Time Series Forecasting

Content

- Taxonomy of Time Series Forecasting Methods
- Time Series Decomposition
- Smoothening Methods
- Average Method, Moving Average smoothing, Time series analysis using linear regression
- ARIMA Model
- Performance Evaluation - mean absolute error,root mean square error, mean absolute percentage error,mean absolute scaled error

What Is Forecasting?

- Process of predicting a future event
- Underlying basis of all business decisions
  - Production
  - Inventory
  - Personnel
  - Facilities

Forecasting Approaches

Qualitative Methods

- Used when situation is vague & little data exist
  - New products
  - New technology
- Involve intuition, experience
- e.g., forecasting sales on Internet

Quantitative Methods

Forecasting Approaches

Qualitative Methods

- Used when situation is vague & little data exist
  - New products
  - New technology
- Involve intuition, experience
- e.g., forecasting sales on Internet

Quantitative Methods

- Used when situation is 'stable' & historical data exist
  - Existing products
  - Current technology
- Involve mathematical techniques
- e.g., forecasting sales of color televisions

Quantitative Forecasting

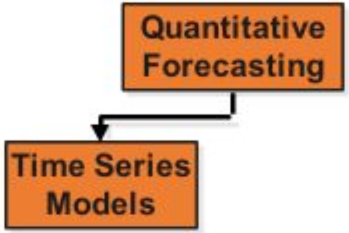
- Select several forecasting methods
- 'Forecast' the past
- Evaluate forecasts
- Select best method
- Forecast the future
- Monitor continuously forecast accuracy

Quantitative Forecasting Methods

Quantitative Forecasting Methods

Quantitative Forecasting

Quantitative Forecasting Methods



Time Series Analysis

- A time series is a series of **observations recorded in the order of time.**
- Time series analysis is the **process of extracting meaningful patterns and information from time series data.**
- Time series forecasting is the process of predicting future values of a time series based on past observations and other inputs.
- Time series forecasting is one of the oldest known predictive analytics techniques and is widely used in every organizational setting.

What is a Time Series?

- Set of evenly spaced numerical data
  - Obtained by observing response variable at regular time periods
- Forecast based only on past values
  - Assumes that factors influencing past, present, & future will continue
- Example

• Year:	1995	1996	1997	1998	1999
• Sales:	78.7	63.5	89.7	93.2	92.1

Time Series Analysis

- Time series data often **exhibit patterns and trends that are not present in other types of data**, such as seasonality, trends, and cyclical patterns.

- When building a time series forecasting model, the goal is to use historical information to make predictions about future values of the same quantity.

Time Series Analysis

- There are a variety of modeling techniques that can be used for time series forecasting, including **ARIMA models, exponential smoothing, and machine learning algorithms.**

- Time series forecasting is a specialized area of **predictive modeling** that requires a deep understanding of time series analysis techniques and statistical modeling methods.

Time Series Data Analysis

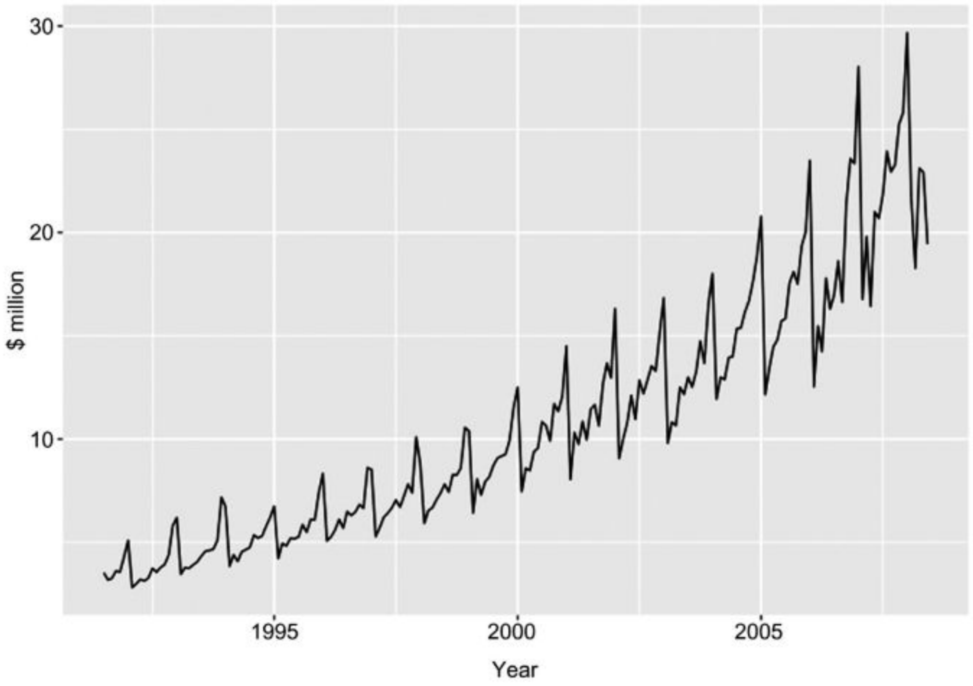


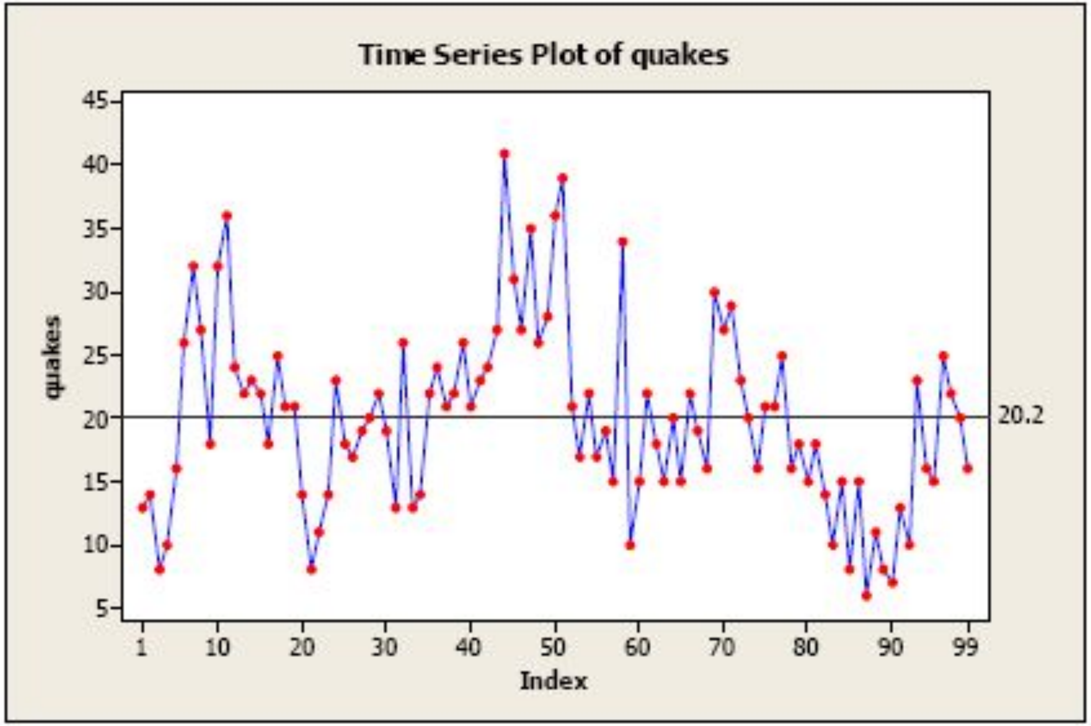
FIGURE 12.1

Time series of monthly antidiabetic drug sales.[1]

Time Series Data Analysis

- Features of the Plot:
- There is **no consistent trend** (upward or downward) over the entire time span.
- The series appears to slowly wander up and down. The horizontal line drawn at quakes = **20.2 indicates the mean** of the series.
- Notice that the **series tends to stay on the same side of the mean** (above or below) for a while and then wanders to the other side.

- There are **no obvious outliers.**
- It's difficult to judge whether the variance is constant or not.



Time Series Data Analysis

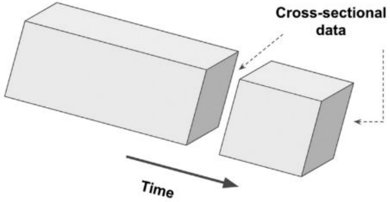


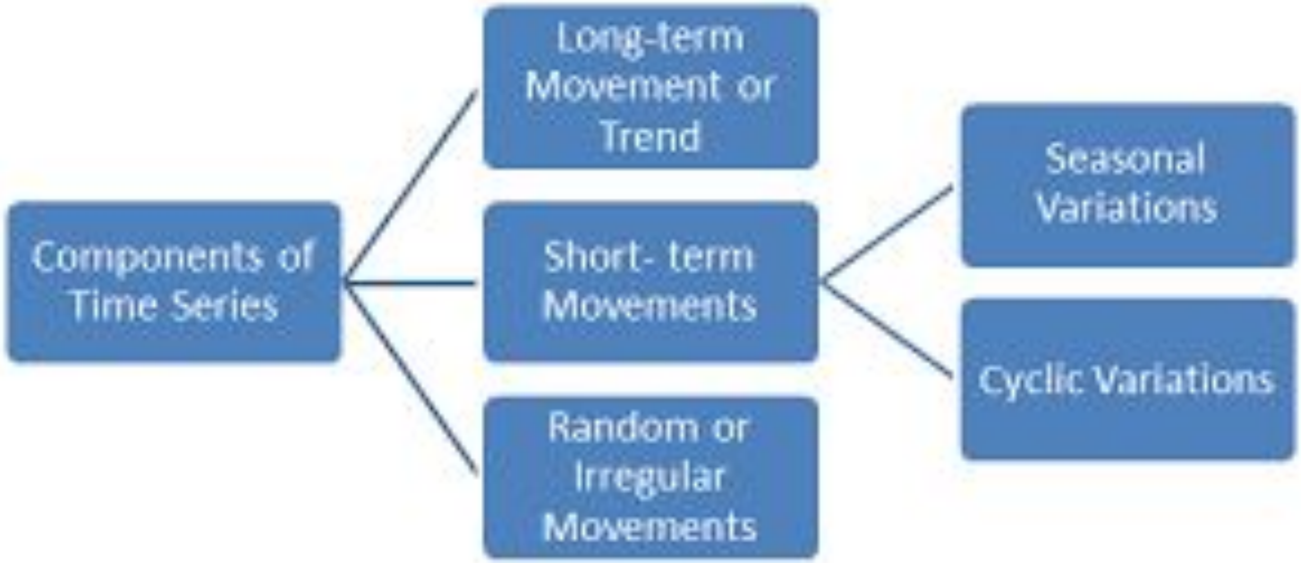
FIGURE 12.2
Cross-sectional data is a subset of time series data.

- In time series analysis, the **focus is on forecasting a specific variable over time, based on its past behavior.**

- For example, **in predicting house prices,** variables such as location, square footage, number of rooms, etc. are observed at a point in time, and the price is predicted at the same point in time.
- it's important to take the time variable into account, as house prices can fluctuate over time due to economic conditions, supply and demand, etc.
- **Time series analysis** is **concerned with understanding and forecasting these fluctuations over time.**

Time Series Decomposition

-Level (Avg)
- Trend (up, down)
- Seasonality
-Cyclic Pattern
-Noise



Time Series Decomposition

- A time series consists of three components:
- trend, seasonal, and random.

Time Series Decomposition

- **Trend**
- A trend is a **long-term pattern in the data** that **shows its overall direction.** It can be **upward, downward, or horizontal**.
- Trends can be **linear or nonlinear.**
- **A linear trend** is a **straight line** that shows a constant increase or decrease in the data.
- **A nonlinear trend** is a **curve that shows a gradual increase** or **decrease** in the data.

Time Series Decomposition

- **Seasonal**

- A seasonal component is a pattern that repeats itself over a fixed period of time.
- For example, **sales of winter clothing are higher in the winter season** than in the summer season.
- The seasonal component is **usually represented by a cycle that repeats each year, quarter, month, or week.**

Time Series Decomposition

- **Random**

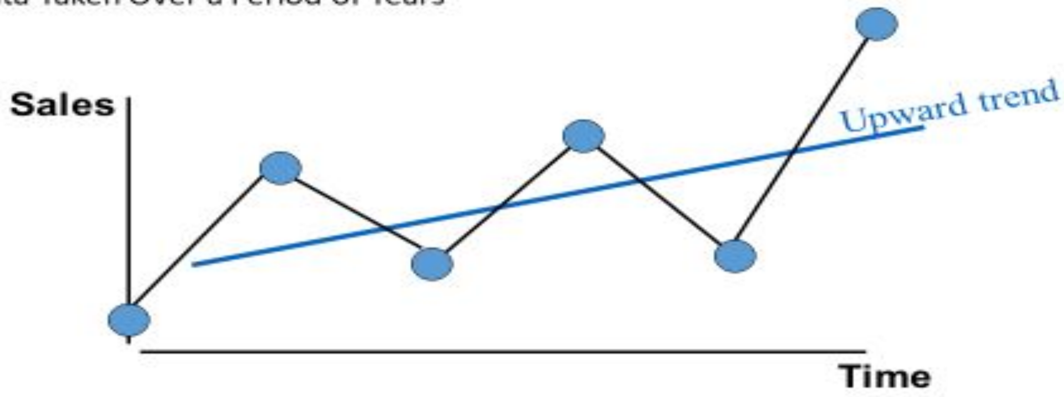
- The random component is **the part of the data that cannot be explained by the trend or seasonal components.**
- It represents the **random fluctuations** in the data that are not predictable.

Trend Component

- Persistent, overall upward or downward pattern
- Due to population, technology etc.
- Several years duration

Trend Component

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

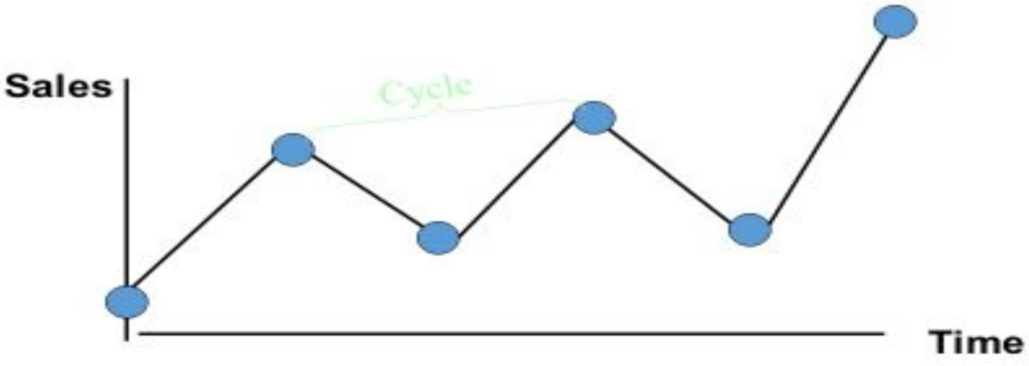


Cyclical Component

- Repeating up & down movements
- Due to interactions of factors influencing economy
- Usually 2-10 years duration

Cyclical Component

- Upward or Downward Swings
- May Vary in Length
- Usually Lasts 2 - 10 Years



Seasonal Component

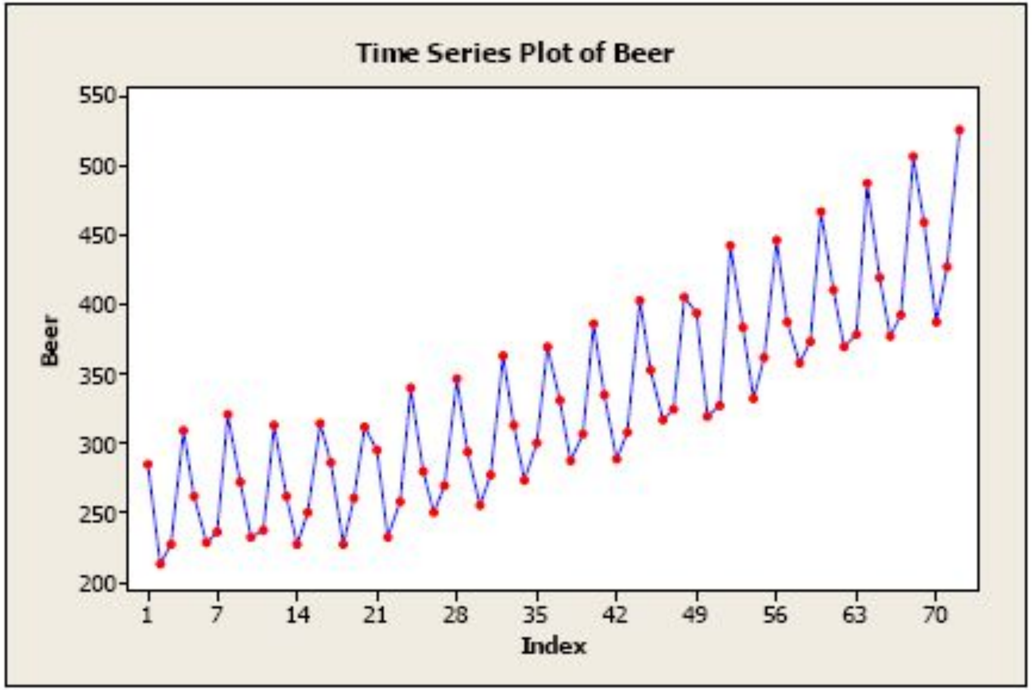
- Regular pattern of up & down fluctuations
- Due to weather, customs etc.
- Occurs within one year

Irregular Component

- Erratic, unsystematic, ‘residual’ fluctuations
- Due to random variation or unforeseen events
  - Union strike
  - War
- Short duration & nonrepeating

Time Series Decomposition

- Features of the Plot:
- **There is an upward trend**, possibly a curved one.
- **There is seasonality** – a regularly repeating pattern of highs and lows related to quarters of the year.
- There are **no obvious outliers.**



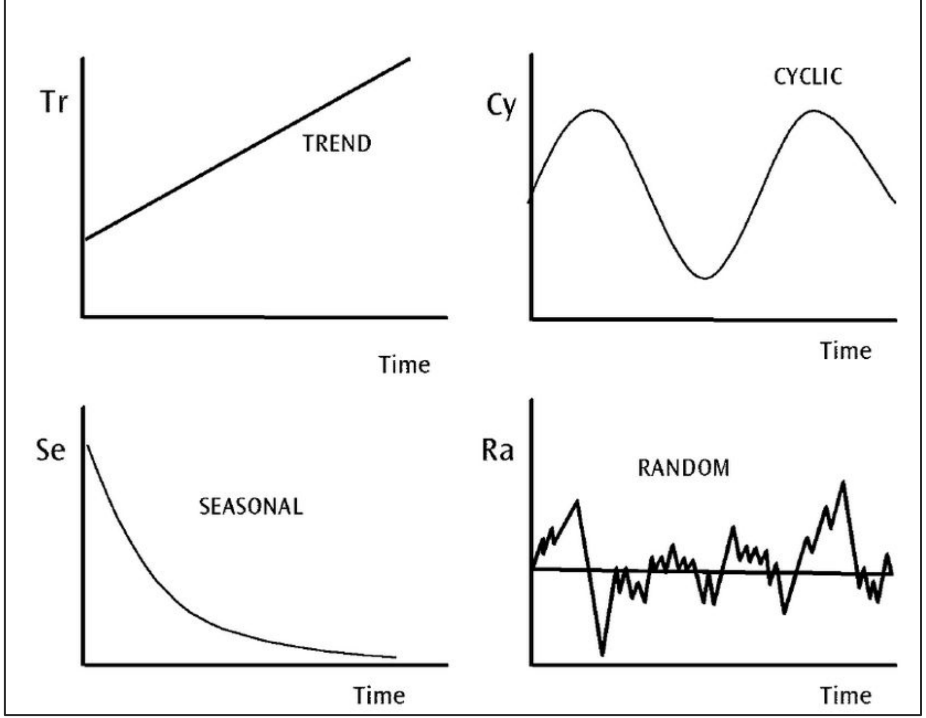
Time Series Decomposition

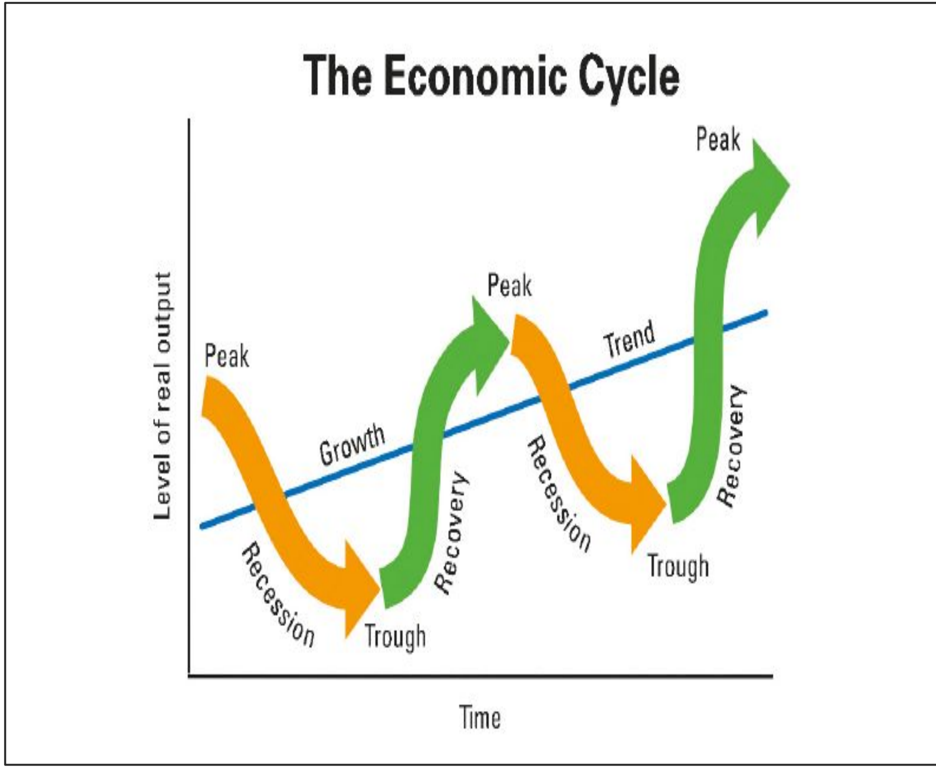
There are four types of time-series patterns:

- ***Trend***: Long-term increase or decrease in the data. The trend can be any function, such as linear or exponential, and can change direction over time.
- ***Seasonality***: Repeating cycle in the series with fixed frequencies (hour of the day, week, month, year, etc.). A seasonal pattern exists of a fixed known period.
- ***Cyclicity*: Occurs when the data rise and fall, but without a fixed frequency and duration caused,** for example, by economic conditions.
- ***Noise***: The random variation in the series.

Time Series Decomposition

The **trend and seasonality** components are predictable (and are called **systematic components**), whereas, the noise, by definition, is random (and is called the non-systematic component).





TAXONOMY OF TIME SERIES FORECASTING

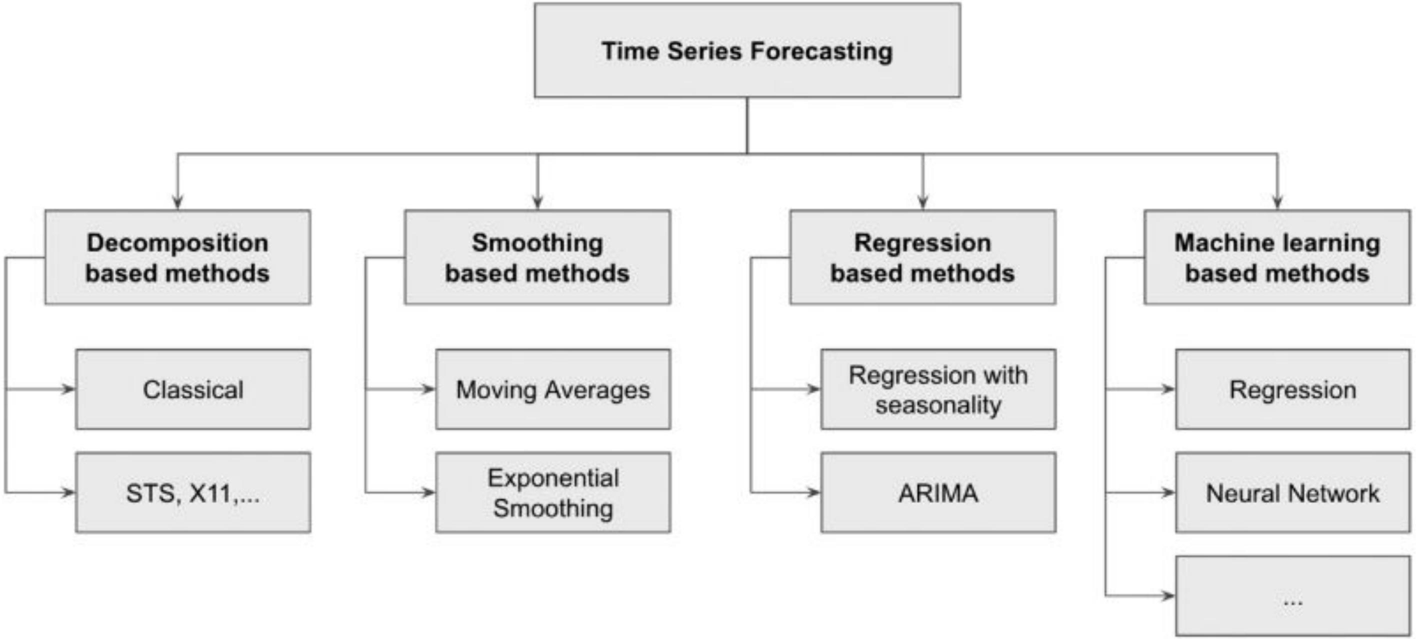
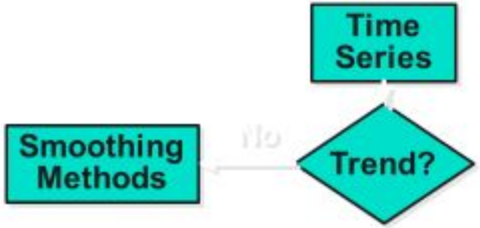


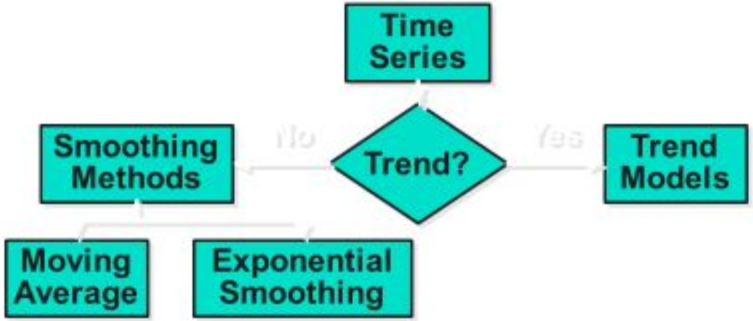
FIGURE 12.3
Taxonomy of time series forecasting techniques.

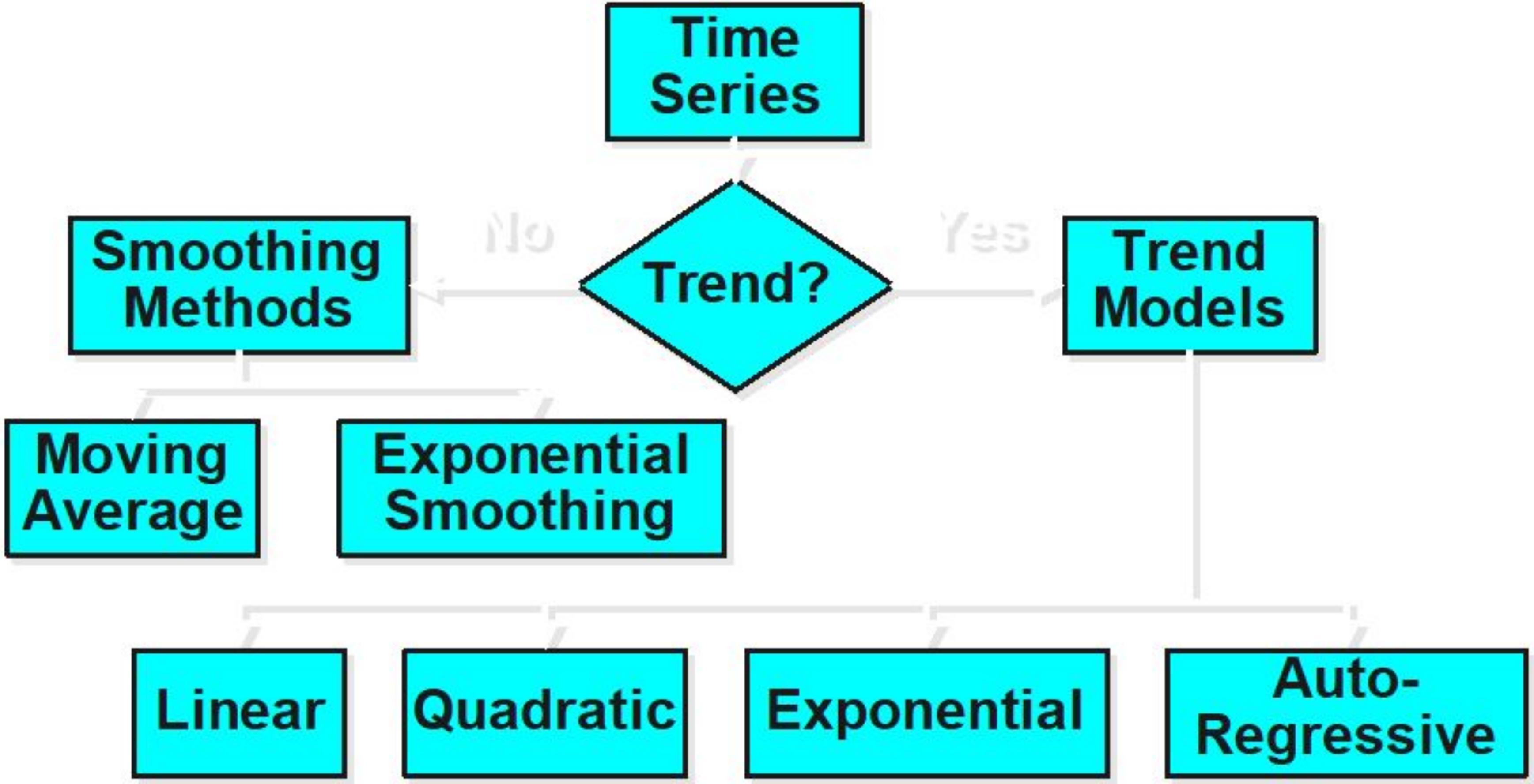
Time Series Forecasting

Time Series Forecasting



Time Series Forecasting




Time Series
Trend?
No
Yes
Smoothing Methods
Trend Models
Moving Average
Exponential Smoothing
Linear
Quadratic
Exponential
Auto-Regressive

1. TIME SERIES FORECASTING - Decomposition based Method

TIME SERIES FORECASTING - Decomposition based Method

- Time series decomposition involves breaking down a time series into its component parts, typically by separating it into its **trend, seasonal, and residual components.**
- The **trend** component represents the **long-term direction** of the time series,
- **Seasonal component** captures the **recurring patterns or cycles** that occur within a time series.
- **The residual component** represents the **random, unpredictable fluctuations** that remain after the trend and seasonal components have been accounted for.

TIME SERIES FORECASTING- Decomposition based Method

- Time series decomposition has numerous applications
- Time series analysis for
  - Forecasting,
  - Anomaly detection,
  - Trend analysis.

TIME SERIES FORECASTING - Decomposition based Method

- **Additive Decomposition Method-**
- Additive decomposition method. This involves simply adding the trend, seasonal, and residual components together to obtain the original time series:
- $Y(t) = T(t) + S(t) + R(t)$
- where **Y(t)** is the **original time series** at time t, T(t) is the trend component at time t, S(t) is the **seasonal component** at time t, and R(t) is the residual component at time t.

TIME SERIES FORECASTING - Decomposition based Method

- To obtain the **trend component,** analysts typically use a moving average or a regression model to smooth out the noise in the data and highlight the underlying long-term trend.
- The **seasonal component** can be obtained by **averaging the values of the time series at each corresponding time point across multiple cycles (e.g., days, weeks, or months).**
- The **residual component** is then obtained by **subtracting the trend and seasonal components** from the original time series.

link

Multiplicative decomposition of time series

- In the case of **multiplicative decomposition** the components are decomposed in the such a way that when they are multiplied together, the original timeseries can be derived back.
- **Time series = Trend X Seasonality X Noise**
- Both additive and multiplicative time series decomposition can be represented by these equations

$$y_t = T_t + S_t + E_t$$

$$y_t = T_t \times S_t \times E_t$$

- where Tt, St, and Et r components respectively.
- The original time se ultiplicative combination of components.
- **If the magnitude or the variation in trend changes with the level of the time series,** then **multiplicative time series decomposition is the better model.**

Additive and Multiplicative decomposition of time series

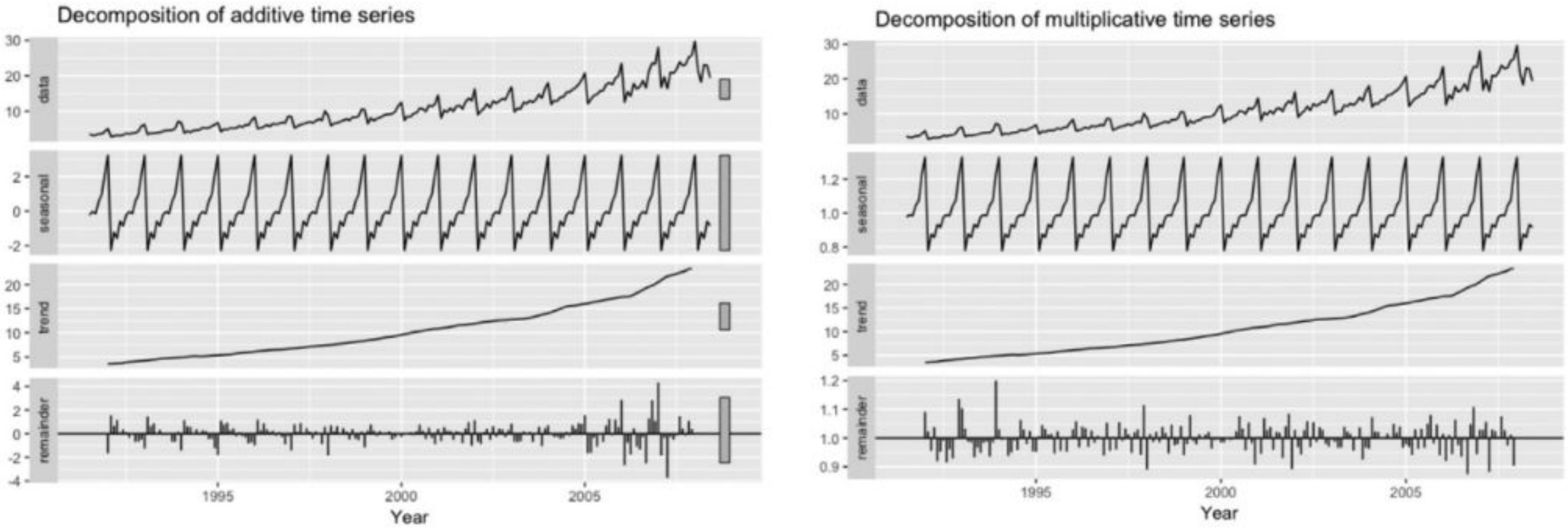


FIGURE 12.6

Additive and multiplicative decomposition of time series.

2. TIME SERIES FORECASTING - Smoothing based forecasting

TIME SERIES FORECASTING - Smoothing based forecasting

- **Smoothing based forecasting** methods involve smoothing the past observations and projecting them to the future by taking a weighted average of past observations.

TIME SERIES FORECASTING - Smoothing based forecasting

- **Smoothing based forecasting** methods involve smoothing the past observations and projecting them to the future by taking a weighted average of past observations.
- It smooth out the noise in a time series and identify its underlying trend.
- The most commonly used smoothing-based forecasting techniques are **moving averages** and **exponential smoothing.**

TIME SERIES FORECASTING - Smoothing based forecasting

In the smoothing based approaches, an observation is a function of past few observations.

- *Time periods:* $t = 1, 2, 3, \ldots, n$. Time periods can be seconds, days, weeks, months, or years depending on the problem.
- *Data series:* Observation corresponding to each time period above: $y_1, y_2, y_3, \ldots y_n$.
- *Forecasts:* F_{n+h} is the forecast for the h^{th} time period following n. Usually $h = 1$, the next time period following the last data point. However h can be greater than 1. h is called the *horizon*.
- *Forecast errors:* $e_t = y_t - F_t$ for any given time, t.

In order to explain the different methods, a simple time series data function will be used, $Y(t)$. Y is the *observed* value of the time series at any time t.

Simple Forecasting Methods

Naïve Method

Probably the simplest forecasting "model." Here one simply assumes that F_{n+1}, the forecast for the next period in the series, is given by the last data point of the series, y_n

$$F_{n+1} = y_n \tag{12.5}$$

Simple Forecasting Methods

Average Method

Moving up a level, one could compute the next data point as an average of all the data points in the series. In other words, this model calculates the forecasted value, F_{n+1}, as:

$$F_{n+1} = \text{Average}(y_n, y_{n-1}, y_{n-2}, \ldots, y_1) \tag{12.7}$$

- E.g one has monthly data from January 2010 to December 2010 and they want to predict the next January 2011 value, they would simply average the values from January 2010 to December 2010.

Simple Forecasting Methods

Moving Average Smoothing

- The obvious problem with a simple average is **figuring out how many points to use in the average calculation**
- One can select a window of the last “k” periods to calculate the average, and as the actual data grows over time, one can always take the last k samples to average, that is, n, n- 1, ..., n - k +1.
- When the actual data from January comes in, the February 2021 value is forecasted using January 2021 (n), December 2020 (n - 1) and November 2020 (n - 3 +1 or n - 2).

Simple Forecasting Methods

Weighted Moving Average Smoothing

- For some cases, the most recent value could have more influence than some of the earlier values.

$$F_{n+1} = (a \times y_n + b \times y_{n-1} + c \times y_{n-k})/(a + b + c) \tag{12.9}$$

Simple Forecasting Methods

Exponential Smoothing

- Exponential smoothing is the **weighted average of the past data**, with the recent data points given more weight than earlier data points.
- The **weights decay exponentially towards the earlier data points**, hence, the name. The exponential smoothing is given by the equation

$$F_{n+1} = \alpha y_n + \alpha(1-\alpha)y_{n-1} + \alpha(1-\alpha)^2 y_{n-2} + \cdots.$$

α is generally between 0 and 1

Simple Forecasting Methods

Exponential Smoothing

To forecast the future values using exponential smoothing, be rewritten as:

$$F_{n+1} = \alpha \times y_n + (1 - \alpha) \times F_n$$

Where Actual Value is yn and **forecasted value** Fn

E.g. **February 2011 forecast** using not only the actual January 2011 value **but also the previously forecasted January 2011 value,** the new forecast would have “learnt” the data better. This is the concept behind basic exponential smoothing .

The model is suited only for time series without clear trend or seasonality.

Simple Forecasting Methods : Exponential Smoothing

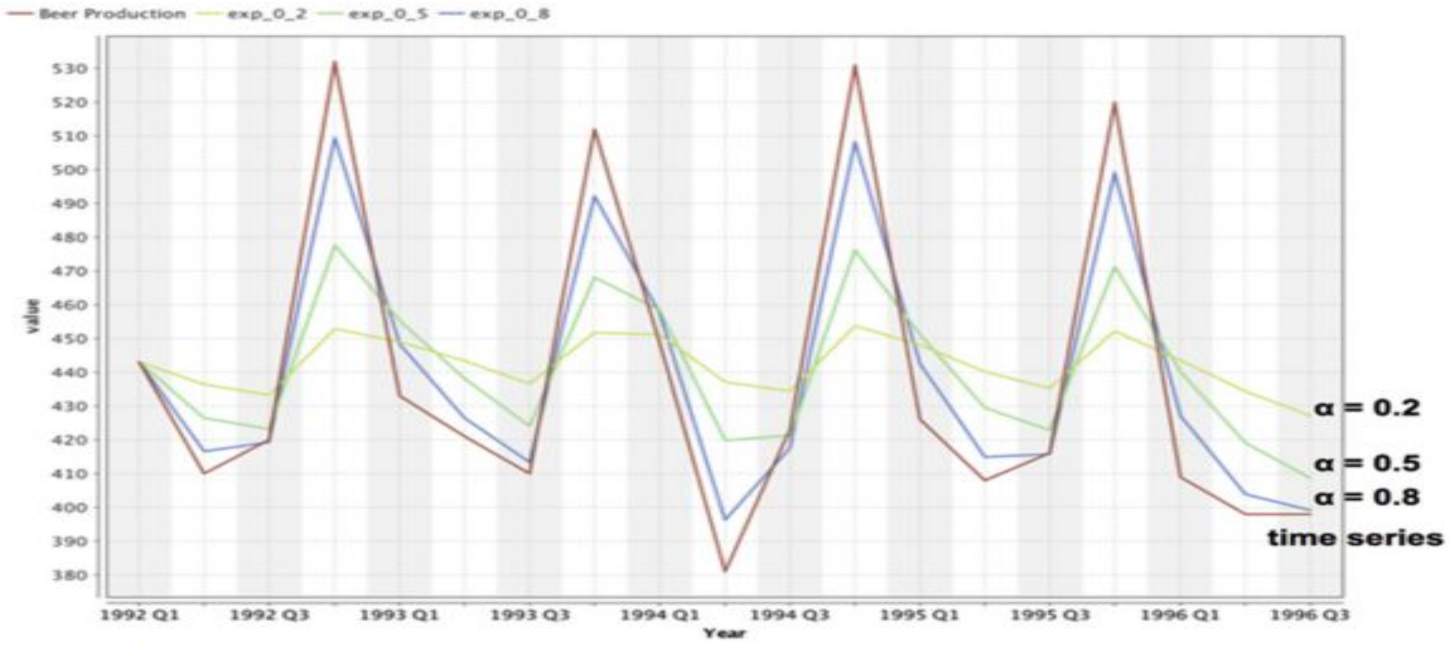


FIGURE 12.10

Fitted time series—exponential smoothing with different α levels.

Simple Forecasting Methods :Exponential Smoothing

- **Limitations:**
  - The f**orecasts cannot be made more than one-step ahead**, because to make a forecast for step (n 1 1), the data for the previous step, n, is needed.
  - It is **not possible to make forecasts several steps ahead,** that is, (n + h), using the methods described (where it was simply assumed that Fn+h = Fn+1), where h is the horizon.
  - For making longer horizon forecasts, that is, where h>=1, the **trend and seasonality information also needs to be considered**.

Simple Forecasting Methods :Exponential Smoothing

- **Holt's Two-Parameter Exponential Smoothing**
  - **A trend** is an **averaged long-term tendency** of a time series.
  - Model described earlier is **not particularly effective at capturing trends.**
  - **If the series also has a trend,** then **an average slope of the series** needs to be estimated as well.
  - Holt's two parameter smoothing does by means of another parameter, **β**. **A smoothing equation** similar to Eq. (12.10) is constructed for the **average trend at n + 1.** With two parameters, **α** and **β,** any time series with a trend can be modeled and, therefore, forecasted.
  - The f**orecast can be expressed as** a sum of these two components, **average value or "level" of the series, Ln, and trend, Tn, recursively as:**

Simple Forecasting Methods :Exponential Smoothing

- **Holt's Two-Parameter Exponential Smoothing**

$$F_{n+1} = L_n + T_n \tag{12.12}$$

where,

$$L_n = \alpha \times y_n + (1-\alpha) \times (L_{n-1} + T_{n-1}) \text{ and } T_n = \beta \times (L_n - L_{n-1}) + (1-\beta) \times T_{n-1} \tag{12.13}$$

To make future a forecast over an horizon, one can modify Equation 12.12 to:

$$F_{n+h} = L_n + h \times T_n \tag{12.14}$$

The values of the parameter can be estimated based on the best fit with the training (past) data

Simple Forecasting Methods :**Exponential Smoothing**

- **Holt-Winters’ Three-Parameter Exponential Smoothing**
  - When a **time series contains seasonality** in addition to a trend, yet another parameter, **γ, will be needed** to estimate the seasonal component of the time series .
  - The estimates for value (or level) are now adjusted by a **seasonal index**, which is computed with a **third equation that includes γ.**

$$F_{t+h} = (L_t + {}_hT_t)\, S_{t+h-p} \tag{12.15}$$

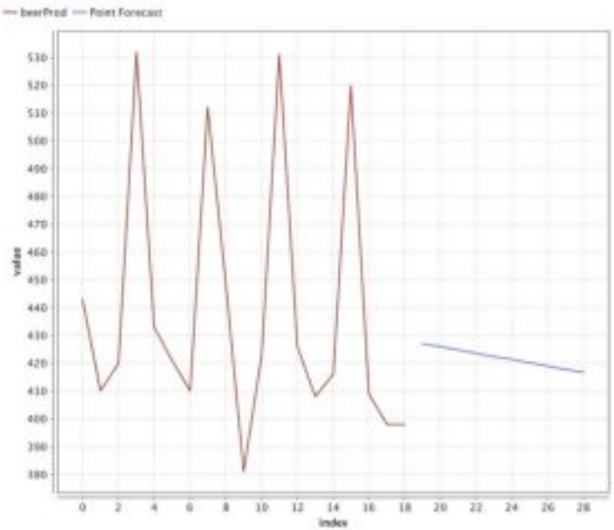
$$L_t = \alpha y_t / S_{t-p} + (1-\alpha)(L_{t-1} + T_{t-1}) \tag{12.16}$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \tag{12.17}$$

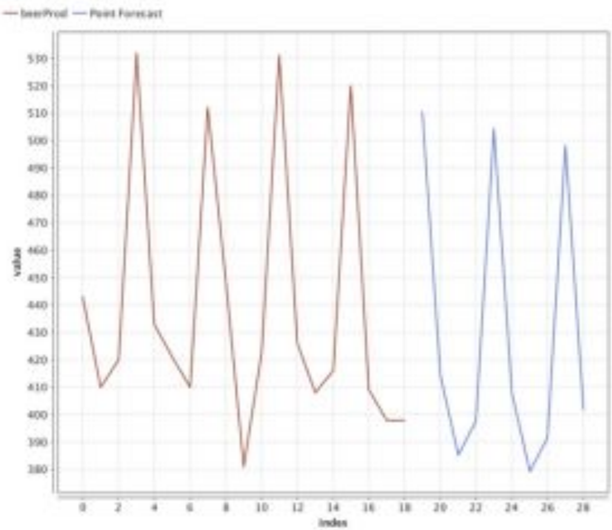
$$S_t = \gamma(y_t / L_t) + (1-\gamma)\, S_{t-p} \tag{12.18}$$

where p is the seasonality period. One can estimate the value of the parameters α, β, and γ from the fitting of the smoothing equation with the training data.

Simple Forecasting Methods :Exponential Smoothing



Holt's Two-Parameter Smoothing



Holt-Winters' Three-Parameter Smoothing

FIGURE 12.12

Holt's and Holt-Winters' smoothing.

TIME SERIES FORECASTING - Smoothing based forecasting

1. **Moving Averages based Smoothing:**

- Moving averages involve **taking the average of a fixed number of consecutive data points** in a time series, and then **using this average to predict future values** of the time series.
- The size of the **moving average window** (i.e., the number of data points included in the average) can be **adjusted to capture different levels of trend or seasonality in the data.**

2. **Exponential Smoothing:**

- Exponential smoothing, assigns exponentially decreasing weights to the historical values of the time series, with more recent values given greater weight.
- This technique is particularly useful for time series with trend and seasonality components that change over time.

TIME SERIES FORECASTING - Smoothing based forecasting

- They are **relatively simple to implement** and do not require extensive knowledge of advanced mathematical concepts.
- Additionally, they **can be applied to a wide range of time series data, including financial data, stock prices, and weather data.**

3. TIME SERIES FORECASTING - Regression based methods

3. TIME SERIES FORECASTING - Regression based methods

- Regression-based forecasting methods are
  - Linear regression,
  - Multiple regression,
  - Autoregressive integrated moving average (ARIMA) models.

TIME SERIES FORECASTING - Regression based method

- **Linear regression** is a simple form of regression-based forecasting, where the **target variable's value at time t (yt) is estimated using a linear function of time (t) with coefficients a and b.** The coefficients are estimated from a training set, and future values can be predicted using this model. yt = a * t + b
- Linear regression involves **fitting a straight line to the historical data o**f the time series, with the slope and intercept of the line representing the trend and intercept of the time series, respectively.
- This method is useful for **time series that have a linear trend over time.**

TIME SERIES FORECASTING - Regression based method

- The **variable time** is the predictor or independent variable and the **time series value** is the **dependent.**
- Regression based methods are generally **preferable when the time series appears to have a global pattern.**

TIME SERIES FORECASTING - Regression based method

- The **linear regression model is able to capture the long-term tendency** of the series, but it does a poor job of fitting the data.
- **Polynomial functions c**an be used to improve the fit. Polynomial regression is similar to linear regression except that higher-degree functions of the independent variable are used

TIME SERIES FORECASTING - Regression based method

- **Regression-based techniques** can become **more complex with the use of exponential, polynomial,** or power law functions to model the relationship between future values and time. These models can capture more complex patterns in the data and produce more accurate predictions.
- It involves **modeling the relationship between a time series and multiple predictor variables**, such as economic indicators or weather patterns. This method is useful for time series that are influenced by external factors.
- **Choice of method** depends on the nature of the data and the specific forecasting problem.

Autoregressive Integrated Moving Average(ARIMA)

- **Autocorrelation: Correlation measures how two variables are dependent on each other or if they have a linear relationship with each other.**

Year	prod	prod-1	prod-2	prod-3	prod-4	prod-5	prod-6
1992 Q1	443	?	?	?	?	?	?
1992 Q2	410	443	?	?	?	?	?
1992 Q3	420	410	443	?	?	?	?
1992 Q4	532	420	410	443	?	?	?
1993 Q1	433	532	420	410	443	?	?
1993 Q2	421	433	532	420	410	443	?
1993 Q3	410	421	433	532	420	410	443
1993 Q4	512	410	421	433	532	420	410
1994 Q1	449	512	410	421	433	532	420
1994 Q2	381	449	512	410	421	433	532
1994 Q3	423	381	449	512	410	421	433
1994 Q4	531	423	381	449	512	410	421
1995 Q1	426	531	423	381	449	512	410

FIGURE 12.18
Lag series and autocorrelation.

- Prod column shows simple time series data
- **1-lag series** - In the third column, data are lagged by one step. 1992 Q1 data is shown in 1992 Q2.
- additional **2-lag, 3-lag, ..., n-lag** series in the dataset.

A strong correlation between the original time series “prod” and 4-lag “prod-4.” They tend to move together. This is Autocorelation

Autoregressive Integrated Moving Average(ARIMA)

1.Autocorrelation:

- As in a **multivariate correlation matrix** , one can measure the **strength of correlation** between the original time series and all the lag series.
- The plot of the resultant correlation matrix is called an **Autocorrelation Function (ACF)** chart.
- The **ACF chart** is **used to study all the available seasonality in the time series.**
- From Fig. it can be concluded that the **time series is correlated with the 4th, 8th, and 12th lagged** quarter due to the yearly seasonality.
- It is also evident than Q1 is negatively correlated with Q2 and Q3.

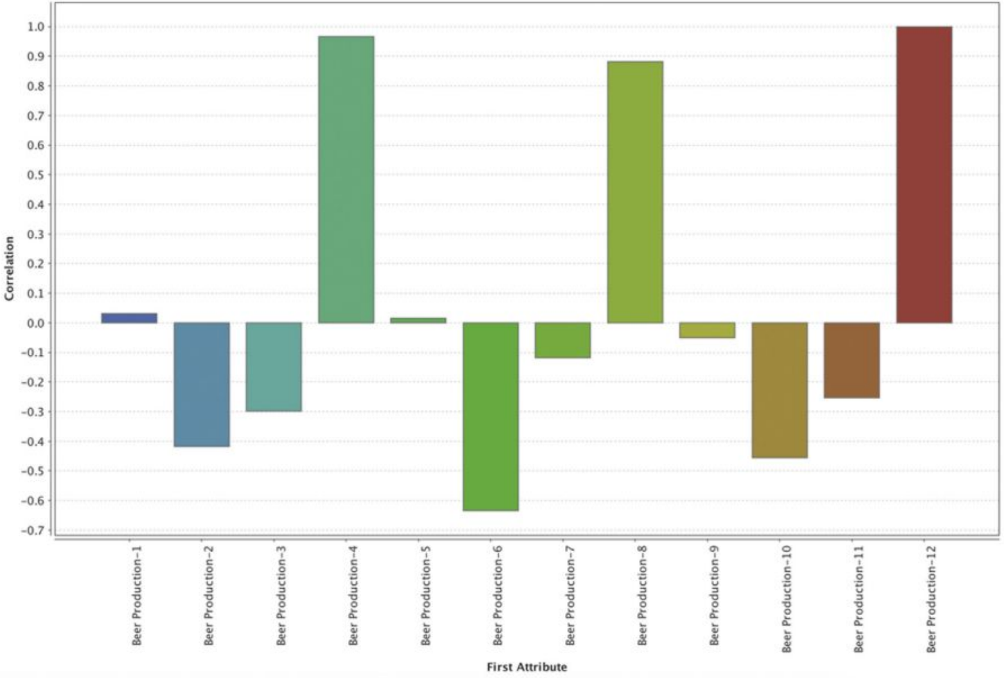


FIGURE 12.19
ACF chart. *ACF*, Autocorrelation Function.

ARIMA

ARIMA is an acronym that stands for AutoRegressive Integrated Moving Average. It is a generalization of the simpler AutoRegressive Moving Average and adds the notion of integration.

Let’s decode the essence of ARIMA:

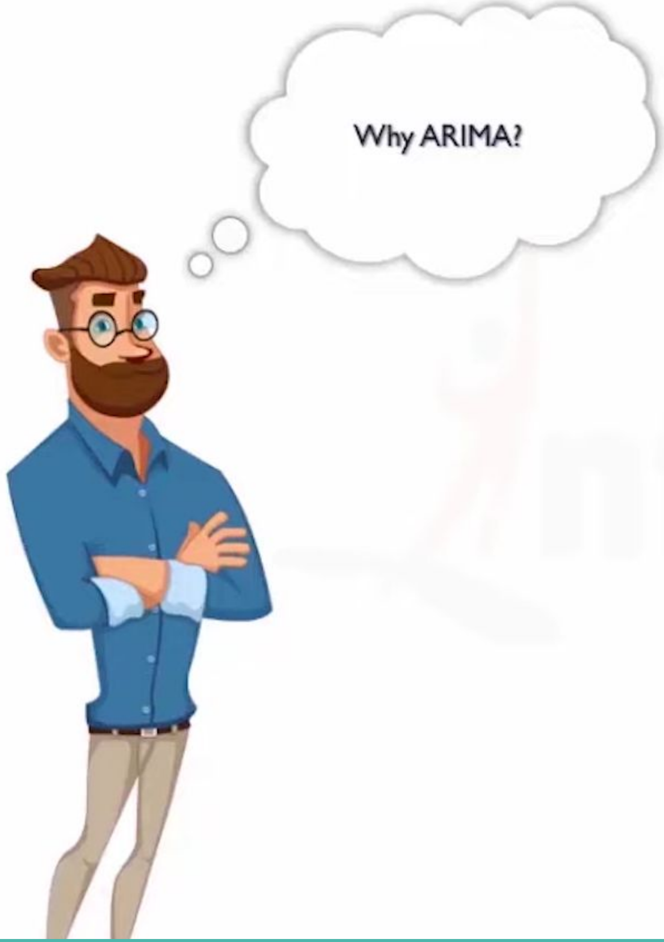
- **AR** (Autoregression): This emphasizes the dependent relationship between an observation and its preceding or ‘lagged’ observations.
- **I** (Integrated): To achieve a stationary time series, one that doesn’t exhibit trend or seasonality, differencing is applied. It typically involves subtracting an observation from its preceding observation.
- **MA** (Moving Average): This component zeroes in on the relationship between an observation and the residual error from a moving average model based on lagged observations.

Each of these components is explicitly specified in the model as a parameter. A standard notation is used for **ARIMA(p,d,q)** where the parameters are substituted with integer values to quickly indicate the specific ARIMA model being used.

ARIMA

The parameters of the ARIMA model are defined as follows:

- **p**: The lag order, representing the number of lag observations incorporated in the model.
- **d**: Degree of differencing, denoting the number of times raw observations undergo differencing.
- **q**: Order of moving average, indicating the size of the moving average window.

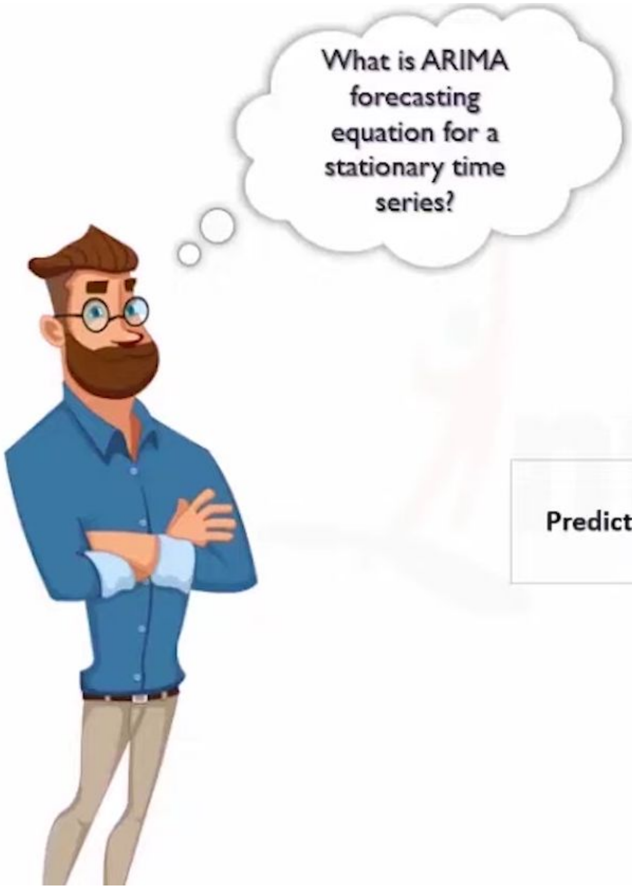


A class of statistical model for analyzing and forecasting time series data

ARIMA - **A**uto**R**egressive **I**ntegrated **M**oving **A**verage

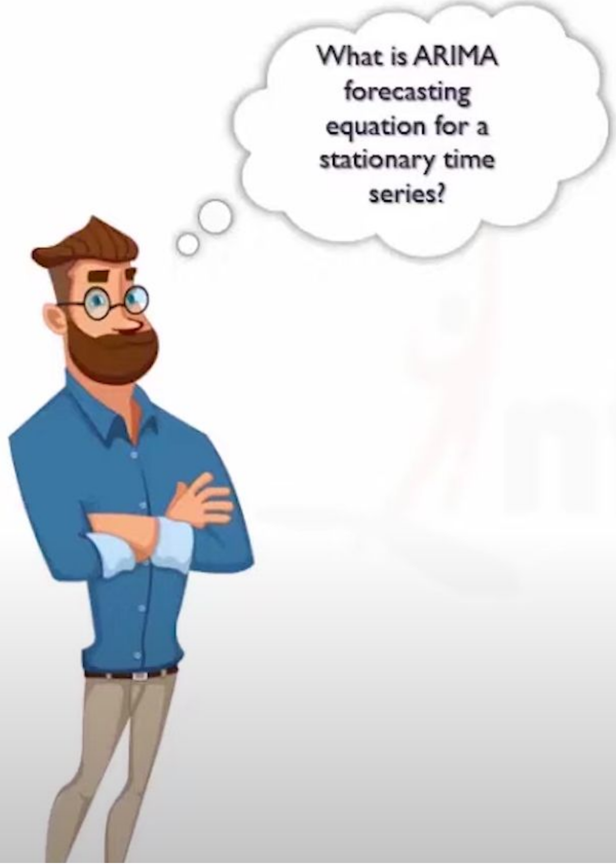
Data shows evidence of non-stationarity

A random variable that is a time series is stationary if its statistical properties are all constant over time.



A linear equation in which the predictors consist of lags of the dependent variable and/or lags of the forecast errors

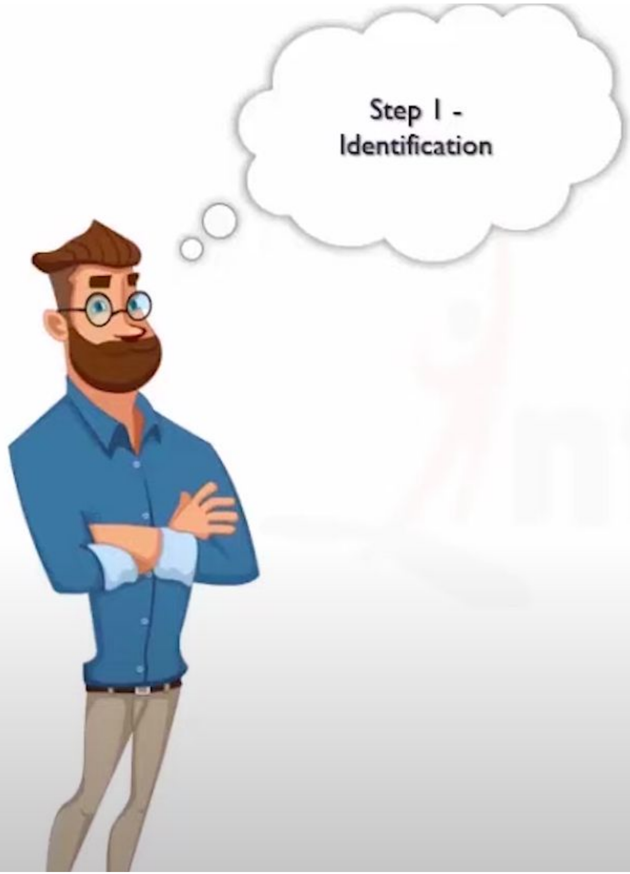
Predicted value of Y = **a constant**
a weighted sum of one or more recent values of Y
a weighted sum of one or more recent values of the errors.



AR = Auto**R**egressive - uses the dependent relationship between an observation and some number of lagged observations.

I = Integrated - The use of differencing of raw observations

MA = Moving **A**verage - uses the dependency between an observation and residual errors from a moving average model applied to lagged observations.



Assess whether the time series is stationary, and if not, how many differences are required to make it stationary

→

Identify the parameters of an ARMA model for the data

Autoregressive Integrated Moving Average(ARIMA)

Autoregression:

- **Autoregressive models** are regression models applied on lag series generated using the original time series.
- In **multiple linear regression**, the output is a linear combination of multiple input variables.
- In the case of autoregression models, the output is the future data point and it can be expressed as a linear combination for past p data points. **p is the lag window**

$$y_t = l + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \cdots + \alpha_p y_{t-p} + e$$

where, **l** is the **level in the dataset** and **e** is the **noise.** α are the coefficients that need to be learned from the data.

This can be referred to as an **autoregressive model with p lags or an AR(p) model**.

In an **AR(p) model**, lag series is a new predictor used to fit the dependent variable, which is still the original series value, Yt.

Autoregressive Integrated Moving Average(ARIMA)

Stationary Data:

In a time series with **trends or seasonality**, the **value is affected by time**

A time series is called stationary when the value of time series is not dependent on time.

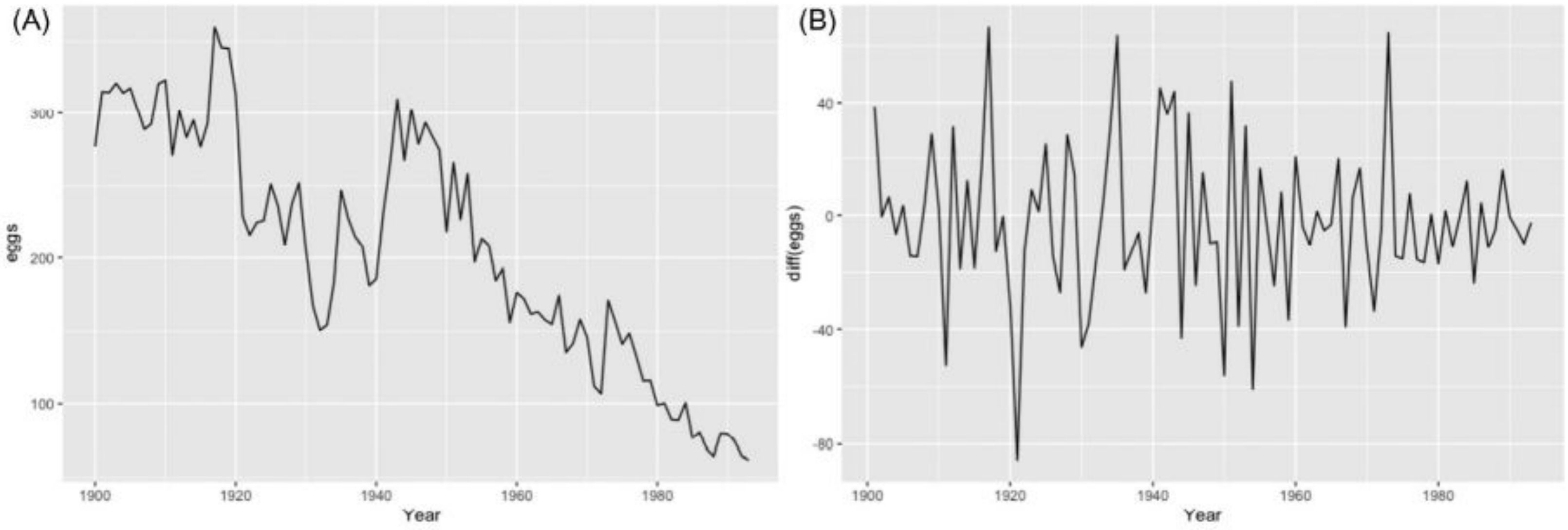


FIGURE 12.20

(A) Non-stationary Time series and (B) Stationary time series.

Random white noise is a stationary time series.

Autoregressive Integrated Moving Average(ARIMA)

Differencing:

- A **non-stationary time series** can be converted to a stationary time series through a technique called **differencing.**
- **Differencing series** is the **change between consecutive data points in the series.**

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

This is called **first order differencing.**

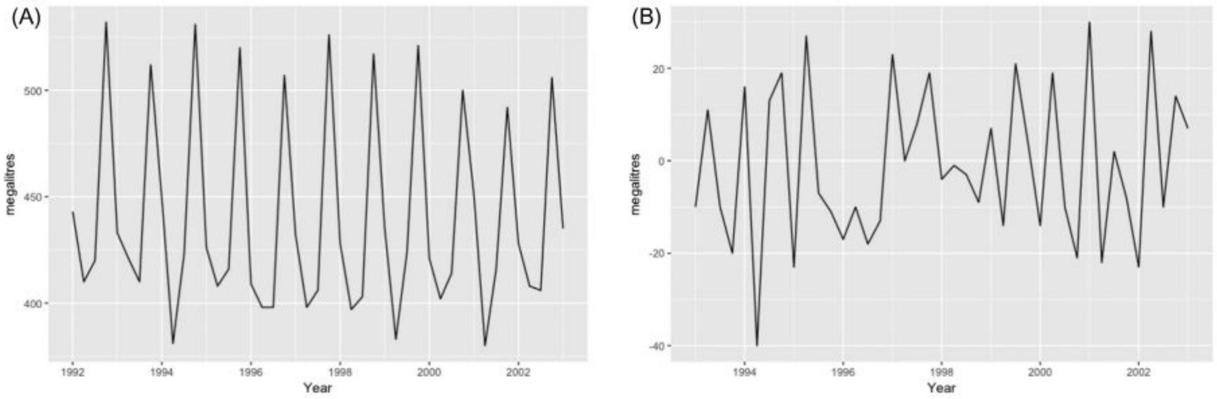


FIGURE 12.21

(A) Time series (B) Seasonal differencing of the time series.

Just differencing once will still yield a **nonstationary time series.**

Autoregressive Integrated Moving Average(ARIMA)

Differencing:

- Second order differencing is the **change between two consecutive data points** in a first order differenced time series.
- Differencing of order **d** is used to convert non stationary time series to stationary time series.
- **Seasonal differencing** is the **change between the same period in two different seasons**. Assume a season has period, m

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

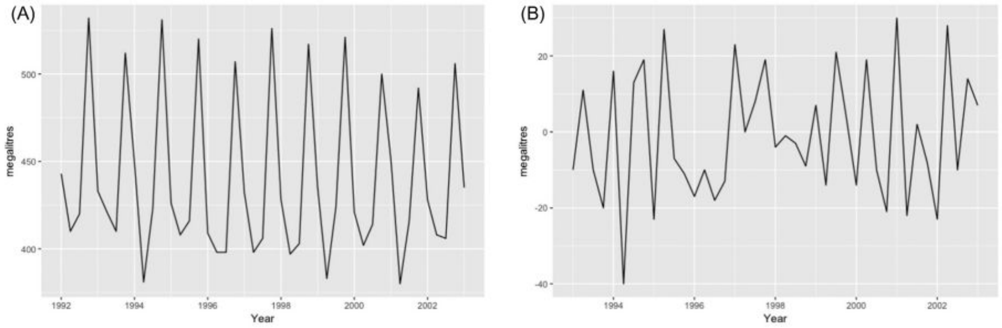


FIGURE 12.21

(A) Time series (B) Seasonal differencing of the time series.

Autoregressive Integrated Moving Average(ARIMA)

Moving Average Error

- Along with **Autoregressive Models** we can also use **forecast errors of past data** to predict the **future values**

$$y_t = l + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \cdots + \alpha_p y_{t-p} + e$$

Autoregressive Models

$$y_t = I + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} \qquad (12.22)$$

where **ei** is the **forecast error of data point i.**
(This makes sense for the past data points but not for data point t because it is still being forecasted.)
et is assumed as **white noise**. The regression equation for yt can be understood as the **weighted (θ) moving average of past q forecast errors**.

This is called **Moving Average with q lags model** or **MA(q)**

Autoregressive Integrated Moving Average(ARIMA)

6.Autoregressive Integrated Moving Average

- The Autoregressive Integrated Moving Average (ARIMA) model is a combination of the differenced autoregressive model with the moving average model. It is expressed as:

$$y_t = I + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \cdots + \alpha_p y_{t-p} + e$$

Autoregressive Models

+

$$y_t = I + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} \qquad (12.22)$$

The **AR part of ARIMA** shows that the **time series is regressed** on its own past data.
The **MA part of ARIMA** indicates that **the forecast error is a linear combination of past respective errors.**
The **I part** of ARIMA shows that the **data values have been replaced with differenced values of d order to obtain stationary data,** which is the requirement of the ARIMA model approach

$$y'_t = I + \alpha_1 y'_{t-1} + \alpha_2 y'_{t-2} + \cdots + \alpha_p y'_{t-p} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} \qquad (12.23)$$

Autoregressive Integrated Moving Average(ARIMA)

6.Autoregressive Integrated Moving Average

The **AR part of ARIMA** shows that the **time series is regressed** on its own past data.
The **MA part of ARIMA** indicates that **the forecast error is a linear combination of past respective errors.**
The **I part** of ARIMA shows that the **data values have been replaced with differenced values of d order to obtain stationary data**, which is the requirement of the ARIMA model approach

$$y'_t = I + \alpha_1 y'_{t-1} + \alpha_2 y'_{t-2} + \cdots + \alpha_p y'_{t-p} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} \qquad (12.23)$$

- The predictors are the **lagged p data points** for the **autoregressive part**
- The **lagged q errors** are for the **moving average part,**
- **The prediction** is the **difference yt** in the **dth order.**
- **ARIMA(p,d,q) model.**
    - Estimating the coefficients α and θ for a given p,d,q is what ARIMA does when it learns from the training data in a time series.

Autoregressive Integrated Moving Average(ARIMA)

- ARIMA is a generalized model.
- Some of the models are special cases of an ARIMA model.
- For example:

- ARIMA (0,1,0) is expressed as $y_t = y_{t-1} + e$. It is the naive model with error, which is called the Random walk model.
- ARIMA (0,1,0) is expressed as $y_t = y_{t-1} + e + c$. It is a random walk model with a constant trend. It is called random walk with drift.
- ARIMA (0,0,0) is $y_t = e$ or white noise
- ARIMA (p,0,0) is the autoregressive model

Seasonal ARIMA

The ARIMA model can be further enhanced to take into account of the seasonality in the time series. Seasonal ARIMA is expressed by the notion $ARIMA(p,d,q)(P,D,Q)_m$ where

p is the order of nonseasonal autoregression
d is the degree of differencing
q is the order of nonseasonal moving average of the error
P is the order of the seasonal autoregression
D is the degree of seasonal differencing
Q is the order of seasonal moving average of the error
m is the number of observations in the year (for yearly seasonality)

Seasonal ARIMA- How to Implement

- Decide : (p,d,q)(P,D,Q)m.
- The optimal parameters for the Beer production dataset is **ARIMA(1,0,0)(1,1,0)4**. The seasonal ARIMA model is used to forecast the future 12 data points using the forecast() function

TIME SERIES FORECASTING - Regression based method

- A more sophisticated technique is **based on the concept of autocorrelation.** Autocorrelation refers to the fact that **data from adjacent time periods are correlated** in a time series. The most well-known among these techniques is **ARIMA,** which stands for Auto Regressive Integrated Moving Average.
- ARIMA models, are a type of time series model that **combines autoregression** (i.e., predicting future values based on past values) and **moving averages (i**.e., smoothing out the noise in the data) to model the underlying trend and seasonality in a time series.
- This method is useful for **time series that exhibit non-linear trends or seasonality.**

TIME SERIES FORECASTING - Regression based method

- **Advantages:**
- They can **capture the influence of external factors on the time series,** making them useful for forecasting in a variety of industries, including finance and economics.
- Additionally, **they can handle time series with non-linear trends or seasonality,** which may not be captured by other forecasting methods.
- **Limitations/Drawbacks:**
- Regression-based forecasting methods also have some limitations. They assume that the relationship between the time series and predictor variables remains constant over time, which may not always be the case.
- Additionally, they can be **computationally intensive, requiring large amounts of data and computational resources**.
- **Applications**

  They are commonly used in **finance to forecast stock prices** and other financial indicators.

  They are also used in **marketing to forecast sales and customer demand**.

4. TIME SERIES FORECASTING-Machine Learning Based Methods

4. TIME SERIES FORECASTING-Machine Learning Based Methods

Machine Learning Models

- The series is **transformed into cross-sectional data** using a technique called **windowing**.
- This technique defines a **set of consecutive time series data** as a window, where the latest record forms the **target** while other series data points, which are lagged compared to the target, form the input variables.

FIGURE 12.27
Machine learning model for time series.

- Sufficient number of **windows are extracted** from the dataset,
- A **supervised model** can be learned based on the inferred relationship between the lagged input variables and the target variable.

Windowing

- The purpose of windowing is **to transform the time series data** into a **generic machine learning** input dataset.

Date	inputYt
Jan 1, 2009	0.709
Feb 1, 2009	1.886
Mar 1, 2009	1.293
Apr 1, 2009	0.822
May 1, 2009	-0.173
Jun 1, 2009	0.552
Jul 1, 2009	1.169
Aug 1, 2009	1.604
Sep 1, 2009	0.949
Oct 1, 2009	0.080
Nov 1, 2009	-0.040
Dec 1, 2009	1.381
Jan 1, 2010	0.761
Feb 1, 2010	2.312
Mar 1, 2010	1.795
Apr 1, 2010	0.586
May 1, 2010	-0.077
Jun 1, 2010	0.613
Jul 1, 2010	1.845

(A) Time series data set

Window id	inputYt + 1 (horizon)	inputYt - 5	inputYt - 4	inputYt - 3	inputYt - 2	inputYt - 1	inputYt - 0
0	1.169	0.709	1.886	1.293	0.822	-0.173	0.552
1	1.604	1.886	1.293	0.822	-0.173	0.552	1.169
2	0.949	1.293	0.822	-0.173	0.552	1.169	1.604
3	0.080	0.822	-0.173	0.552	1.169	1.604	0.949
4	-0.040	-0.173	0.552	1.169	1.604	0.949	0.080
5	1.381	0.552	1.169	1.604	0.949	0.080	-0.040
6	0.761	1.169	1.604	0.949	0.080	-0.040	1.381
7	2.312	1.604	0.949	0.080	-0.040	1.381	0.761
8	1.795	0.949	0.080	-0.040	1.381	0.761	2.312
9	0.586	0.080	-0.040	1.381	0.761	2.312	1.795
10	-0.077	-0.040	1.381	0.761	2.312	1.795	0.586
11	0.613	1.381	0.761	2.312	1.795	0.586	-0.077
12	1.845	0.761	2.312	1.795	0.586	-0.077	0.613
13	1.984	2.312	1.795	0.586	-0.077	0.613	1.845
14	1.861	1.795	0.586	-0.077	0.613	1.845	1.984
15	0.661	0.586	-0.077	0.613	1.845	1.984	1.861

(B) Cross-sectional data set

FIGURE 12.28

Windowing process. (A) original time series and (B) cross-sectional data set with consecutive windows.

Windowing

1. *Window Size*: Number of lag points in one window excluding the target data point.
2. *Step*: Number of data points between the first value of the two consecutive windows. If the step is 1, maximum number of windows can be extracted from the time series dataset.
3. *Horizon width*: The prediction horizon controls how many records in the time series end up as the target variable. The common value for the horizon width is 1.
4. *Skip*: Offset between the window and horizon. If the skip is zero, the consecutive data point(s) from the window is used for horizon.

In Fig. 12.28, the window size is 6, step is 1, horizon width is 1, and skip is 0.

Windowing

- The series data are now converted into a **generic cross-sectional dataset** that can be predicted with **learning algorithms** like **regression, neural networks, or support vector machines.**

How to Implement

- (1) conversion to cross-sectional data,
- (2) training an machine learning model, and
- (3) forecasting one data point at a time in a loop.

Step 1: Set Up Windowing

- The operator must be informed that one of the columns in the dataset is a date and should be considered as an "id." - Set Role operator.
- If the input data has multiple time series, Select
- Attributes operator can be used to select the one to be forecasted.

1. *Window size:* Determines how many "attributes" are created for the cross-sectional data. Each row of the original time series within the window size will become a new attribute. In this example, $w = 6$ was chosen.
2. *Step size:* Determines how to advance the window. $s = 1$ was used.
3. *Horizon width:* Determines how far out to make the forecast. If the window size is 6 and the horizon is 1, then the seventh row of the original time series becomes the first sample for the "label" variable. $h = 1$ was used.

Step 1: Set Up Windowing

1. *Window size:* Determines how many "attributes" are created for the cross-sectional data. Each row of the original time series within the window size will become a new attribute. In this example, $w = 6$ was chosen.
2. *Step size:* Determines how to advance the window. $s = 1$ was used.
3. *Horizon width:* Determines how far out to make the forecast. If the window size is 6 and the horizon is 1, then the seventh row of the original time series becomes the first sample for the "label" variable. $h = 1$ was used.

Step 2: Train the Model

- When training any supervised model using this data, the attributes labeled input Yt-5 through input Yt-0 form the independent variables.
- linear regression is used to fit the dependent variable called label, using the independent variables input Yt-5 through input Yt-0.

$$\text{label} = 0.493 \times \text{input } Y_{t-5} + 0.258 \times \text{input } Y_{t-4} + 0.107 \times \text{input } Y_{t-3} - 0.098 \times \text{input } Y_{t-2} - 0.073 \times \text{input } Y_{t-1} + 0.329 \times \text{input } Y_{t-0} + 0.135$$

Step 3: Generate the Forecast in a Loop

- Note that given this configuration of the window size and horizon, one

can now only make the forecast for the next step.
The regression equation is be used to predict December 2011 value.

Step 3: Generate the Forecast in a Loop

Date	prediction(l...	inputYt-5	inputYt-4	inputYt-3	inputYt-2	inputYt-1	inputYt-0
Dec 1, 2011...	1.694	1.201	2.466	2.497	2.245	1.179	1.119
Jan 1, 2012 ...	2.597	2.466	2.497	2.245	1.179	1.119	1.694
Feb 1, 2012...	2.693	2.497	2.245	1.179	1.119	1.694	2.597
Mar 1, 2012...	2.196	2.245	1.179	1.119	1.694	2.597	2.693
Apr 1, 2012...	1.457	1.179	1.119	1.694	2.597	2.693	2.196
May 1, 201...	1.457	1.119	1.694	2.597	2.693	2.196	1.457
Jun 1, 2012 ...	2.087	1.694	2.597	2.693	2.196	1.457	1.457
Jul 1, 2012 ...	2.784	2.597	2.693	2.196	1.457	1.457	2.087
Aug 1, 2012...	2.807	2.693	2.196	1.457	1.457	2.087	2.784
Sep 1, 2012...	2.265	2.196	1.457	1.457	2.087	2.784	2.807
Oct 1, 2012...	1.720	1.457	1.457	2.087	2.784	2.807	2.265
Nov 1, 2012...	1.816	1.457	2.087	2.784	2.807	2.265	1.720
Dec 1, 2012...	2.433	2.087	2.784	2.807	2.265	1.720	1.816
Jan 1, 2013 ...	2.974	2.784	2.807	2.265	1.720	1.816	2.433
Feb 1, 2013...	2.911	2.807	2.265	1.720	1.816	2.433	2.974

FIGURE 12.30

Forecasting one step ahead.

Neural Network Autoregressive

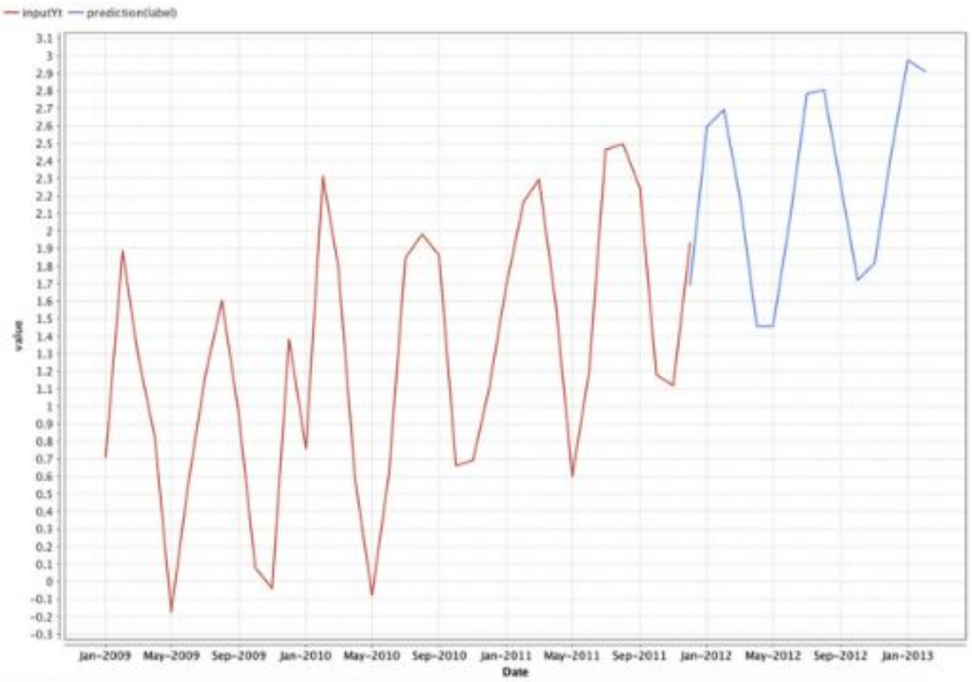


FIGURE 12.32
Forecasted time series.

Neural Network Autoregressive

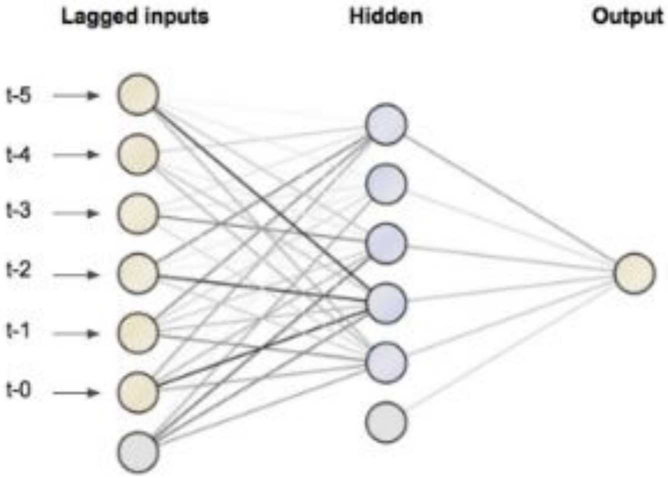


FIGURE 12.33
Neural network autoregressive model.

Evaluation of Time Series Models

between the actual value $\mathbf{y_i}$ and the forecasted value $\hat{\mathbf{y}}_\mathbf{i}$. The error or the residue for the i^{th} data point is given by Eq. (12.24)

$$e_i = y_i - \hat{y}_i \tag{12.24}$$

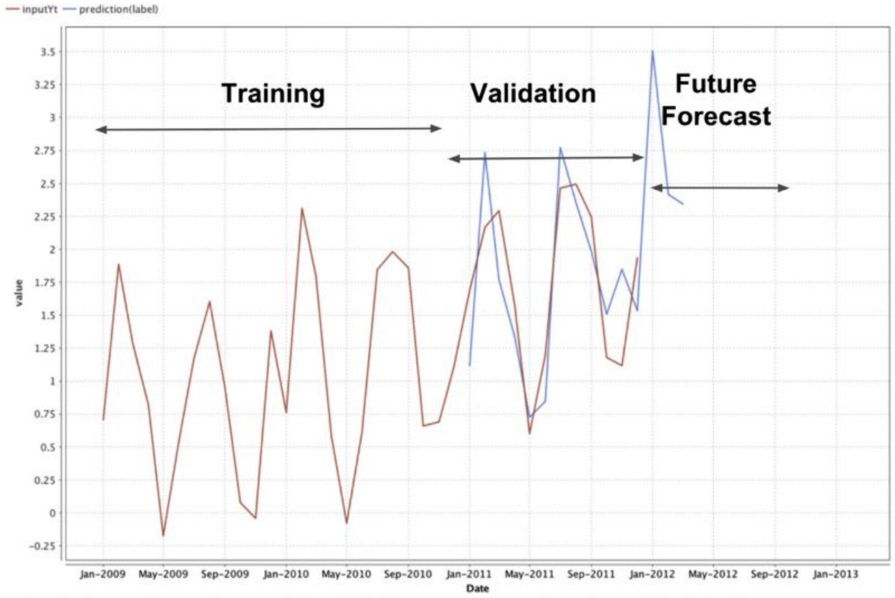


FIGURE 12.36
Validation dataset.

Mean Absolute Error

- The error of the individual data point may be positive or negative and may cancel each other out. To derive the overall forecast for the model, calculate the absolute error to aggregate all the residuals and average it.
- $$\text{Mean absolute error} = \text{mean}(|e_i|) \qquad (12.25)$$

Root Mean Squared Error

- In some cases it is **advantageous to penalize the individual point error with higher residues.** Even though two models have the same MAE, one might have consistent error and the other might have low errors for some points and high error for other points. RMSE penalizes the latter.
- $$\text{Root mean squared error} = \sqrt{\text{mean}\,(e^2)} \tag{12.26}$$

Mean Absolute Percentage Error

Percentage error of a data point is $p_i = 100\ \frac{e_i}{y_i}$. It is a scale independent error that can be aggregated to form mean absolute percentage error.

$$\text{Mean absolute percentage error} = \text{mean}(|p_i|) \qquad (12.27)$$

MAPE is useful to compare against multiple models across the different forecasting applications. For example, the quarterly revenue forecast, measured in USD, for a car brand might be $\pm 5\%$ and the forecast for world-wide car demand, measured in quantity, might be $\pm 3\%$. The firm's ability to forecast the car demand is higher than the revenue forecast for one brand. Even

Mean Absolute Scaled Error

MASE is scale independent and overcomes the key limitations of MAPE by comparing the forecast values against a naive forecast. Naive forecast is a simple forecast where the next data point has the same value as the previous data point (Hyndman & Koehler, 2006). Scaled error is defined as:

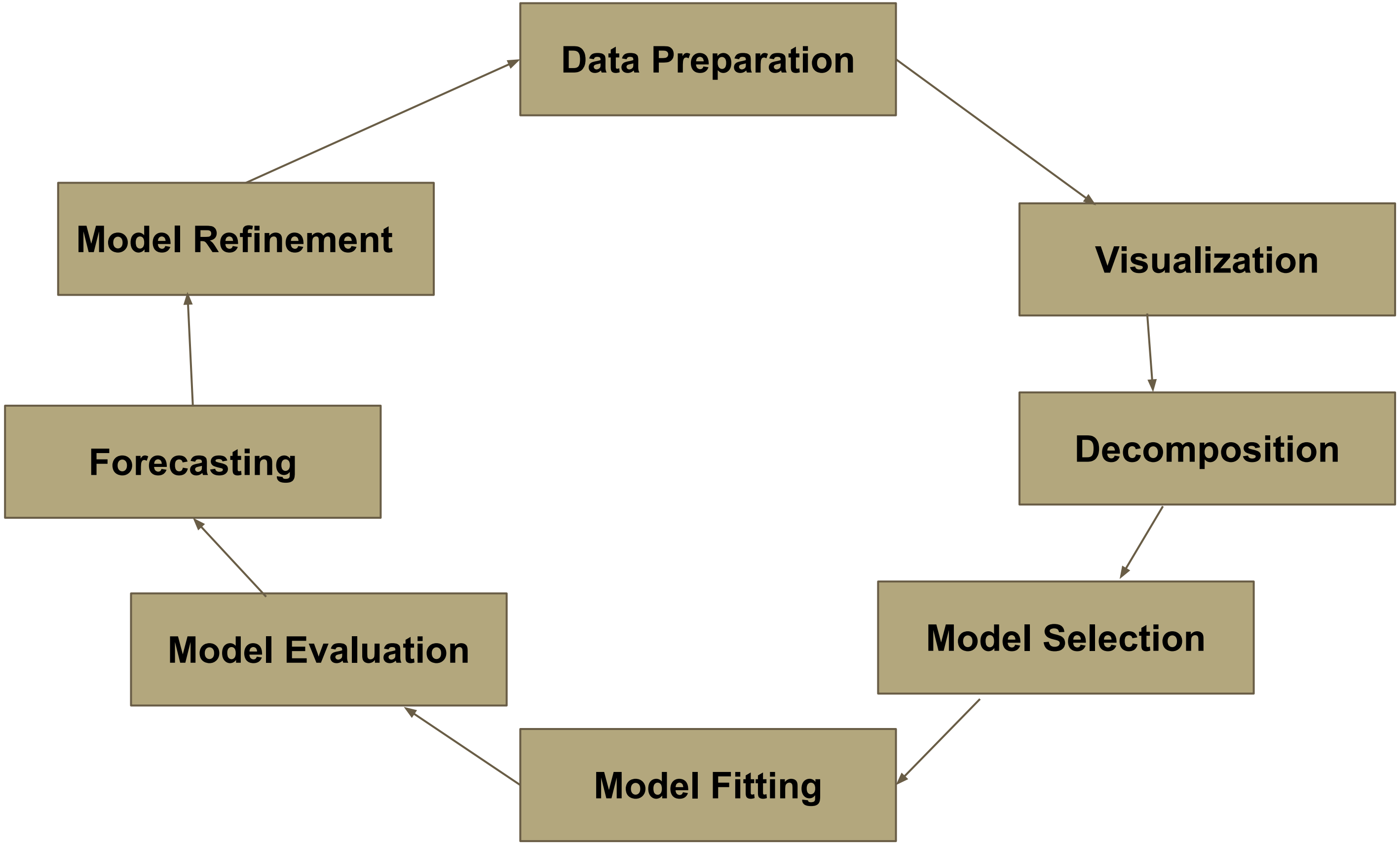
$$\text{MASE} = \frac{\sum_{i=1}^{T} |e|}{\frac{T}{T-1}\sum_{i=2}^{T} |\overline{y}_i - \overline{y}_{i-1}|} \tag{12.28}$$

T is the total number of data points. Scaled error is less than one if the forecast is better than naive forecast and greater than one if it is worse than naive

Sliding Window Validation

- Sliding window validation is a process of **backtesting time series models** built through machine learning based methods.
- The whole cross-sectional dataset is divided into different training windows by specifying the **window width.**
- A model is trained using a **training window** and **applied on the testing window** to compute the performance for the first run.
  For the next run, the training
- **window is slid to new set of training records** and the process is repeated until all the training windows are used.
- By this technique, **an average performance metric can be calculated** across the entire dataset.
- The performance metric derived through sliding window validation is generally more robust than split validation technique.

Revision - steps to analyze time series dataset



Revision - steps to analyze time series dataset

- **Data Preparation -** It involves **cleaning, transforming and organizing** data including handling missing values and outliers.
- **Visualization-** It involves **creating graphs and plots** of the data to help identify trends, patterns , relationships, stability and stationarity of the data.
- **Decomposition -** breaking down the data into its **component parts such as trend, seasonality and residuals** to better understand the structure of underlying data.
- **Model Selection -** selecting appropriate model to use for the analysis such as simple moving average, exponential smoothing model, ARIMA , ML models etc.

Revision - steps to analyze time series dataset

- **Model Fitting -** This involves fitting the selected model to the data, adjusting its parameters to maximize the accuracy of the forecasts.
- **Model evaluation -** This involves evaluating the performance of the model using measures such as mean absolute error etc.
- **Forecasting-** This involves using the fitted model to make predictions about future values of the variable based on past patterns and relationships.
- **Model Refinement -** This involves fine tuning the model, to improve the accuracy and reliability and to incorporate new information as it becomes available.

Forecasting Best Practices:

1. *Understand the metric*: Investigate how the time series metric is derived. Is the metric influenced by other metrics or phenomenon that can be better candidates for the forecasting? For example, instead of forecasting profit, both revenue and cost can be forecasted, and profit can be calculated. This is particularly suitable when profit margins are low and can go back and forth between positive and negative values (loss).
2. *Plot the time series*: A simple time series line chart reveals a wealth of information about the metric being investigated. Does the time series have a seasonal pattern? Long-term trends? Are the seasonality and trend linear or exponential? Is the series stationary? If the trend is exponential, can one derive log() series? Aggregate daily data to weeks and months to see the normalized trends.
3. *Is it forecastable*: Check if the time series is forecastable using stationary checks.

Forecasting Best Practices:

4. *Decompose*: Identify trends and seasonality using decomposition methods. These techniques show how the time series can be split into multiple meaningful components.
5. *Try them all*: Try several different methods mentioned in the forecasting taxonomy in Fig. 12.3, after splitting, training, and validation samples. For each method:
   a. Perform residual checks using MAE or MAPE metric.
   b. Evaluate forecasts using the validation period.
   c. Select the best performing method and parameters using optimization functions.
   d. Update the model using the full dataset (training + validation) for future forecasts.
6. *Maintain the models*: Review models on a regular basis. Time series forecast models have a limited shelf life. Apart from feeding the latest data to the model, the model should be refreshed to make it relevant for the latest data. Building a model daily is not uncommon.

Questions

- What are the features of time series data set?
- What is difference between univariate and Multivariate time series dataset?
- Given a time series dataset, how to perform time series decomposition ?
- Explain the steps for time series data analysis.
- What are the evaluation measures used for evaluating performance of time series models?