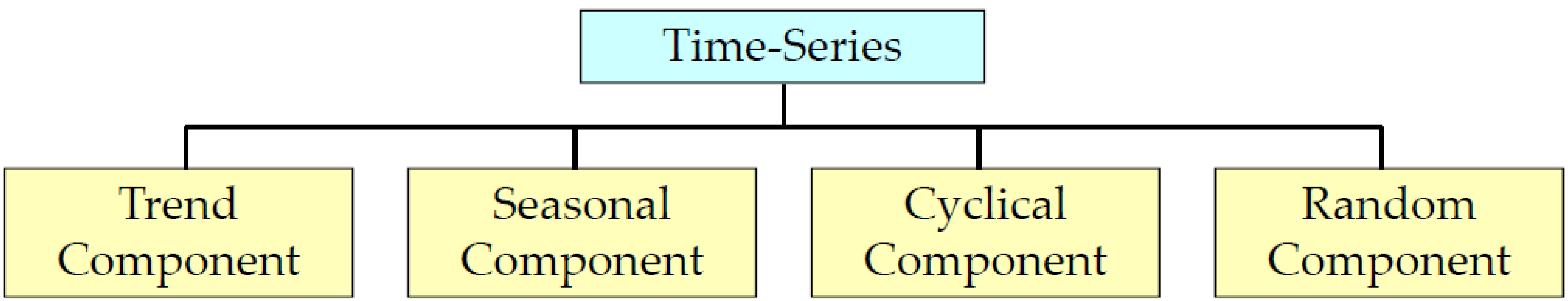
M7
Predictive Analytics

ENİS KAYIŞ

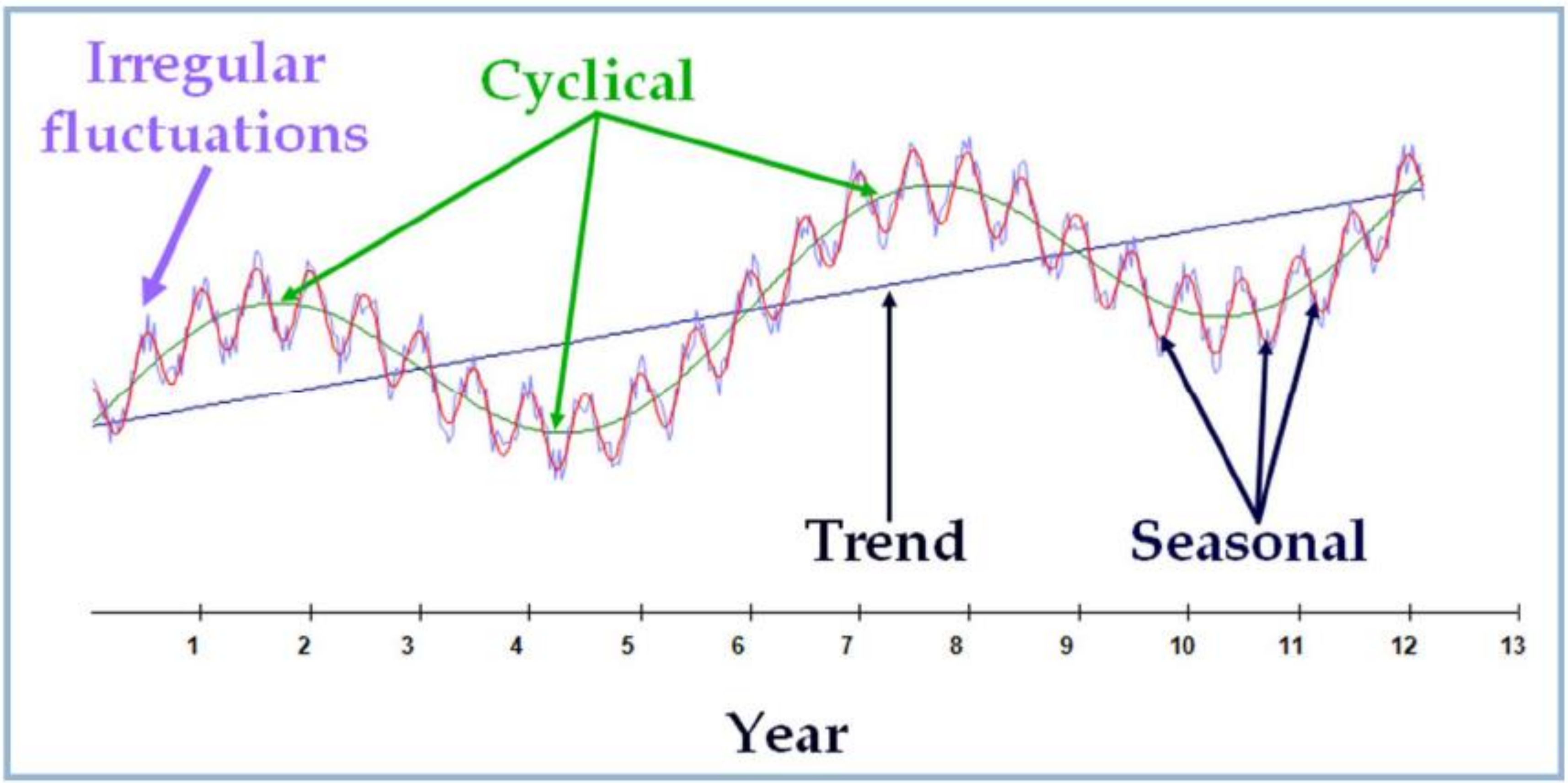
Time Series and Their Components

- Decomposition
- It is an approach to the analysis of time series data involves an attempt to identify the component factors that influence each of the values in a series.
- The components of time series are:



Time Series and Their Components

- Components of a time series



Time Series Decomposition

$$y_t = f(S_t, T_t, R_t)$$

where

- y_t = data at period t
- T_t = trend-cycle component at period t
- S_t = seasonal component at period t
- R_t = remainder component at period t

- **Additive decomposition:** $y_t = S_t + T_t + R_t$
- **Multiplicative decomposition:** $y_t = S_t * T_t * R_t$

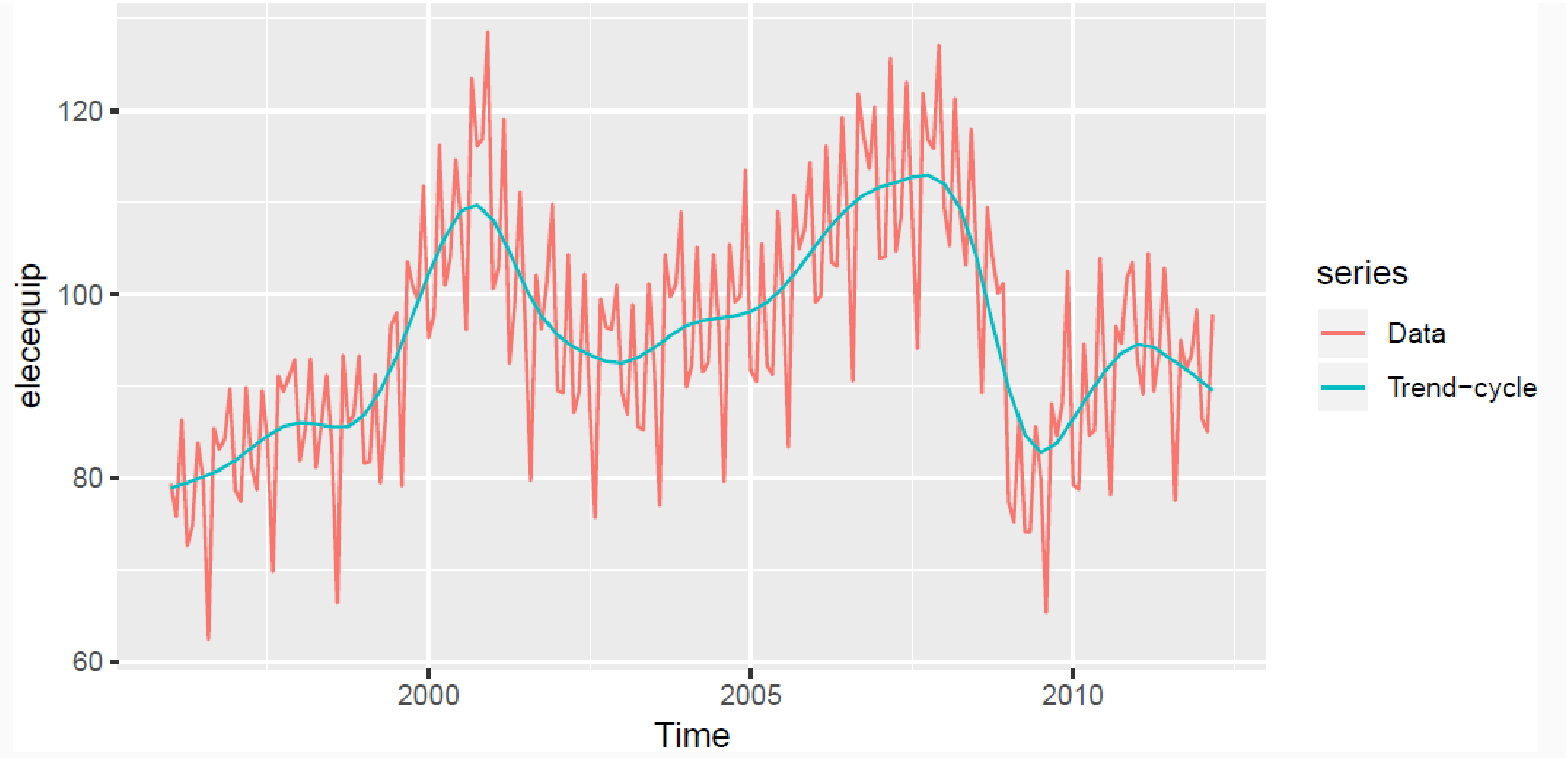
Time Series Decomposition

- Additive model appropriate if magnitude of seasonal fluctuations does not vary with level.
- If seasonal are proportional to level of series, then multiplicative model appropriate.
- Multiplicative decomposition more prevalent with economic series
- Alternative: use a Box-Cox transformation, and then use additive decomposition.
- Logs turn multiplicative relationship into an additive relationship:

$$y_t = S_t \times T_t \times E_t \quad \Rightarrow \quad \log y_t = \log S_t + \log T_t + \log R_t.$$

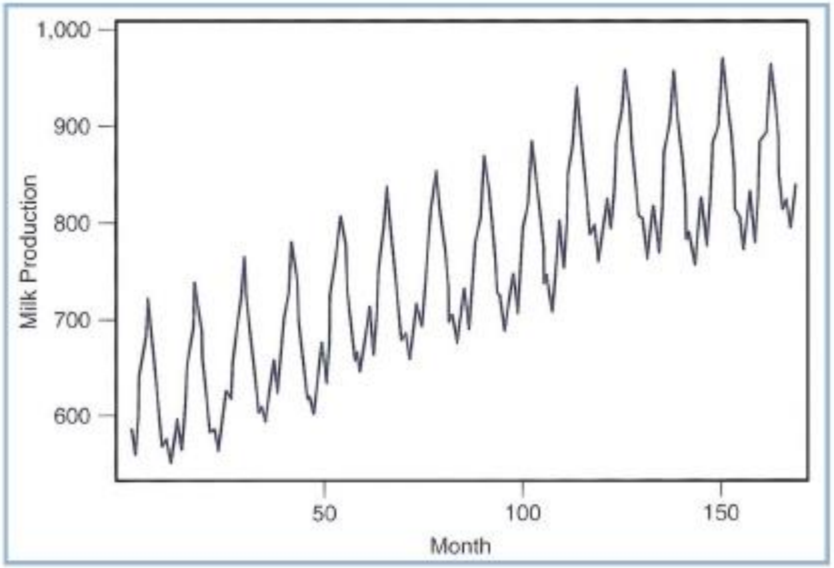
Time Series Decomposition

- Euro electrical equipment

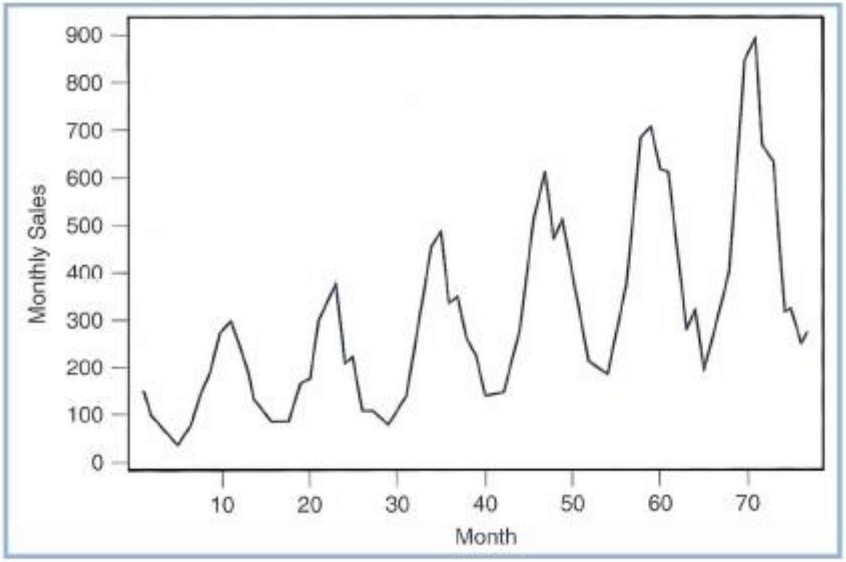


Time Series and Their Components

- Compare these two time series:



Time series with constant variability



Time series with increasing variability

History of time series decomposition

- Classical method originated in 1920s.
- Census II method introduced in 1957. Basis for X-11 method and variants (including X-12-ARIMA, X-13-ARIMA)
- STL method introduced in 1983.
- TRAMO/SEATS introduced in 1990s.

Time Series and Their Components

Estimation of Time Series Components

- Estimation of Trend Component
  - Trends are long term movements in a time series that can be sometimes be described by a straight line or a smooth curve.

Remark

- Fitting a trend curve helps us in providing some indication of the general direction of the observed series, and in getting a clear picture of the seasonality after removing the trend from the original series.

Time Series and Their Components

- The Linear Trend

$$\hat{T}_t = b_0 + b_1 t$$

- The Quadratic Trend

$$\hat{T}_t = b_0 + b_1 t + b_2 t^2$$

- The Exponential Trend

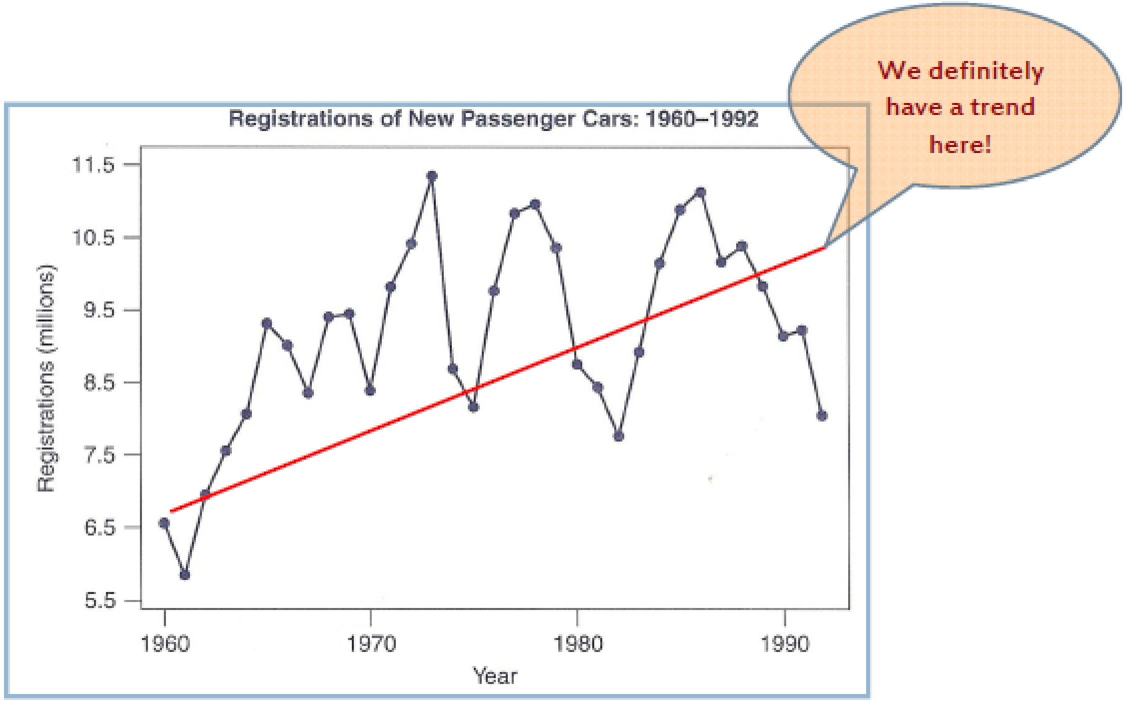
$$\hat{T}_t = b_0 b_1{}^t$$

  - $\hat{T}_t$: predicted value of the trend at time t
  - b_0, b_1, b_2: model parameters
- We can forecast the trend using the above models as and so on.
- Model parameters are selected so as to minimize SSE.

Time Series and Their Components

Example:

- Data on annual registrations of new passenger cars in the United States from 1960 to 1992.

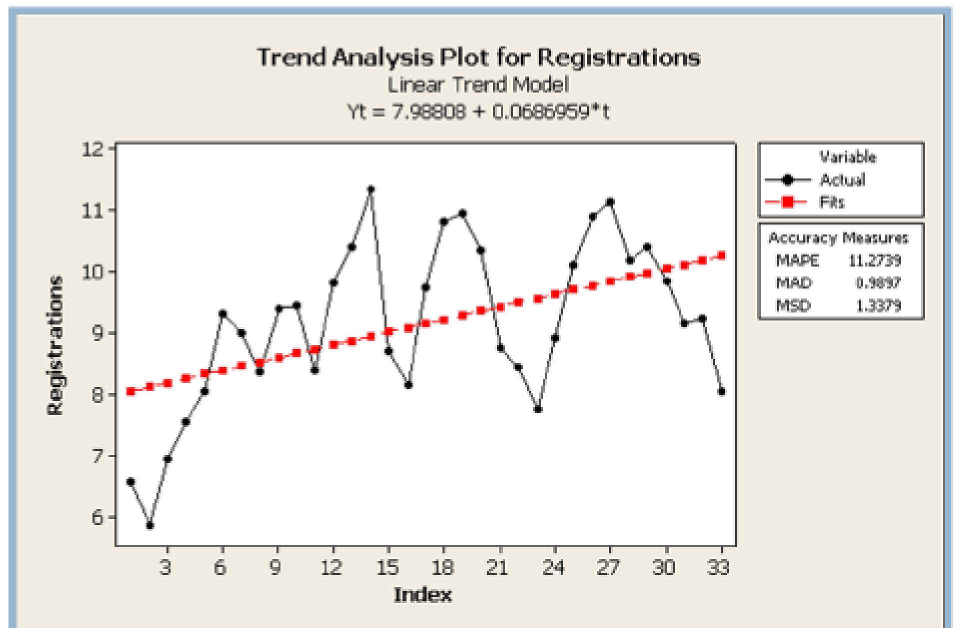


Time Series and Their Components

- The values from 1960 to 1992 are used to develop the trend equation.
- Registrations is the dependent variable, and the independent variable is time t coded as 1960 = 1, 1961 = 2, and so on. The fitted trend line has the equation:

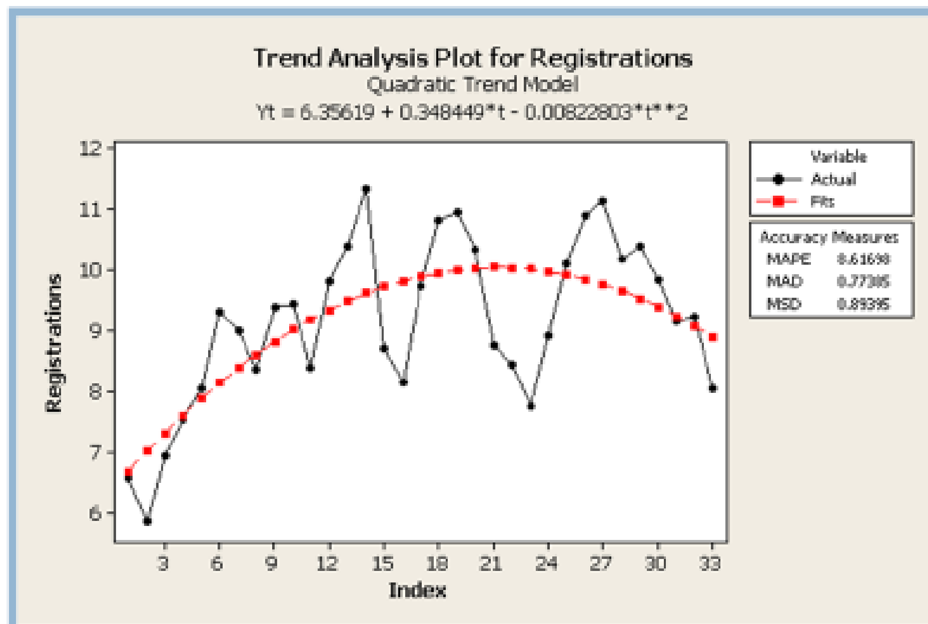
$$\hat{T}_t = 7.988 + 0.0687t$$

- The slope of the trend equation indicates that registrations are estimated to increase an average of 68,700 each year.



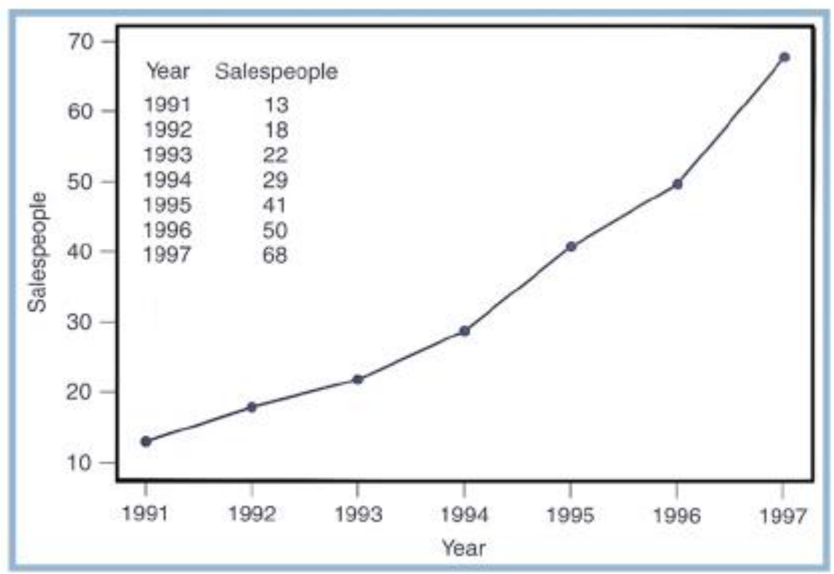
Time Series and Their Components

- Forecasting a Trend
  - Which trend model is appropriate?
    - Linear, quadratic or exponential
  - Linear models assume that a variable is increasing (or decreasing) by a constant amount each time period.
  - A quadratic curve is needed to model the trend.
  - Based on the accuracy measures, a quadratic trend appears to be a better representation of the general direction of the data.



Time Series and Their Components

- When a time series starts slowly and then appears to be increasing at an increasing rate such that the percentage difference from observation to observation is constant, an exponential trend can be fitted.
- The coefficient b_1 is related to the growth rate.
- If the exponential trend is fit to annual data, the annual growth rate is estimated to be $100(b_1 — 1)\%$.
- The figure next contains the number of mutual fund salespeople for several consecutive years.
- The increase in the number of salespeople is not constant.
- It appears as if increasingly larger numbers of people are being added in the later years.

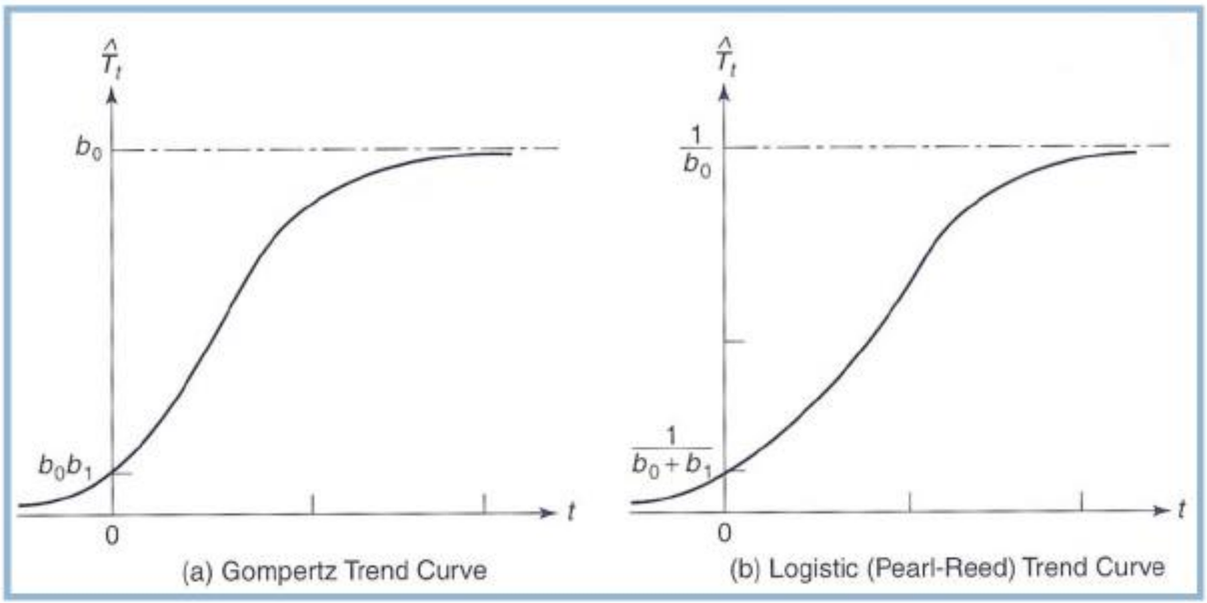


Time Series and Their Components

- A linear trend fit to the salespeople data would indicate a constant average increase of about nine salespeople per year.
- This trend overestimates the actual increase in the earlier years and underestimates the increase in the last year.
- It does not model the apparent trend in the data as well as the exponential curve.
- It is clear that extrapolating an exponential trend with a 31 % growth rate will quickly result in some very big numbers.
- This is a potential problem with an exponential trend model.
- What happens when the economy cools off and stock prices begin to retreat?
- The demand for mutual fund salespeople will decrease and the number of salespeople could even decline.
- The trend forecast by the exponential curve will be much too high.

Time Series and Their Components

- Growth curves of the Gompertz and logistic types reflect a situation in which sales begin low, then increase as the product catches on, and finally ease off as saturation is reached.
- Judgment and common sense are very important in selecting the right approach.



Time Series and Their Components

Estimation of Seasonal Component

- A seasonal pattern is one that repeats itself year after year.
- For annual data, seasonality is generally not an issue because there is little chance to model a within year pattern with data recorded once per year.
- Time series consisting of weekly, monthly, or quarterly observations often exhibit seasonality.
- Several methods for measuring seasonal variation have been developed.
- The basic idea in these methods is to first estimate and remove the trend from the original series and then smooth out the irregular component.

Time Series and Their Components

- The seasonal values are collected and summarized to produce a number (generally an index number) for each observed interval of the year (week, month, quarter, and so on).
- If an additive decomposition is used, estimates of the trend, seasonal, and irregular components are added together to produce the original series.
- If a multiplicative decomposition is used, the individual components must be multiplied together to reconstruct the original series, and in this formulation, the seasonal component is represented by a collection of index numbers.
- These numbers show which periods within the year are relatively low and which periods are relatively high.

Remark

- In the rest of this part, we study the multiplicative model.
- In multiplicative decomposition model, the ratio to moving average is a popular method for measuring seasonal variation.

Classical Decomposition

Finding Seasonal Indexes

- Ratio-to-moving average method:
  - Begin by removing the seasonal and irregular components (S_t and I_t), leaving the trend and cyclical components (T_t and C_t)
- Example: Four-quarter moving average
- First average:

$$Moving\ average_1 = \frac{Q1 + Q2 + Q3 + Q4}{4}$$

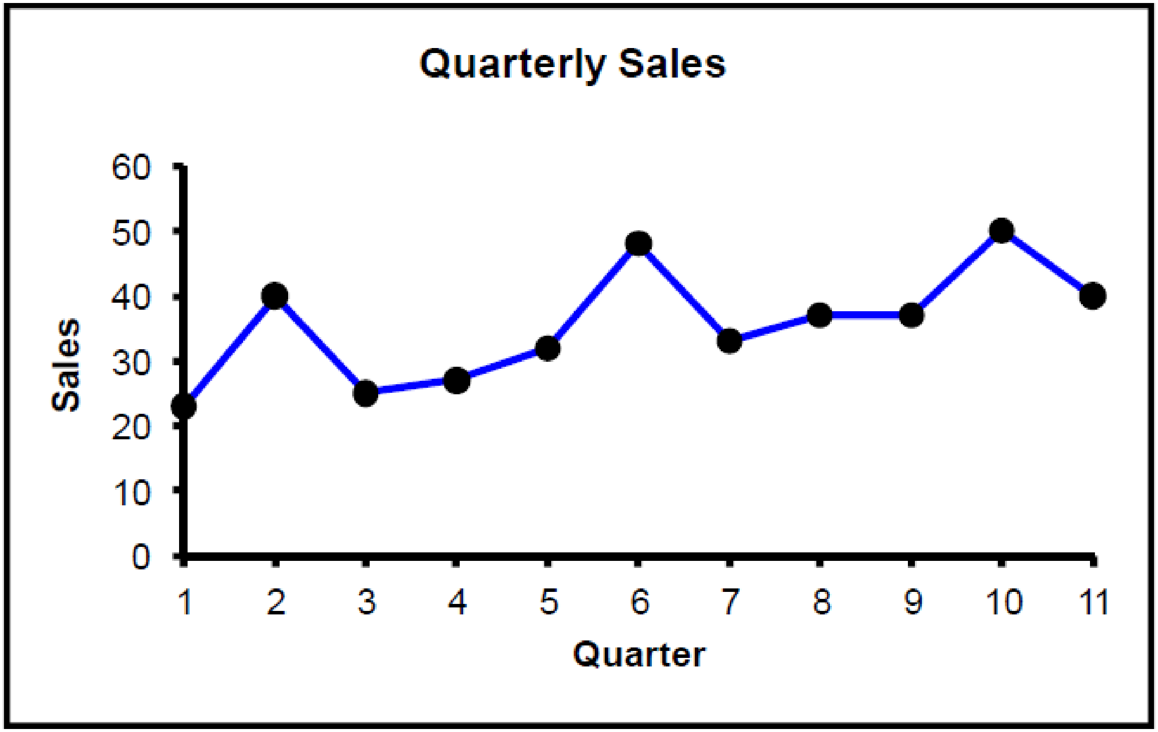
- Second average:

$$Moving\ average_1 = \frac{Q2 + Q3 + Q4 + Q5}{4}$$

- Etc…

Classical Decomposition

Quarter	Sales
1	23
2	40
3	25
4	27
5	32
6	48
7	33
8	37
9	37
10	50
11	40
etc…	etc…



Classical Decomposition

- Centered Seasonal Index

Quarter	Sales
1	23
2	40
3	25
4	27
5	32
6	48
7	33
8	37
9	37
10	50
11	40

etc…

Average Period	4-Quarter Moving Average
2.5	28.75
3.5	31.00
4.5	33.00
5.5	35.00
6.5	37.50
7.5	38.75
8.5	39.25
9.5	41.00

$$2.5 = \frac{1+2+3+4}{4}$$

$$28.75 = \frac{23+40+25+27}{4}$$

Each moving average is for a consecutive block of 4 quarters

Classical Decomposition

- Average periods of 2.5 or 3.5 don't match the original quarters, so we average two consecutive moving averages to get centered moving averages.

Average Period	4-Quarter Moving Average
2.5	28.75
3.5	31.00
4.5	33.00
5.5	35.00
6.5	37.50
7.5	38.75
8.5	39.25
9.5	41.00

etc…

Centered Period	Centered Moving Average
3	29.88
4	32.00
5	34.00
6	36.25
7	38.13
8	39.00
9	40.13

- Now estimate the $S_t \times I_t$ value by dividing the actual sales value by the centered moving average for that quarter.

Classical Decomposition

- Ratio-to-Moving Average formula: $S_t \; x \; It = \dfrac{Y_t}{T_t \; x \; Ct}$

Quarter	Sales	Centered Moving Average	Ratio-to-Moving Average
1	23		
2	40		
3	25	29.88	0.837
4	27	32.00	0.844
5	32	34.00	0.941
6	48	36.25	1.324
7	33	38.13	0.865
8	37	39.00	0.949
9	37	40.13	0.922
10	50	etc…	etc…
11	40	…	…
…	…	…	…

Example

$$0.837 = \frac{25}{29.88}$$

Classical Decomposition

	Quarter	Sales	Centered Moving Average	Ratio-to-Moving Average
	1	23		
	2	40		
Fall	3	25	29.88	$\boxed{0.837}$
	4	27	32.00	0.844
	5	32	34.00	0.941
	6	48	36.25	1.324
Fall	7	33	38.13	$\boxed{0.865}$
	8	37	39.00	0.949
	9	37	40.13	0.922
	10	50	etc…	etc…
Fall	11	40	…	$\boxed{\ldots}$
	…	…	…	…

Average all of the Fall values to get Fall's seasonal index

Do the same for the other three seasons to get the other seasonal indexes

Classical Decomposition

- Suppose we get these seasonal indices:

Season	Seasonal Index
Spring	0.825
Summer	1.310
Fall	0.920
Winter	0.945

Interpretation:

Spring sales average 82.5% of the annual average sales

Summer sales are 31.0% higher than the annual average sales

etc…

$\Sigma = 4.000$ -- four seasons, so must sum to 4

Classical Decomposition

ÖZYEĞİN UNIVERSITY

- The data is deseasonalized by dividing the observed value by its seasonal index

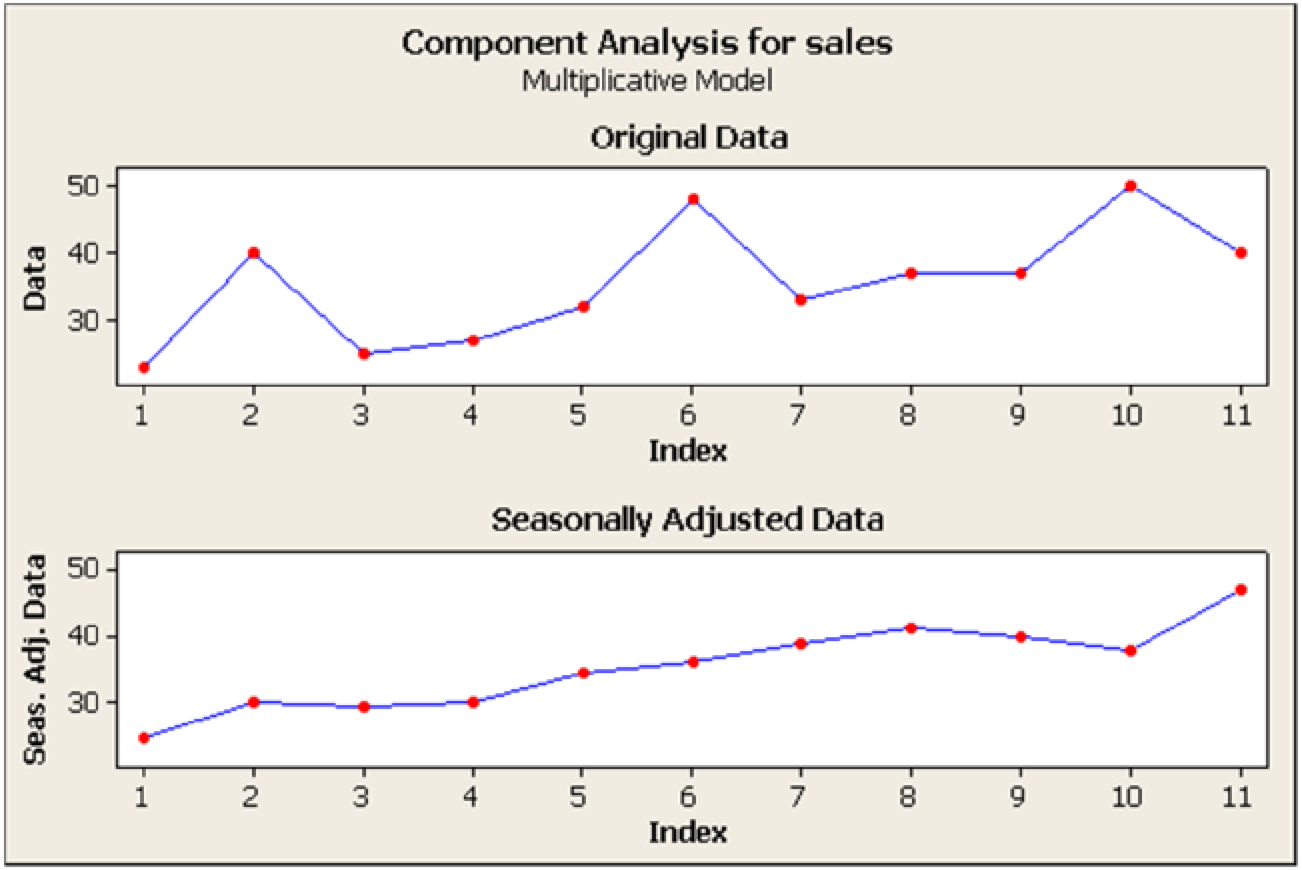
$$T_t \, x \, Ct \, x \, It = \frac{Y_t}{S_t}$$

- This smoothes the data by removing seasonal variation

Quarter	Sales	Seasonal Index	Deseasonalized Sales
1	23	0.825	27.88
2	40	1.310	30.53
3	25	0.920	27.17
4	27	0.945	28.57
5	32	0.825	38.79
6	48	1.310	36.64
7	33	0.920	35.87
8	37	0.945	39.15
9	37	0.825	44.85
10	50	1.310	38.17
11	40	0.920	43.48
...		...	...

$$27.88 = \frac{23}{0.825}$$

Time Series and Their Components



STL decomposition

- STL: "Seasonal and Trend decomposition using Loess"
- Very versatile and robust.
- Unlike many others, STL will handle any type of seasonality (not just monthly or quarterly like SEATS and X-11).
- Seasonal component allowed to change over time, and rate of change controlled by user.
- Smoothness of trend-cycle also controlled by user.
- Robust to outliers.
- Works only for additive decomposition!!! Take logs to get multiplicative decomposition.

STL decomposition