Time Series Forecasting

Dr. D. Narayana

Objective

- Understand the importance of forecasting and its impact on the effectiveness of the supply chain and overall performance of an organization
- Learn various components of time-series data such as trend, seasonality, cyclical components and random components
- Learn different techniques such as moving average, exponential smoothing methods
- Learn practical challenges associated with forecasting models using case studies

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
- Example: Non-availability of a 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

Introduction to Forecasting

- Every organization prepares long-range and short-range planning for the organization
- Forecasting demand for product and service is an important input for both long-range and short-range planning

Forecasting Methodologies

- There are many different time series techniques
- It is usually impossible to know which technique will be best for a particular data set
- It is customary to try out several different techniques and select the one that seems to work best
- To be effective time series modeler, you need to keep several time series techniques in your “tool box”
- Simple ideas
  - Moving averages
  - AR
- Complex statistical concepts
  - Box-Jenkins methodology

Time Series Data

- Time series data is a sequence of observations collected from a process with **equally** spaced time periods
- Examples:
  - Daily closing stock prices
  - Daily data on sales
  - Monthly inventory
  - Daily customers
  - Monthly unemployment rates
  - GDP

Time Series Data

- Why do we study time series analysis?
    - Time series are analyzed to understand the past and to predict the future, enabling managers or policy makers to make properly informed decisions.
    - Time series analysis quantifies the main features in data and the random variation.
    - These reasons, combined with improved computing power, have made time series methods widely applicable in government, industry, and commerce.

Typical Time Series

$$\hat{y}_{t+1} = f(y_t, y_{t-1}, y_{t-2} \dots) + g(x_t, x_{t-1}, x_{t-2} \dots)$$

f and g can be linear or nonlinear

Time Series Data: Components

Trend

Gradual long term movement(up or down). Easiest to detect

Eg: Population growth in India

Seasonality

Results from events that are periodic and recurrent in nature. An up-and-down repetitive movement within a trend occurring periodically.

Eg: Sales in festive seasons

Cyclical Patterns

Results from events recurrent but not periodic in nature. An up-and-down repetitive movement in demand repeats itself over a long period of time

Eg: Recession in US economy

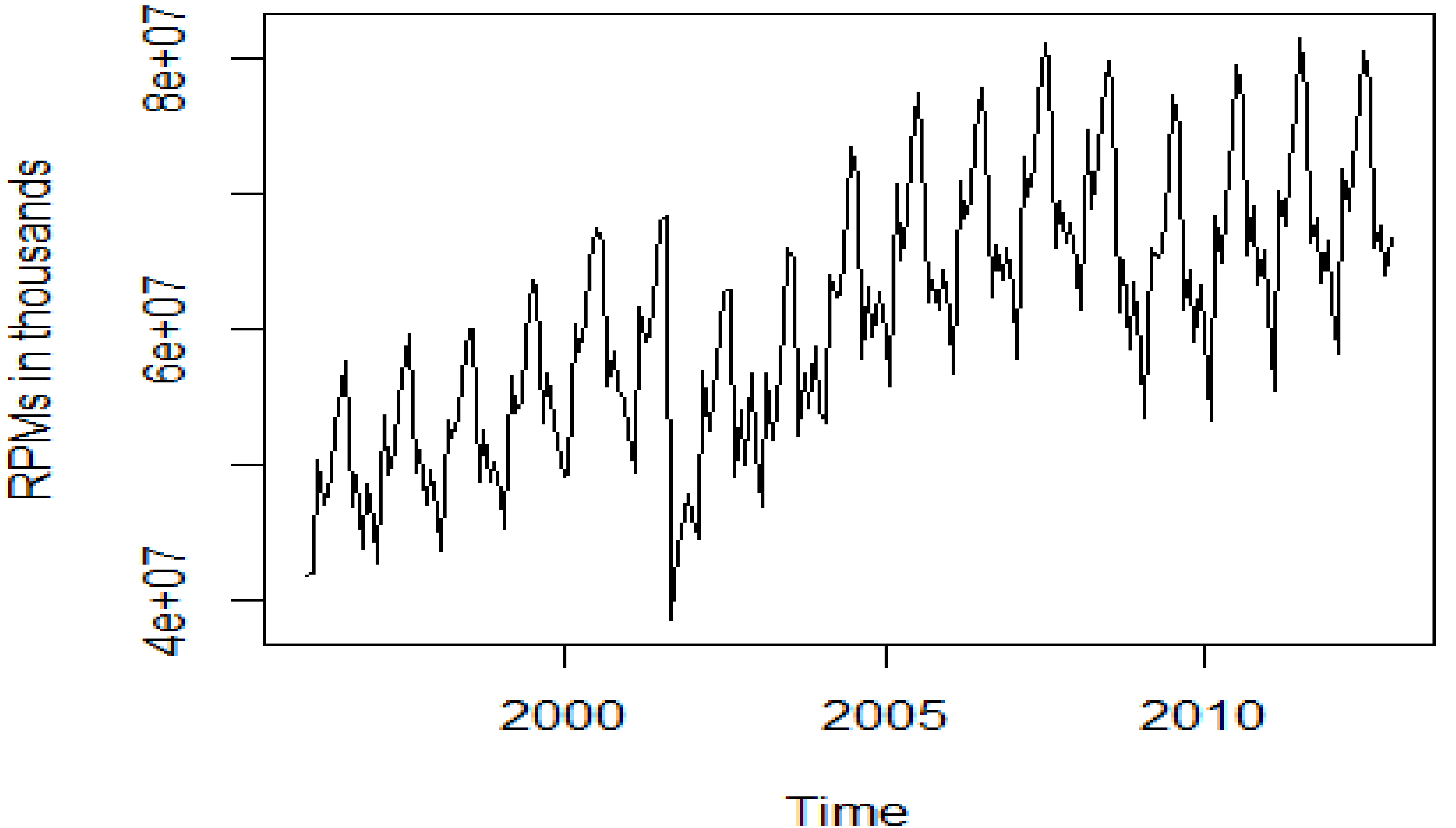
Random component / Irregular component

Disturbances or residual variation that remain after all the other behaviors have been accounted for. Erratic movements that are not predictable because they do not follow a pattern

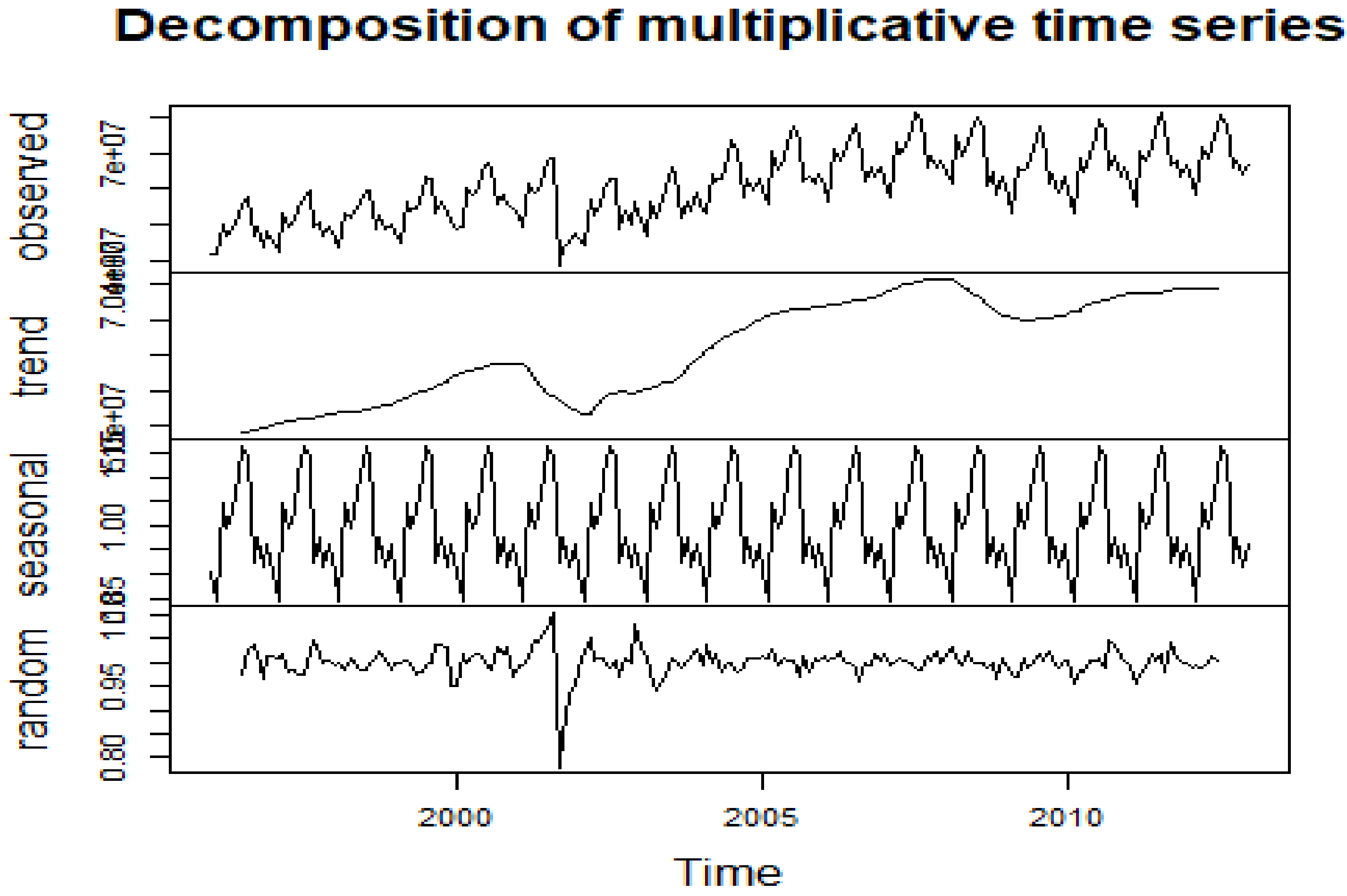
Eg: Earthquake

Components of Time Series

Revenue Passenger Miles
Jan 1996 to Dec 2012 (17 years)



Trend, Seasonality and Randomness



Forecasting Techniques

- Different forecasting techniques developed based on different logics
- Simple techniques such as moving average and exponential smoothing predict the future value of a time-series data as a function of the past observations
- Regression based models such as auto-regressive(AR), moving average(MA), auto-regressive and moving-average(ARMA), auto-regressive integrated moving average(ARIMA), and ARIMA with X (ARMAX) use more sophisticated regression models to forecast the future value of a time-series data
- Important: Using complex mathematical models does not guarantee more accurate forecast
- Simple moving average technique may outperform complex ARIMA models in few cases

Forecasting Accuracy

- Different forecasting techniques such as moving average, exponential smoothing, and ARIMA will be used for forecasting before selecting the best model
- The model selection may depend on the chosen forecasting accuracy measure
- Frequently used forecasting accuracy measures:
    - Mean absolute error
    - Mean absolute percentage error
    - Mean squared error
    - Root mean square error

Forecasting Accuracy

- Let Y_t is the actual value of Y at time t and F_t is the corresponding forecasted value
- Assume that there are n (for example $n = 100$) observations in total
  - Mean absolute error(MAE)

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |Y_t - F_t|$$

    - MAE is the average absolute error and should be calculated on the test data set
  - Mean absolute percentage error(MAPE)

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \frac{|Y_t - F_t|}{Y_t} \times 100\%$$

    - MAPE is one of the popular forecasting accuracy measures used by practioners since it expresses the average error in percentage terms and is easy to interpret
    - Since MAPE is dimensionless it can be used for comparing different models with varying scales

Forecasting Accuracy

- Mean squared error(MSE)

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(Y_t - F_t)^2$$

  - Lower MSE implies better prediction
  - However, it depends on the range of the time-series data
- Root mean square error(RMSE)

$$\text{RMS}E = \sqrt{(\frac{1}{n}\sum_{t=1}^{n}(Y_t - F_t)^2)}$$

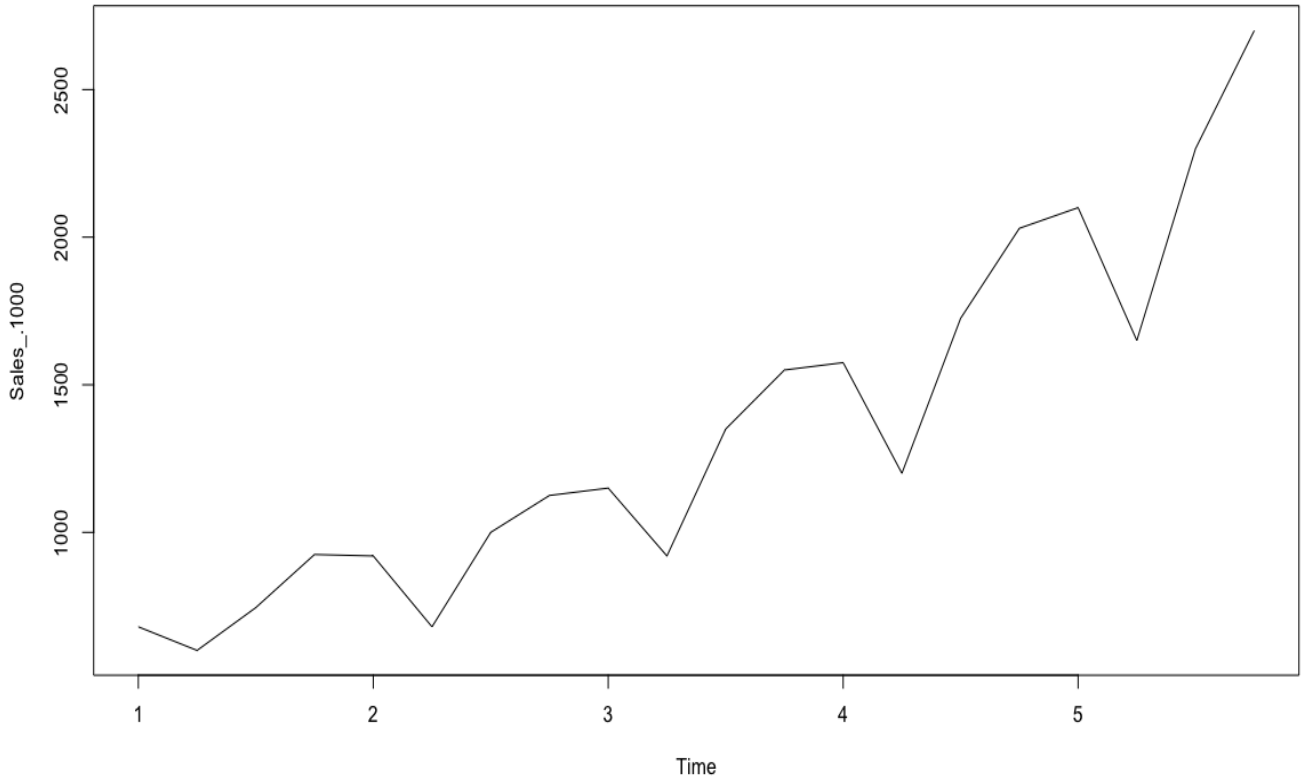
  - RMSE and MAPE two most popular accuracy measures of forecasting
  - RMSE is the standard deviation of errors or residuals

- Example: In 2006, Netflix, the movie portal, announced a competition with a prize money worth one million dollars to predict the rating on a 5-point scale likely to be given by a customer for a movie. The participants were given a target RMSE of 0.8572 to qualify for the prize (source: https://en.wikipedia.org/wiki/Netflix_Prize)

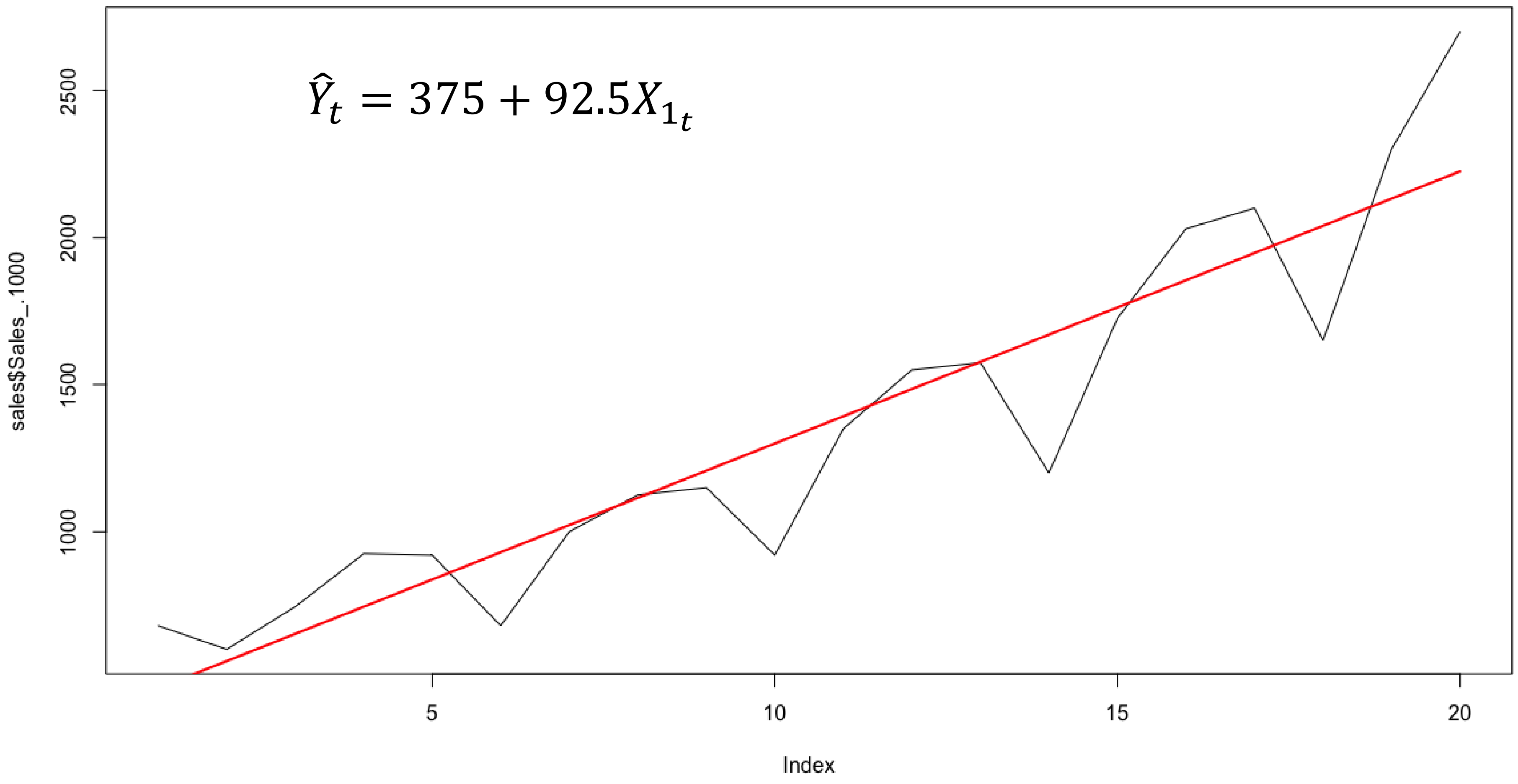
Regression on Time

Use when trend is the most pronounced

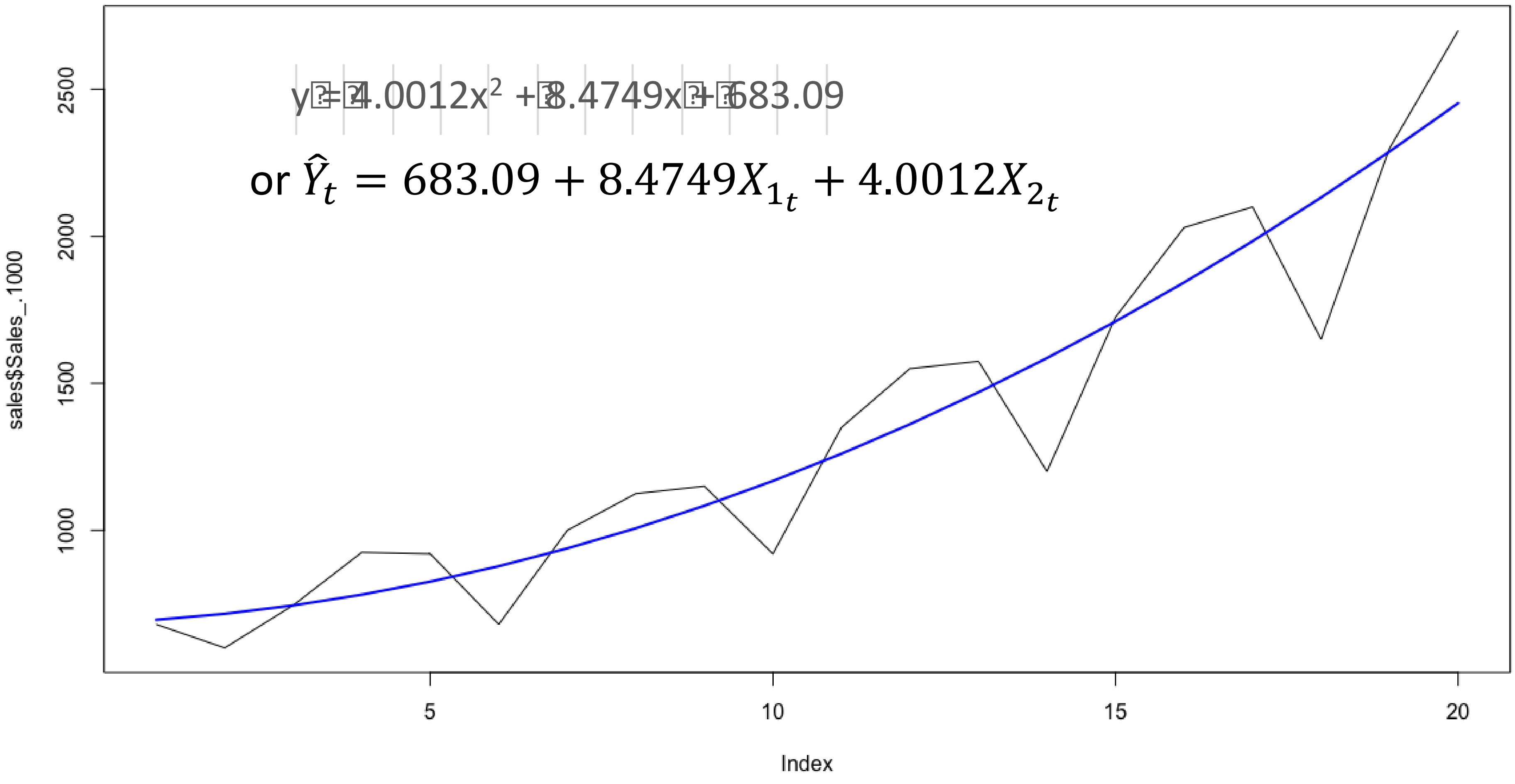
Quarter	Sales_$1000
1	680
2	600
3	745
4	925
5	920
6	680
7	1000
8	1125
9	1150
10	920
11	1350
12	1550
13	1575
14	1200
15	1725
16	2030
17	2100
18	1650
19	2300
20	2700



Regression Analysis



Quadratic Trend

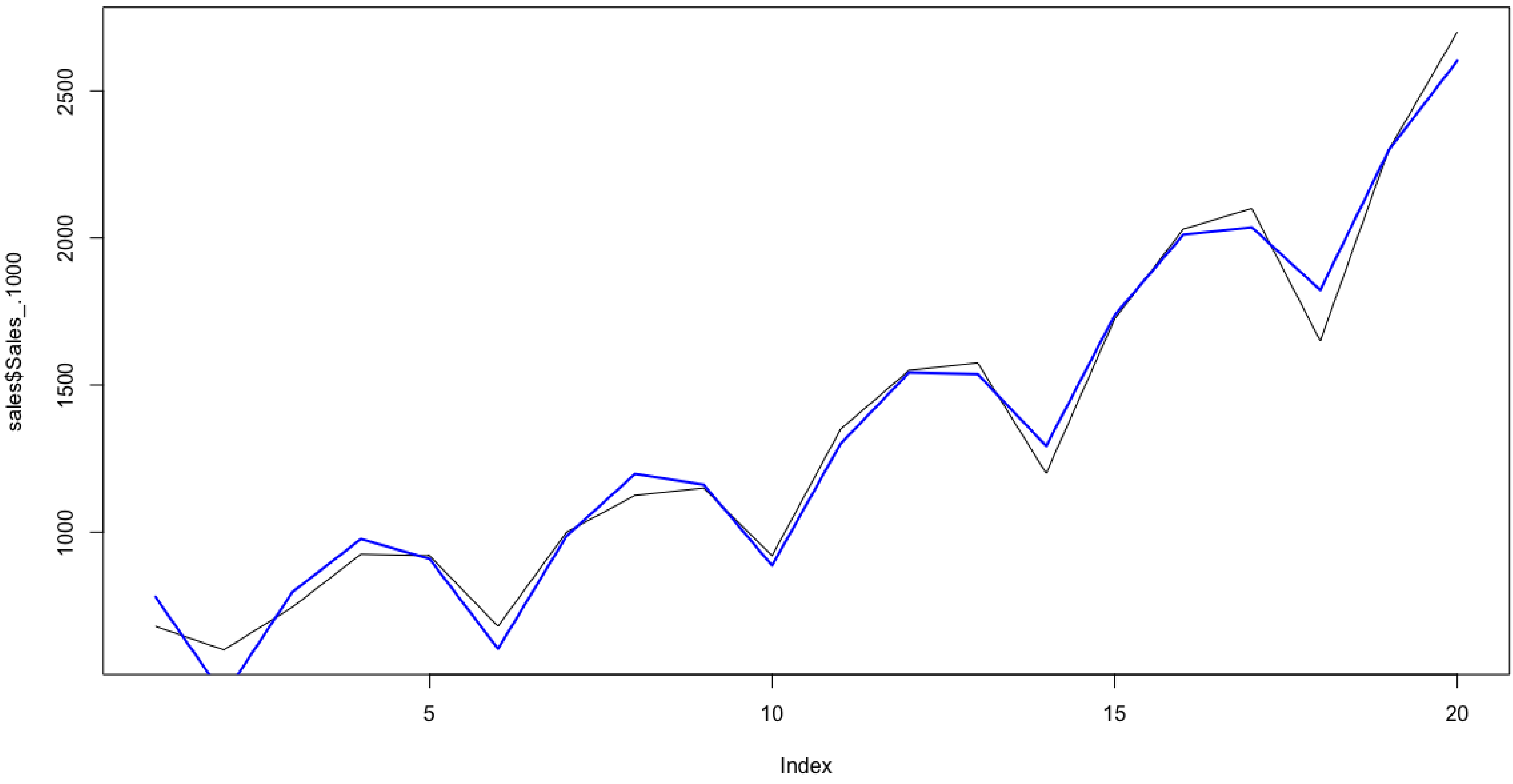


Seasonal Regression Models

	Value of		
Quarter	X_{3_t}	X_{4_t}	X_{5_t}
1	1	0	0
2	0	1	0
3	0	0	1
4	0	0	0

$$\hat{Y}_t = \beta_0 + \beta_1 X_{1_t} + \beta_2 X_{2_t} + \beta_3 X_{3_t} + \beta_4 X_{4_t} + \beta_5 X_{5_t} + \varepsilon_t$$

Seasonal Regression Models



Another Way of Incorporating Seasonality

Take the trend prediction and actual value.

Depending on additive or multiplicative model compute the deviation and map it as seasonality effect for each prediction.

Take averages of the seasonality value. Use this to make future predictions.

Case

Year	Quarter	Time variable (this is created)	Revenues (in $M)
2008	I	1	10.2
	II	2	12.4
	III	3	14.8
	IV	4	15
2009	I	5	11.2
	II	6	14.3
	III	7	18.4
	IV	8	18

What is the Regression equation?

$$y = 10.0393 + 0.9440x$$

```
Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max
-3.5595 -0.9384  0.4405  1.3265  1.9286

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  10.0393     1.5531   6.464  0.00065 ***
x             0.9440     0.3076   3.069  0.02196 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.993 on 6 degrees of freedom
Multiple R-squared: 0.6109,     Adjusted R-squared: 0.5461
F-statistic: 9.422 on 1 and 6 DF,  p-value: 0.02196
```

Seasonality: Multiplicative

Time	Observed values TSI (assuming no impact of cyclicality)	Predicted values (per the regression) T	SI = TSI/T
1	10.2	10.983	0.929
2	12.4	11.927	1.040
3	14.8	12.871	1.150
4	15.0	13.815	1.086
5	11.2	14.759	0.759
6	14.3	15.703	0.911
7	18.4	16.647	1.105
8	18.0	17.591	1.023

T: Trend; S: Seasonal; I: Irregular

Quarterly Seasonality

Time	Average seasonality factor
Q1	$0.844\left(=\frac{0.929+0.759}{2}\right)$
Q2	0.975
Q3	1.127
Q4	1.054

Time	Observed values TSI* (assuming no impact of cyclicality)	Predicted values (per the regression)	SI* = TSI/T
1	10.2	10.983	0.929
2	12.4	11.927	1.040
3	14.8	12.871	1.150
4	15.0	13.815	1.086
5	11.2	14.759	0.759
6	14.3	15.703	0.911
7	18.4	16.647	1.105
8	18.0	17.591	1.023

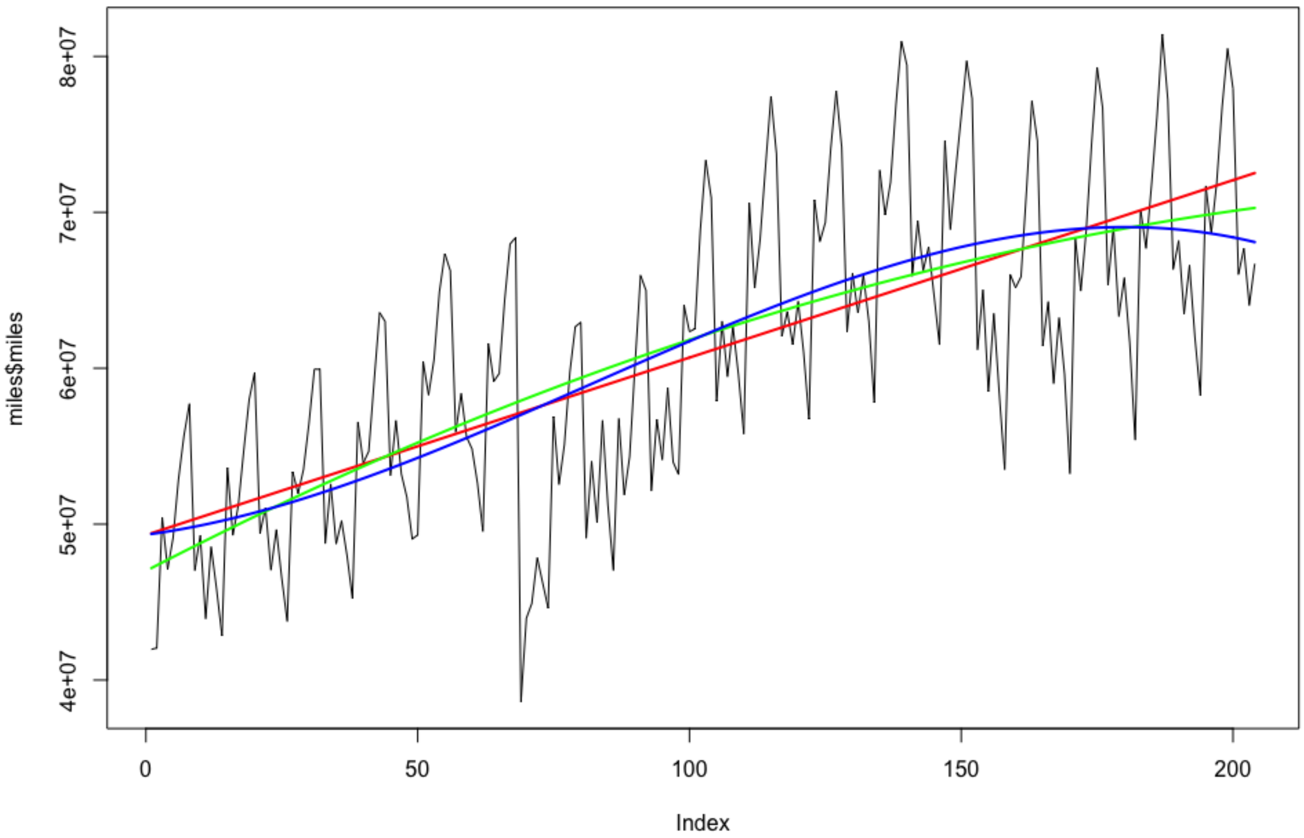
Computations

Trend Y_9 =10.039 + 0.944(9) = 18.535

Corrected for seasonality and randomness: 18.535 * 0.844 = 15.643

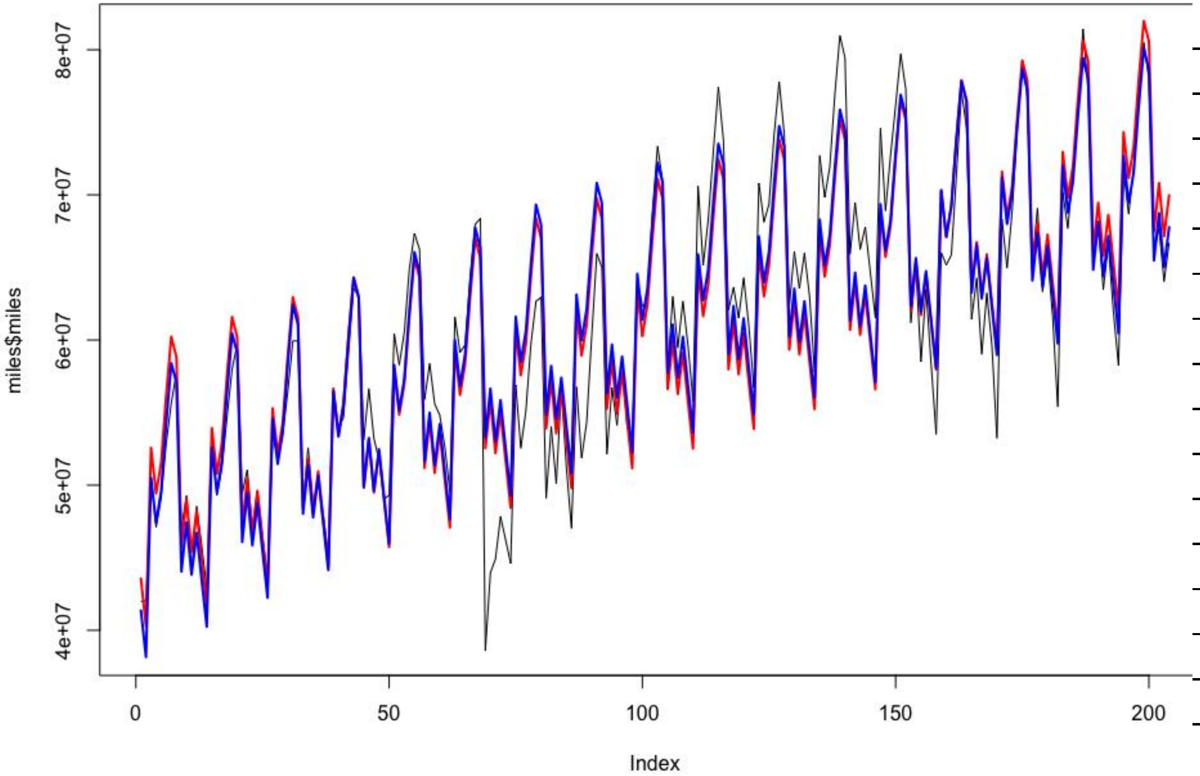
Hands-On

Seasonal Regression Models - RPM



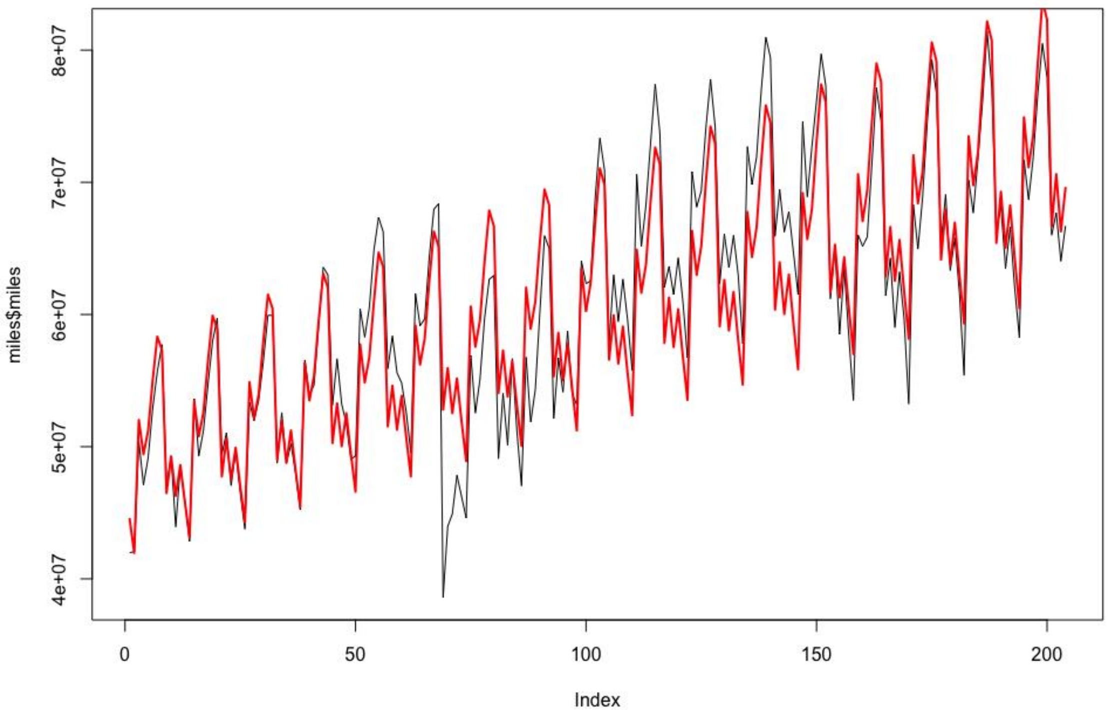
	miles	time	var3
1	41972194	1	
2	42054796	2	
3	50443045	3	
4	47112397	4	
5	49118248	5	
6	52880510	6	
7	55664750	7	
8	57723208	8	
9	47035464	9	
10	49263120	10	
11	43937074	11	
12	48539606	12	
13	45850623	13	
14	42838949	14	
15	53620994	15	

Seasonal Regression Models - RPM



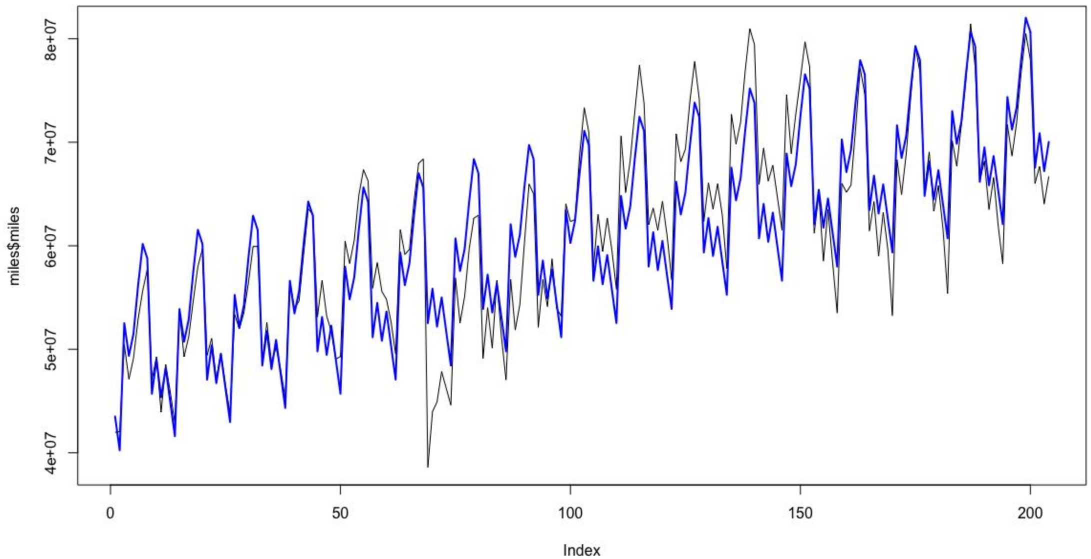
	miles	time	seasonal
1	41972194	1	1
2	42054796	2	2
3	50443045	3	3
4	47112397	4	4
5	49118248	5	5
6	52880510	6	6
7	55664750	7	7
8	57723208	8	8
9	47035464	9	9
10	49263120	10	10
11	43937074	11	11
12	48539606	12	12
13	45850623	13	1
14	42838949	14	2
15	53620994	15	3

Seasonality: Multiplicative



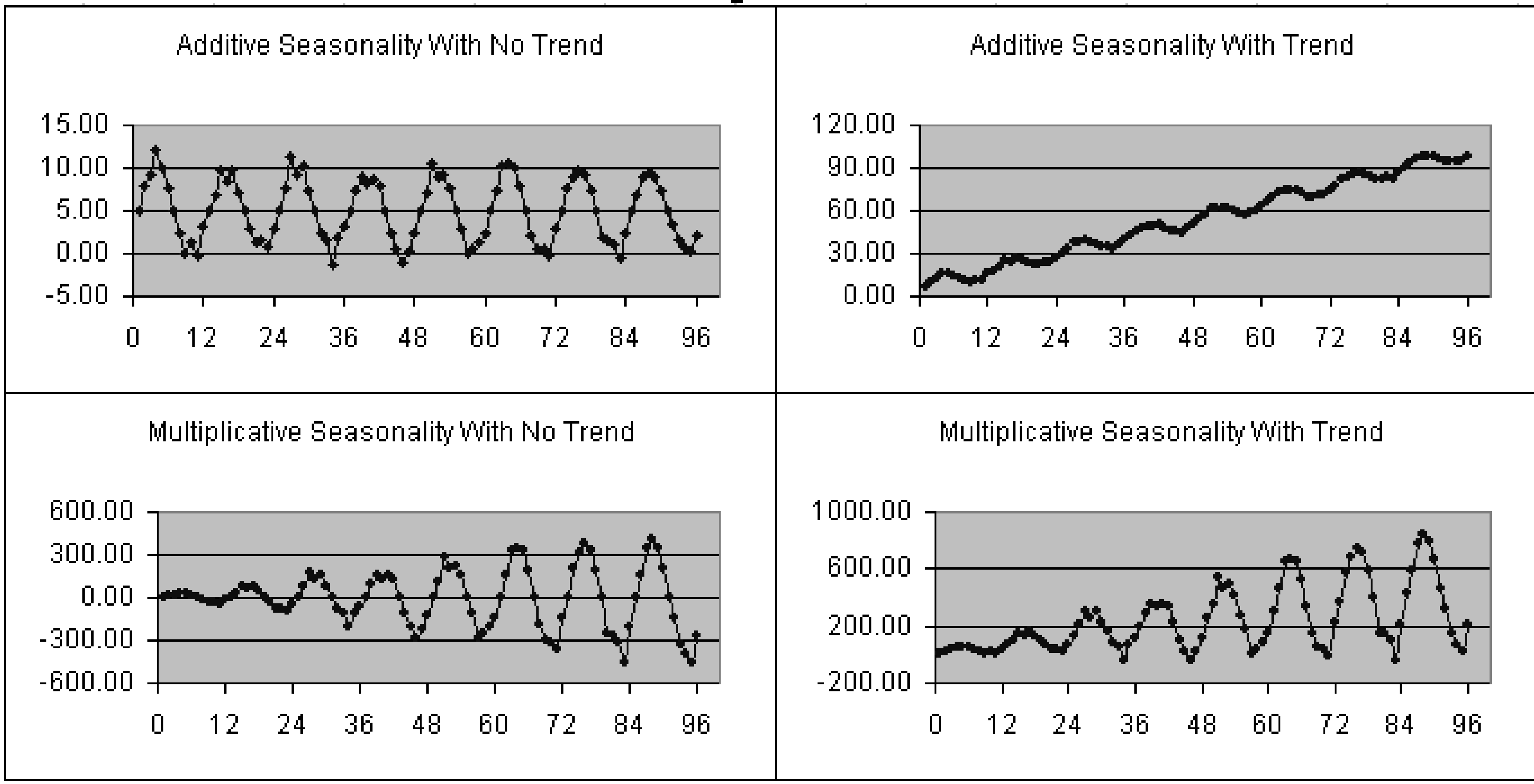
	miles	time	seasonal	mae
1	41972194	1	1	0.849386
2	42054796	2	2	0.8491019
3	50443045	3	3	1.016129
4	47112397	4	4	0.946865
5	49118248	5	5	0.9849257
6	52880510	6	6	1.057953
7	55664750	7	7	1.111125
8	57723208	8	8	1.149602
9	47035464	9	9	0.9346292
10	49263120	10	10	0.9766855
11	43937074	11	11	0.8691307
12	48539606	12	12	0.9580177
13	45850623	13	1	0.9029174
14	42838949	14	2	0.8417232
15	53620994	15	3	1.051224

Seasonality: Additive



	miles	time	seasonal	mae
1	41972194	1	1	-7442550
2	42054796	2	2	-7473763
3	50443045	3	3	800670.5
4	47112397	4	4	-2643793
5	49118248	5	5	-751757.1
6	52880510	6	6	2896690
7	55664750	7	7	5567114
8	57723208	8	8	7511757
9	47035464	9	9	-3289802
10	49263120	10	10	-1175962
11	43937074	11	11	-6615823
12	48539606	12	12	-2127106
13	45850623	13	1	-4929905
14	42838949	14	2	-8055394
15	53620994	15	3	2612836

Additive or Multiplicative



Issues with Regressing on Time

If there is no trend or if seasonality and fluctuations are more important than trend, then the coefficients behave weirdly

Moving Average (MA) Method

- MA is one of the simplest forecasting techniques which forecasts the future value of a time-series data using average (or weighted average) of the past N observations
- Mathematically, a simple moving average is calculated using the formula

$$\mathrm{F}_{\mathrm{t+1}} = \frac{1}{N} \sum_{k=t+1-N}^{t} Y_k$$

- The above formula is called simple moving average(SMA) since N past observations are given equal weights, i.e., $\frac{1}{N}$
- In a weighted moving average, past observations are given differential weights(usually the weights decrease as the data becomes older)
- Weighted moving average is given by

$$\mathrm{F}_{\mathrm{t+1}} = \sum_{k=t+1-N}^{t} W_k \times Y_k$$

where W_k is the weight given to value of Y at time k (Y_k) and $\sum_{k=t+1-N}^{t} W_k = 1$

Single Exponential Smoothing(ES)

- One of the drawbacks of simple moving average technique is that it gives equal weight to all the previous obserations used in forecasting the future value
- This can be overcome by assigning different weights to the past observations
- One easier way to assign differential weight is achieved by using single exponential smoothing (SES) technique
- Just like the moving average, SES assumes a fairly steady time-series data with no significant trend, seasonal or cyclical component
- Here, the weights assigned to past data decline exponentially with the most recents observations assigned higher weights

Single Exponential Smoothing(ES)

- (Winters, 1960): In single ES, the forecast at time t+1 is given by

$$F_{t+1} = \alpha Y_t + (1-\alpha)F_t$$

The parameter α in the above equation is called the smoothing consant and its value lies between 0 and 1

- Since the model uses one smoothing constant, it is called single exponential smoothing
- Substituting for F_t recursively in the above equation, we get

$$F_{t+1} = \alpha Y_t + \alpha(1-\alpha)Y_{t-1} + \alpha(1-\alpha)^2 Y_{t-2} + \cdots + \alpha(1-\alpha)^{t-1}Y_1 + (1-\alpha)^t F_1$$

It is evident that the weights assigned to older observations decrease exponentially

Single Exponential Smoothing(ES)

- Exponential smoothing uses the entire historical data
- To begin exponential smoothing, we will need the forecast for the F_t
- We can use $F_t = Y_t$ or use moving average to forecast the initial forecast F_t
- The forecasted value for period 2 is given by

$$F_2 = \alpha\, Y_1 + (1 - \alpha)F_1$$

- We assume F_1 same as Y_1. Thus the value of F_2 will be same as Y_1, i.e., 3002666
- Question: What are the RMSE and MAPE for SMA and SES with $\alpha = 0.2, \alpha = 0.6, \alpha = 0.8, \alpha = 0.9$

SMA Disadvantages SES Advantages

- SMA Disadvantages:
  - Increasing n makes forecast less sensitive to changes in data
  - It always lags behind trend as it is based on past observations. The longer the time period n, the greater the lag as it is slow to recognize the shifts in the level of the data points
  - Forecast bias and systematic errors occur when the observations exhibit strong trend or seasonal patterns
- SES Advantages:
  - It uses all the historic data unlike the moving average where only the past few observations are considered to predict the future value
  - It assigns progressively decreasing weights to older data

Optimal smoothing constant in a SES

- Choosing optimal smoothing constant α is important for accurate forecast.
- Whenever the data is smooth(without much fluctuations), we may choose higher value of α
- When the data is highly fluctuating, then it is better to choose lower value of α
- We can find the optimal value of the smoothing constant by solving a non-linear optimization problem
- Example: Assume that we have to find the optimal α that will give the minimum RMSE. This can be achieved by solving the following optimization problem:

$$Min_{\alpha} \left[\sqrt{\frac{1}{n}\sum_{t}(Y_t - F_t)^2} \right]$$

  Subject to the constraint: $0 < \alpha < 1$

  - For WSB slaes data, the optimal value of α that minimizes the RMSE is 0.1574 and corresponding RMSE is 739399.76
- Question: Forecast sales values with $\alpha = 0.1574$

Moving Average (MA) Method

- WBS sales case study

Double Exponential Smoothing: Holt's method

- One of the drawbacks of SES is that the model does not do well in the presence of trend
- This can be improved by introducing an additional equation for capturing the trend in the time-series data
- Double exponential smoothing uses two equations to forecast the future values of the time series, one for forecasting the level (short-term average value) and another for capturing the trend

Double Exponential Smoothing: Holt's method

- Equation1: Level (or intercept): $L_t = \alpha Y_t + (1-\alpha)F_t$
- Equation2: Trend: $T_t = \beta \times (L_t \ - L_{t-1}) + (1-\beta) \times T_{t-1}$

where α and β are the smoothing constants for level and trend, $0 < \alpha < 1$ and $0 < \beta < 1$

- The forecast at time t+1 is given by
  - $F_{t+1} = L_t + T_t$
  - $F_{t+n} = L_t + nT_t$

  where L_t is the level which represents the smoothed value upto and including the last data, T_t is the slope of the line or the rate of increase or decrease at period t, n is the number of time periods into the future

Double Exponential Smoothing: Holt's method

- Initial values: $L_t = Y_t$ and $T_t = (Y_t - Y_{t-1})$ or $T_t = \frac{(Y_t - Y_1)}{t-1}$
- For WSB sales data
  - $L_1 = Y_1 = 3002666$
  - $T_1 = \frac{Y_{36} - Y_1}{35} = \frac{4732677 - 3002666}{35} = 49428.8857$
  - $F_2 = L_1 + T_1 = 3002666 + 49428.8857 = 3052095$
  - Calculate forecasting values from 37 to 48 months by using $\alpha = 0.457$ and $\beta = 0.7287$

Triple Exponential Smoothing

- Moving averaging and single and double exponential smoothing techniques discussed so far can handle data as long as the data do not have any seasonal component associated with it
- However, when there is seasonality in the time series data, techniques such as moving average, exponential smoothing, and double exponential smoothing are no longer appropriate
- In most cases, the fitted error values(actual demand minus forecast) associated with simple exponential smoothing and Holt’s method will indicate systematic error patterns that reflect the existence of seasonality
- For example, presence of seasonality may result in all positive errors, except for negative values that occur at fixed intervals
- Such pattern in error would imply existence of seasonality
- Such time series data requires the use of a seasonal method to eliminate the systematic patterns in error

Triple Exponential Smoothing

- Triple exponential smoothing is used when the data has trend as well as seasonality
- The following three equations which account for level, trend and seasonality are used for forecasting(for multiplicative model)
  - Equation1: Level(or Intercept) equation: $L_t = \alpha \frac{Y_t}{S_{t-c}} + (1-\alpha)[L_{t-1} + T_{t-1}]$
  - Equation2: Trend equation: $T_t = \beta \times (L_t - L_{t-1}) + (1-\beta)T_{t-1}$
  - Equation3: Seasonal equation: $S_t = \gamma \frac{Y_t}{L_t} + (1-\gamma)S_{t-c}$

- The forecast, F_{t+1}, using triple exponential smoothing is given by

$$F_{t+1} = [L_t + T_t] \times S_{t+1-c}$$

Where c is the number of seasons(if it is monthly seasonality, then c=12; in case of quarterly seasonality c = 4; and in case of daily data c = 7)

Triple Exponential Smoothing

- Initial values for L_t and T_t
  - $L_t = Y_t$ Alternatively $L_t = \frac{1}{c}(Y_1 + Y_2 + \cdots + Y_c)$
  - $T_t = \frac{1}{c}[\frac{Y_t - Y_{t-c}}{12} + \frac{Y_{t-1} - Y_{t-c-1}}{12} + \frac{Y_{t-2} - Y_{t-c-2}}{12} + \cdots + \frac{Y_{t-c+1} - Y_{t-2c+1}}{12}]$
  - Several techniques exist to calculate the initial seasonality index
  - The initial seasonality index can be calculated using a technique called method of simple averages

Forecasting using triple exponential smoothing

- Question: Forecast WSB sales for the period 37 to 48
- Question: Compare SMA, SES, Double Exponential and triple exponential methods performance

Summary

- Overview of time series
- Trend, seasonality, cyclical irregular components
- Regression based forecasts
- SMA, Single EMA, Double EMA, Holt-Winters
- Case studies

For the next class on Time series…

- Come prepared with the following topics
    - Auto-regression
    - Correlation
    - Auto-correlation
    - Partial auto-correlation (will be revised in the class room also)
- Practice the above topics with R language before next class
- We assume the above topics for the next class

Thanks