PES UNIVERSITY ONLINE

DATA ANALYTICS

Unit 3: Introduction to Time series data

Jyothi R.
Department of Computer Science and Engineering

INTRODUCTION TO FORECASTING

- Forecasting is one of the most important and frequently addressed problems in analytics.
- Inaccurate forecasting can have significant impact on both top line and bottom line of an organization.
- For example, non-availability of product in the market can result in customer dissatisfaction, whereas, too much inventory can erode the organization's profit.
- Thus, it becomes necessary to forecast the demand for a product and service as accurately as possible.
- Every organization prepares long-range and short-range planning for the organization and forecasting demand for product and service is an important input for both long-range and short-range planning

INTRODUCTION TO FORECASTING contd.

- Different capacity planning problems such as manpower planning, machine capacity, warehouse capacity, materials requirements planning (MRP) will depend on the forecasted demand for the product/ service.
- Budget allocation for marketing promotions and advertisement are usually made based on forecasted demand for the product.

INTRODUCTION TO FORECASTING

- Forecasting can be very challenging due to several factors that 428 Business Analytics influence the demand and scale of business with stock keeping units (SKUs) running into several millions. For example:

1. Boeing 747-400 has more than 6 million parts and several thousand unique parts (Hill, 2011). Forecasting demand for spare parts is important since non-availability of mission critical parts can result in aircraft on ground (AOG) which can be very expensive for airlines.

INTRODUCTION TO FORECASTING Example contd.

2. Amazon.com sells more than 350 million products through its E-commerce portal. Amazon itself sells about 13 million SKUs and has more (about 2 million) retailers selling their products through Amazon (Ali, 2017). Predicting demand for these products is important since overstocking can impact the bottom line and under stocking can result in customer dissatisfaction. Amazon.com may not stock all SKUs they sell through their portal since most of them are sold by their suppliers (online marketplace) directly to the customers, but even if they have to predict demand for products directly sold by them, then the number of SKUs is 13 million.

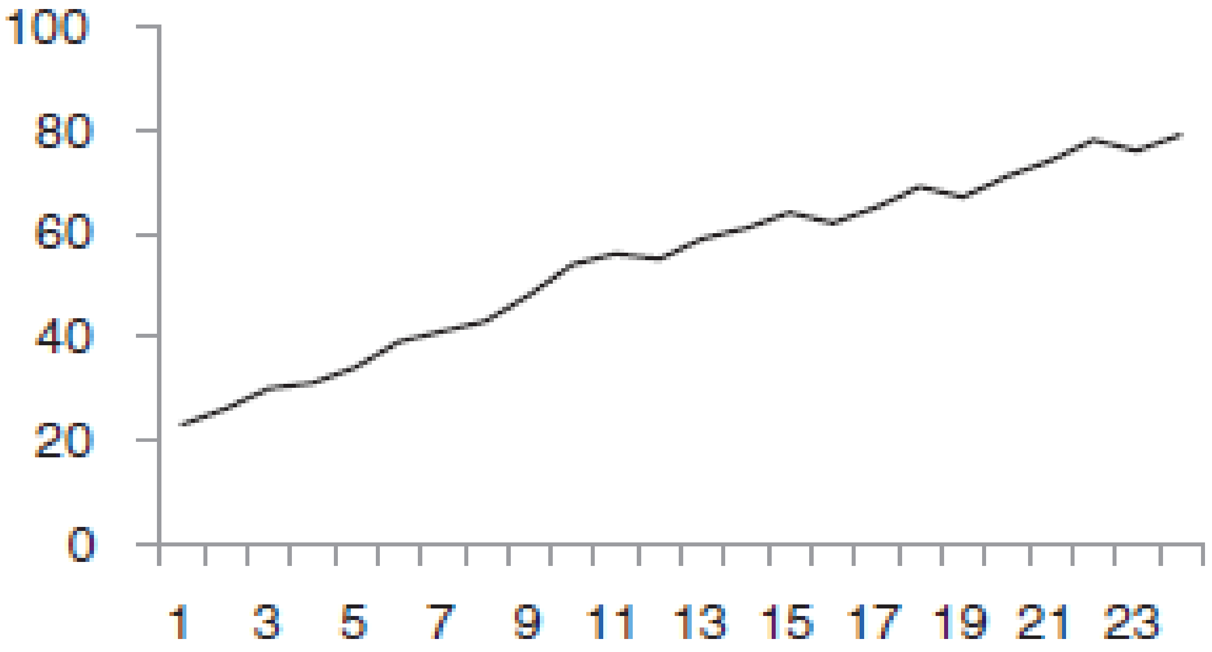
INTRODUCTION TO FORECASTING Example contd.

3. Walmart sells more than 142,000 products through their supercenters (source: Walmart website1). Being a brick-and-mortar retail store, Walmart does not have the advantages of Amazon.com (being also a market place, Amazon do not have to predict demand for all the products sold through their portal). They have to maintain stock for each and every product sold by Walmart and predict demand for the products as accurately as possible.

COMPONENTS OF TIME-SERIES DATA

From a forecasting perspective, a time-series data can be broken into the following components

1. **Trend Component (Tt):** Trend is the consistent long-term upward or downward movement of the data over a period of time.



(a) Trend

COMPONENTS OF TIME-SERIES DATA Contd.

2. Seasonal Component (St): Seasonal component is the repetitive upward or downward movement (or fluctuations) from the trend that occurs within a calendar year such as seasons, quarters, months, days of the week, etc.

- The upward or downward fluctuation may be caused due to festivals, customs within a society, school holidays, business practices within the market such as ‘end of season sale’, and so on.
- For example, in India demand for many products surge during the festival months of October and November.
- A similar pattern exists during December in many countries due to Christmas.

Usually, for a given context seasonal fluctuation occurs at fixed intervals (such as months, quarters) known as periodicity of seasonal variation and repeats over time.

Seasonal Component (St): Contd.

The seasonal component consists of effects that are reasonably stable with respect to timing, direction and magnitude. It arises from systematic, calendar related influences such as:

- **Natural Conditions: W**eather fluctuations that are representative of the season (uncharacteristic weather patterns such as snow in summer would be considered irregular influences)

- **Business and Administrative procedures:** Start and end of the school term

- **Social and Cultural behaviour** Christmas

Seasonal Component (St):

It also includes calendar related systematic effects that are not stable in their annual timing or are caused by variations in the calendar from year to year, such as:

- **Trading Day Effects**

the number of occurrences of each of the day of the week in a given month will differ from year to year

- There were 4 weekends in March in 2000, but 5 weekends in March of 2002

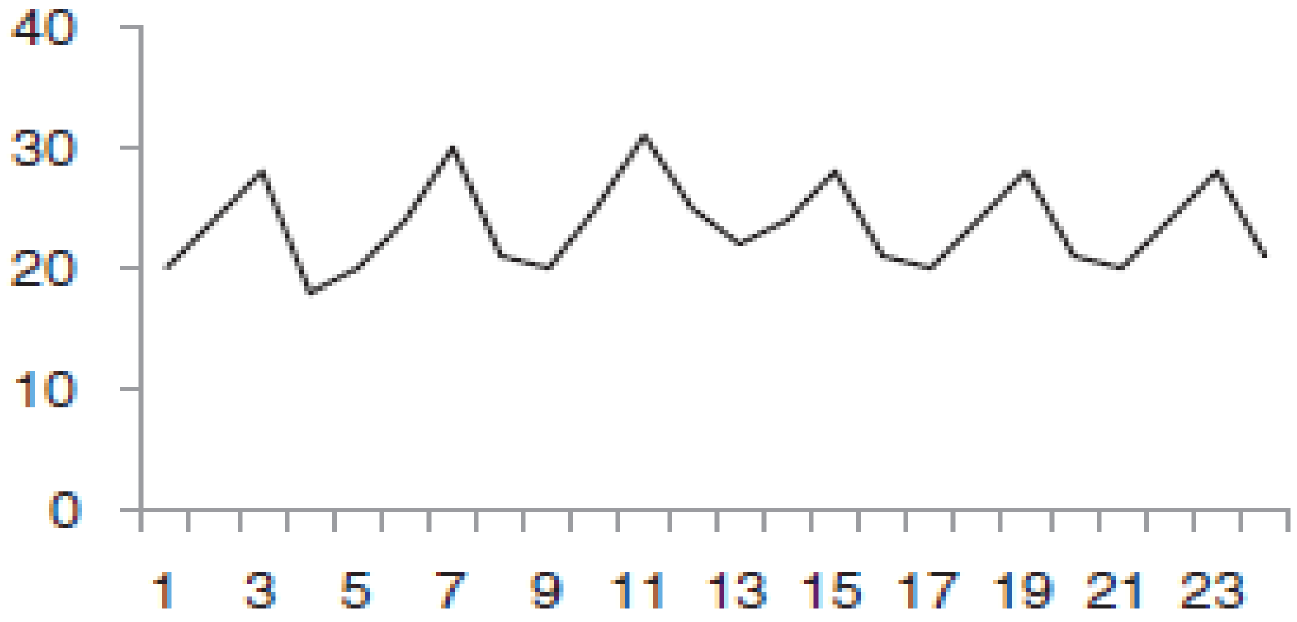
- **Moving Holiday Effects**

holidays which occur each year, but whose exact timing shifts

- Easter, Chinese New Year

COMPONENTS OF TIME-SERIES DATA

Seasonal Component (St):



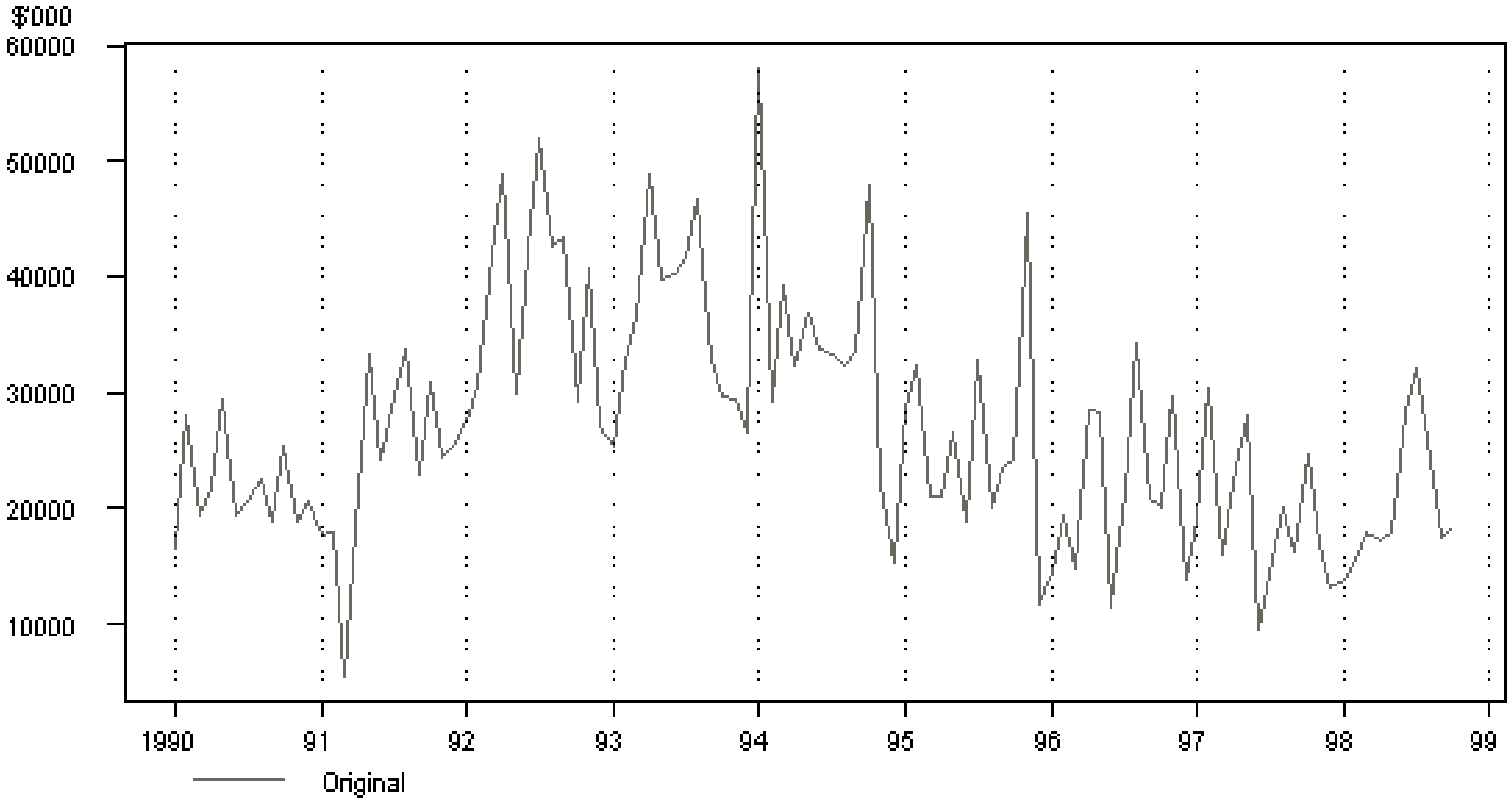
(b) Seasonality (fixed periodicity)

HOW DO WE IDENTIFY SEASONALITY?

- Seasonality in a time series can be identified by regularly spaced peaks and troughs which have a consistent direction and approximately the same magnitude every year, relative to the trend.
- The following diagram depicts a strongly seasonal series.
- There is an obvious large seasonal increase in December retail sales in New South Wales due to Christmas shopping.
- In this example, the magnitude of the seasonal component increases over time, as does the trend.

HOW DO WE IDENTIFY SEASONALITY?

- **Figure 1: Monthly Retail Sales in New South Wales (NSW) Retail Department Stores**

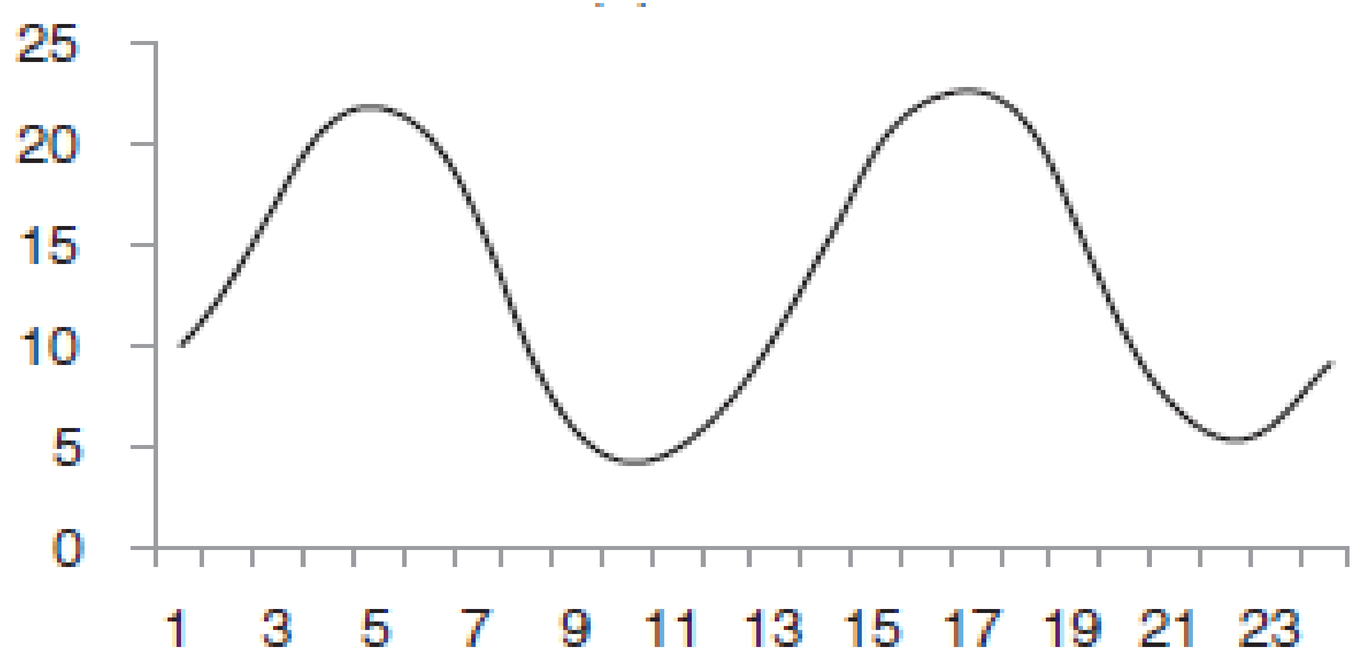


COMPONENTS OF TIME-SERIES DATA

3. Cyclical Component (Ct): Cyclical component is fluctuation around the trend line that happens due to macro-economic changes such as recession, unemployment, etc.

- Cyclical fluctuations have repetitive patterns with a time between repetitions of more than a year.
- Whereas in case of seasonality, the fluctuations are observed within a calendar year and are driven by factors such as festivals and customs that exist in a society.
- A major difference between seasonal fluctuation and cyclical fluctuation is that seasonal fluctuation occurs at fixed period within a calendar year, whereas cyclical fluctuations have random time between fluctuations.
- That is, periodicity of seasonal fluctuations is constant, whereas the periodicity of cyclical fluctuations is not constant.

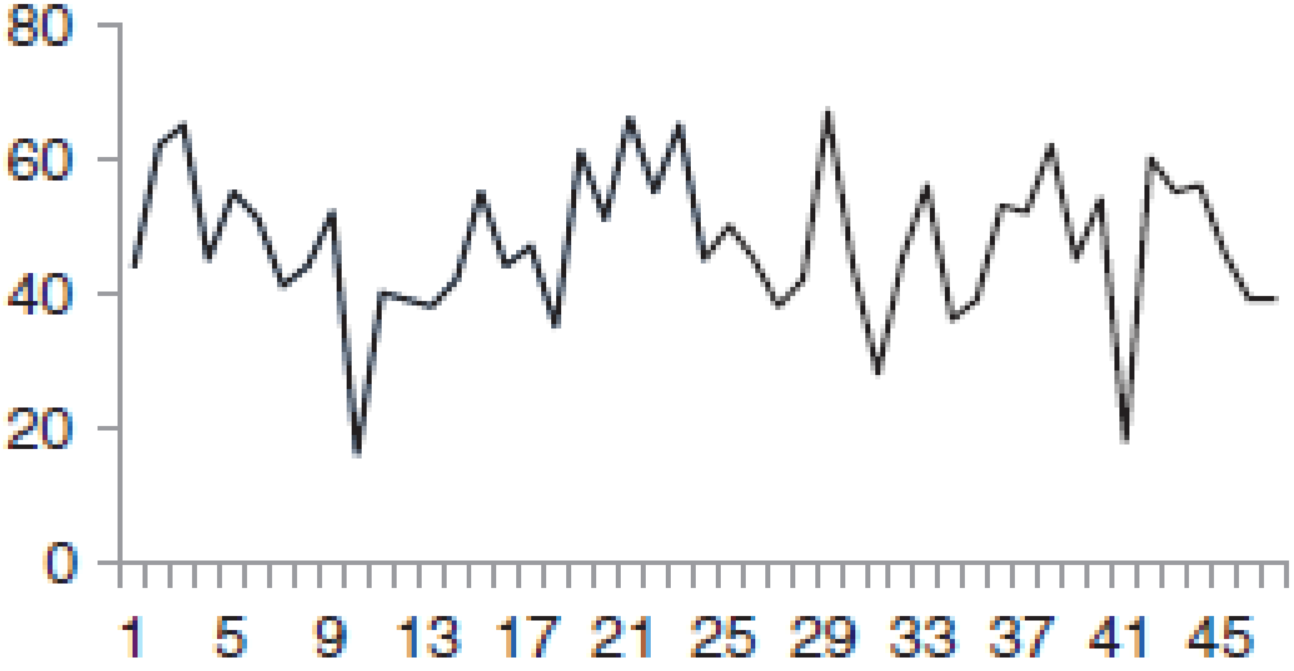
Cyclical Component (Ct):



(c) Cyclicality

COMPONENTS OF TIME-SERIES DATA contd.

4. Irregular Component (It): Irregular component is the white noise or random uncorrelated changes that follow a normal distribution with mean value of 0 and constant variance.



(d) Irregular

COMPONENTS OF TIME-SERIES DATA contd.

- The time-series data can be modelled as an addition of the above components or product of the above components.
- The additive time-series model is given by

$$Y_t = T_t + S_t + C_t + I_t \qquad \textit{Eqn-1}$$

- The additive models assume that the seasonal and cyclical components are independent of the trend component.
- Additive models are not very common since in many cases the seasonal component may not be independent of trend.

COMPONENTS OF TIME-SERIES DATA contd.

- The multiplicative time-series model is given by
- $Y_t = T_t \times S_t \times C_t \times I_t$ Eqn-2
- Multiplicative models are more common and are a better fit for many data sets. In many cases, we will use the form $Y_t = T_t \times S_t$ which is simpler form of Eq. 2
- To estimate the cyclical component we will need a large data set.
- The additive model is appropriate if the seasonal component remains constant about the level (or mean) and does not vary with the level of the series.
- The multiplicative model is more appropriate, if seasonal variation is correlated with level (local mean).

COMPONENTS OF TIME-SERIES DATA contd.

- Decomposition models are typically additive or multiplicative, but can also take other forms such as pseudo-additive.

Additive Decomposition

In some time series, the amplitude of both the seasonal and irregular variations do not change as the level of the trend rises or falls. In such cases, an additive model is appropriate.

In the additive model, the observed time series (O_t) is considered to be the sum of three independent components: the seasonal S_t, the trend T_t and the irregular I_t. Observed series = Trend + Seasonal + Irregular i.e

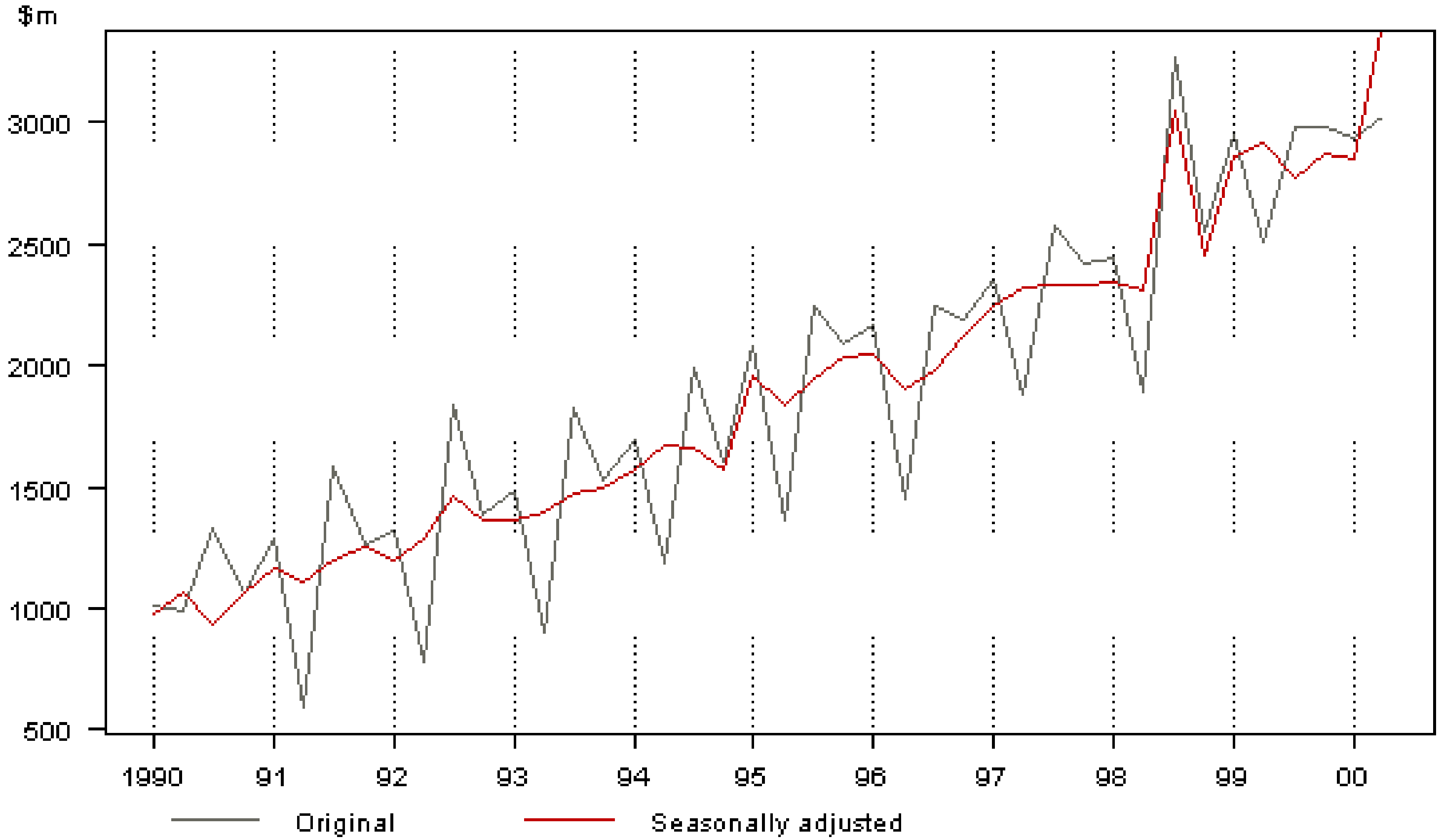
$$O_t = T_t + S_t + I_t$$

COMPONENTS OF TIME-SERIES DATA contd.

The following figure depicts a typically additive series.

The underlying level of the series fluctuates but the magnitude of the seasonal spikes remains approximately stable.

Figure 5: General Government and Other Current Transfers to Other Sectors



COMPONENTS OF TIME-SERIES DATA contd.

- The multiplicative time-series model is given by
- $Y_t = T_t \times S_t \times C_t \times I_t$ Eqn-2
- Multiplicative models are more common and are a better fit for many data sets. In many cases, we will use the form $Y_t = T_t \times S_t$ which is simpler form of Eq. 2
- To estimate the cyclical component we will need a large data set.
- The additive model is appropriate if the seasonal component remains constant about the level (or mean) and does not vary with the level of the series.
- The multiplicative model is more appropriate, if seasonal variation is correlated with level (local mean).

COMPONENTS OF TIME-SERIES DATA contd.

- **Multiplicative Decomposition**

  In many time series, the amplitude of both the seasonal and irregular variations increase as the level of the trend rises. In this situation, a multiplicative model is usually appropriate.

  In the multiplicative model, the original time series is expressed as the product of trend, seasonal and irregular components.

- Observed series = Trend x Seasonal x Irregular

$$O_t = T_t + S_t + I_t$$

COMPONENTS OF TIME-SERIES DATA contd.

- **Multiplicative Decomposition**

  The seasonally adjusted data then becomes:

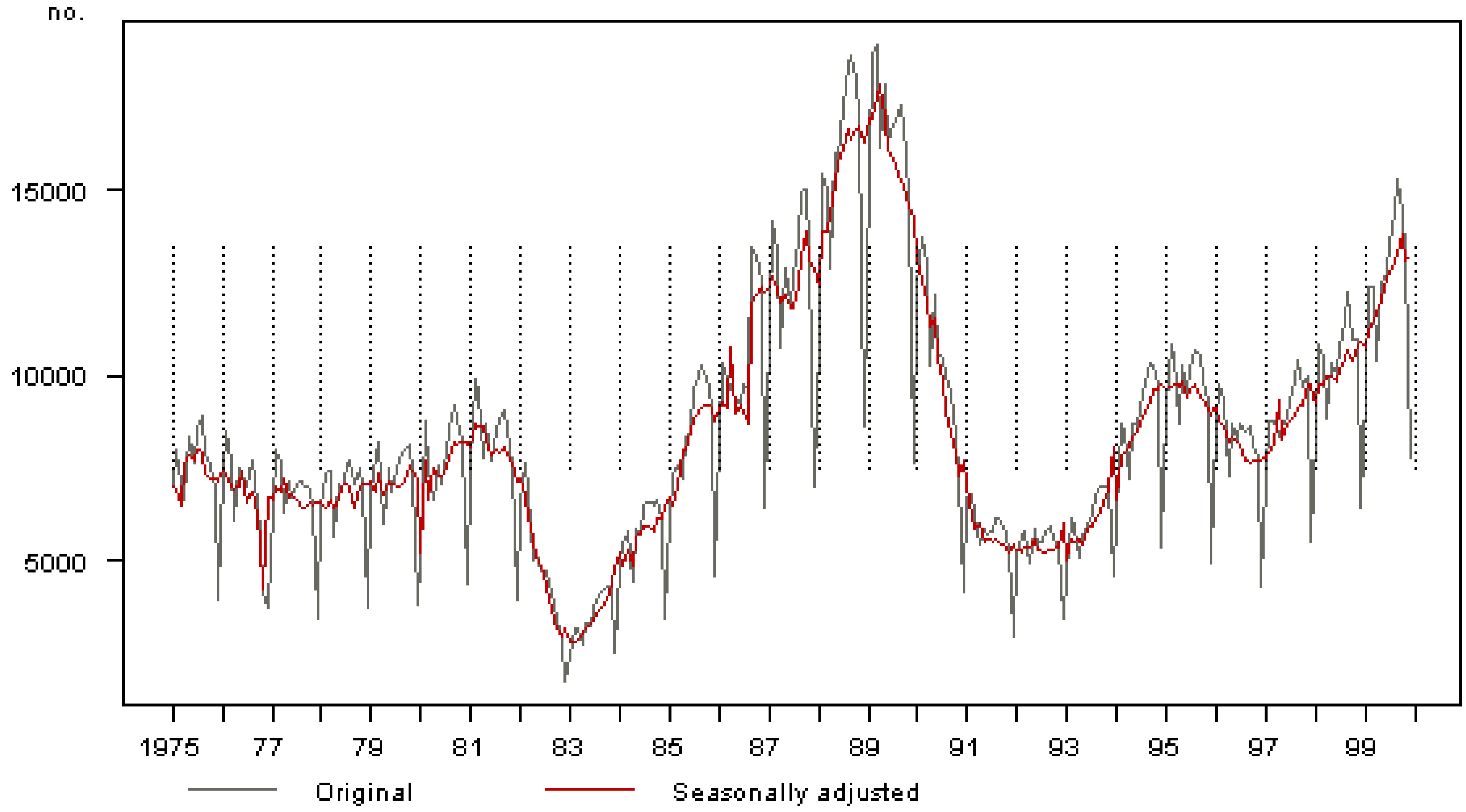
  Seasonally adjusted series = Observed / Seasonal i.e. $SA = O_t / S_t$

  = Trend x Irregular $= T_t \times I_t$

- Under this model, the trend has the same units as the original series, but the seasonal and irregular components are unitless factors, distributed around 1.
- Most of the series analyzed by the ABS show characteristics of a multiplicative model. As the underlying level of the series changes, the magnitude of the seasonal fluctuations varies as well.

COMPONENTS OF TIME-SERIES DATA contd.

Figure 6: Monthly NSW ANZ Job Advertisements



COMPONENTS OF TIME-SERIES DATA contd.

Pseudo-Additive Decomposition:

- The multiplicative model cannot be used when the original time series contains very small or zero values.
- This is because it is not possible to divide a number by zero.
- In these cases, a pseudo additive model combining the elements of both the additive and multiplicative models is used.
- This model assumes that seasonal and irregular variations are both dependent on the level of the trend but independent of each other.

The original data can be expressed in the following form:

$$O_t = T_t + T_t \text{ x } (S_t – 1) + T_t \text{ x } (I_t - 1)$$

$$= T_t \times (S_t + I_t – 1)$$

COMPONENTS OF TIME-SERIES DATA contd.

- The pseudo-additive model continues the convention of the multiplicative model to have both the seasonal factor S_t and the irregular factor I_t centred around one.
- Therefore we need to subtract one from S_t and I_t to ensure that the terms T_t x (S_t - 1) and T_t x (I_t - 1) are centred around zero.
- These terms can be interpreted as the additive seasonal and additive irregular components respectively and because they are centred around zero the original data O_t will be centred around the trend values T_t .

The seasonally adjusted estimate is defined to be:

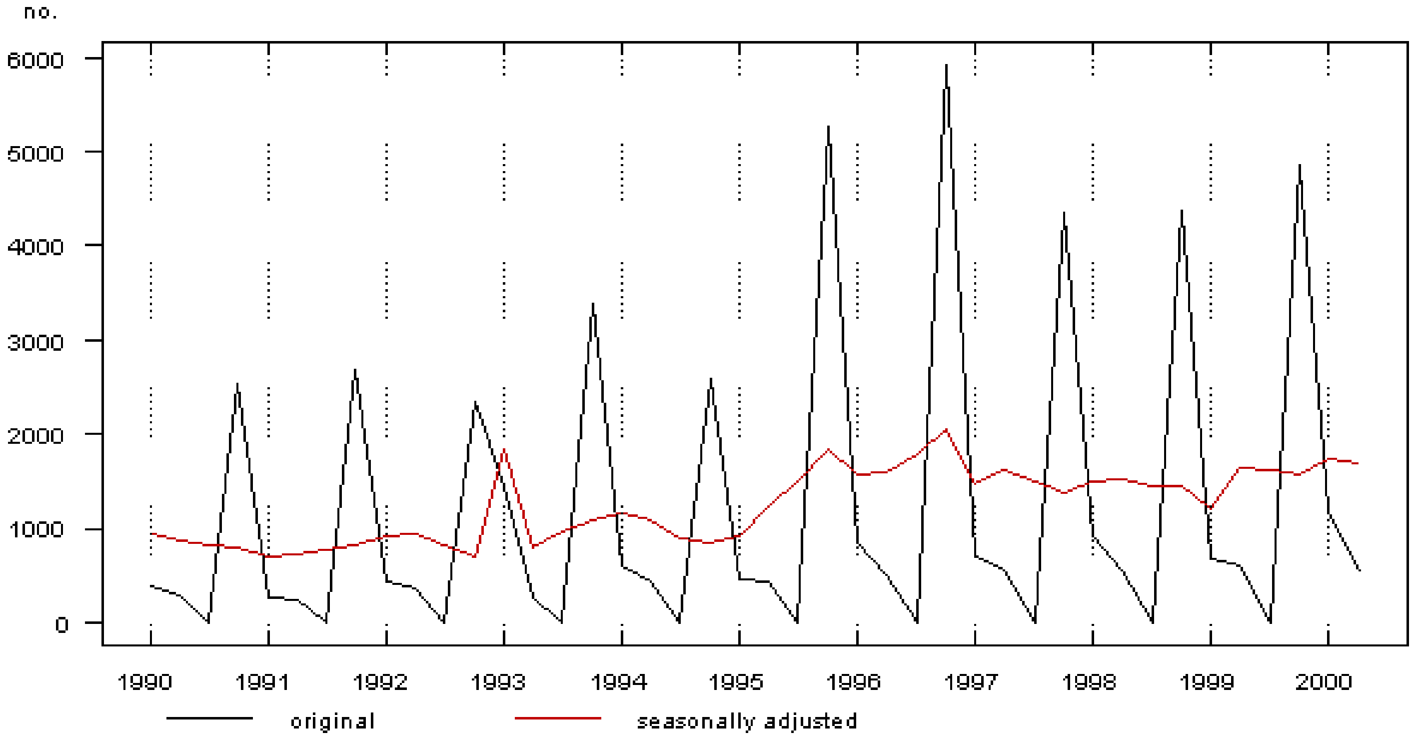
$$SA = O_t - T_t \times (S_t - 1)$$

$$= T_t \times I_t$$

Where T_t and S_t are the trend and seasonal component estimates

COMPONENTS OF TIME-SERIES DATA contd.

- An example of series that requires a pseudo-additive decomposition model is shown below.
- This model is used as cereal crops are only produced during certain months, with crop production being virtually zero for one quarter each year.
- **Figure 7: Quarterly Gross Value for the Production of Cereal Crops**



References

Text Book:

"Business Analytics, The Science of Data-Driven Making", U. Dinesh Kumar, Wiley 2017

Chapter-13

Image Courtesy

https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Time+Series+Analysis:+The+Basics

THANK YOU

Jyothi R
Assistant Professor,Department of Computer Science
jyothir@pes.edu