Chapter 5

Forecasting

What is Forecasting

- Forecasting is the scientific methodology for predicting what will happen in the future based on the data in the past.

Types of Forecasting Models

- Time-series models
  - Assuming future is a function of the past
  - Moving average, exponential smoothing, regression, ...
- Causal models
  - Using the influential variables to predict future.
  - Regression, ...
- Qualitative models
  - Incorporating subjective factors
  - Delphi method, Jury of executive opinion, sales force composite, ...

Measure of Accuracy of Forecast

$$MAD \quad \frac{|\textit{forecast errors}|}{n}$$

where

forecast error = actual value – forecast value;

n = number of pieces of data observed.

Example, Table 5.1, page 153

Year	Actual Sales	Forecast sales	Absolute values of errors
1	110	-	
2	100	110	
3	120	100	
4	140	120	
5	170	140	
6	150	170	
7	160	150	
8	190	160	
9	200	190	
10	190	200	
11	-	190	

Meanings Contained in Data

- **Trend**: - general tendency (direction) of movement or course.
- **Seasonal variation**: - changes periodically occurred in a year.
- **Cycle**: - changes with economy over years.
- **Random variation or noise**: - no explainable meanings.

The Quest of Forecasting Methods

- A good forecast method should reflect the trend, seasonal effect, and cycle, while filter out random variations.

Moving Average Method

Use the average of last n-periods as the forecast of the next period.

$$n \quad period \;\; moving \;\; average$$

$$\text{avg. of actual demands of last n periods}$$

$$\frac{actual \;\; demands \;\; in \;\; previous \;\; n \;\; periods}{n}$$

where *n* is the number of past period counted in the average. n can be 1, 2, 3, 4, 5, ...

Moving Average Formulas

- Particularly, if n=2 (i.e. 2-period moving average), then

$$F_t \qquad \frac{A_{t\ 1} \qquad A_{t\ 2}}{2}$$

where F_t = forecast for period t,

A_i =actual data for period i.

- If n=3 (i.e. 3-period moving average), then

$$F_t \qquad \frac{A_{t\ 1} \qquad A_{t\ 2} \qquad A_{t\ 3}}{3}$$

Example, Table 5.2, p.157

Month	Actual Sales	3-month moving average (n=3)
Jan	10	
Feb	12	
Mar	13	
Apr	16	
May	19	
Jun	23	
Jul	26	
Aug	30	
Sep	28	
Oct	18	
Nov	16	
Dec	14	
Jan		

Why “Moving Average”?

- “Average” is to average off the “noises” in data.
- “Moving” is to pick up trend if there is.

Effects of *n*

- A small *n* makes forecasts pick the trend, but is not good in smoothing out random fluctuations (noises) in data.
- A large *n* is effective in smoothing out noises in data, but is not good in picking real changes such as trend.

Selecting *n*

- We do "experiments" on the past periods for which we know the actual values.
- For a particular n, we do forecasting for the past periods. Calculate MAD for that n.
- Such "experiments" can be done on a few n's. The n with lowest MAD is our choice.

Select n

Month	Actual Sales	Moving Avg. n=3		Moving Avg. n=2	
		Forecasts	Abs. errors	Forecasts	Abs. errors
Jan	10				
Feb	12				
Mar	13			11	2
Apr	16	11.67	4.33	12.5	3.5
May	19	13.67	5.33	14.5	4.5
Jun	23	16	7	17.5	5.5
Jul	26	19.33	6.67	21	5
Aug	30	22.67	7.33	24.5	5.5
Sep	28	26.33	1.67	28	0
Oct	18	28	10	29	11
Nov	16	25.33	9.33	23	7
Dec	14	20.67	6.67	17	3
Jan		16		15	
MAD			**6.481**		**4.7**

Using Computer

- Computer can help us do forecasting and select appropriate *n* for moving average by doing “experiments”.
- Use QM for Windows (page 191-193).
  - QM for Windows is available on campus.
  - How to use it: (class demonstration).

Weighted Moving Average

- n-period weighted moving average:

$$F_t \qquad \frac{w_1 A_{t \;\; 1} \quad w_2 A_{t \;\; 2} \quad \ldots \quad w_n A_{t \;\; n}}{w_1 \quad w_2 \quad \ldots \quad w_n}$$

where

(w_1, w_2, …, w_n) are weights for data in the past,

F_t = forecast for period t,

A_i = actual data for period i.

.

Why “Weighted”

- What will happen in future may be more related to some pieces of past data than others.
- Sales next month, for example, may be more related to last months sales than sales three months ago.

Example, p.158

- n=3; weights=(3, 2, 1). That is:
  - weight for the most recent past period = 3.
  - weight for 2nd most recent past period = 2.
  - weight for 3rd most recent past period = 1.
- Calculations in Table 5.3.

Example, Table 5.3, p.158

Month	Actual Sales	3-month weighted moving average (n=3, weights 3,2,1)
Jan	10	
Feb	12	
Mar	13	
Apr	16	
May	19	
Jun	23	
Jul	26	
Aug	30	
Sep	28	
Oct	18	
Nov	16	
Dec	14	
Jan		

Effect of “Weights”

- If you want your forecast to be more responsive to a period, you put a larger weight for that period.
- Usually, people put larger weights on the more recent past periods.

Select "Weights"

- To select a set of weights:
  - Try various sets of weights on historical data by doing forecasts and calculating MAD (with help of computer);
  - Pick the set of weights that generated lowest MAD.

Exponential Smoothing

- New forecast

  = last period's forecast + error adjustment

- $F_t = F_{t-1} + \quad (A_{t-1} \quad F_{t-1})$

  where

  = smoothing constant (0 1)

  F_t = forecast for period t

  A_{t-1} = actual value in period t 1

Exponential Smoothing, Table 5.4, page 160

Quarter	Actual	Forecast (=0.1)
1	180	175
2	168	
3	159	
4	175	
5	190	
6	205	
7	180	
8	182	
9		

Exponential Smoothing, Table 5.4, page 160

Quarter	Actual	Forecast (=0.5)
1	180	175
2	168	
3	159	
4	175	
5	190	
6	205	
7	180	
8	182	
9		

Effects of (1)

- Larger makes forecast F_t closer to last period's actual occurrence.
- Smaller makes forecast F_t closer to last period's forecast.
- If =1, then $F_t=A_{t-1}$.
- If =0, then $F_t=F_{t-1}$.

Effects of (2)

- A large is good for forecasts to pick up the trend, but not good in smoothing off the noises.
- A small is good in smoothing off the noises, but not good in picking up the trend.

Selecting

- Do “experiments” on the past periods for which we know the actual values.
- For a particular , do forecasting for the past periods. Calculate MAD for that .
- Such “experiments” can be done on a few ’s. The with lowest MAD is our choice.
- QM helps calculations.

Compare MADs for =0.1 and =0.5

Table 5.5, page 160

Quarter	Actual	Forecast, =0.1	Absolute error, **=0.1**	Forecast, =0.5	Absolute error, **=0.5**
1	180	175	5	175	5
2	168	175.5	7.5	177.5	9.5
3	159	174.75	15.75	172.75	13.75
4	175	173.18	1.82	165.88	9.12
5	190	173.36	16.64	170.44	19.56
6	205	175.02	29.98	180.22	24.78
7	180	178.02	1.98	192.61	12.61
8	182	178.22	3.78	186.3	4.3
Total			82.45		98.63
Average			**10.31**		**12.33**

Forecast of Starting Period

- To start exponential smoothing, we must know (or assume) the forecast of the 1st period. For example, we may assume $F_1=A_1$.

Using Computer in Forecasting

- Computer can help select a most appropriate forecasting model:
  - Moving average? n=?
  - Weighted moving average? n=? weights=?
  - Exponential smoothing? =?
- By doing experiments on historical data and using MAD as the criterion
- Do forecasting by using the selected model.

Linear Regression

- Relationship between a dependent variable Y and a couple of independent variables X_i's is represented in a linear equation:

  $$Y=a+b_1X_1+b_2X_2+\ldots+b_nX_n$$

- The values of coefficients a, b_1, b_2, …, b_n are derived from the past data with the *least square method* by computers.

Two Steps for Forecasting with Regression

- **Step 1.** Run QM to get regression equation.
- **Step 2.** For a period, determine its values of independent variable(s), plug them into the regression equation and calculate the forecast.

Simple Regression for Forecasting

- The fundamental model of regression ('*simple regression*') for forecasting:

  Y = the amount to be predicted (sales, demands, for example);

  X = time period (1, 2, 3, …)

Example p.164-167 for trend projection

	Year	**Number of units of generators sold**
	2007	74
	2008	79
	2009	80
	2010	90
	2011	105
	2012	142
	2013	122

What is the forecast number of units sold in 2014?

Define Regression Variables

- Y= number of units sold in a year
- X = time periods, 1 for 2007, 2 for 2008, …
- Input data into QM. Run QM to find the coefficients a and b in the equation

$$Y = a + bX.$$

Entering Data into QM

X	Year	Y (Number of units of generators sold)
1	2007	74
2	2008	79
3	2009	80
4	2010	90
5	2011	105
6	2012	142
7	2013	122

Regression Equation and forecast

- QM calculates the regression equation for us:

  Y = 56.71 + 10.54X

- For year 2011, X=8. Plug it into the equation, we have the forecast for 2014:

  Y = 56.71 + 10.54*8 = 141.03 (units)

Trend Projection

- The simple regression line $Y=a+bX$ can be viewed as a trend line, in which $X=$ time periods.
- So, $Y=a+bX$ can be used to forecast not only next period's Y, but also a few Y's in future (compared to 'moving average' and 'exponential smoothing' which do forecasting for only next period.)

Seasonal Variations

- Seasonal variations recur at certain seasons of a year.
- Casinos revenues vary with the four seasons;
- Demands on tourism and for sandals, sweaters, electricity, gas, lawn fertilizer, Christmas items, road service, and stationery are all seasonal variations.

Multiple Regression

- In multiple regression, there are two or more independent variables X's:

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n$$

- The coefficients a, b_1, b_2, … are calculated with the least square method by computer from the past data.

Seasonal Consideration and Multiple Regression

- Let X_1 = period series number (1, 2, 3, …)
- For the other X's, each represents a "season" with 0 and 1 (winter or summer, a quarter, low or high season, a month, for examples).

Definitions of X's

- X_1 = period series number (1, 2, 3, …)
- X_2 = 1 if current period is season 2,
- = 0 if current period is <u>not</u> season 2.
- X_3 = 1 if current period is season 3,

  = 0 if current period is <u>not</u> season 3.
- …

What about Season 1?

- What if the current period is Season 1?
  - All X's, except X_1, are 0; that is
  - $X_2=0$, $X_3=0$, $X_4=0$, …

Example, data on p.169, regression on p.174

Year	Quarter	Sales ($ million)
Year 1	1	108
	2	125
	3	150
	4	141
Year 2	1	116
	2	134
	3	159
	4	152
Year 3	1	123
	2	142
	3	168
	4	165

Example (continuing)

- By observation, we can see a seasonal reoccurring with quarters:
  - Season 1 is composed of Quarter 1;
  - Season 2 is composed of Quarter 2;
  - Season 3 is composed of Quarter 3;
  - Season 4 is composed of Quarter 4.

Define Variables

- Let Y = Sales in million dollars.
- Let X_1=time periods number, 1, 2, 3, …, 12;
- Let X_2 =1 if the current quarter is season 2, X_2=0 otherwise,
- Let X_3 =1 if the current quarter is season 3, X_3=0 otherwise,
- Let X_4 =1 if the current quarter is season 4, X_4=0 otherwise.

Example (continuing). **Data Input**

Y, Sales	X_1, periods	X_2=1 if Season 2	X_3=1 if Season 3	X_4=1 if Season 4
108	1	0	0	0
125	2	1	0	0
150	3	0	1	0
141	4	0	0	1
116	5	0	0	0
134	6	1	0	0
159	7	0	1	0
152	8	0	0	1
123	9	0	0	0
142	10	1	0	0
168	11	0	1	0
165	12	0	0	1

Example (continuing)

- QM gives the values of coefficients:

  a = 104.1042,

  b_1 = 2.3125,

  b_2 = 15.6875,

  b_3 = 38.7083,

  b_4 = 30.0625.

- That is, the regression equation is:

$$\boldsymbol{Y=104.1+2.3X_1+15.7X_2+38.7X_3+30.1X_4}$$

Regression Result

$$Y=104.1+2.3X_1+15.7X_2+38.7X_3+30.1X_4$$

MAD = 1.0278

Year	Quarter	X1	X2	X3	X4	Quarter in	Actual	Forecast
1	1-1	1	0	0	0	1	108	106.4
	1-2	2	1	0	0	2	125	124.4
	1-3	3	0	1	0	3	150	149.7
	1-4	4	0	0	1	4	141	143.4
2	2-1	5	0	0	0	5	116	115.6
	2-2	6	1	0	0	6	134	133.6
	2-3	7	0	1	0	7	159	158.9
	2-4	8	0	0	1	8	152	152.6
3	3-1	9	0	0	0	9	123	124.8
	3-2	10	1	0	0	10	142	142.8
	3-3	11	0	1	0	11	168	168.1
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Example (continuing), **Calculating Forecasts Using the Regression Equation**

- Next period (1st qtr of yr 4):

  $X_1=13,\ X_2=X_3=X_4=0$

  Forecast: $Y = 104.1+2.3(13) = 134$

- 2nd qtr of yr 4:

  $X_1=14,\ X_2=1,\ X_3=X_4=0$

  Forecast: $Y = 104.1+2.3(14) +15.7 = 152$

- 3rd qtr of yr 4:
- 4th qtr of yr 4:

Example (continuing)

What if Two Seasons?

- We may put a year as two seasons, low season and high season, instead of four seasons.
- Composition of the two seasons:

  Season 1 (low): Qtr. 1 and Qtr. 2,

  Season 2 (high): Qtr. 3 and Qtr. 4.

Example (continuing)

Definitions of Variables for Two Seasons

- Let Y = Sales in million dollars.
- Let X_1=time periods number, 1, 2, 3, …, 12;
- Let X_2 =1 if the current quarter is season 2 (high season), X_2=0 if not.

Data for Two Seasons to Enter into QM

Y, Sales	X_1, periods	X_2, Season 2
108	1	0
125	2	0
150	3	1
141	4	1
116	5	0
134	6	0
159	7	1
152	8	1
123	9	0
142	10	0
168	11	1
165	12	1

Example (continuing)
Regression Equation from QM

- Enter the data into QM and click “Solve” button, we get the regression equation:

  $Y = 111.50 + 2.39X_1 + 26.38X_2$

- MAD = 6.0833

Calculating Forecasts Using the Regression Equation

- Next period (1st qtr of yr 4):

  X_1=13, X_2=0 (since qtr 1 is in low season),

  Forecast: Y = 111.50+2.39(13)+26.38(0)
  = 142.57

- 3rd qtr of yr 4:

  X_1=15, X_2=1 (since Qtr 3 is in high season),

  Forecast: Y=111.50+2.39(15)+26.38(1)

  = 173.73

Example (continuing)
Four Seasons or Two Seasons?

- Which is better? - Using four seasons or using two seasons for this example?
- Compare their MADs:
  - MAD = 1.0278 if using four seasons,
  - MAD = 6.0833 if using two seasons.
- Using four seasons is better since its MAD is smaller.

What if No-Season?

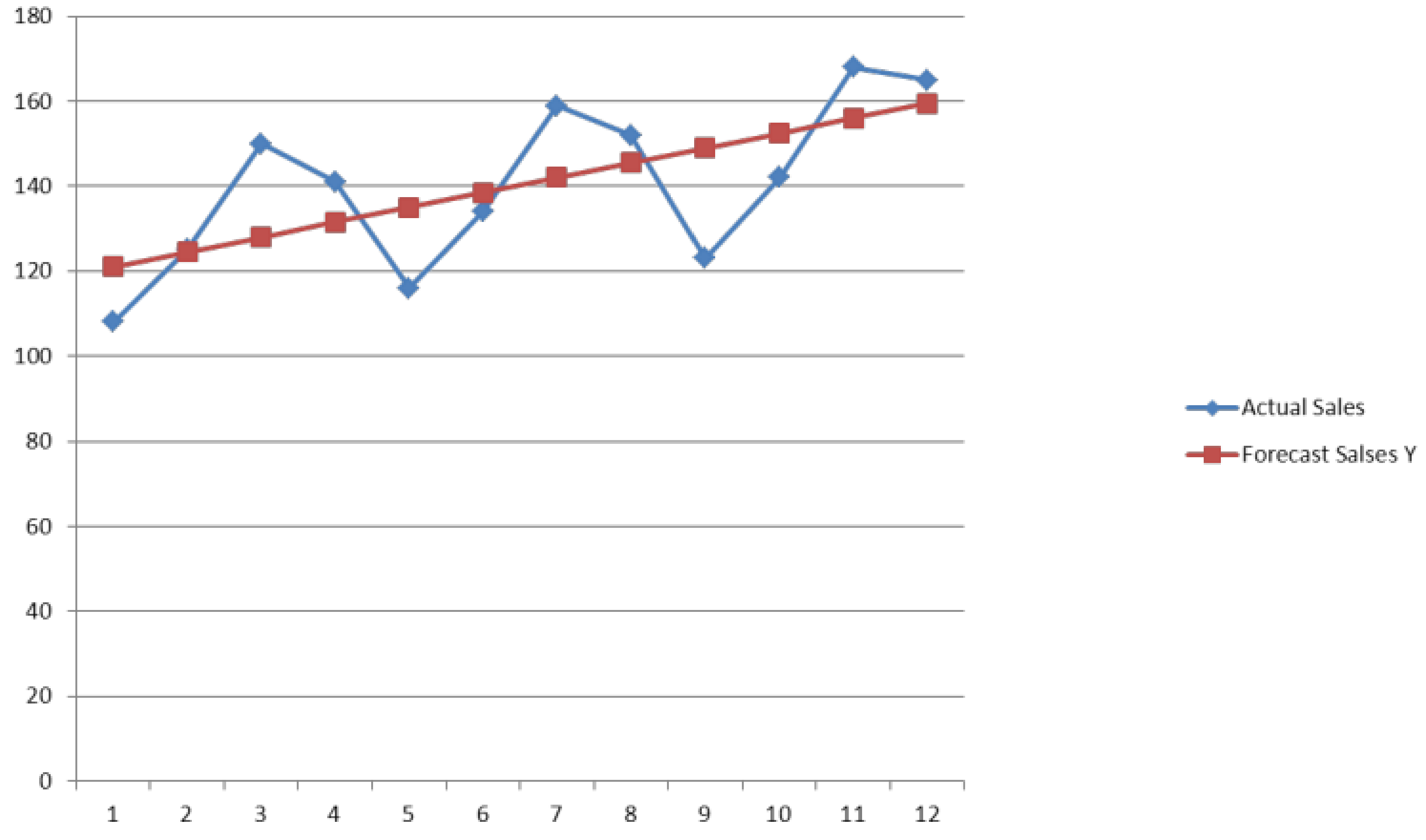
- If no seasonal variation is considered, then we’ll have one X, which represents period series number.
- Regression equation would be: $Y=a+bX$, which is simple regression!

Data for No Season to Enter into QM

Y, Sales	X_1, periods
108	1
125	2
150	3
141	4
116	5
134	6
159	7
152	8
123	9
142	10
168	11
165	12

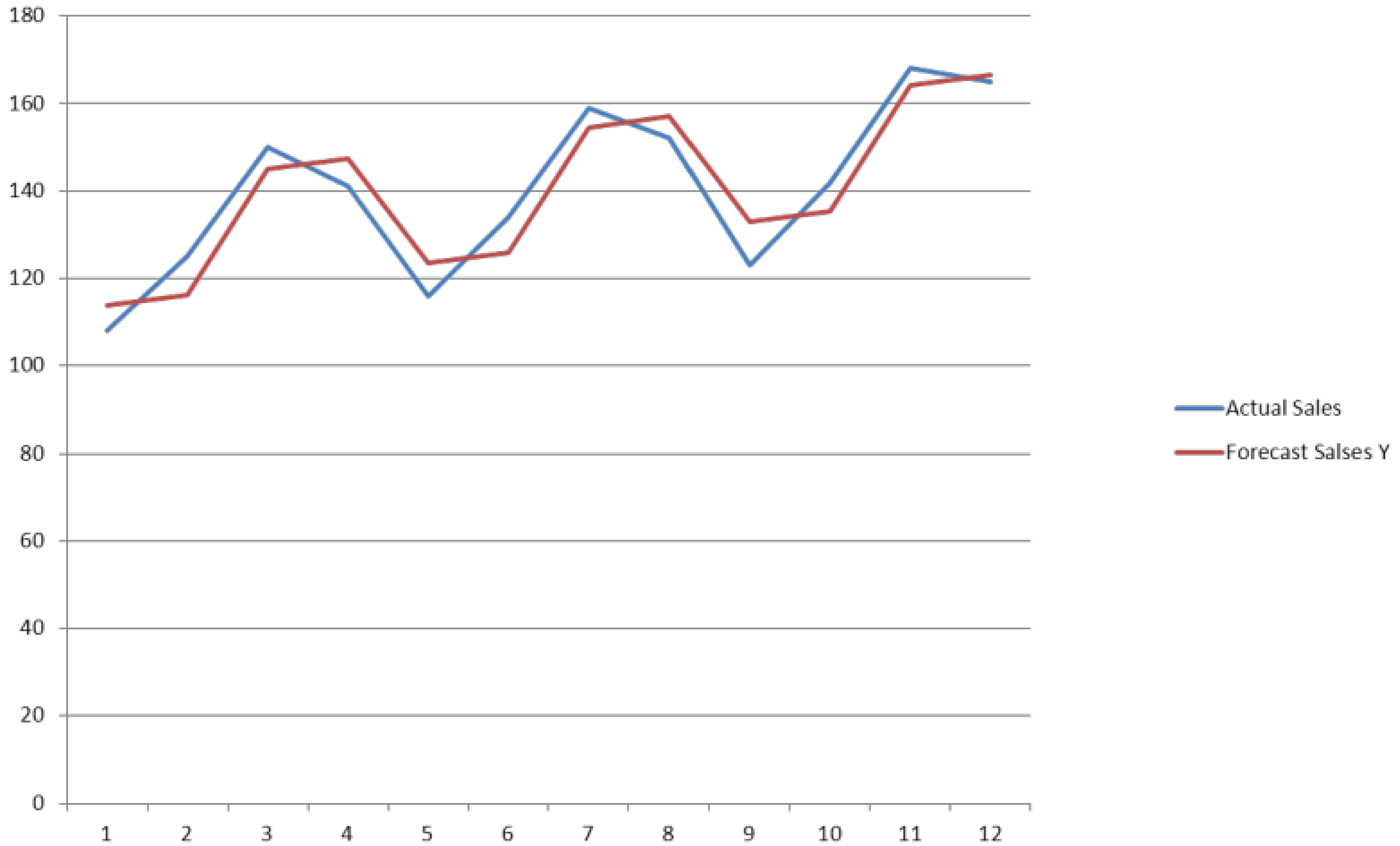
Forecasting by Regression without Considering Seasonal Effects, i.e. Using One Season.

MAD = 12.167



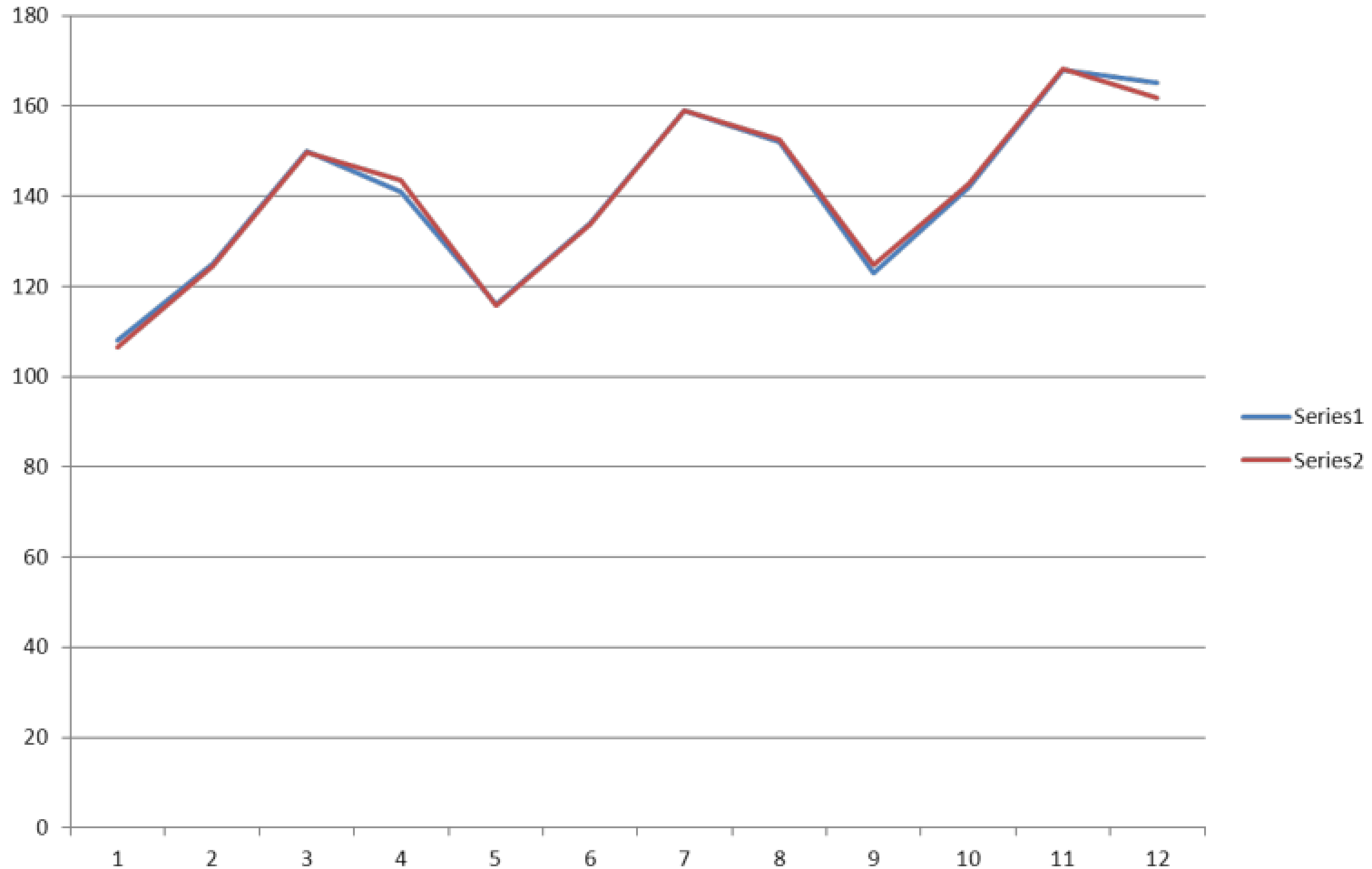
Using Two Seasons

MAD=6.0833



Using Four Seasons

MAD=1.0278



Summary on Independent Variables

- X_1 is always the variable for period series number (1, 2, 3, 4, …), which is used to pick up trend;
- For picking up seasonal variations, 0-1 dummy variables are used, and number of dummy variables is one less than number of seasons. Particularly, $X_2=1$ for 2^{nd} season, $X_3=1$ for 3^{rd} season, …, and zero for each dummy variable represents 1^{st} season.

Regression in General

- A regression equation shows the relationship between Y and (X_1, X_2, …, X_n).
- If X_1=area family income, X_2=avg property tax, X_3=avg. number of kids in a family, Y=area sales revenue of cars, then the regression equation between them can be used for forecasting car sales on changed income, property tax and family size.

Other Approaches for Seasonal Variations

- Other than the multiple regression with dummy variables as we have just learned, there are alternative methods taking care of seasonal variations, such as the *Decomposition Method* as in p.175-177.
- But the regression method is simple and straightforward, and it forecasts as good as the other methods do.