Chapter 2

Forecasting

Past date | Demand

D_1

D_2

$\vdots$

D_N

approximated value | Error

SE MSE

Methods

MV ES Linear Regression

Introduction to Forecasting

- What is forecasting?
  - Primary Function is to Predict the Future
- Why are we interested?
  - Affects the decisions we make today
- Examples: who uses forecasting in their jobs?
  - forecast demand for products and services
  - forecast availability of manpower
  - forecast inventory and materiel needs daily

what is

Subjective Forecasting Methods

1. Sales Force Composites
   - Aggregation of sales personnel estimates by sales manager for each geographic region or product group.
2. Customer Surveys
3. Jury of Executive Opinion

Define

4. The Delphi Method
   - Individual opinions are compiled and reconsidered. Repeat until and overall group consensus is (hopefully) reached.

what is the

Objective Forecasting Methods

- Two primary methods: ***causal models*** (1) and ***time series*** (2) *methods*

MV ES Linear

Explain

Causal Models

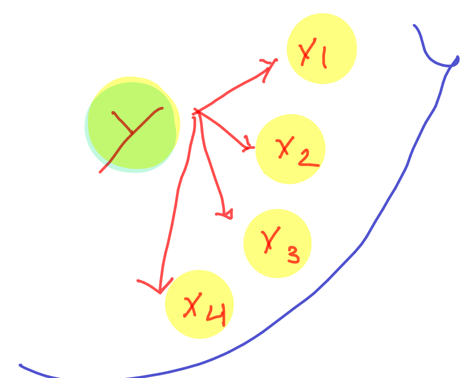
w E Example

Let Y be the quantity to be forecasted and

$(X_1, X_2, \ldots, X_n)$ be n variables that have predictive power for Y.

A causal model is $Y = f(X_1, X_2, \ldots, X_n)$

A typical relationship is a linear one. That is,



Linear

Quadratic

$$Y = a_0 + a_1X_1 + \ldots + a_n X_n$$

Causal Models

Example

Let us consider a simple example of a causal forecasting model. A realtor is trying to estimate his income for the succeeding year. In the past he has found that his income is close to being proportional to the total number of housing sales in his territory. He also has noticed that there has typically been a close relationship between housing sales and interest rates for home mortgages. He might construct a model of the form

$$Y_t = \alpha_0 + \alpha_1 X_{t-1},$$

where Y_t is the number of sales in year t and X_{t-1} is the interest rate in year $t-1$. Based on past data he would then determine the least squares estimators for the constants α_0 and α_1. Suppose that the values of these estimators are currently $\alpha_0 = 385.7$ and $\alpha_1 = -1{,}878$. Hence, the estimated relationship between home sales and mortgage rates is

$$Y_t = 385.7 - 1{,}878 X_{t-1},$$

where X_{t-1}, the previous year's interest rate, is expressed as a decimal. Then if the current mortgage interest rate is 10 percent, the model would predict that the number of sales the following year in his territory would be $385.7 - 187.8 = 197.9$, or about 198 houses sold.

Home

Time Series Methods

MA
ES
Linear

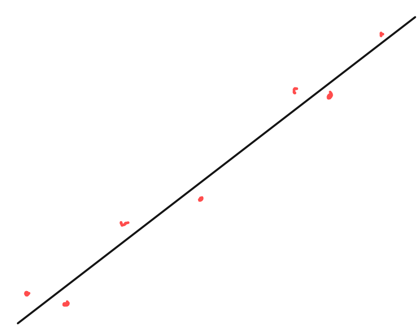
Given Past date

A time series is just collection of past values of the variable being predicted. Also, known as naïve methods.

Past data may exhibit the following patterns:

- Trend
- Seasonality
- Cycles
- Randomness

Past date



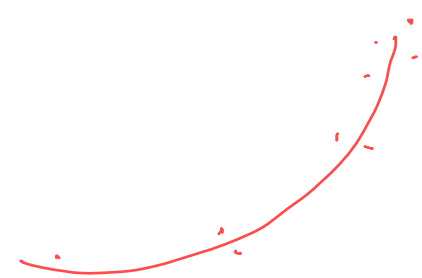
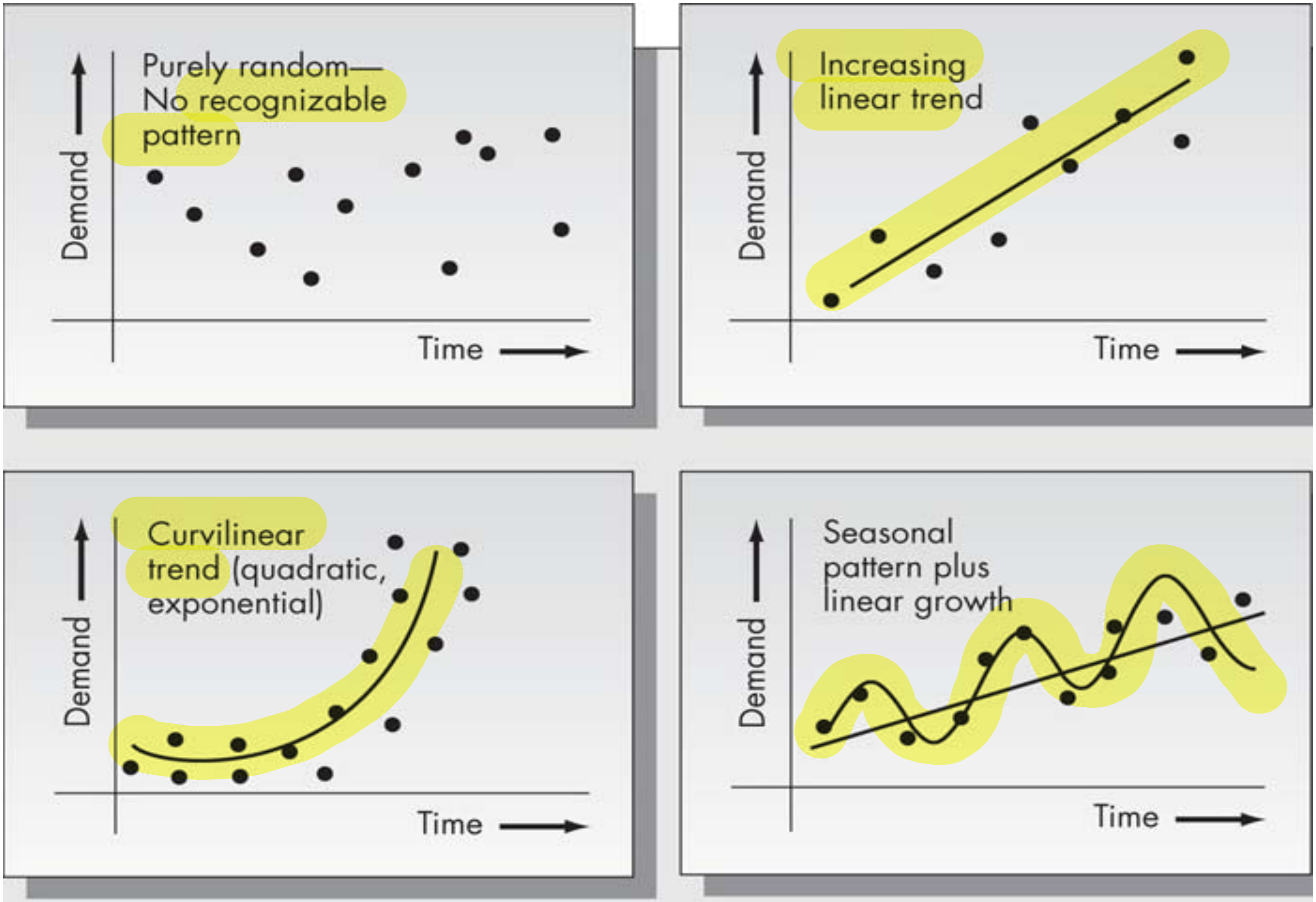


Figure 2.2


Purely random—
No recognizable
pattern
Demand
Time
Increasing
linear trend
Demand
Time
Curvilinear
trend (quadratic,
exponential)
Demand
Time
Seasonal
pattern plus
linear growth
Demand
Time

Notation Conventions

D_1
D_2
$\vdots$
D_n

Past values of the series (actual Demand)

$F_{t,t+1}$ = Forecast made in t for Demand in ~~t+1~~

$t = 1, 2, 3$

$F_{t-1,t}$ = // // // $t-1$ // // // t

F_t = Forecast for Demand in t

F_5 = // // // // 5

F_{10} = // // // // 10

Evaluation of Forecasts

The forecast error in period *t*, e_t, is the difference between the forecast for demand in period *t* and the actual value of demand in *t*.

approximate time

true value

For one step ahead forecast: $|e_t| = |F_t - D_t|$

Mean absolute deviation (MAD): $MAD = \frac{\sum |e_i|}{n}$

Mean squared error (MSE): $MSE = \frac{\sum e_i^2}{n}$

Mean absolute percentage error (MAPE): $= \frac{\sum \left(e_i / D_i \right)}{n} * 100$

Example 2.1:

Artel, a manufacturer of static random access memories (SRAMs), has production plants in Austin, Texas, and Sacramento, California. The managers of these plants are asked to forecast production yields (measured in percent) one week ahead for their plants. Based on six weekly forecasts, the firm's management wishes to determine which manager is more successful at predicting his plant's yields. The results of their predictions are given in the following table.

$E_1 = |F_1 - D_1|$ $\qquad$ $E_2 = |F_2 - D_2|$

Week	P1 $= F_1$	O1 $= D_1$	|E1|	E_1^2	|E1/O1|	P2 $= F_2$	O2 $= D_2$	|E2|	E_2^2	|E2/O2|
1	92	88	4	16	.0455	96	91	5	25	.0549
2	87	88	1	1	0.0114	89	89	0	0	0
3	95	97	2	4	0.0206	92	90	2	4	0.022
4	90	83	7	49	0.0843	93	90	3	9	0.0333
5	88	91	3	9	.0330	90	86	4	16	.0465
6	93	93	0	0	.0000	85	89	4	16	.0449
			17	79	0.1948			18	70	0.2018

$n = 6$

Forecast of Plant 1 $\qquad$ Actual demand of Plant 1 $\qquad$ Absolute error

Compare the performance of manager 1 and 2 using MAD, MSE, MAPE.

$|E_1| = |F_1 - D_1|$

$|E_2| = |F_2 - D_2|$

Based on the MADs :-

$$MAD = \frac{\sum |e_i|}{n} \quad \left.\begin{aligned} MAD_1 &= \frac{17}{6} = 2.83 \\ MAD_2 &= \frac{18}{6} = 3 \end{aligned}\right\}$$

The First Manger is better

Based on the MSE :-

$$MSE = \frac{\sum E_i^2}{n} \quad \left.\begin{aligned} MSE_1 &= \frac{79}{6} = 13.17 \\ MSE_2 &= \frac{70}{6} = 11.67 \end{aligned}\right\}$$

The Second Manger is better

Based on the MAPE :-

$$MAPE = \frac{\sum \left(\frac{e_i}{D_i}\right)}{n} * 100 \quad \left.\begin{aligned} MAPE_1 &= \frac{0.1948}{6} * 100 = 3.25\,\% \\ MAPE_2 &= \frac{0.2018}{6} * 100 = 3.36\,\% \end{aligned}\right\}$$

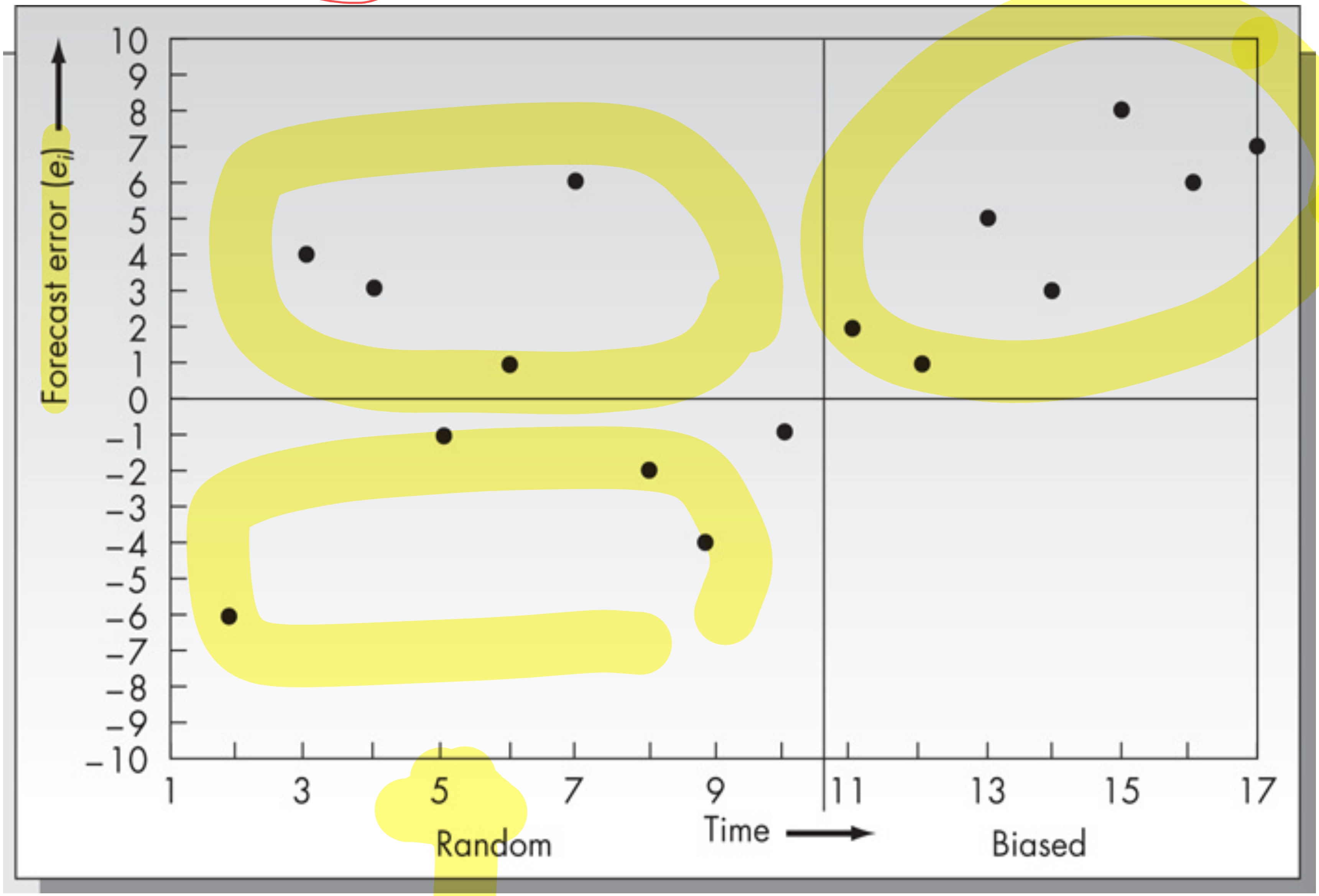
The Firest Manger is better

Biases in Forecasts

- A bias occurs when the ***average* value** of a forecast error tends to be positive or negative.

- Mathematically an unbiased forecast is one in which $E(e_i) = 0$.

Forecast Errors Over Time



A time series is just collection of past values of the variable being predicted. Also, known as naïve methods.

Explain

Forecasting for Stationary Series

and its type

A stationary time series (similar to the graph in the previous slide) has the form:

$D_t = m + e_t$ where m is a constant and e_t is a random variable with mean 0 and var s^2 .

Two common methods for forecasting stationary series are ***moving averages*** and ***exponential smoothing***.

MA

ES

MA

Moving Averages

F_t using D_{t-1}, D_{t-2}, D_{t-3} - - - -

The arithmetic average of the ***N*** **most recent** observations.

1 - step ahead

For a one-step-ahead forecast: MA_3

$$F_{100} = \frac{D_{99} + D_{98} + D_{97}}{3}$$

$$F_5 = \frac{D_4 + D_3 + D_2}{3}$$

$$F_t = (1/N)\,(D_{t-1} + D_{t-2} + \ldots + D_{t-n})$$

$$F_{49} = \frac{D_{48} + D_{47} + D_{46}}{3}$$

$$F_8 = \frac{D_7 + D_6 + D_5}{3}$$

two step ahead forecast $\boxed{MA}_3$

$$F_6 = \frac{D_4 + D_3 + D_2}{3}$$

Example 2.2:

Quarterly data for the failures of certain aircraft engines at a local military base during the last two years are $\overset{D_1}{200}$, $\overset{D_2}{250}$, $\overset{D_3}{175}$, $\overset{D_4}{186}$, $\overset{D_5}{225}$, $\overset{D_6}{285}$, $\overset{D_7}{305}$, $\overset{D_8}{190}$. Both three-quarter (MA3) and six-quarter (MA6) moving averages are used to forecast the numbers of engine failures. Determine the one-step-ahead forecasts for periods 4 through 8 using three-period moving averages, and the one-step-ahead forecasts for periods 7 and 8 using six-period moving averages.

The three-period moving-average forecast for period 4 is obtained by averaging the first three data points.

$$F_4 = \frac{D_3 + D_2 + D_1}{3} = \frac{175 + 250 + 200}{3} = 208$$

F_6

F_7

F_8

$$F_8 = \frac{\overset{305}{D_7} + \overset{285}{D_6} + \overset{225}{D_5}}{3} = 272$$

The three-period moving-average forecast for period 5 is

$$F_5 = \frac{D_4 + D_3 + D_2}{3} = \frac{186 + 175 + 250}{3} = 204$$

The six-period moving-average forecast for period 7 is

$$F_7 = \frac{D_6 + D_5 + D_4 + D_3 + D_2 + D_1}{6} = \frac{285 + 225 + 186 + 175 + 250 + 200}{6} = 220$$

Other forecasts are computed in a similar fashion. Arranging the forecasts and the associated forecast errors in a table, we obtain

$$F_8 = \frac{D_7 + D_6 + D_5 + D_4 + D_3 + D_2}{6} = 238$$

Example 2.2:

Quarter	Engine Failures	MA(3)	Error	MA(6)	Error
1	200				
2	250				
3	175				
4	186	208	22		
5	225	204	−21		
6	285	195	−90		
7	305	232	−73	220	−85
8	190	272	82	238	48

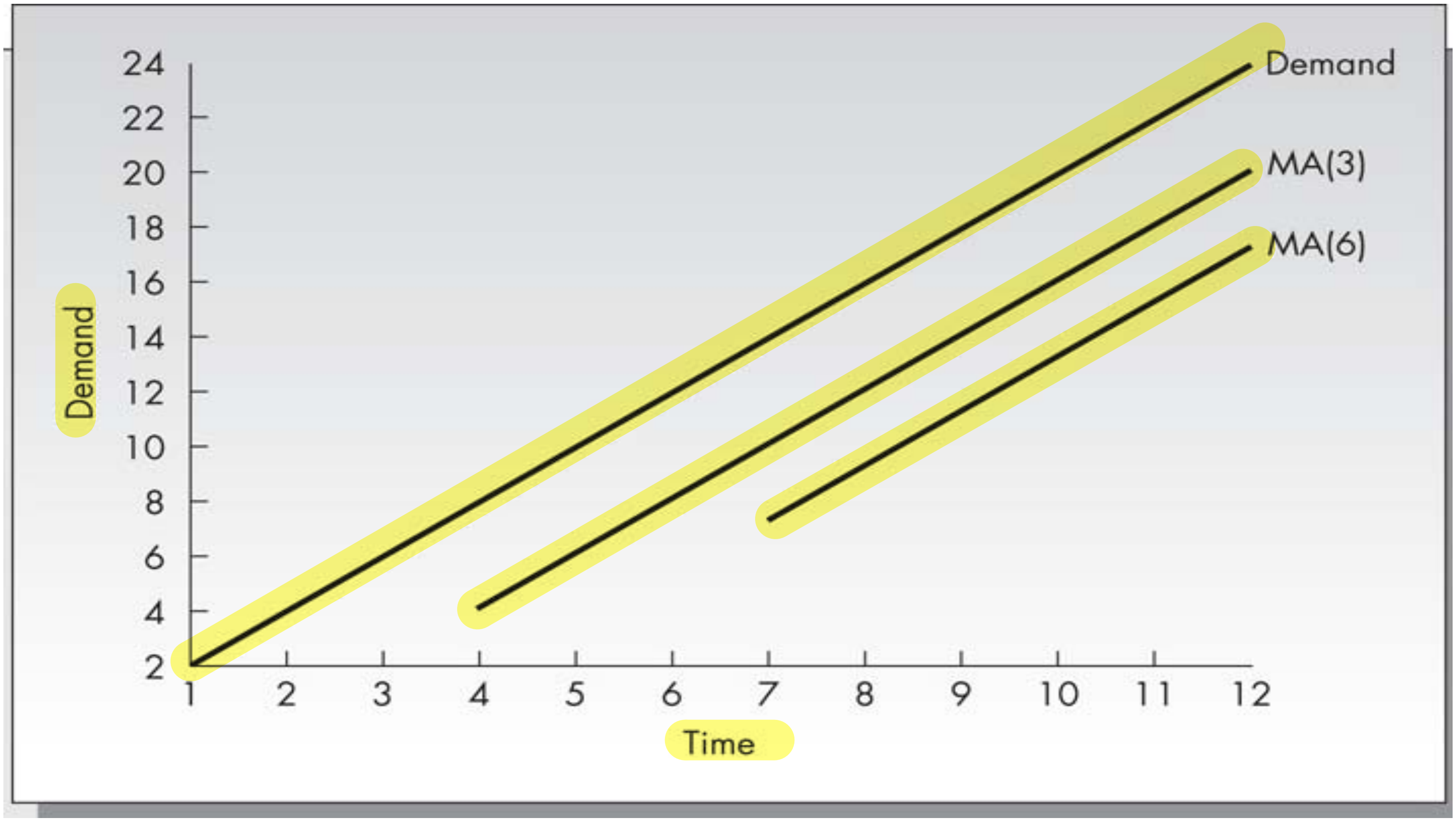
Example 2.2:

An interesting question is, how does one obtain multiple-step-ahead forecasts? For example, suppose in Example 2.2 that we are interested in using MA(3) in period 3 to forecast for period 6. Because the moving-average method is based on the assumption that the demand series is stationary, the forecast made in period 3 for *any* future period will be the same. That is, the multiple-step-ahead and the one-step-ahead forecasts are identical (although the one-step-ahead forecast will generally be more accurate). Hence, the MA(3) forecast made in period 3 for period 6 is 208. In fact, the MA(3) forecast made in period 3 for any period beyond period 3 is 208 as well.

An apparent disadvantage of the moving-average technique is that one must recompute the average of the last N observations each time a new demand observation becomes available. For large N this could be tedious. However, recalculation of the full

Reading

Moving Average Lags a Trend



Exponential Smoothing Method

$\alpha \Rightarrow$ given

Demand

last Forecast

A type of weighted moving average that applies declining weights to past data.

New Forecast = α (Most recent observation)
+ (1 - α) (Last forecast)

$$F(t+1) = \alpha D_t + (1-\alpha) F(t)$$

or

New Forecast = last forecast - α (last forecast error)

where $0 < \alpha < 1$ and generally is small for stability of forecasts (around 0.1 to 0.2)

$$0 < \alpha < 1$$

Example 2.3:

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

Consider Example 2.2 (on page 65), in which moving averages were used to predict aircraft engine failures. The observed numbers of failures over a two-year period were 200, 250, 175, 186, 225, 285, 305, 190. We will now forecast using exponential smoothing. In order to get the method started, let us assume that the forecast for period 1 was 200. Suppose that $\alpha = .1$. The one-step-ahead forecast for period 2 is

F_1

$$F_2 = \alpha D_1 + (1-\alpha) F_1 = 0.1(200) + (1-0.1)(200) = 200$$

$$F_3 = \alpha D_2 + (1-\alpha) F_2 = 0.1(250) + (1-0.1)(200) = 205$$

$$F_4 = \alpha D_3 + (1-\alpha) F_3 = 0.1(175) + (1-0.1)(205) = 202$$

$$F_5 = 201$$

Other one-step-ahead forecasts are computed in the same fashion. The observed numbers of failures and the one-step-ahead forecasts for each quarter are the following:

$$F_6 = 203$$

$$F_7 = 211$$

$$F_8 = \alpha D_7 + (1-\alpha) F_7 = 0.1(305) + (1-0.1)(211) = 220$$

Example 2.3:

Quarter	Failures	Forecast
1	200	200 (by assumption)
2	250	200
3	175	205
4	186	202
5	225	201
6	285	203
7	305	211
8	190	220

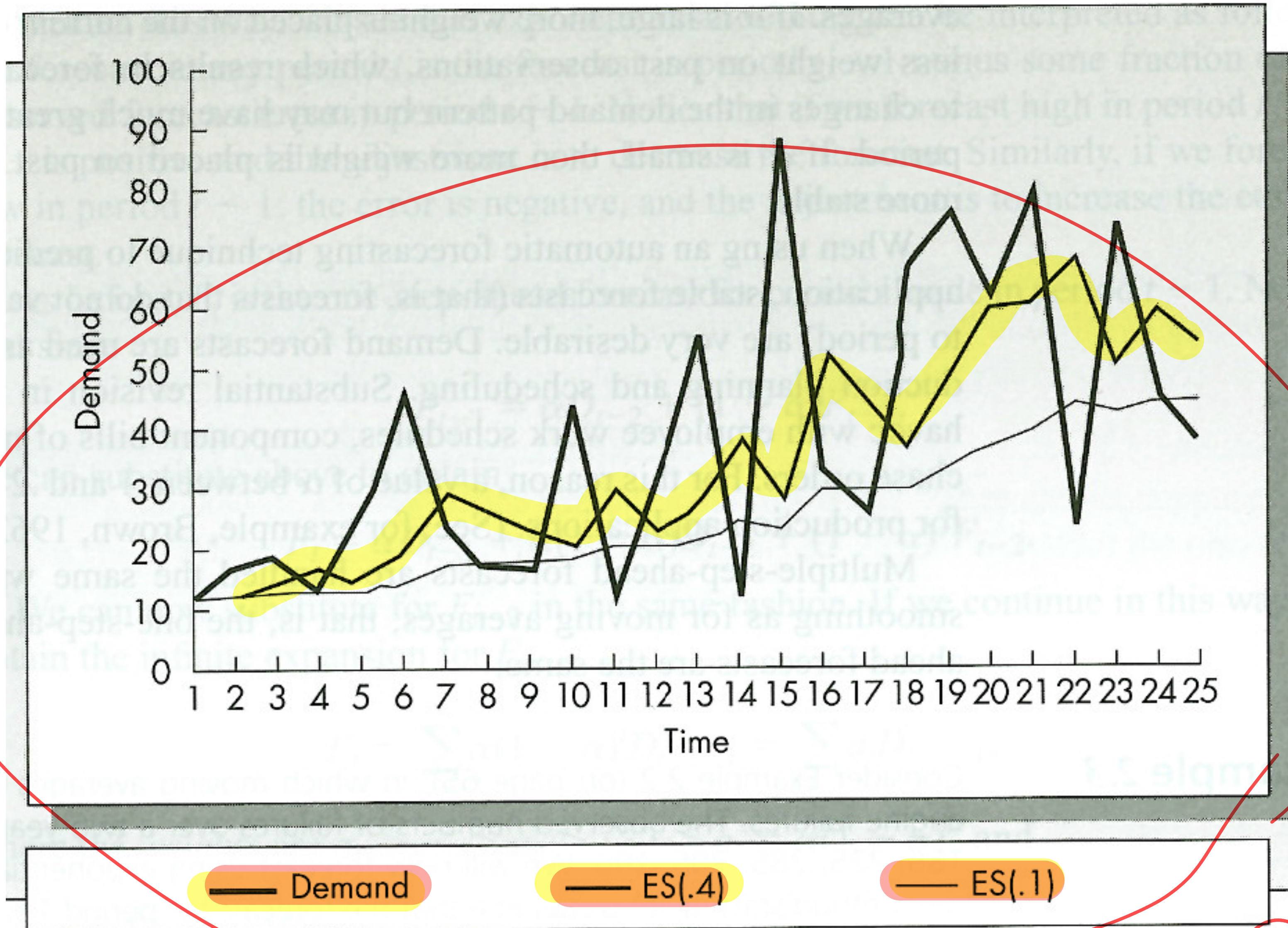
Notice the effect of the smoothing constant. Although the original series shows high variance, the forecasts are quite stable. Repeat the calculations with a value of $\alpha = .4$. There will be much greater variation in the forecasts.

α is smoothing constant which control forecast stability

Midterm 1

$0 < \alpha < 1$

Effect of the value o alpha on the forecast stability



Midterm 1

Forecasting For Stationary Series

Three MA

Moving averages

one-step-head

$$F_4 = \frac{D_3 + D_2 + D_1}{3}$$

$$F_{10} = \frac{D_9 + D_8 + D_7}{3}$$

exponential

smooth Method

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

$$F_4 = \alpha D_3 + (1-\alpha) F_3$$

$$F_{10} = \alpha D_9 + (1-\alpha) F_9$$

$$|e_t| = |F_t - D_t|$$

MAD

MSE

MAPE

Using Regression for Times Series Forecasting

$D_t = a + bt$

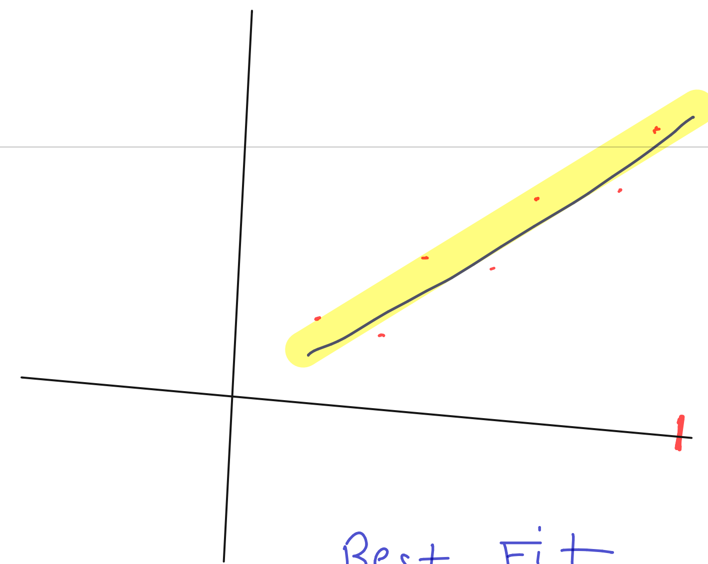
a = ??

b = ??

x	y
t	D Given
1	D_1
2	D_2
3	D_3
⋮	⋮

S_{xy} = ??

S_{xx} = ??



Best Fit

Using Regression for Times Series Forecasting

Set $S_{xy} = \left(n \sum i * D_i \right) - \left(\left(n(n+1)/2\right) \sum D_i\right)$

Set $S_{xx} = \left(n^2(n+1)(2n+1)/6\right) - \left(n(n+1)/2\right)^2$

Let $b = \dfrac{S_{xy}}{S_{xx}}$ $\quad a = \bar{D} - \dfrac{b(n+1)}{2}$

$$\bar{D} = \frac{D_1 + D_2 \cdots D_n}{n}$$

These values of ***a*** and ***b*** *provide the "best" fit of the data in a least squares sense.*

Example 2.4:

We will apply regression analysis to the problem, treated in Examples 2.2 and 2.3, of predicting aircraft engine failures. Recall that the demand for aircraft engines during the last eight quarters was 200, 250, 175, 186, 225, 285, 305, 190. Suppose that we use the first five periods as a baseline in order to estimate our regression parameters. Then

$$\Rightarrow S_{xy} = \left(n \sum i \times D_i\right) - \left(\left(n(n+1)/2\right) \sum D_i\right)$$

$$= \left(5\left(1*200 + 2*250 + 3*175 + 4*186 + 5*225\right)\right) - \left(\frac{5(6)}{2}\left(200+250+175+186+225\right)\right) = -70$$

$$\Rightarrow S_{xx} = \left(n^2(n+1)(2n+1)/6\right) - \left(\frac{n(n+1)}{2}\right)^2$$

$$= \left(25(6)(11)/6\right) - \left(\frac{5(6)}{2}\right)^2 = 50$$

$$\Rightarrow \bar{D} = \frac{200+250+175+186+225}{5} = 207.2$$

$$b = \frac{S_{xy}}{S_{xx}} = \frac{-70}{50} = -7/5$$

$$a = \bar{D} - \frac{b(n+1)}{2} = 207.2 - \frac{(-7/5)(6)}{2}$$

$$= 211.4$$

$$D_t = 211.4 - 7/5\, t$$

$$D_6 = 211.4 - 7/5(6) = \checkmark$$

$$D_9 = 211.4 - 7/5(9) = \checkmark$$

$\boxed{80\%}$

$(*)$