Time series analysis

Lecture 3. Coefficient estimation in ARMA (p,q) processes. Box-Jenkins' approach

Dr. Khamidov Obidjon

Introduction

- Autoregressive Integrated Moving Average models (ARIMA models) were popularized by George Box and Gwilym Jenkins in the early 1970s.
- ARIMA models are a class of linear models that is capable of representing stationary as well as non-stationary time series.
- ARIMA models do not involve independent variables in their construction. They make use of the information in the series itself to generate forecasts.

Introduction

- ARIMA models rely heavily on autocorrelation patterns in the data.
- ARIMA methodology of forecasting is different from most methods because it does not assume any particular pattern in the historical data of the series to be forecast.
- It uses an interactive approach of identifying a possible model from a general class of models. The chosen model is then checked against the historical data to see if it accurately describe the series.

Introduction

- Recall that, a time series data is a sequence of numerical observations naturally ordered in time
  - Daily closing price of IBM stock
  - Weekly automobile production by the Pontiac division of general Motors.
  - Hourly temperatures at the entrance to Grand central Station.

Introduction

- Two question of paramount importance When a forecaster examines a time series data are:
  - Do the data exhibit a discernible pattern?
  - Can this be exploited to make meaningful forecasts?

Introduction

- The Box-Jenkins methodology refers to a set of procedures for identifying, fitting, and checking ARIMA models with time series data.Forecasts follow directly from the form of fitted model.
- The basis of BOX-Jenkins approach to modeling time series consists of three phases:
  - Identification
  - Estimation and testing
  - Application

Introduction

- Identification
  - Data preparation
    - Transform data to stabilize variance
    - Differencing data to obtain stationary series
  - Model selection
    - Examine data, ACF and PACF to identify potential models

Introduction

- Estimation and testing
  - Estimation
    - Estimate parameters in potential models
    - Select best model using suitable criterion
  - Diagnostics
    - Check ACF/PACF of residuals
    - Do portmanteau test of residuals
    - Are the residuals white noise?

Introduction

- Application
  - Forecasting: use model to forecast

Examining correlation in time series data

- The key statistic in time series analysis is the autocorrelation coefficient (the correlation of the time series with itself, lagged 1, 2, or more periods.)
- Recall the autocorrelation formula:

$$r_k = \frac{\sum_{t=k+1}^{n} (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{n} (y_t - \bar{y})^2}$$

Examining Correlation in Time Series Data

- Recall r_1 indicates how successive values of Y relate to each other, r_2 indicates how Y values two periods apart relate to each other, and so on.
- The auto correlations at lag 1, 2, …, make up the autocorrelation function or ACF.
- Autocorrelation function is a valuable tool for investigating properties of an empirical time series.

A white noise model

- A white noise model is a model where observations Y_t is made of two parts: a fixed value and an uncorrelated random error component.

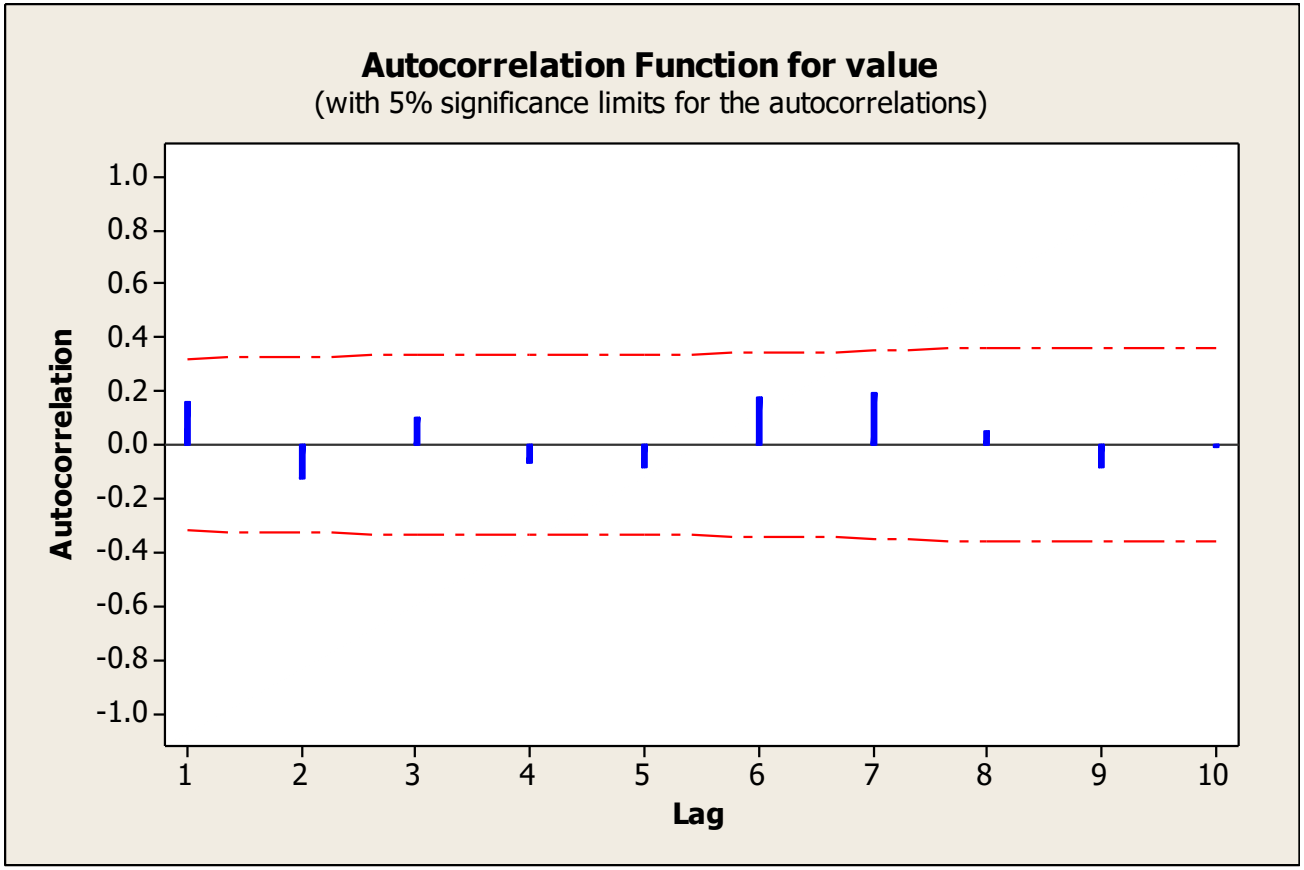
$$y_t = C + e_t$$

- For uncorrelated data (a time series which is white noise) we expect each autocorrelation to be close to zero.
- Consider the following white noise series.

White noise series

period	value	period	value
1	23	21	50
2	36	22	86
3	99	23	90
4	36	24	65
5	36	25	20
6	74	26	17
7	30	27	45
8	54	28	9
9	59	29	73
10	17	30	33
11	36	31	17
12	89	32	3
13	77	33	29
14	86	34	30
15	33	35	68
16	90	36	87
17	74	37	44
18	7	38	5
19	54	39	26
20	98	40	52

ACF for the white noise series



Sampling distribution of autocorrelation

- The autocorrelation coefficients of white noise data have a sampling distribution that can be approximated by a normal distribution with mean zero and standard error $1/\sqrt{n}$. where n is the number of observations in the series.
- This information can be used to develop tests of hypotheses and confidence intervals for ACF.

Sampling distribution of autocorrelation

- For example
  - For our white noise series example, we expect 95% of all sample ACF to be within

$$\pm 1.96 \frac{1}{\sqrt{n}} = \pm 1.96 \frac{1}{\sqrt{40}} = \pm .3099$$

  - If this is not the case then the series is not white noise.
  - The sampling distribution and standard error allow us to distinguish what is randomness or white noise from what is pattern.

Portmanteau tests

- Instead of studying the ACF value one at a time, we can consider a set of them together, for example the first 10 of them (r_1 through r_{10}) all at one time.
- A common test is the Box-Pierce test which is based on the Box-Pierce Q statistics

$$Q = n\sum_{k=1}^{h} r_k^2$$

  - Usually h ≈ 20 is selected

Portmanteau tests

- This test was originally developed by Box and Pierce for testing the residuals from a forecast model.
- Any good forecast model should have forecast errors which follow a white noise model.
- If the series is white noise then, the Q statistic has a chi-square distribution with (h-m) degrees of freedom, where m is the number of parameters in the model which has been fitted to the data.
- The test can easily be applied to raw data, when no model has been fitted , by setting m = 0.

Example

- Here is the ACF values for the white noise example.

Lag	ACF
1	0.159128
2	-0.12606
3	0.102384
4	-0.06662
5	-0.08255
6	0.176468
7	0.191626
8	0.05393
9	-0.08712
10	-0.01212
11	-0.05472
12	-0.22745
13	0.089477
14	0.017425
15	-0.20049

Example

- The box-Pierce Q statistics for h = 10 is

$$Q = n\sum_{k=1}^{h} r_k^2 = 40[(.159)^2 + (-.126)^2 + \cdots + (-.0121)^2] = 5.66$$

- Since the data is not modeled m =0 therefore df = 10.
- From table C-4 with 10 df, the probability of obtaining a chi-square value as large or larger than 5.66 is greater than 0.1.
- The set of 10 r_k values are not significantly different from zero.

Portmanteau tests

- An alternative portmanteau test is the Ljung-Box test.

$$Q^* = n(n+2)\sum_{k=1}^{h}(n-k)^{-1}r_k^2$$

- Q^* has a Chi-square distribution with (h-m) degrees of freedom.
- In general, the data are not white noise if the values of Q or Q^* is greater than the the value given in a chi square table with α = 5%.

The Partial autocorrelation coefficient

- Partial autocorrelations measures the degree of association between y_t and y_{t-k}, when the effects of other time lags 1, 2, 3, …, k-1 are removed.
- The partial autocorrelation coefficient of order k is evaluated by regressing y_t against $y_{t-1},\ldots y_{t-k}$:

$$y_t = b_0 + b_1 y_{t-1} + b_2 y_{t-2} + \cdots + b_k y_{t-k}$$

  - α_k (partial autocorrelation coefficient of order k) is the estimated coefficient b_k.

The Partial autocorrelation coefficient

- The partial autocorrelation functions (PACF) should all be close to zero for a white noise series.
- If the time series is white noise, the estimated PACF are approximately independent and normally distributed with a standard error $1/\sqrt{n}$.
- Therefore the same critical values of

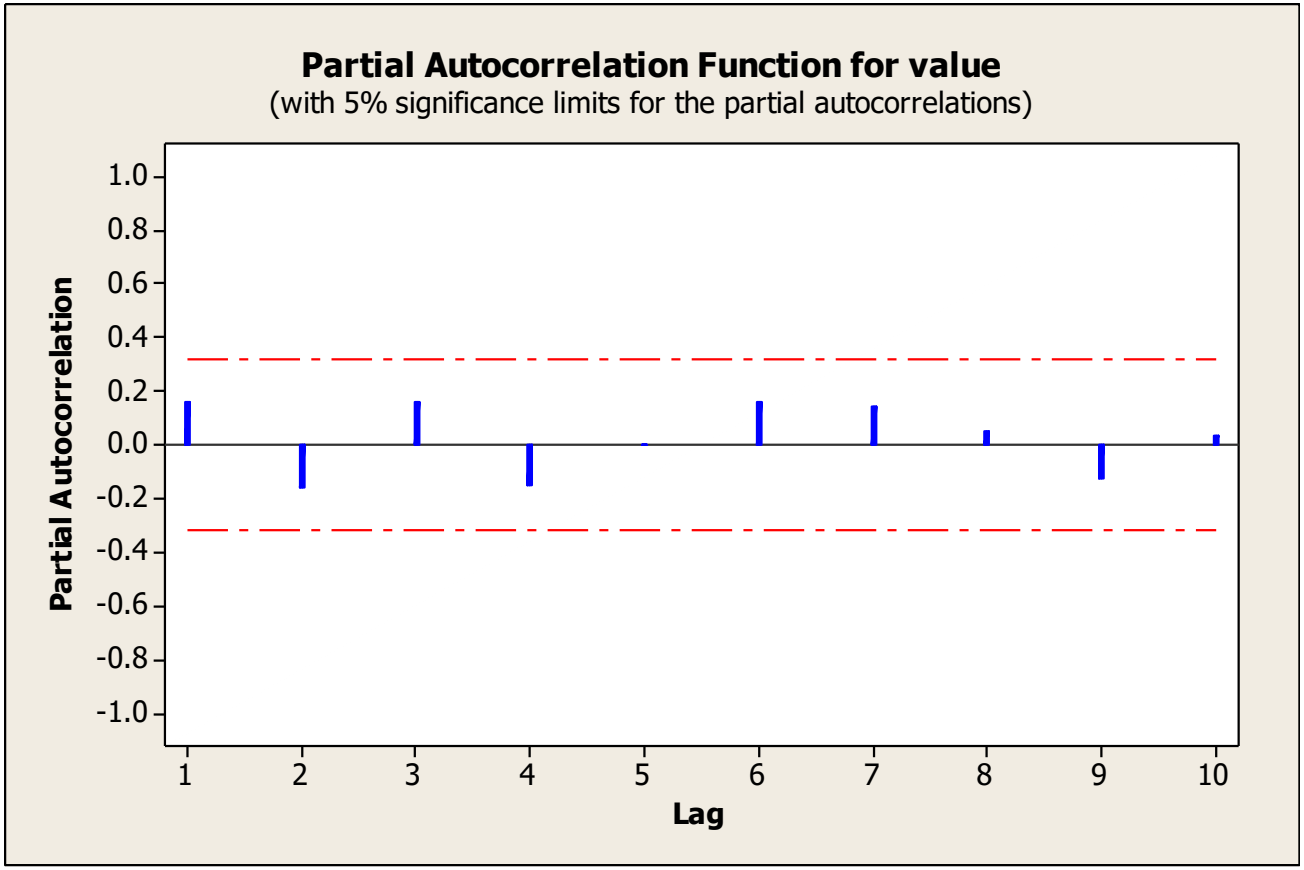
$$\pm 1.96 \frac{1}{\sqrt{n}}$$

Can be used with PACF to asses if the data are white noise.

The Partial autocorrelation coefficient

- It is usual to plot the partial autocorrelation function or PACF.
- The PACF plot of the white noise data is presented in the next slide.

PACF plot of the white noise series.



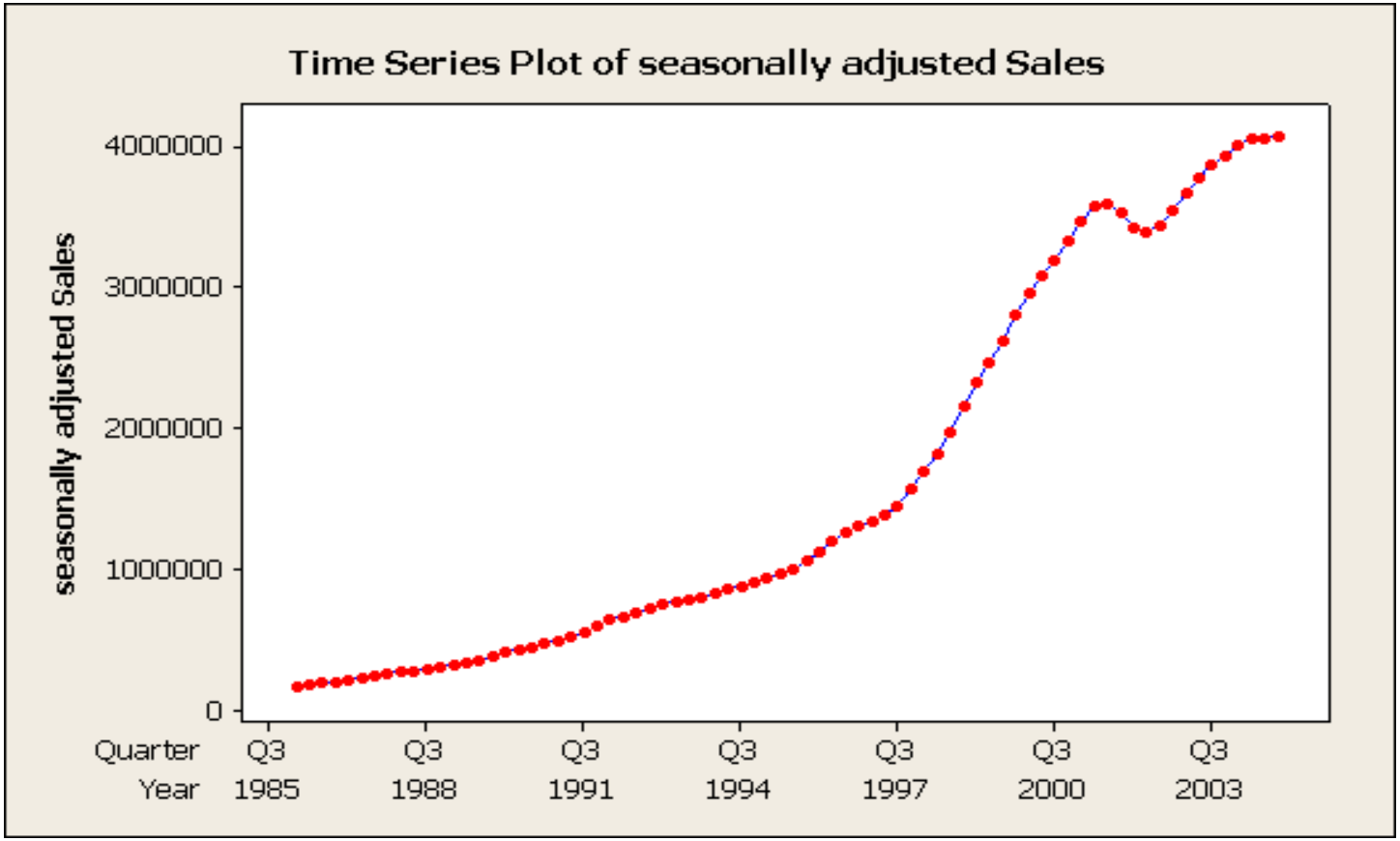
Examining stationarity of time series data

- Stationarity means no growth or decline.
- Data fluctuates around a constant mean independent of time and variance of the fluctuation remains constant over time.
- Stationarity can be assessed using a time series plot.
  - Plot shows no change in the mean over time
  - No obvious change in the variance over time.

Examining stationarity of time series data

- The autocorrelation plot can also show non-stationarity.
  - Significant autocorrelation for several time lags and slow decline in r_k indicate non-stationarity.
- The following graph shows the seasonally adjusted sales for Gap stores from 1985 to 2003.

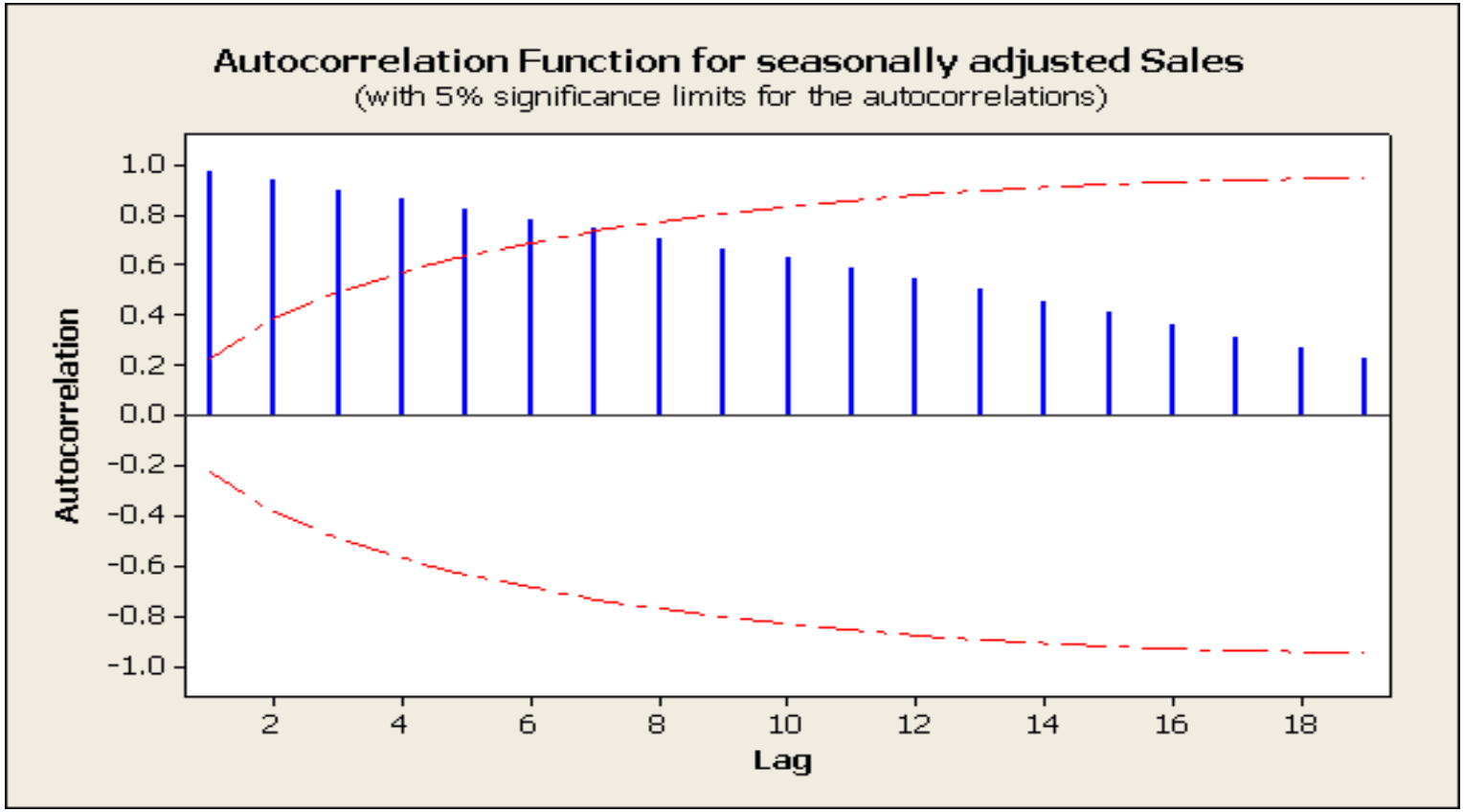
Examining stationarity of time series data



Examining stationarity of time series data

- The time series plot shows that it is non-stationary in the mean.
- The next slide shows the ACF plot for this data series.

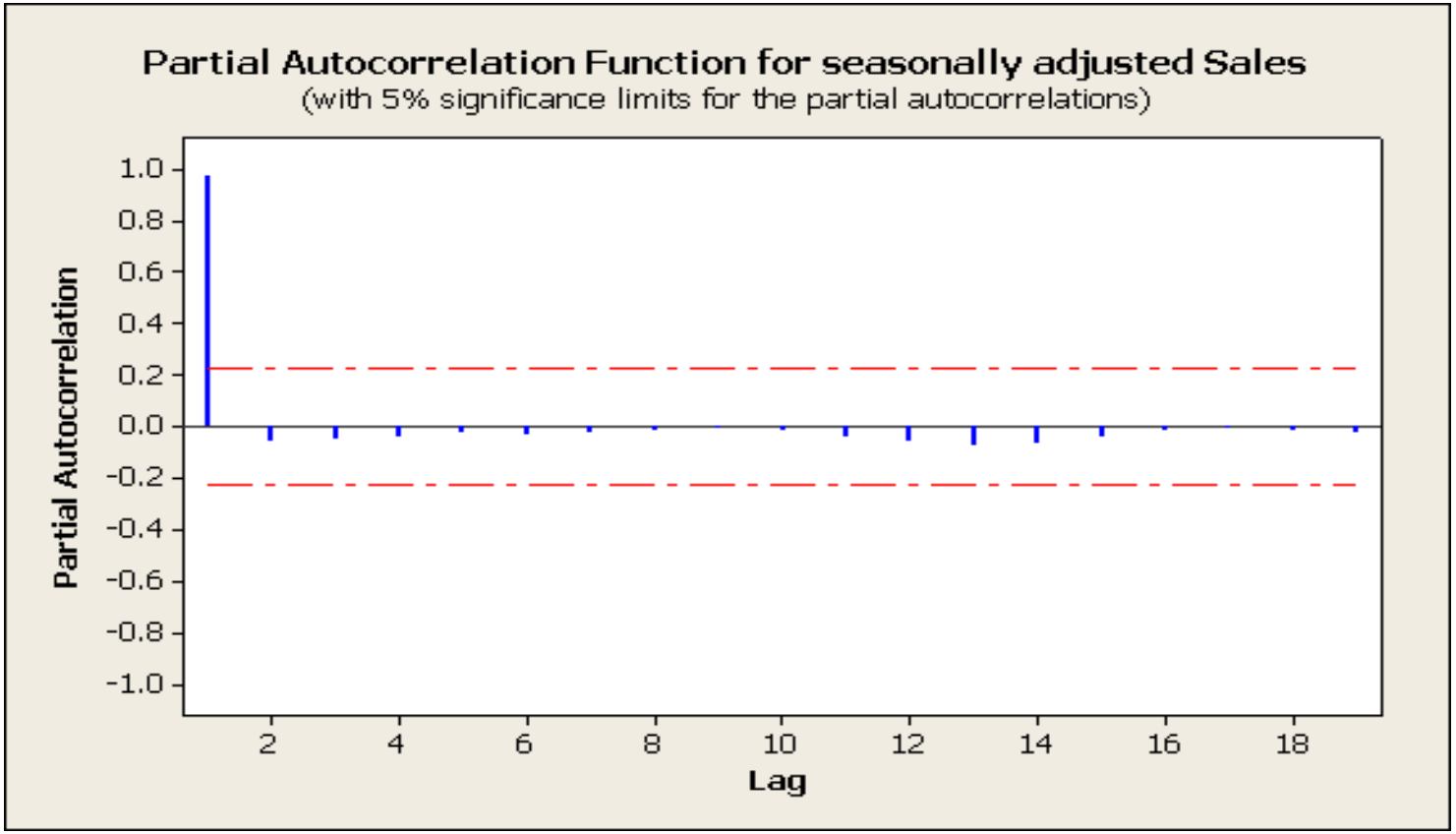
Examining stationarity of time series data



Examining stationarity of time series data

- The ACF also shows a pattern typical for a non-stationary series:
  - Large significant ACF for the first 7 time lag
  - Slow decrease in the size of the autocorrelations.
- The PACF is shown in the next slide.

Examining stationarity of time series data



Examining stationarity of time series data

- This is also typical of a non-stationary series.
  - Partial autocorrelation at time lag 1 is close to one and the partial autocorrelation for the time lag 2 through 18 are close to zero.

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