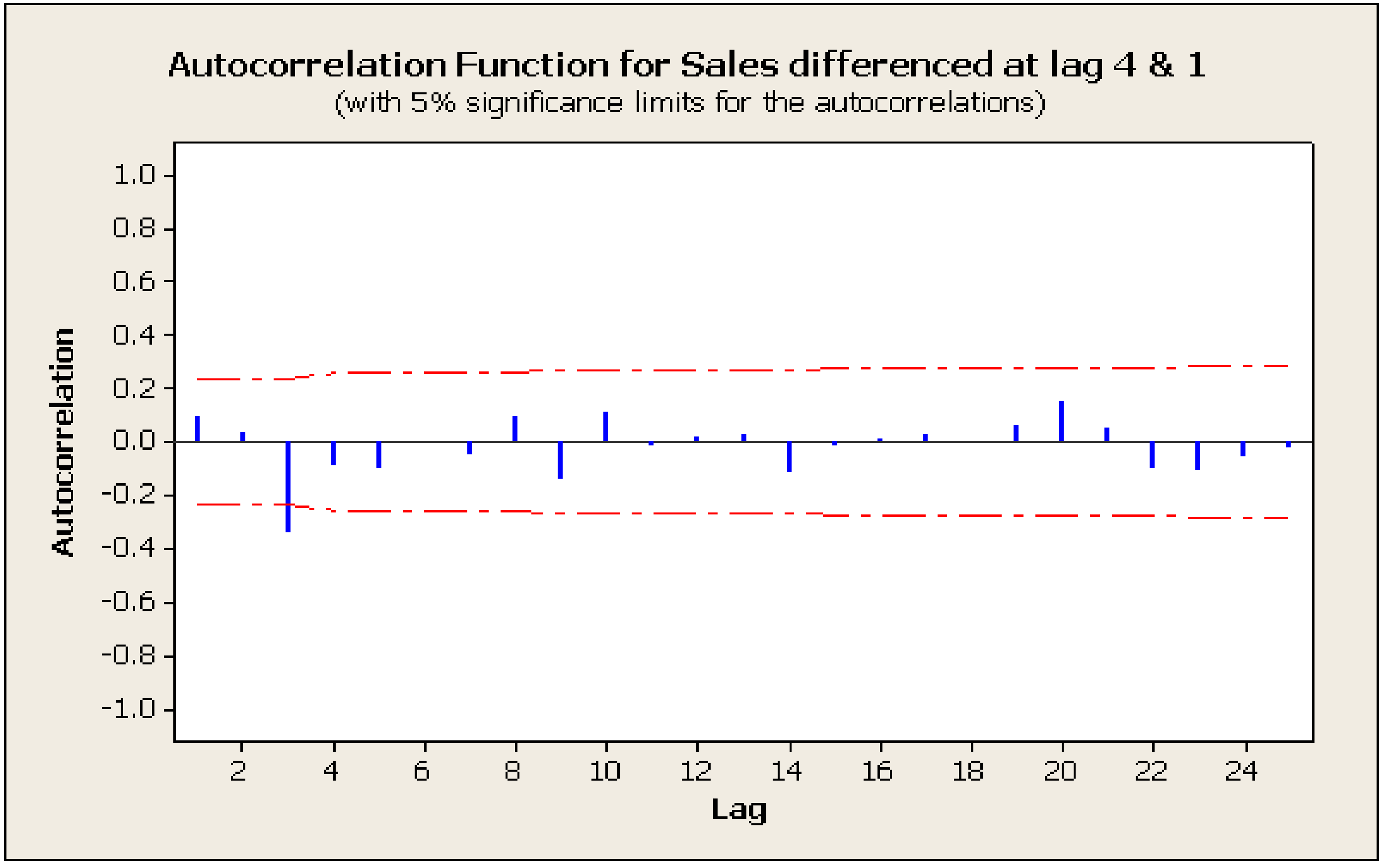
Seasonal differencing

- The series resulted from first difference of seasonally differenced Gap's quarterly sales data is reported in the following three slides.
- Is the resulting series white noise?

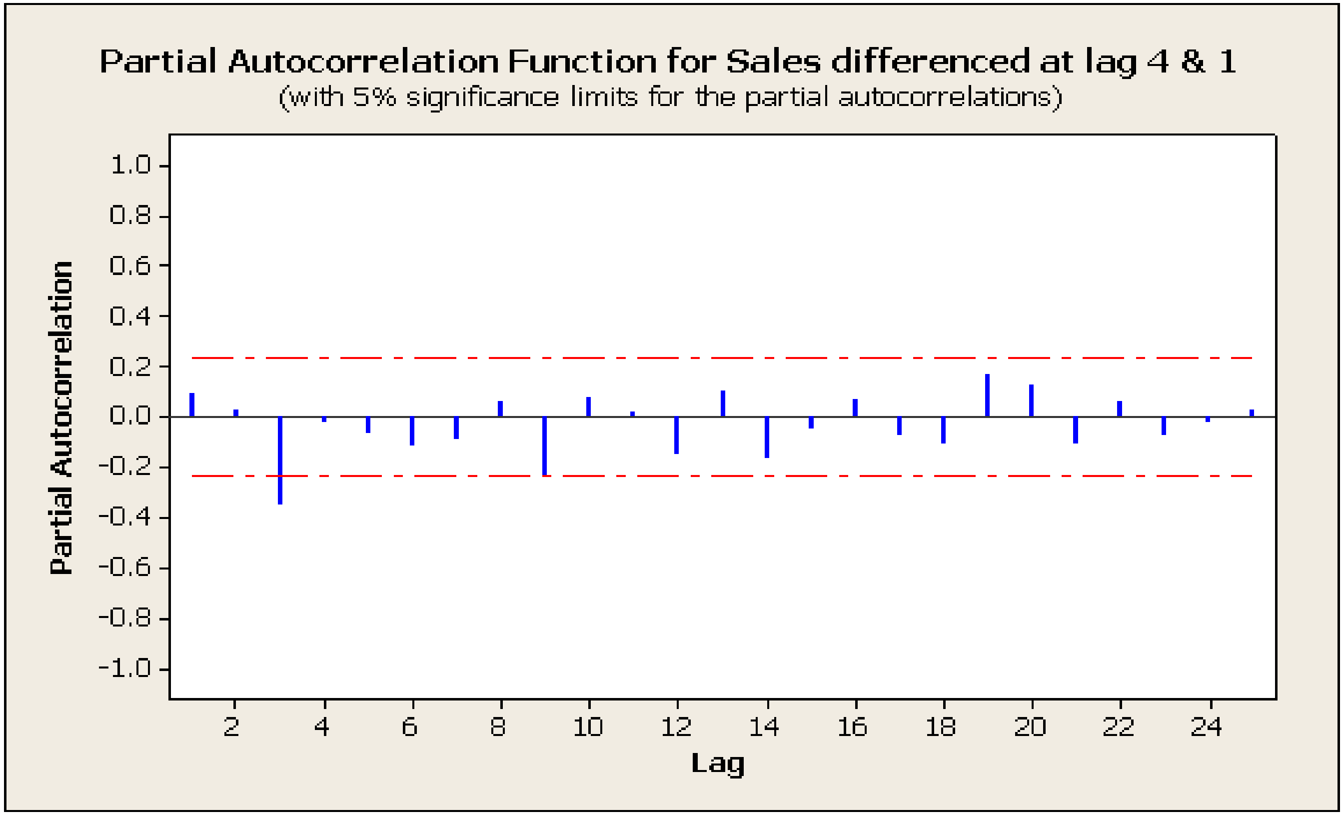
Seasonal differencing



Seasonal differencing



Seasonal differencing



Tests for stationarity

- Several statistical tests has been developed to determine if a series is stationary.
- These tests are also known as unit root tests.
- One of the widely used such test is the Dickey-fuller test.

Tests for stationarity

- To carry out the test, fit the regression model

$$y'_t = \phi y_{t-1} + b_1 y'_{t-1} + b_2 y'_{t-2} + \cdots + b_p y'_{t-p}$$

- Where

y'_t represents the differenced series $y_t - y_{t-1}$

The number of lagged terms p, is usually set to 3.

Tests for stationarity

- The value of ϕ is estimated using ordinary least squares.
- If the original series y_t needs differencing, the estimated value of ϕ will be close to zero.
- If y_t is already stationary, the estimated value of ϕ will be negative.

Reference and sourse:

1. Multivariate Time Series Analysis: With R and Financial Applications by Ruey S. Tsay
2. Time Series Analysis by James Douglas Hamilton
3. The Analysis of Time Series: An Introduction with R Chapman & Hall/CRC Texts in Statistical Science)
4. Machine Learning for Time Series Forecasting with Python by Francesca Lazzeri
5. Time Series Analysis for the Social Sciences (Analytical Methods for Social Research) Part of: Analytical Methods for Social Research (14 Books)
6. Introduction to Probability, Statistics, and Random Processes by Hossein Pishro-Nik
7. Introduction to Time Series and Forecasting (Springer Texts in Statistics) Part of: Springer Texts in Statistics