Time series analysis

Lecture 6. The unit root problem

Dr. Khamidov Obidjon

Mixtures ARMA models

- Basic elements of AR and MA models can be combined to produce a great variety of models.
- The following is the combination of MA(1) and AR(1) models

$$y_t = C + \phi_1 y_{t-1} + e_t - \theta_1 e_{t-1}$$

- This is model called ARMA(1, 1) or

  ARIMA (1, 0, 1)
- The series is assumed stationary in the mean and in the variance.

Mixtures ARIMA models

- If non-stationarity is added to a mixed ARMA model, then the general ARIMA (p, d, q) is obtained.
- The equation for the simplest ARIMA (1, 1, 1) is given below.

$$y_t = C + \phi_1 y_{t-1} - \phi_1 y_{t-2} + e_t - \theta_1 e_{t-1}$$

Mixtures ARIMA models

- The general ARIMA (p, d, q) model gives a tremendous variety of patterns in the ACF and PACF, so it is not practical to state rules for identifying general ARIMA models.
- In practice, it is seldom necessary to deal with values p, d, or q that are larger than 0, 1, or 2.
- It is remarkable that such a small range of values for p, d, or q can cover such a large range of practical forecasting situations.

Seasonality and ARIMA models

- The ARIMA models can be extended to handle seasonal components of a data series.
- The general shorthand notation is

$$ARIMA\ (p, d, q)(P, D, Q)_s$$

  - Where s is the number of periods per season.

Seasonality and ARIMA models

- The general ARIMA(1,1,1)$(1,1,1)_4$ can be written as

$$y_t = (1+\phi_1)y_{t-1} - \phi_1 y_{t-2} + (1+\Phi_1)y_{t-4} - (1+\phi_1+\Phi_1+\phi_1\Phi_1)y_{t-6}$$
$$-\Phi_1 y_{t-8} + (\Phi_1+\phi_1\Phi_1)y_{t-9} - \phi_1\Phi_1 y_{t-10} + e_t - \theta_1 e_{t-1} - \Theta_1 e_{t-4} + \theta_1\Theta_1 e_{t-5}$$

- Once the coefficients ϕ_1, Φ_1, θ_1, and Θ_1 have been estimated from the data, the above equation can be used for forecasting.

Seasonality and ARIMA models

- The seasonal lags of the ACF and PACF plots show the seasonal parts of an AR or MA model.
- Examples:
  - Seasonal MA model:
    - $ARIMA(0,0,0)(0,0,1)_{12}$
      - will show a spike at lag 12 in the ACF but no other significant spikes.
      - The PACF will show exponential decay in the seasonal lags i.e. at lags 12, 24, 36,…

Seasonality and ARIMA models

Seasonal AR model:

- $ARIMA(0,0,0)(1,0,0)_{12}$
  - will show exponential decay in seasonal lags of the ACF.
  - Single significant spike at lag 12 in the PACF.

Implementing the model –Building Strategy

- The Box –Jenkins approach uses an iterative model-building strategy that consist of
  - Selecting an initial model (model identification)
  - Estimating the model coefficients (parameter estimation)
  - Analyzing the residuals (model checking)

Implementing the model –Building Strategy

- If necessary, the initial model is modified and the process is repeated until the residual indicate no further modification is necessary. At this point the fitted model can be used for forecasting.

Model identification

- The following approach outlines an approach to select an appropriate model among a large variety of ARIMA models possible.
  - Plot the data
    - Identify any unusual observations
    - If necessary, transform the dat to stabilize the variance

Model identification

- Check the time series plot, ACF, PACF of the data (possibly transformed) for stationarity.
- IF
  - Time plot shows the data scattered horizontally around a constant mean
  - ACF and PACF to or near zero quickly
- Then, the data are stationary.

Model identification

- Use differencing to transform the data into a stationary series
  - For no-seasonal data take first differences
  - For seasonal data take seasonal differences
- Check the plots again if they appear non-stationary, take the differences of the differenced data.

Model identification

- When the stationarity has been achieved, check the ACF and PACF plots for any pattern remaining.
- There are three possibilities
  - AR or MA models
    - No significant ACF after time lag q indicates MA(q) may be appropriate.
    - No significant PACF after time lag p indicates that AR(p) may be appropriate.

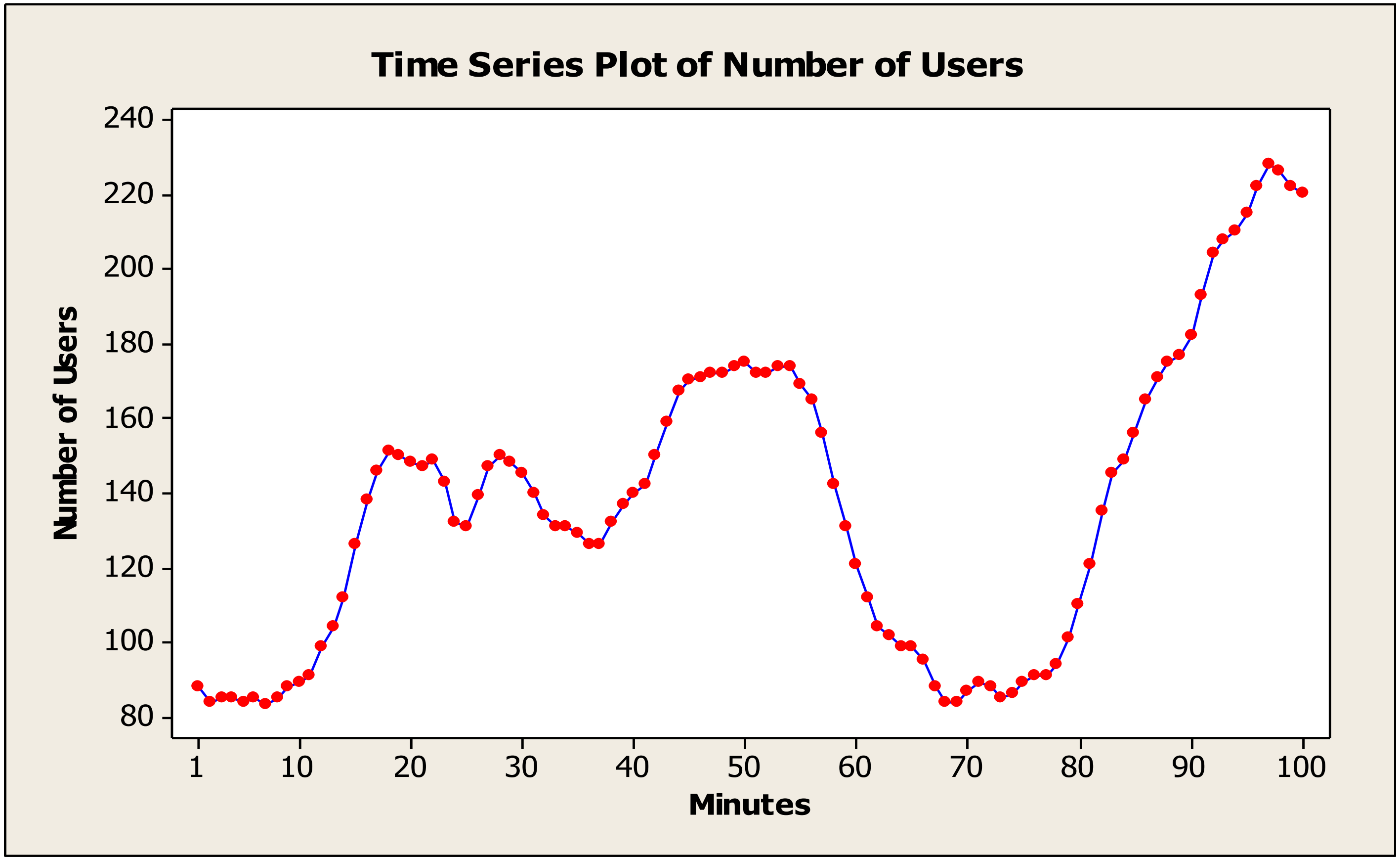
Model identification

- Seasonality is present if ACF and/or PACF at the seasonal lags are large and significant.
- If no clear MA or AR model is suggested, a mixture model may be appropriate.

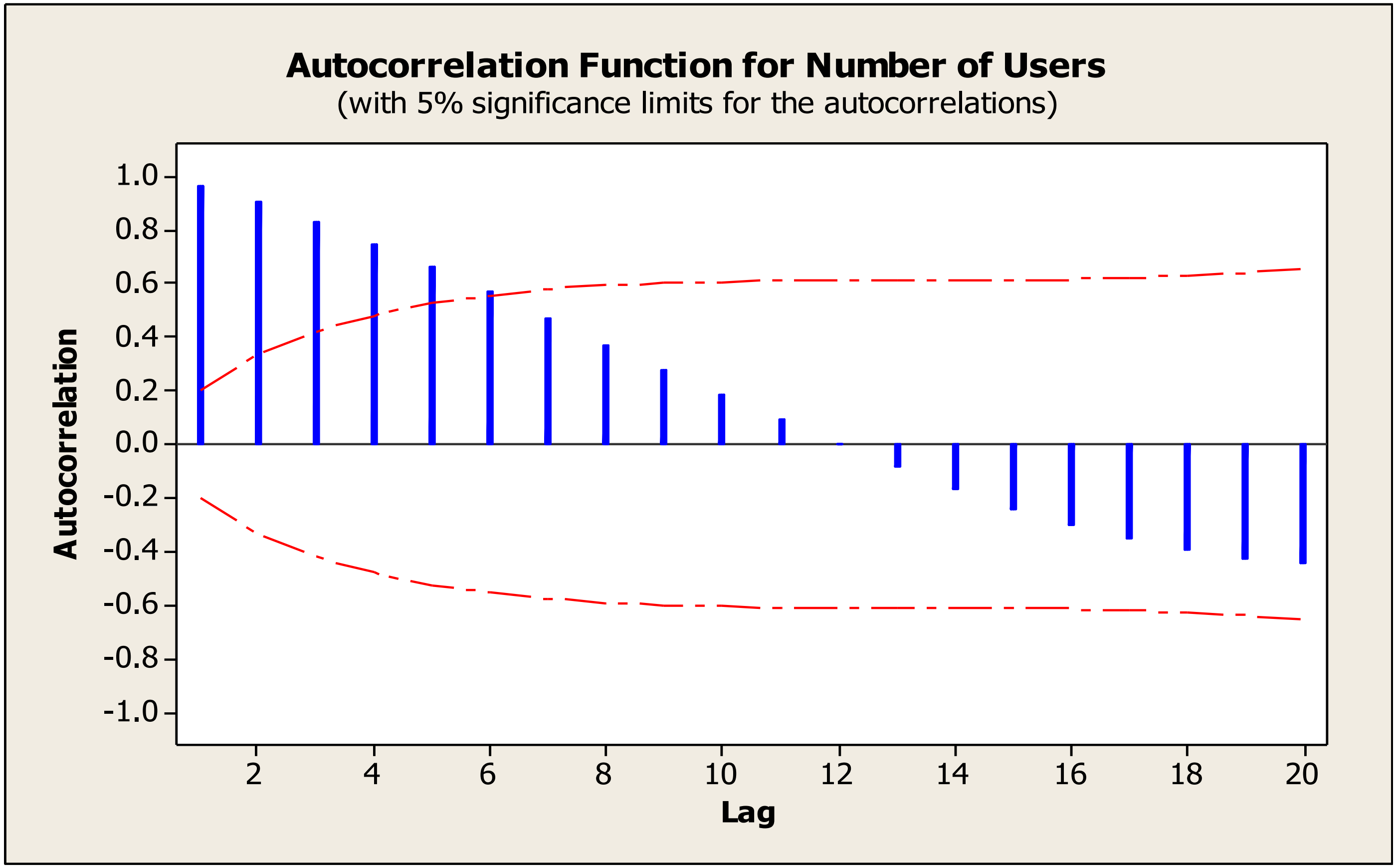
Model identification

- Example
  - Non seasonal time series data.
    - The following example looks at the number of users logged onto an internet server over a 100 minutes period.
    - The time plot, ACF and PACF is reported in the following three slides.

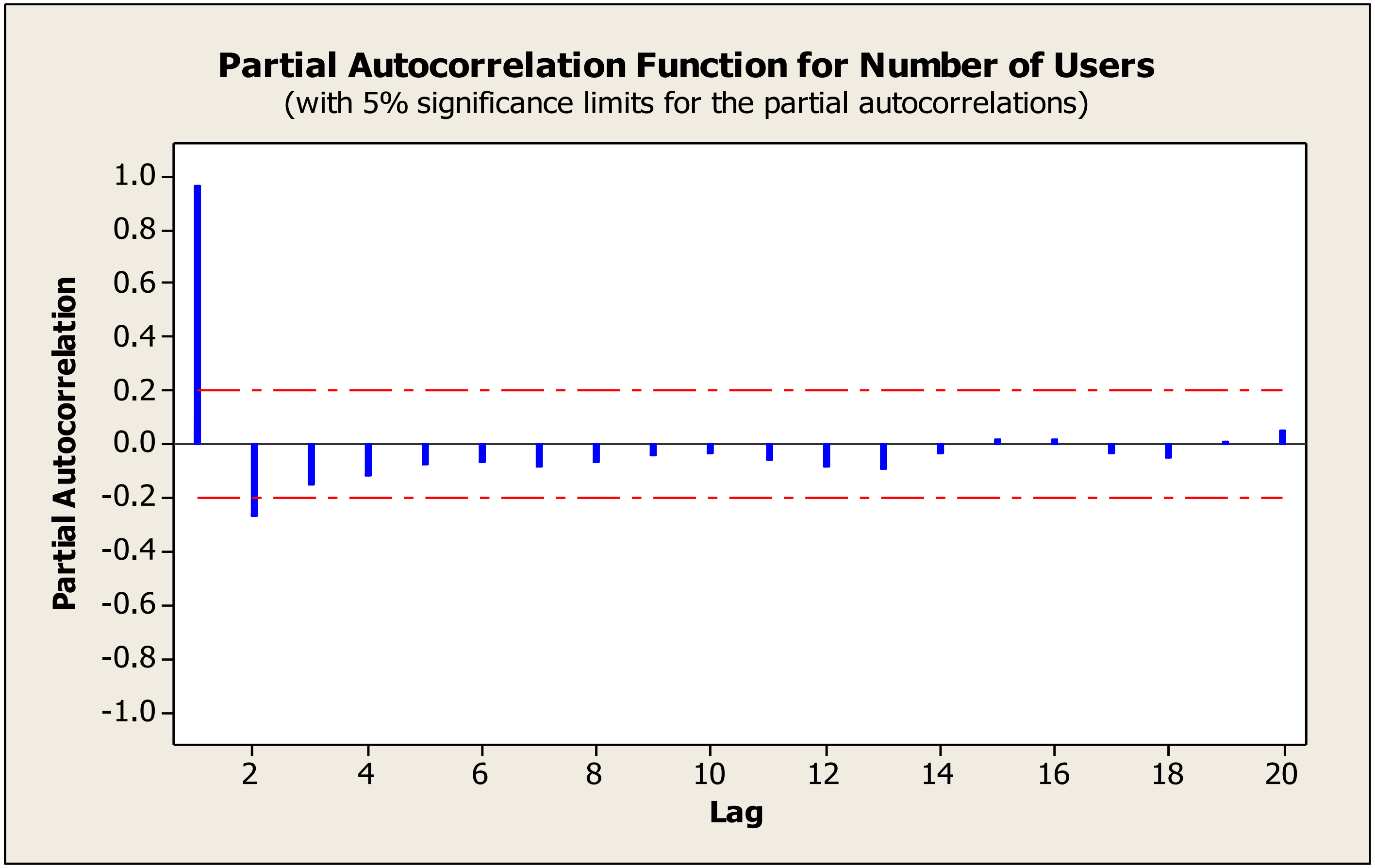
Model identification



Model identification



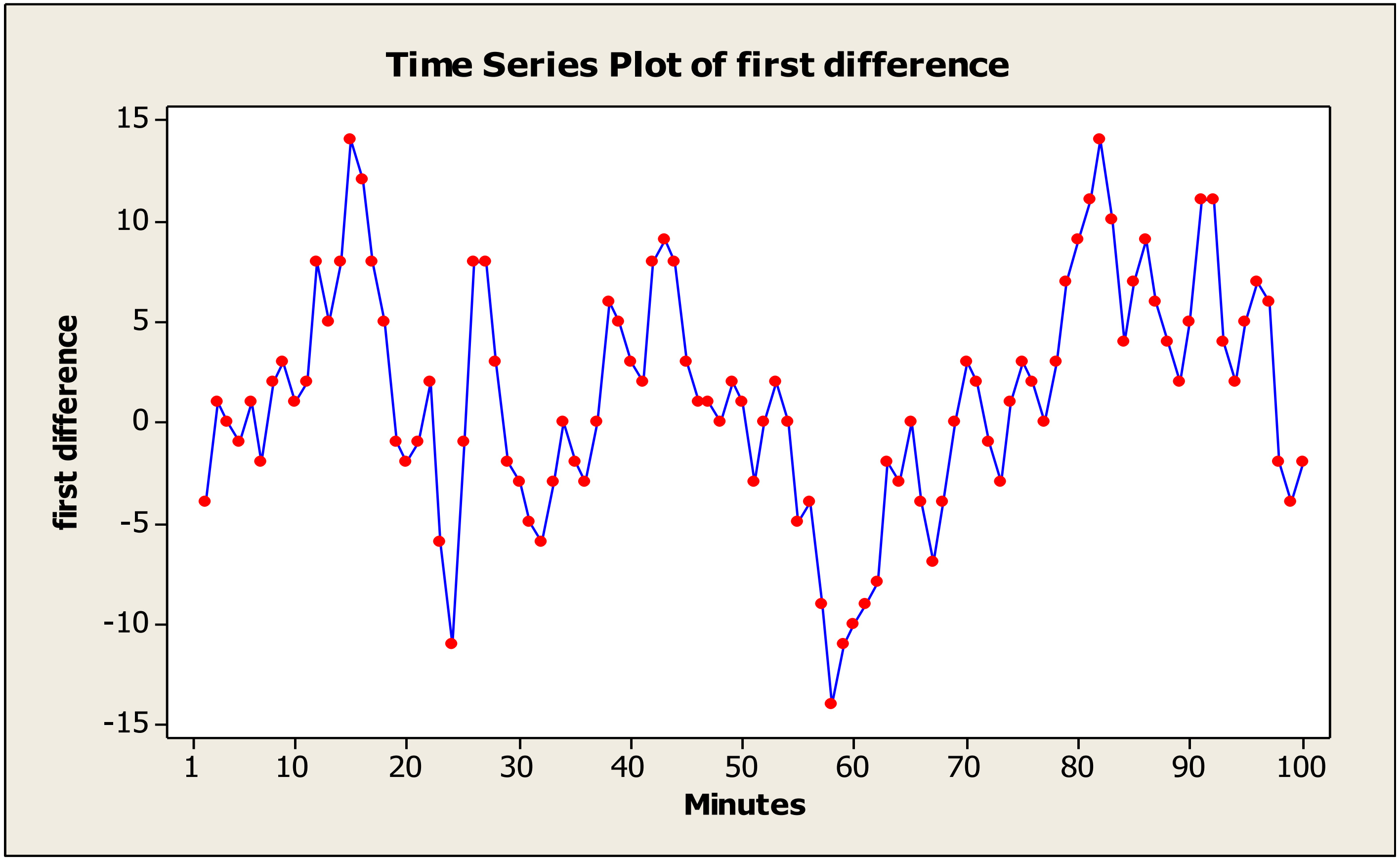
Model identification



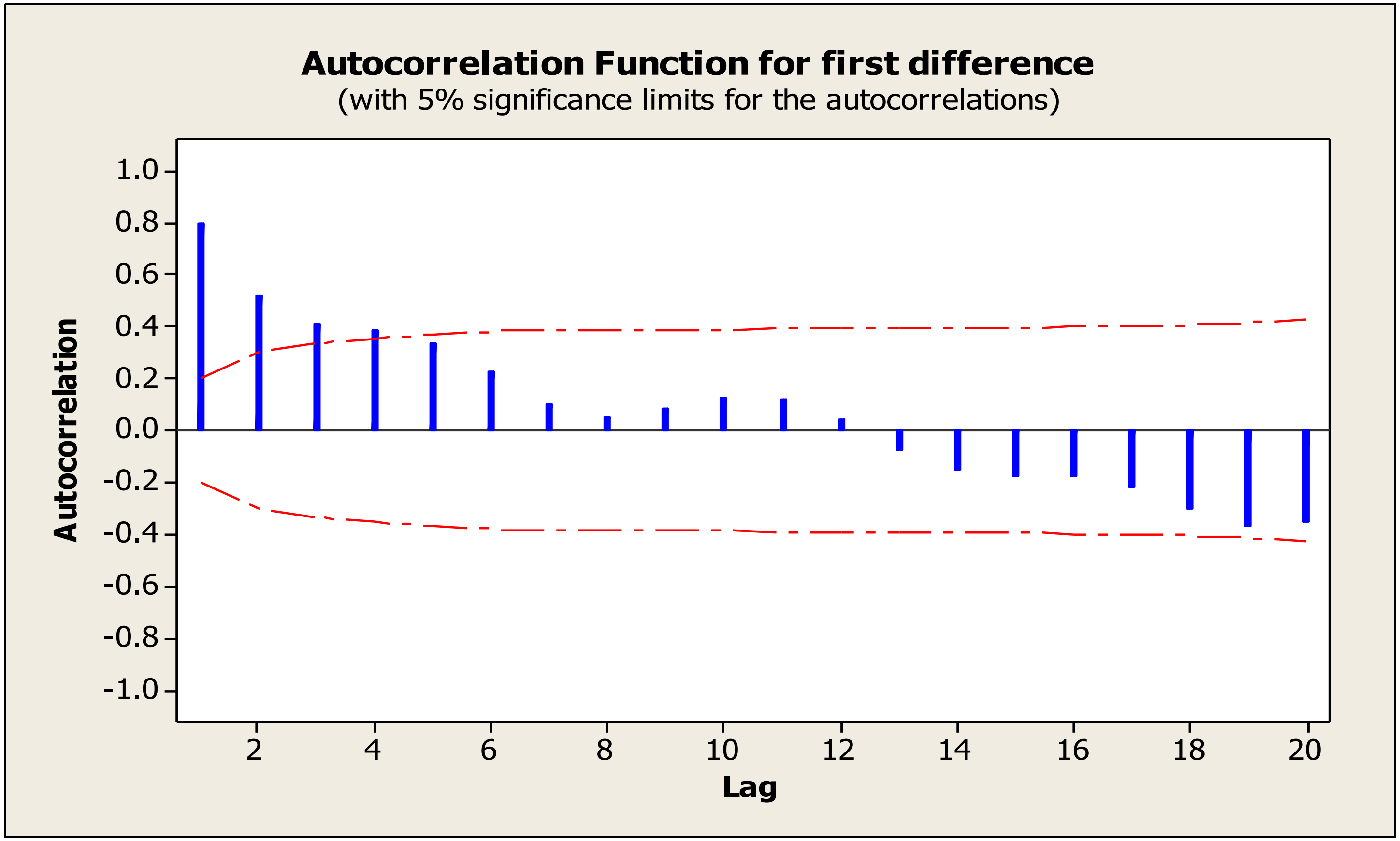
Model identification

- The gradual decline of ACF values indicates non-stationary series.
- The first partial autocorrelation is very dominant and close to 1, indicating non-stationarity.
- The time series plot clearly indicates non-stationarity.
- We take the first differences of the data and reanalyze.

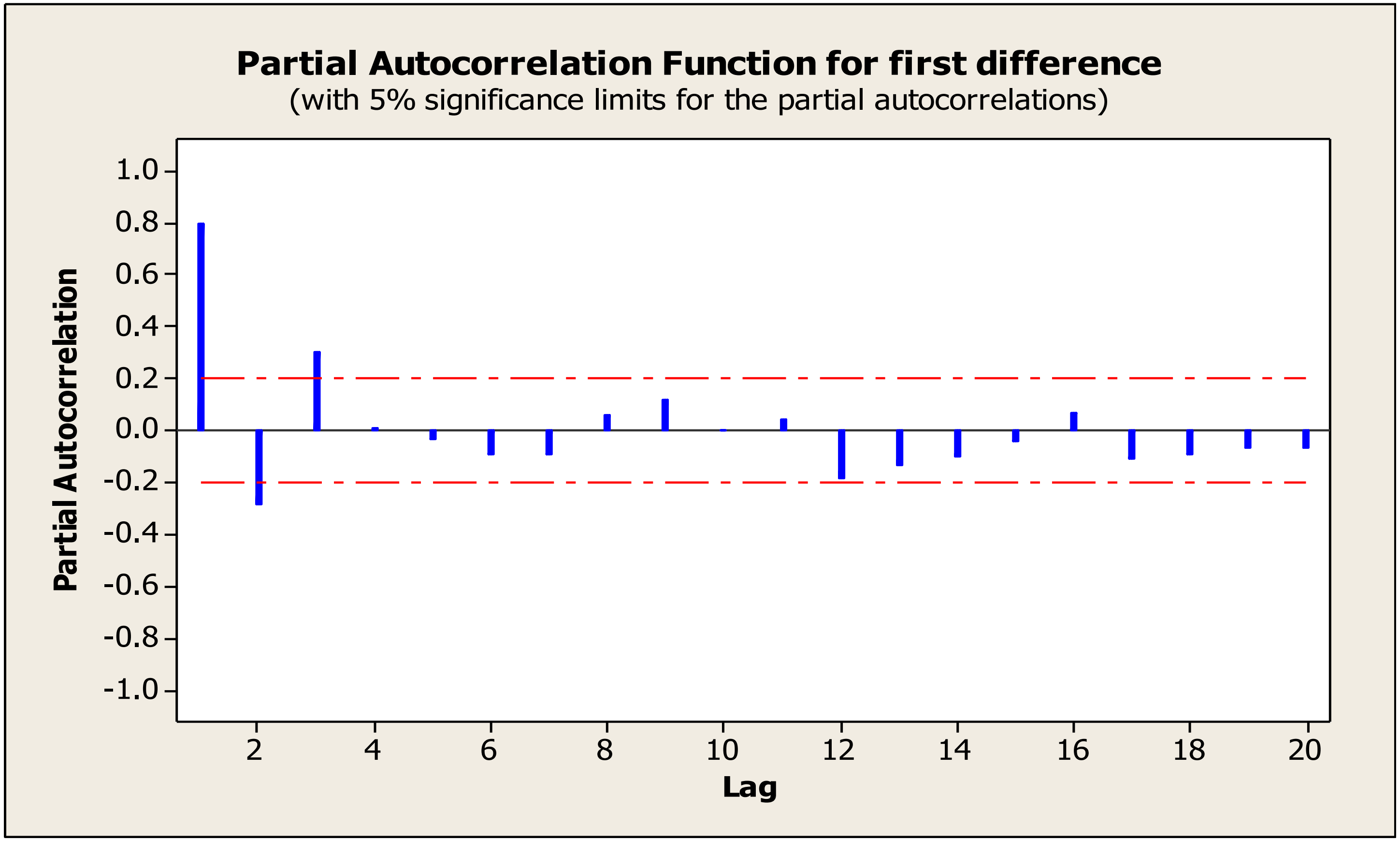
Model identification



Model identification



Model identification



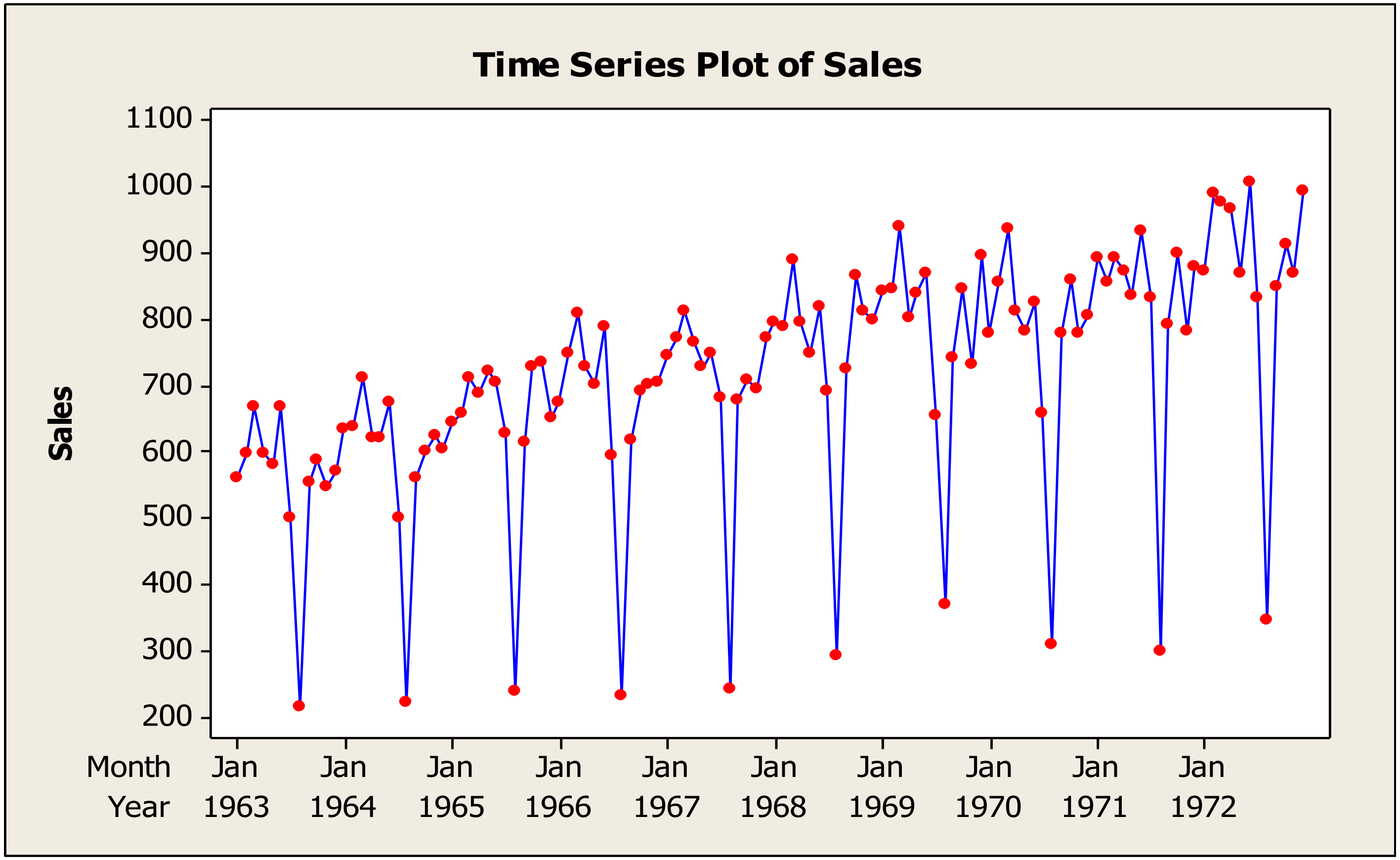
Model identification

- ACF shows a mixture of exponential decay and sine-wave pattern
- PACF shows three significant PACF values.
- This suggests an AR(3) model.
- This identifies an ARIMA(3,1,0).

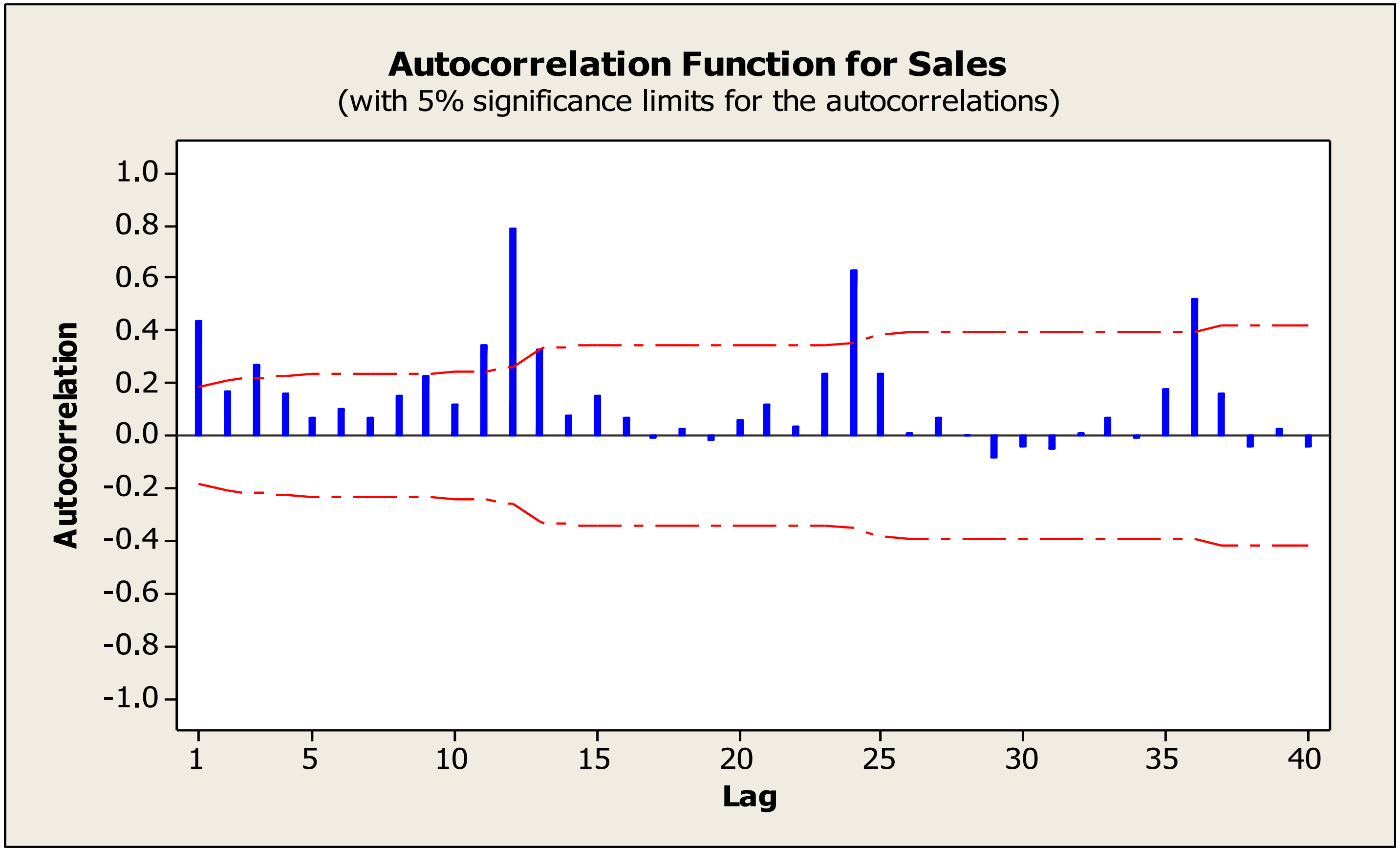
Model identification

- Example
  - A seasonal time series.
    - The following example looks at the monthly industry sales (in thousands of francs) for printing and writing papers between the years 1963 and 1972.
    - The time plot, ACF and PACF shows a clear seasonal pattern in the data.
    - This is clear in the large values at time lag 12, 24 and 36.

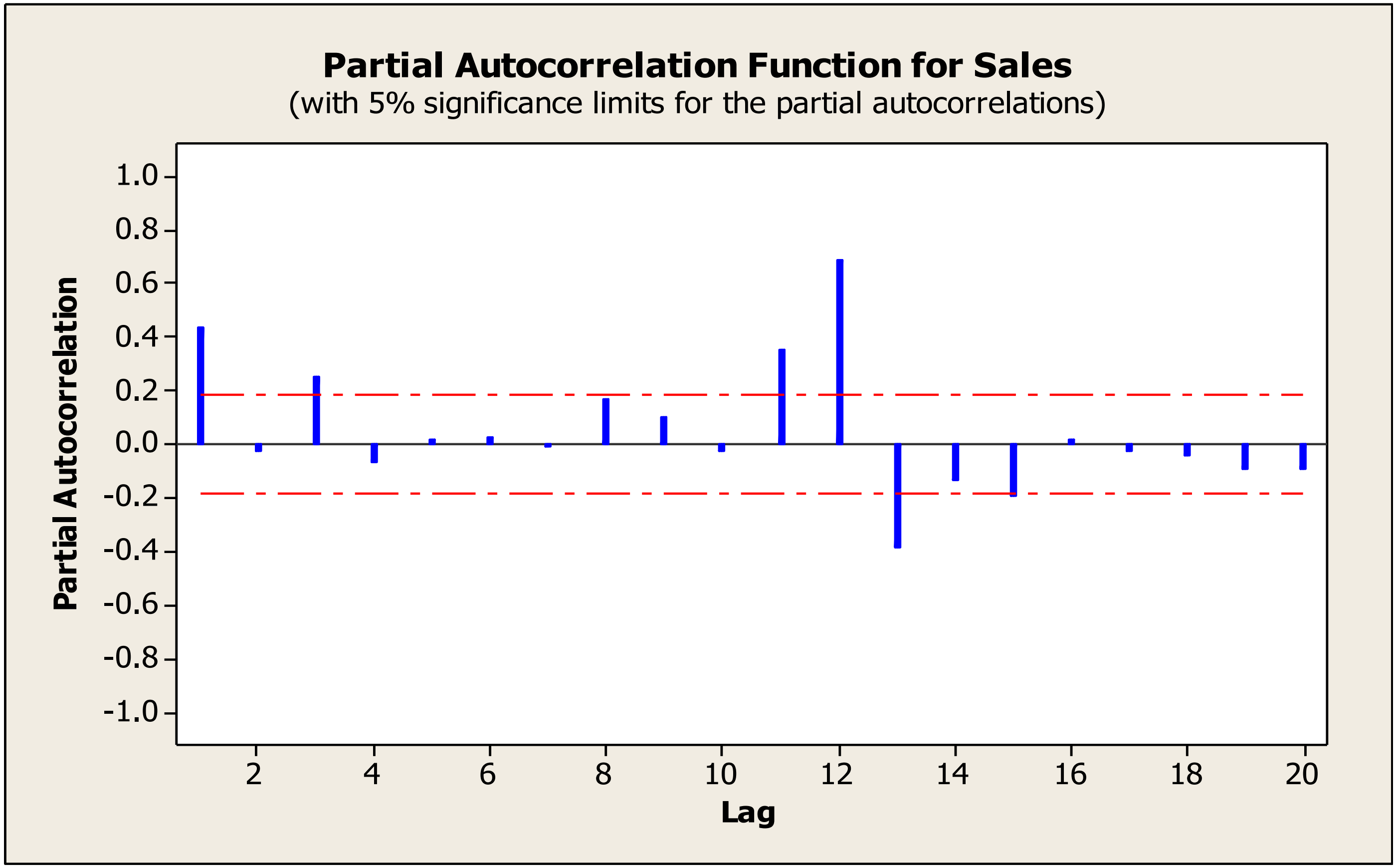
Model identification



Model identification



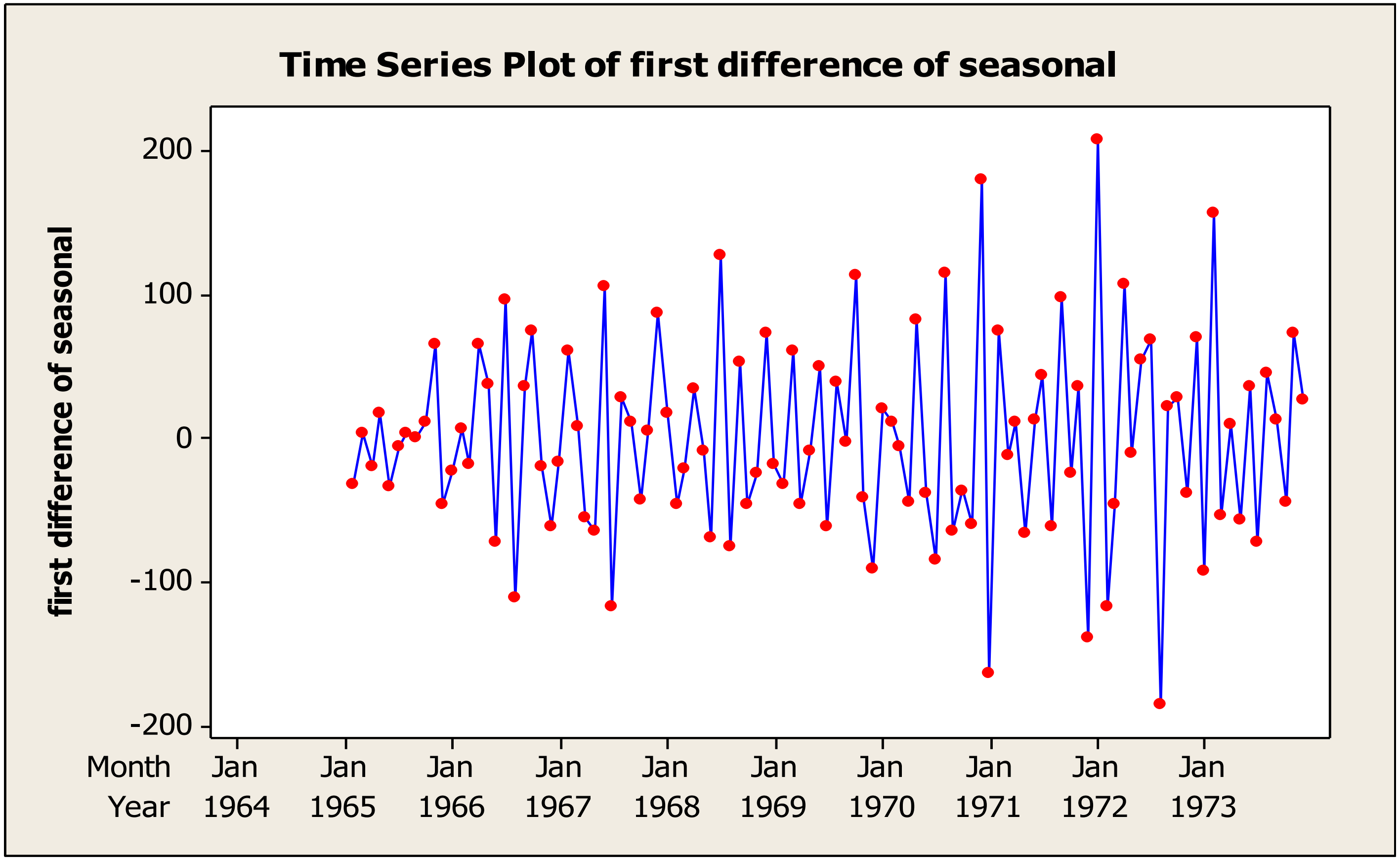
Model identification



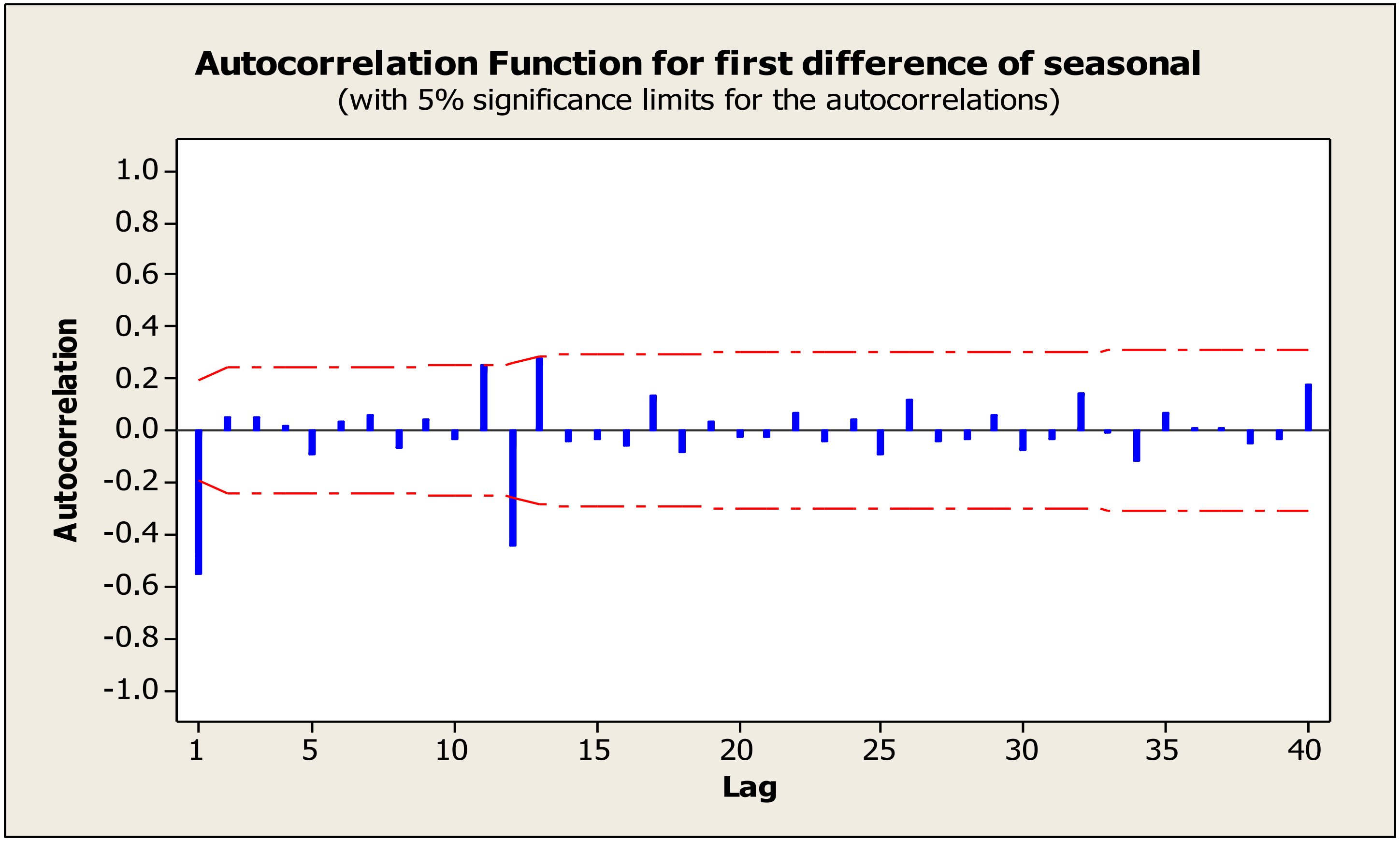
Model identification

- We take a seasonal difference and check the time plot, ACF and PACF.
- The seasonally differenced data appears to be non-stationary (the plots are not shown), so we difference the data again.
- the following three slides show the twice differenced series.

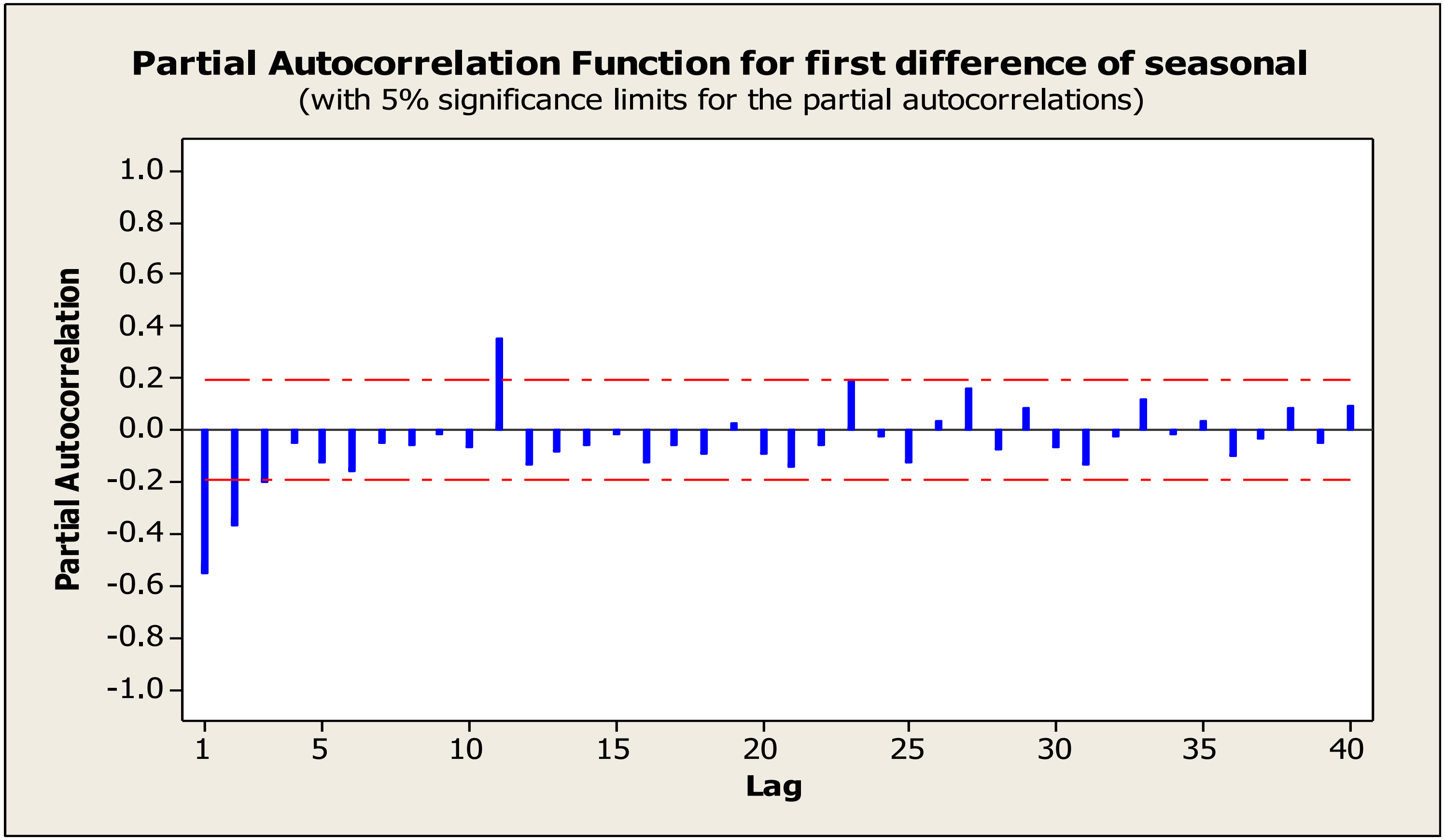
Model identification



Model identification



Model identification



Model identification

- The PACF shows the exponential decay in values.
- The ACF shows a significant value at time lag 1.
  - This suggest a MA(1) model.
- The ACF also shows a significant value at time lag 12
  - This suggest a seasonal MA(1).

Model identification

- Therefore, the identifies model is

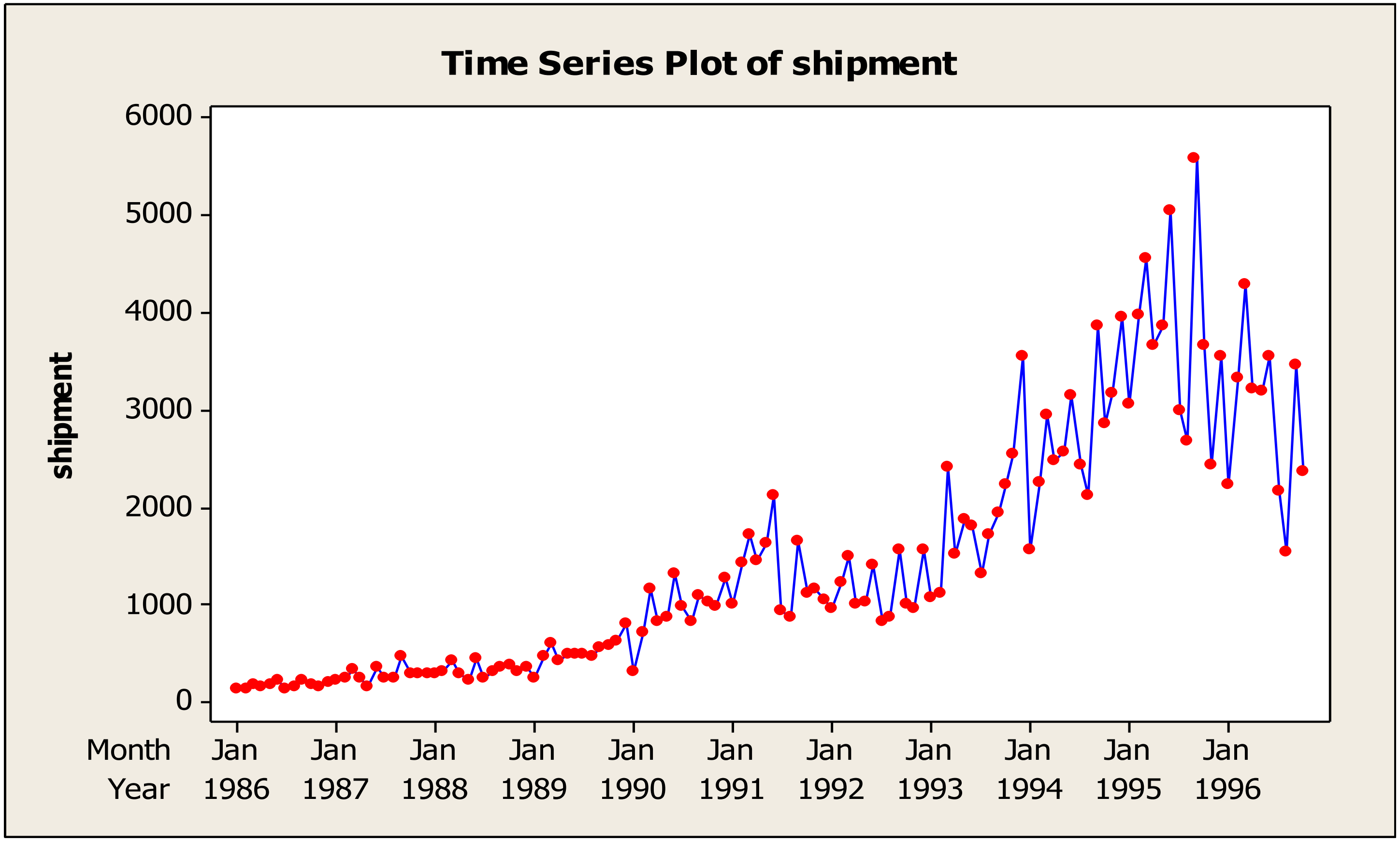
  $$\text{ARIMA }(0,1,1)(0,1,1)_{12}.$$

- This model is sometimes is called the "airline model" because it was applied to international airline data by Box and Jenkins.
- It is one of the most commonly used seasonal ARIMA model.

Model identification

- Example 3
  - A seasonal data needing transformation
    - In this example we look at the monthly shipments of a company that manufactures pollution equipments
    - The time plot shows that the variability increases as the time increases. This indicate that the data is non-stationary in the variance.

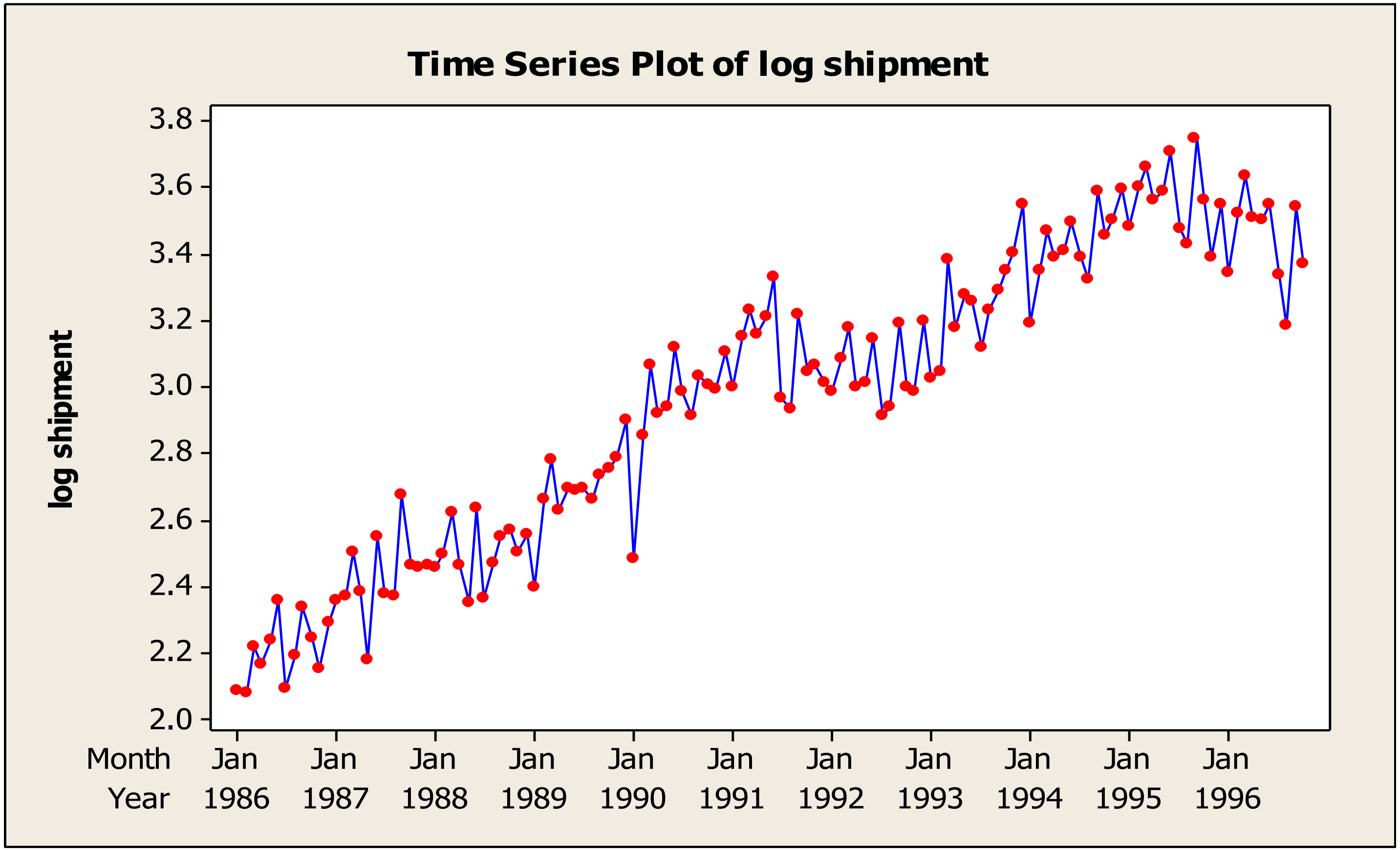
Model identification



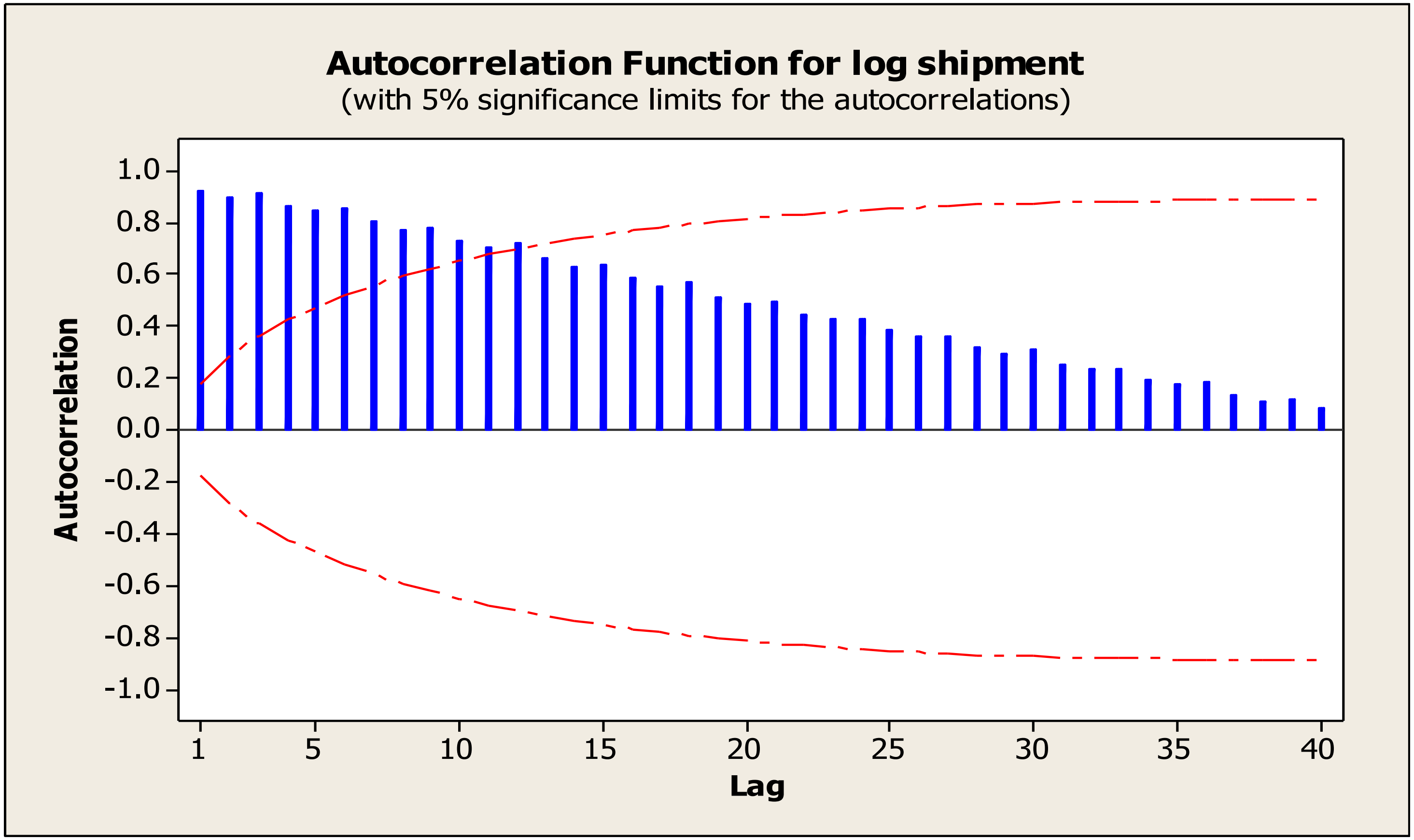
Model identification

- We need to stabilize the variance before fitting an ARIMA model.
- Logarithmic or power transformation of the data will make the variance stationary.
- The time plot, ACF and PACF for the logged data is reported in the following three slides.

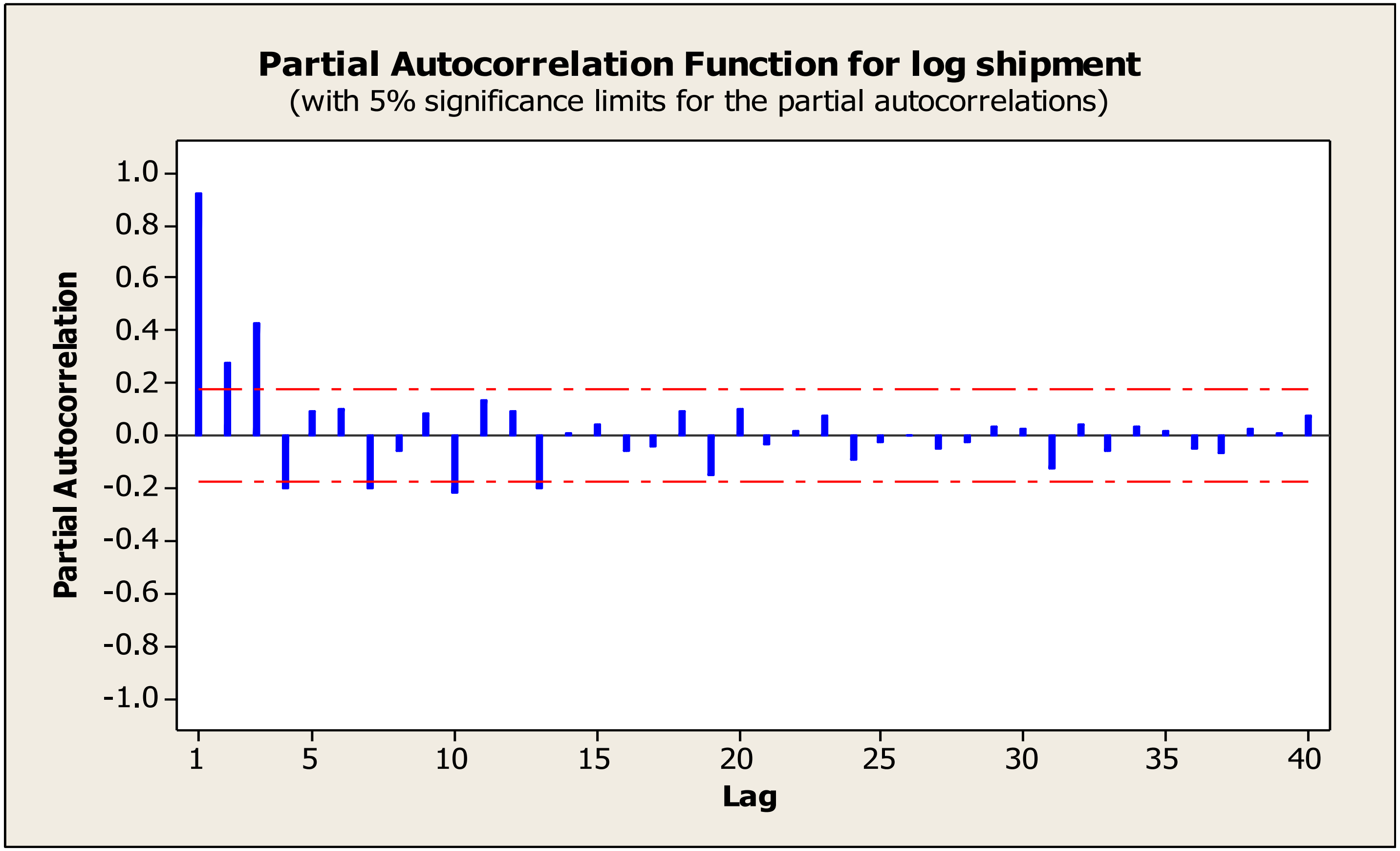
Model identification



Model identification



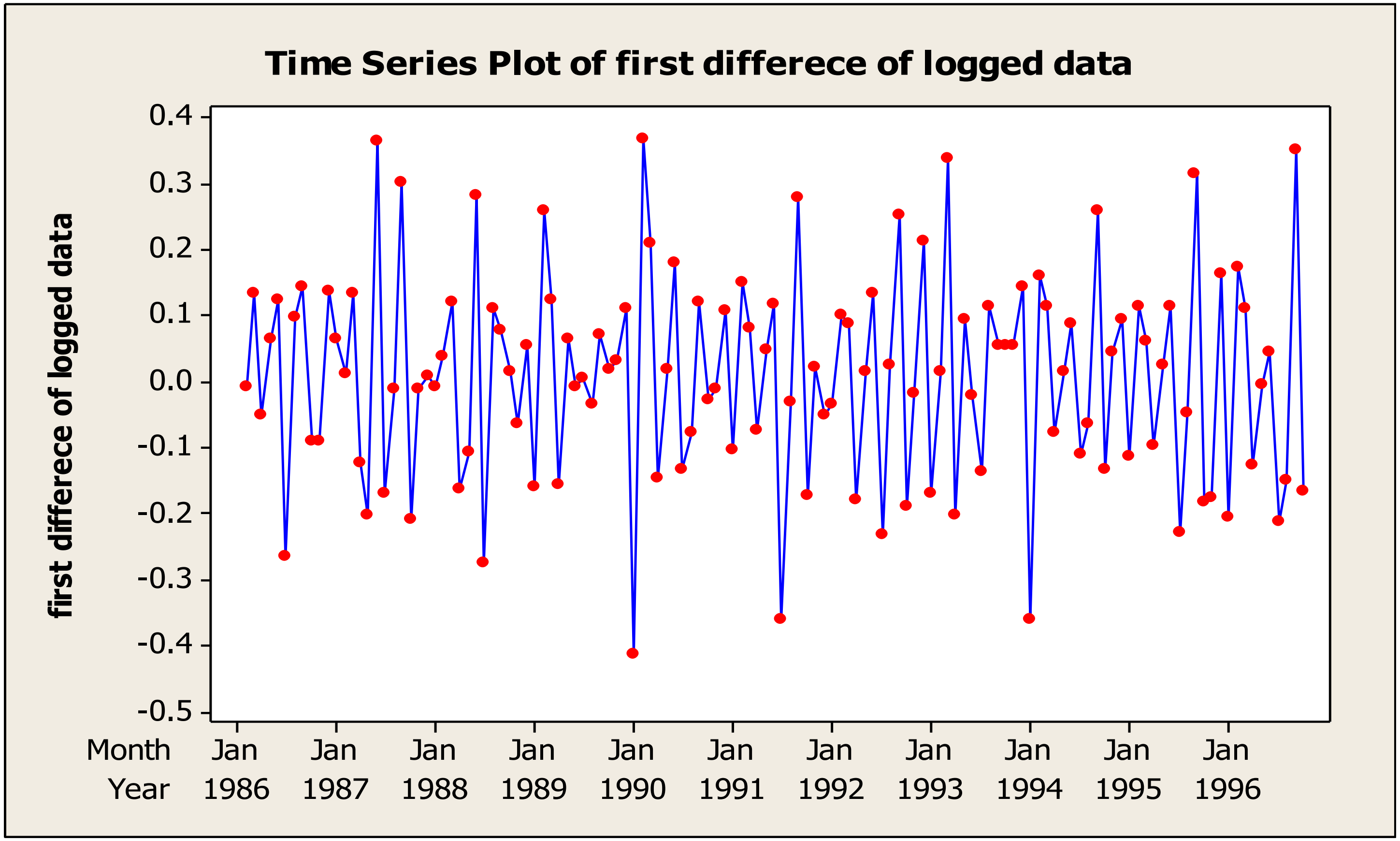
Model identification



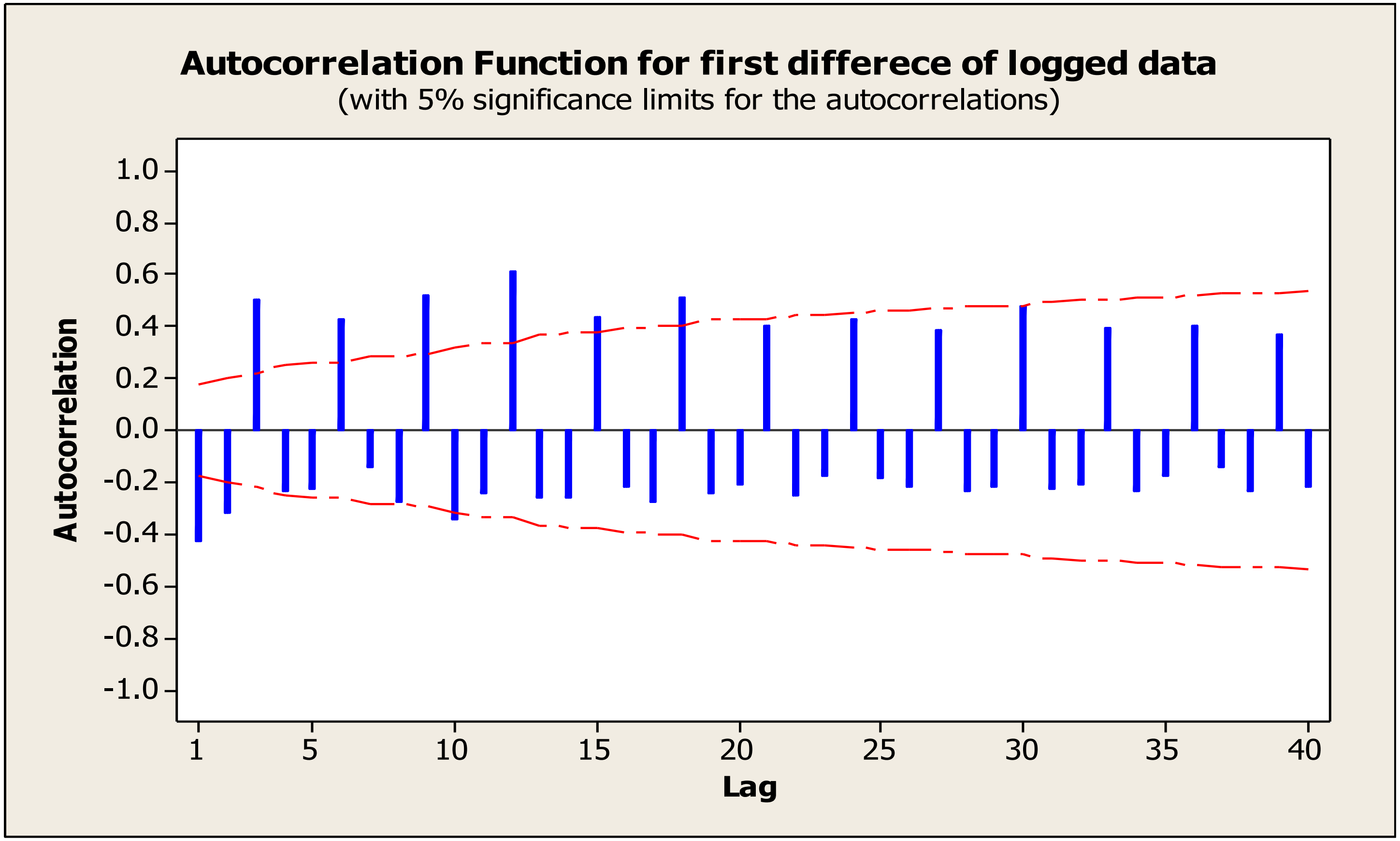
Model identification

- The time plot shows that the magnitude of the fluctuations in the log-transformed data does not vary with time.
- But, the logged data are clearly non-stationary.
  - The gradual decay of the ACF values.
- To achieve stationarity, we take the first differences of the logged data.
- The plots are reported in the next three slides.

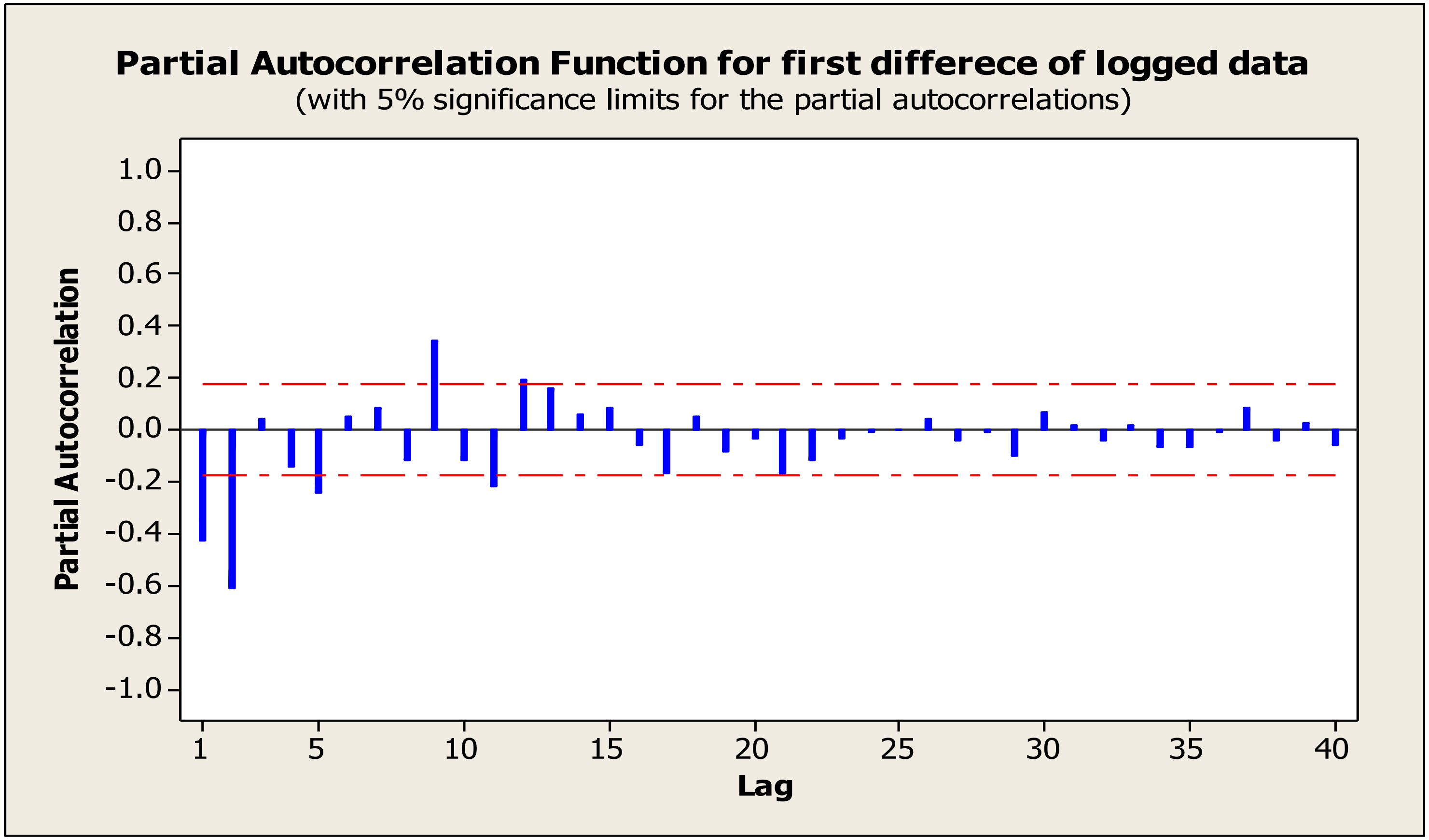
Model identification



Model identification



Model identification



Model identification

- There are significant spikes at time lag 1 and 2 in the PACF, indicating an AR(2) might be appropriate.
- The single significant spike at lag 12 of the PACF indicates a seasonal AR(1) component.
- Therefore for the logged data a tentative model would be

$$ARIMA(2,1,0)(1,0,0)_{12}$$

Summary

- The process of identifying an ARIMA model requires experience and good judgment.The following guidelines can be helpful.
  - Make the series stationary in mean and variance
    - Differencing will take care of non-stationarity in the mean.
    - Logarithmic or power transformation will often take care of non-stationarity in the variance.

Summary

– Consider non-seasonal aspect
  - The ACF and PACF of the stationary data obtained from the previous step can reveal whether MA of AR is feasible.
    – Exponential decay or damped sine-wave. For ACF, and spikes at lags 1 to p then cut off to zero, indicate an AR(P) model.
    – Spikes at lag1 to q, then cut off to zero for ACF and exponential decay or damped sine-wave for PACF indicates MA(q) model.

Summary

- Consider seasonal aspect
  - Examination of ACF and PACF at the seasonal lags can help to identify AR and MA models for the seasonal aspect of the data.
    - For example, for quarterly data the pattern of r_4, r_8, r_{12}, r_{16}, and so on.

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