Machine Learning 520
Advanced Machine Learning

Lesson 9: Forecasting

Today's Agenda

- What is forecasting?
- Autocorrelation, seasonality and noise
- Auto regression
- Moving average
- Autoregressive Integrated Moving Average

Learning Objectives

- Decompose time series into autocorrelation, seasonality, trend, and noise.
- Explain the effects of exponential smoothing models and differentiate them from other models.
- Apply and evaluate the results of an autoregressive model.
- Apply and evaluate the results of a moving average model.
- Apply and evaluate the results of an autoregressive integrated moving average model.
- Apply and evaluate the results of ARIMA model for forecasting (time series prediction).

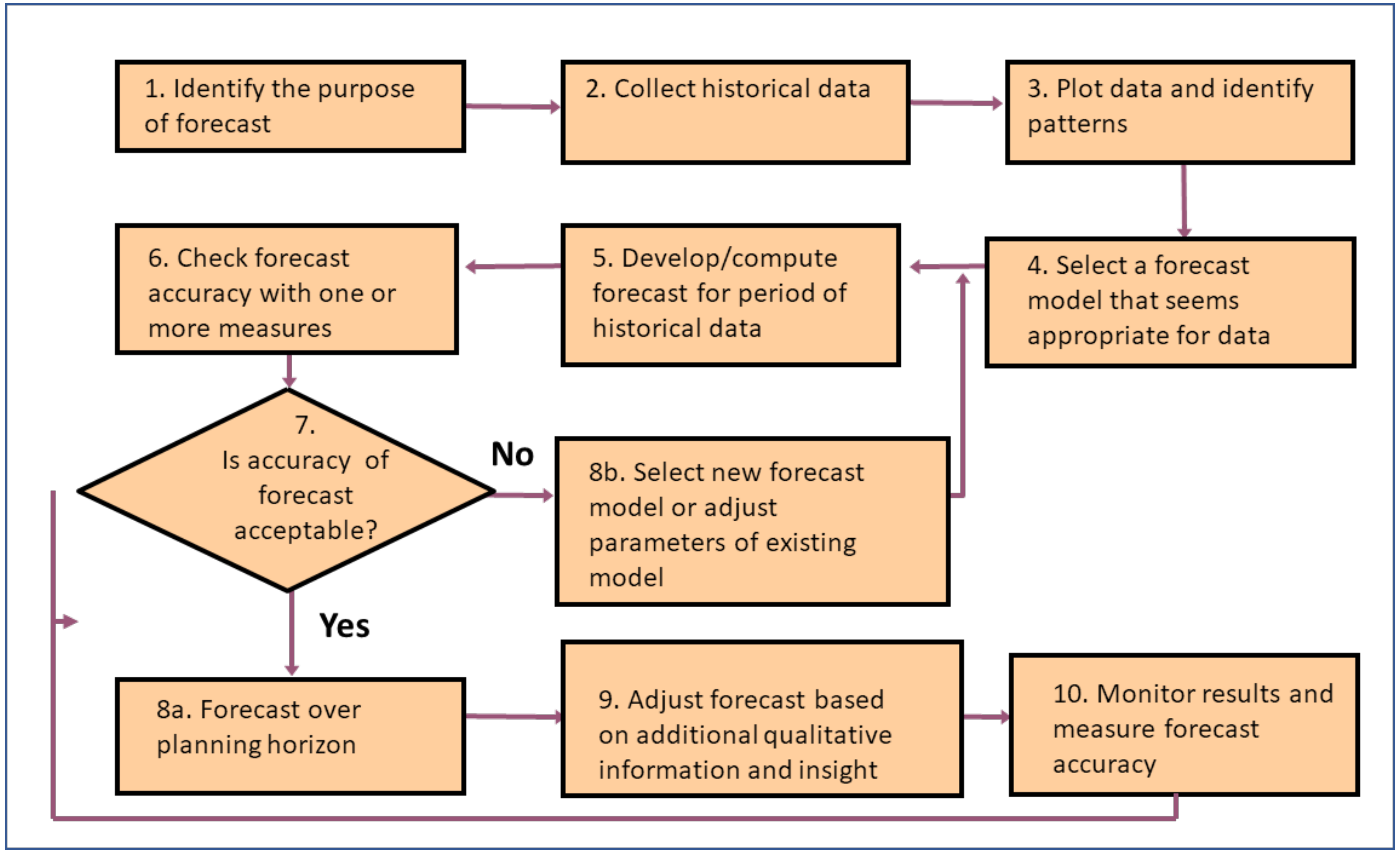
Prediction is very difficult, especially about the future.

- Niels Bohr

Types of Forecasting Methods

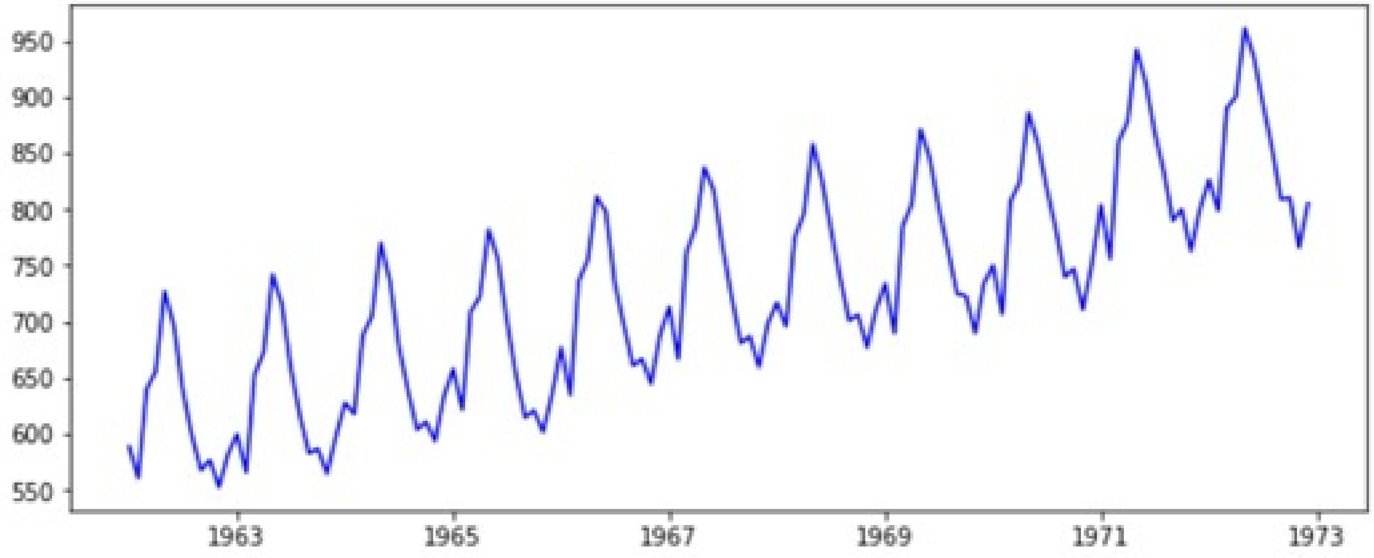
- Decide what needs to be forecast
  - Level of detail, units of analysis & time horizon required
- Evaluate and analyze appropriate data
  - Identify needed data & whether it's available
- Select and test the forecasting model
  - Cost, ease of use & accuracy
- Generate the forecast
- Monitor forecast accuracy over time

Forecasting In Practice



Time Series Data

- Series of data observed over time
- E.g.: Daily Amazon stock prices, Annual population data, Quarterly sales figures, ...



- Forecasting is estimating how the sequence of observations will continue into the future.

Time Series Data

> **Order matters**

There is a dependency on time and changing the order could change the meaning of the data.

Quiz

> A time series is a sequence of data points with the follow characteristics (choose all that apply):

1- Random intervals of time between each data point
2- Covers a continuous time interval
3- Equal spacing between every two consecutive measurements
4- Each time unit within the time interval has at most one data point.

Quiz

> We would use time series forecasting to predict values for which following business situations (select all that apply):

1- Monthly Seattle bike rentals
2- A stock's daily closing value
3- Amount of time to close a sales cycle
4- Retail trade area analysis and demographic profiling
5- Annual Orca population

Time Series Pattern

> Trend:
  – Long-term movement in data

> Seasonality:
  – Short-term regular variations in data

> Cycles:
  – Wavelike variations of long-term

> Irregular variations:
  – Caused by unusual circumstances

> Random variations
  – Caused by chance

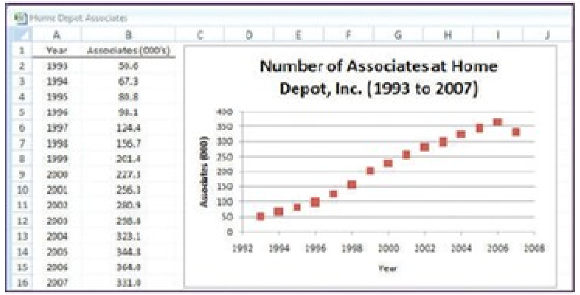
Time Series Components

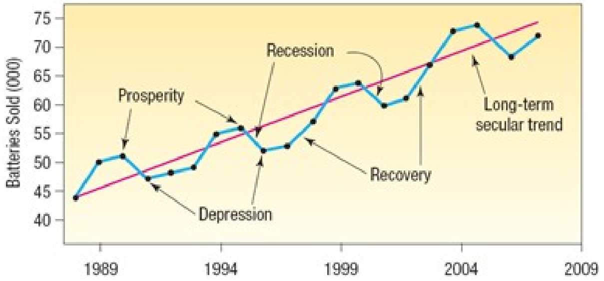
1. Secular Trend – the smooth long term direction of a time series

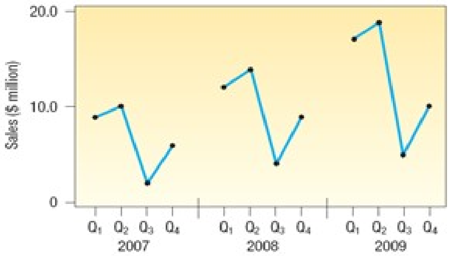
2. Cyclical Variation – the rise and fall of a time series over periods longer than one year

3. Seasonal Variation – Patterns of change in a time series within a year which tends to repeat each year

4. Irregular Variation – classified into:
   Episodic – unpredictable but identifiable
   Residual – also called chance fluctuation and unidentifiable



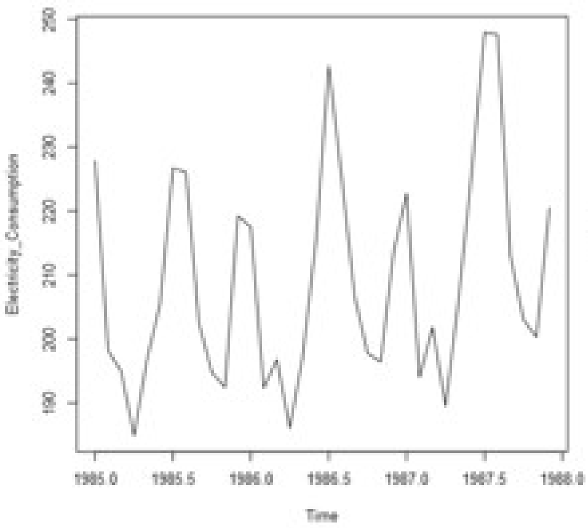


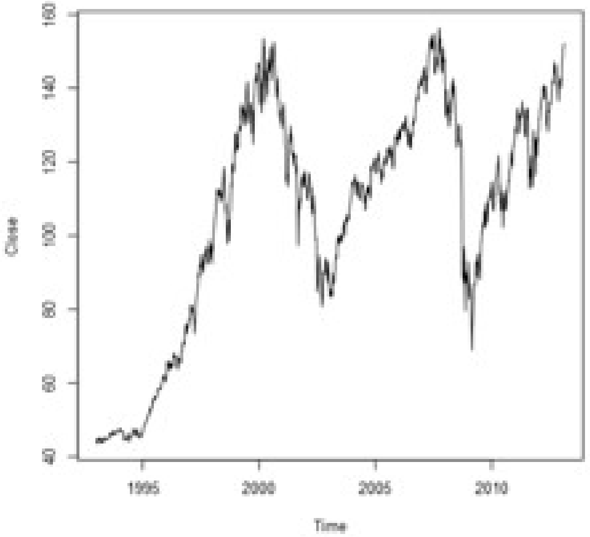


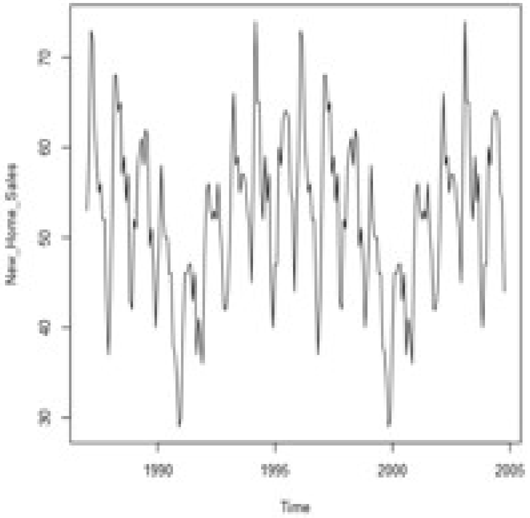
Sales of Baseball and Softball Equipment, Hercher Sporting Goods, 2007–2009 by Quarter

Quiz

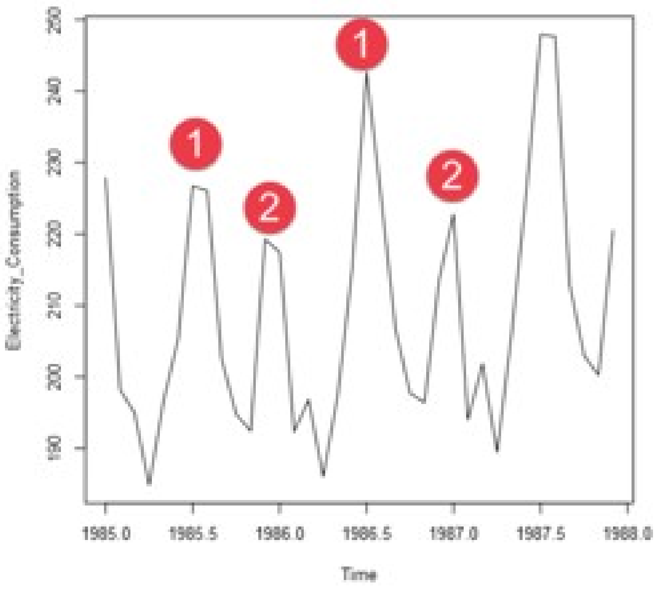
> Mark each time series plots as seasonal, cyclical, both or neither.

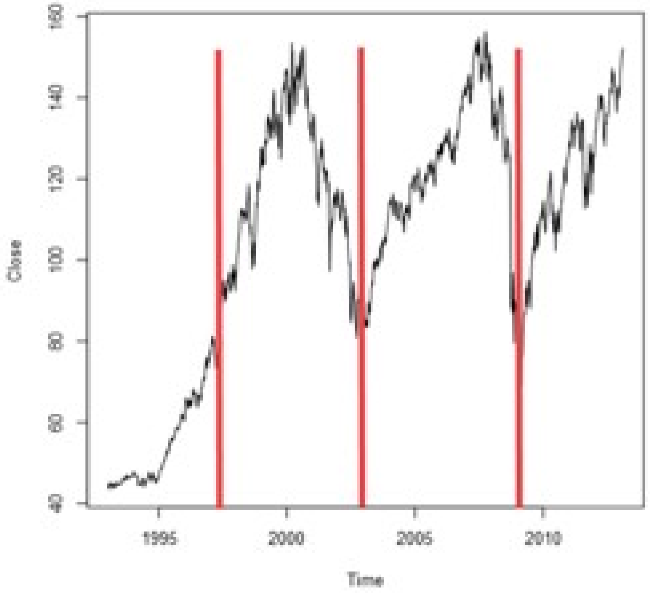


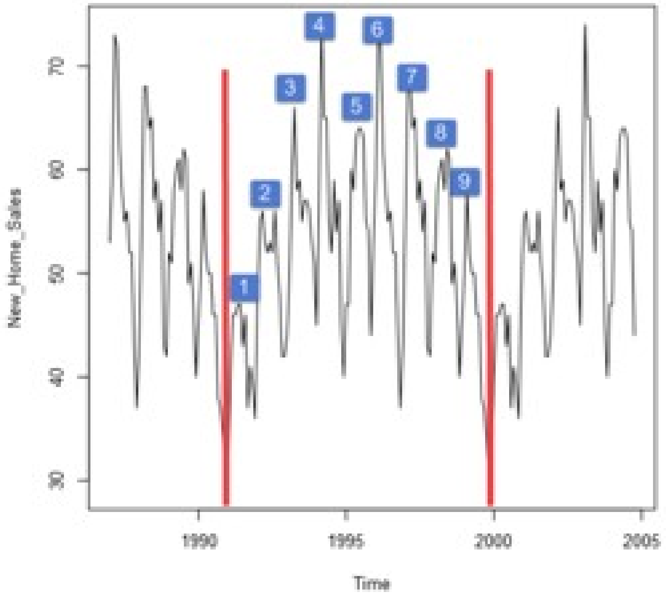




Quiz





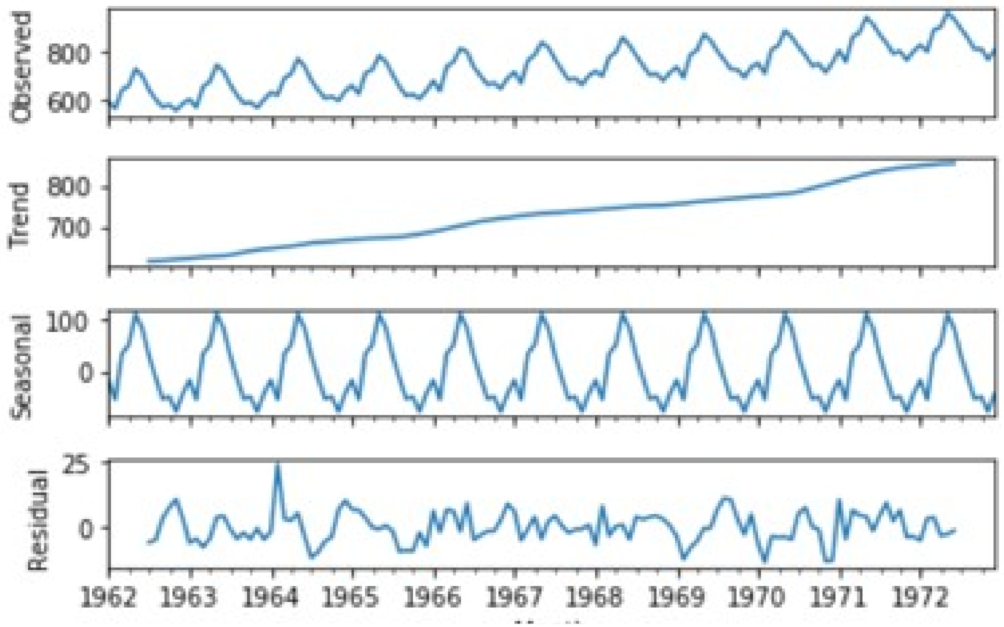


Seasonal - Peak electricity in summer and winter months repeats every year

Cyclical - shows pattern of downtrend every few years. The years are not fixed though.

Both - Seasonal patterns each year with cyclical pattern showing about every 9 years

Time Series Decomposition Plot



> The first plot shows the actual time series.
> The trend line indicates the general tendency of the time series
> The seasonal portion shows that there’s a seasonal pattern
> The residual is the error in the model that calculates the difference between the observed value and the trendline estimate

Simple Forecasting Methods

- **Average method**: Forecast of all future values is equal to the mean of historical data

- **Naïve method**: Forecasts equal to last observed value

- **Seasonal naïve method**: Forecasts equal to last value from same season

- **Drift method**: Forecasts equal to last value plus average change

Moving Average: Naïve Approach

MONTH	ORDERS PER MONTH	FORECAST
Jan	120	-
Feb	90	120
Mar	100	90
Apr	75	100
May	110	75
June	50	110
July	75	50
Aug	130	75
Sept	110	130
Oct	90	110
Nov	-	90

Moving Average

- Simple moving average
  - uses average demand for a fixed sequence of periods
  - stable demand with no pronounced behavioral patterns
- Weighted moving average
  - weights are assigned to most recent data

Simple Moving Average

$$MA_n = \frac{\sum_{i=1}^{n} D_i}{n}$$

where

n = number of periods in the moving average

D_i = demand in period i

3-month Simple Moving Average

MONTH	ORDERS PER MONTH	MOVING AVERAGE
Jan	120	–
Feb	90	–
Mar	100	–
Apr	75	103.3
May	110	88.3
June	50	95.0
July	75	78.3
Aug	130	78.3
Sept	110	85.0
Oct	90	105.0
Nov	-	110.0

$$MA_3 = \frac{\sum_{i=1}^{3} D_i}{3}$$

$$= \frac{90 + 110 + 130}{3}$$

= 110 orders for Nov

5-month Simple Moving Average

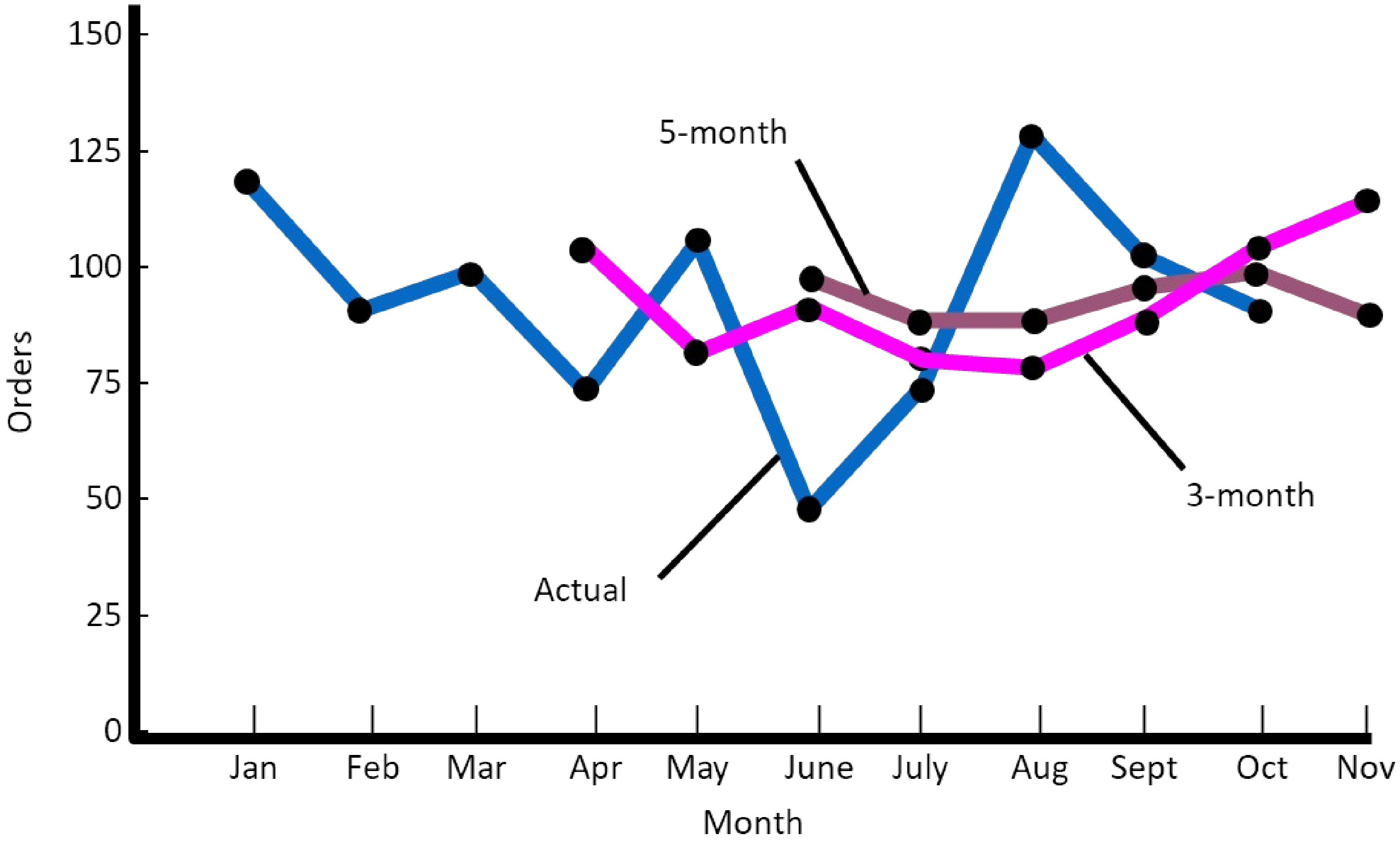
MONTH	ORDERS PER MONTH	MOVING AVERAGE
Jan	120	–
Feb	90	–
Mar	100	–
Apr	75	–
May	110	–
June	50	99.0
July	75	85.0
Aug	130	82.0
Sept	110	88.0
Oct	90	95.0
Nov	-	91.0

$$MA_5 = \frac{\sum_{i=1} D_i}{5}$$

$$= \frac{90 + 110 + 130+75+50}{5}$$

= 91 orders for Nov

Smoothing Effects



Weighted Moving Average

- Adjusts moving average method to more closely reflect data fluctuations

$$WMA_n = \sum_{i=1}^{n} W_i D_i$$

where

W_i = the weight for period i, between 0 and 100 percent

$\sum W_i = 1.00$

Weighted Moving Average Example

MONTH	*WEIGHT*	*DATA*
August	17%	130
September	33%	110
October	50%	90

November Forecast $WMA_3 = \sum_{i=1}^{3} W_i D_i$

$$= (0.50)(90) + (0.33)(110) + (0.17)(130)$$

$$= 103.4 \text{ orders}$$

Exponential Smoothing

- Averaging method
- Weights most recent data more strongly
- Reacts more to recent changes
- Widely used, accurate method

Exponential Smoothing

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

where:

F_{t+1} = forecast for next period

D_t = actual demand for present period

F_t = previously determined forecast for present period

α = weighting factor, smoothing constant

Effect of Smoothing Constant

$$0.0 \le \alpha \le 1.0$$

If $\alpha = 0.20$, then $F_{t+1} = 0.20\, D_t + 0.80\, F_t$

If $\alpha = 0$, then $F_{t+1} = 0\, D_t + 1\, F_t = F_t$

Forecast does not reflect recent data

If $\alpha = 1$, then $F_{t+1} = 1\, D_t + 0\, F_t = D_t$

Forecast based only on most recent data

Exponential Smoothing (α=0.30)

PERIOD	MONTH	DEMAND
1	Jan	37
2	Feb	40
3	Mar	41
4	Apr	37
5	May	45
6	Jun	50
7	Jul	43
8	Aug	47
9	Sep	56
10	Oct	52
11	Nov	55
12	Dec	54

$$F_2 = \alpha D_1 + (1 - \alpha)F_1$$
$$= (0.30)(37) + (0.70)(37)$$
$$= 37$$

$$F_3 = \alpha D_2 + (1 - \alpha)F_2$$
$$= (0.30)(40) + (0.70)(37)$$
$$= 37.9$$

$$F_{13} = \alpha D_{12} + (1 - \alpha)F_{12}$$
$$= (0.30)(54) + (0.70)(50.84)$$
$$= 51.79$$

Exponential Smoothing (α=0.30)

PERIOD	MONTH	DEMAND
1	Jan	37
2	Feb	40
3	Mar	41
4	Apr	37
5	May	45
6	Jun	50
7	Jul	43
8	Aug	47
9	Sep	56
10	Oct	52
11	Nov	55
12	Dec	54

$$F_2 = \alpha D_1 + (1 - \alpha)F_1$$
$$= (0.30)(37) + (0.70)(37)$$
$$= 37$$

$$F_3 = \alpha D_2 + (1 - \alpha)F_2$$
$$= (0.30)(40) + (0.70)(37)$$
$$= 37.9$$

$$F_{13} = \alpha D_{12} + (1 - \alpha)F_{12}$$
$$= (0.30)(54) + (0.70)(50.84)$$
$$= 51.79$$

Exponential Smoothing

PERIOD	MONTH	DEMAND	FORECAST, F_{t+1} ($\alpha = 0.3$)	($\alpha = 0.5$)
1	Jan	37	–	–
2	Feb	40	37.00	37.00
3	Mar	41	37.90	38.50
4	Apr	37	38.83	39.75
5	May	45	38.28	38.37
6	Jun	50	40.29	41.68
7	Jul	43	43.20	45.84
8	Aug	47	43.14	44.42
9	Sep	56	44.30	45.71
10	Oct	52	47.81	50.85
11	Nov	55	49.06	51.42
12	Dec	54	50.84	53.21
13	Jan	–	51.79	53.61

Adjusted Exponential Smoothing

$$AF_{t+1} = F_{t+1} + T_{t+1}$$

where

T = an exponentially smoothed trend factor

$$T_{t+1} = \beta(F_{t+1} - F_t) + (1 - \beta)\, T_t$$

where

T_t = the last period trend factor

β = a smoothing constant for trend

$0 \leq \beta \leq 1$

Adjusted Exponential Smoothing (β=0.30)

PERIOD	MONTH	DEMAND
1	Jan	37
2	Feb	40
3	Mar	41
4	Apr	37
5	May	45
6	Jun	50
7	Jul	43
8	Aug	47
9	Sep	56
10	Oct	52
11	Nov	55
12	Dec	54

$$T_3 = \beta(F_3 - F_2) + (1 - \beta)\,T_2$$
$$= (0.30)(38.5 - 37.0) + (0.70)(0)$$
$$= 0.45$$

$$AF_3 = F_3 + T_3 = 38.5 + 0.45$$
$$= 38.95$$

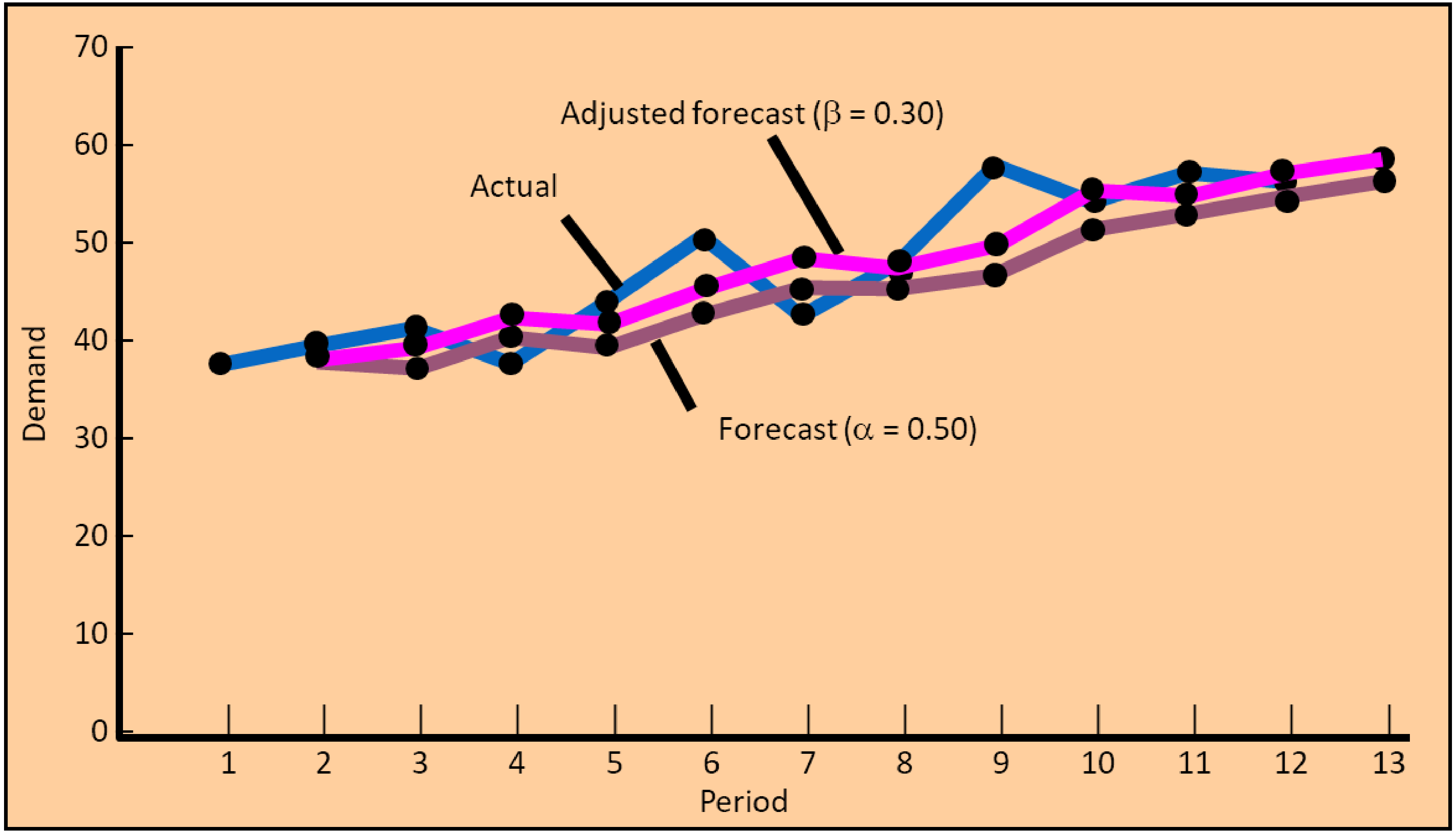
$$T_{13} = \beta(F_{13} - F_{12}) + (1 - \beta)\,T_{12}$$
$$= (0.30)(53.61 - 53.21) + (0.70)(1.77)$$
$$= 1.36$$

$$AF_{13} = F_{13} + T_{13} = 53.61 + 1.36 = 54.97$$

Adjusted Exponential Smoothing

PERIOD	MONTH	DEMAND	FORECAST F_{t+1}	TREND T_{t+1}	ADJUSTED FORECAST AF_{t+1}
1	Jan	37	37.00	–	–
2	Feb	40	37.00	0.00	37.00
3	Mar	41	38.50	0.45	38.95
4	Apr	37	39.75	0.69	40.44
5	May	45	38.37	0.07	38.44
6	Jun	50	38.37	0.07	38.44
7	Jul	43	45.84	1.97	47.82
8	Aug	47	44.42	0.95	45.37
9	Sep	56	45.71	1.05	46.76
10	Oct	52	50.85	2.28	58.13
11	Nov	55	51.42	1.76	53.19
12	Dec	54	53.21	1.77	54.98
13	Jan	–	53.61	1.36	54.96

Adjusted Exponential Smoothing Forecasts



Linear Trend Line

$y = a + bx$

where

a = intercept

b = slope of the line

x = time period

y = forecast for demand for period x

$$b = \frac{\sum xy - n\bar{x}\bar{y}}{\sum x^2 - n\bar{x}^2}$$

$$a = \bar{y} - b\bar{x}$$

where

n = number of periods

$$\bar{x} = \frac{\sum x}{n} = \text{mean of the } x \text{ values}$$

$$\bar{y} = \frac{\sum y}{n} = \text{mean of the } y \text{ values}$$

Least Squares Example

x(PERIOD)	y(DEMAND)	xy	x^2
1	73	37	1
2	40	80	4
3	41	123	9
4	37	148	16
5	45	225	25
6	50	300	36
7	43	301	49
8	47	376	64
9	56	504	81
10	52	520	100
11	55	605	121
12	54	648	144
78	557	3867	650

Least Squares Example

$$\bar{x} = \frac{78}{12} = 6.5$$

$$\bar{y} = \frac{557}{12} = 46.42$$

$$b = \frac{\sum xy - n\bar{x}\bar{y}}{\sum x^2 - n\bar{x}^2} = \frac{3867 - (12)(6.5)(46.42)}{650 - 12(6.5)^2} = 1.72$$

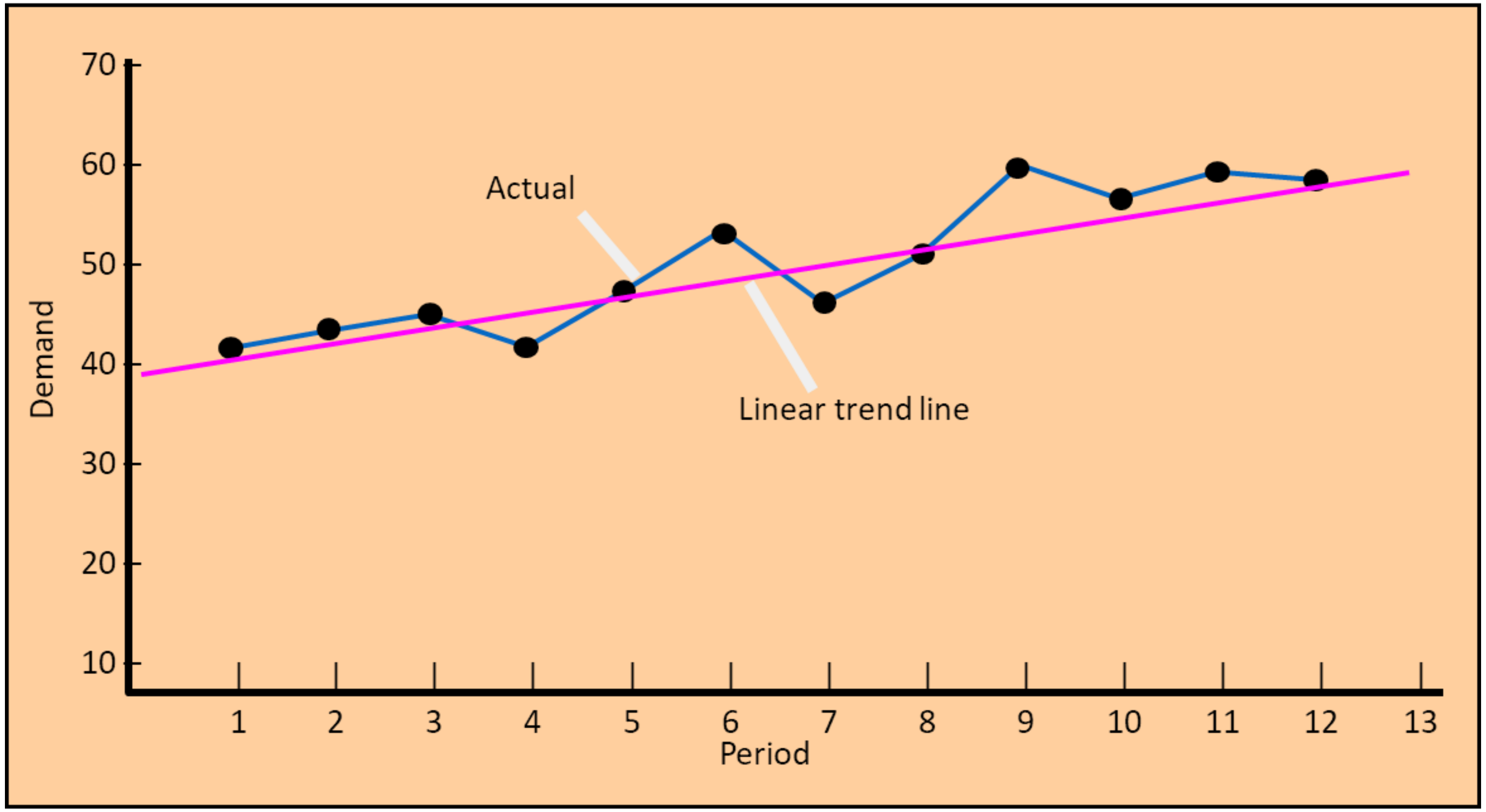
$$a = \bar{y} - b\bar{x}$$

$$= 46.42 - (1.72)(6.5) = 35.2$$

Adjusted Exponential Smoothing Forecasts

Linear trend line $y = 35.2 + 1.72x$

Forecast for period 13 $y = 35.2 + 1.72(13) = 57.56$ units



Seasonal Adjustments

- Repetitive increase/ decrease in demand
- Use seasonal factor to adjust forecast

$$\text{Seasonal factor} = S_i = \frac{D_i}{\sum D}$$

Seasonal Adjustment

DEMAND (1000'S PER QUARTER)

YEAR	*1*	*2*	*3*	*4*	*Total*
2002	12.68.6	6.3	17.5	45.0	
2003	14.110.3	7.5	18.2	50.1	
2004	15.310.6	8.1	19.6	53.6	
Total	42.029.5	21.9	55.3	148.7	

$$S_1 = \frac{D_1}{\sum D} = \frac{42.0}{148.7} = 0.28 \qquad S_3 = \frac{D_3}{\sum D} = \frac{21.9}{148.7} = 0.15$$

$$S_2 = \frac{D_2}{\sum D} = \frac{29.5}{148.7} = 0.20 \qquad S_4 = \frac{D_4}{\sum D} = \frac{55.3}{148.7} = 0.37$$

Seasonal Adjustment

For 2005

$$y = 40.97 + 4.30x \;\; = 40.97 + 4.30(4) = 58.17$$

$$SF_1 = (S_1)\,(F_5) = (0.28)(58.17) = 16.28$$
$$SF_2 = (S_2)\,(F_5) = (0.20)(58.17) = 11.63$$
$$SF_3 = (S_3)\,(F_5) = (0.15)(58.17) = 8.73$$
$$SF_4 = (S_4)\,(F_5) = (0.37)(58.17) = 21.53$$

Forecast Accuracy

- Forecast error
  - difference between forecast and actual demand
- MAD
  - mean absolute deviation
- MAPD
  - mean absolute percent deviation
- Cumulative error
- Average error or bias

Mean Absolute Deviation (MAD)

$$\text{MAD} = \frac{\Sigma|D_t - F_t|}{n}$$

where

t = period number

D_t = demand in period t

F_t = forecast for period t

n = total number of periods

MAD Example

PERIOD	DEMAND, D_t	F_t (α =0.3)	(D_t - F_t)	$\lvert D_t - F_t \rvert$
1	37	37.00	–	–
2	40	37.00	3.00	3.00
3	41	37.90	3.10	3.10
4	37	38.83	-1.83	1.83
5	45	38.28	6.72	6.72
6	50	40.29	9.69	9.69
7	43	43.20	-0.20	0.20
8	47	43.14	3.86	3.86
9	56	44.30	11.70	11.70
10	52	47.81	4.19	4.19
11	55	49.06	5.94	5.94
12	54	50.84	3.15	3.15
	557		49.31	53.39

MAD Calculation

$$\text{MAD} = \frac{\sum |D_t - F_t|}{n}$$

$$= \frac{53.39}{11}$$

$$= 4.85$$

Comparison of Forecasts

FORECAST	MAD	MAPD	E	$(\bar{E})$
Exponential smoothing ($\alpha = 0.30$)	4.85	9.6%	49.31	4.48
Exponential smoothing ($\alpha = 0.50$)	4.04	8.5%	33.21	3.02
Adjusted exponential smoothing ($\alpha = 0.50$, $\beta = 0.30$)	3.81	7.5%	21.14	1.92
Linear trend line	2.29	4.9%	–	–

Regression Methods

- Linear regression
  - mathematical technique that relates a dependent variable to an independent variable in the form of a linear equation
- Correlation
  - a measure of the strength of the relationship between independent and dependent variables

Linear Regression

$$y = a + bx \qquad a = \bar{y} - b\bar{x}$$

$$b = \frac{\sum xy - n\bar{x}\bar{y}}{\sum x^2 - n\bar{x}^2}$$

where

a = intercept

b = slope of the line

$\bar{x} = \frac{\sum x}{n}$ = mean of the x data

$\bar{y} = \frac{\sum y}{n}$ = mean of the y data

Linear Regression Example

x (WINS)	y (ATTENDANCE)	xy	x^2
4	36.3	145.2	16
6	40.1	240.6	36
6	41.2	247.2	36
8	53.0	424.0	64
6	44.0	264.0	36
7	45.6	319.2	49
5	39.0	195.0	25
7	47.5	332.5	49
49	346.7	2167.7	311

Linear Regression Example

$$\bar{x} = \frac{49}{8} = 6.125$$

$$\bar{y} = \frac{346.9}{8} = 43.36$$

$$b = \frac{\sum xy - n\bar{x}\bar{y}}{\sum x^2 - n\bar{x}^2}$$

$$= \frac{(2{,}167.7) - (8)(6.125)(43.36)}{(311) - (8)(6.125)^2}$$

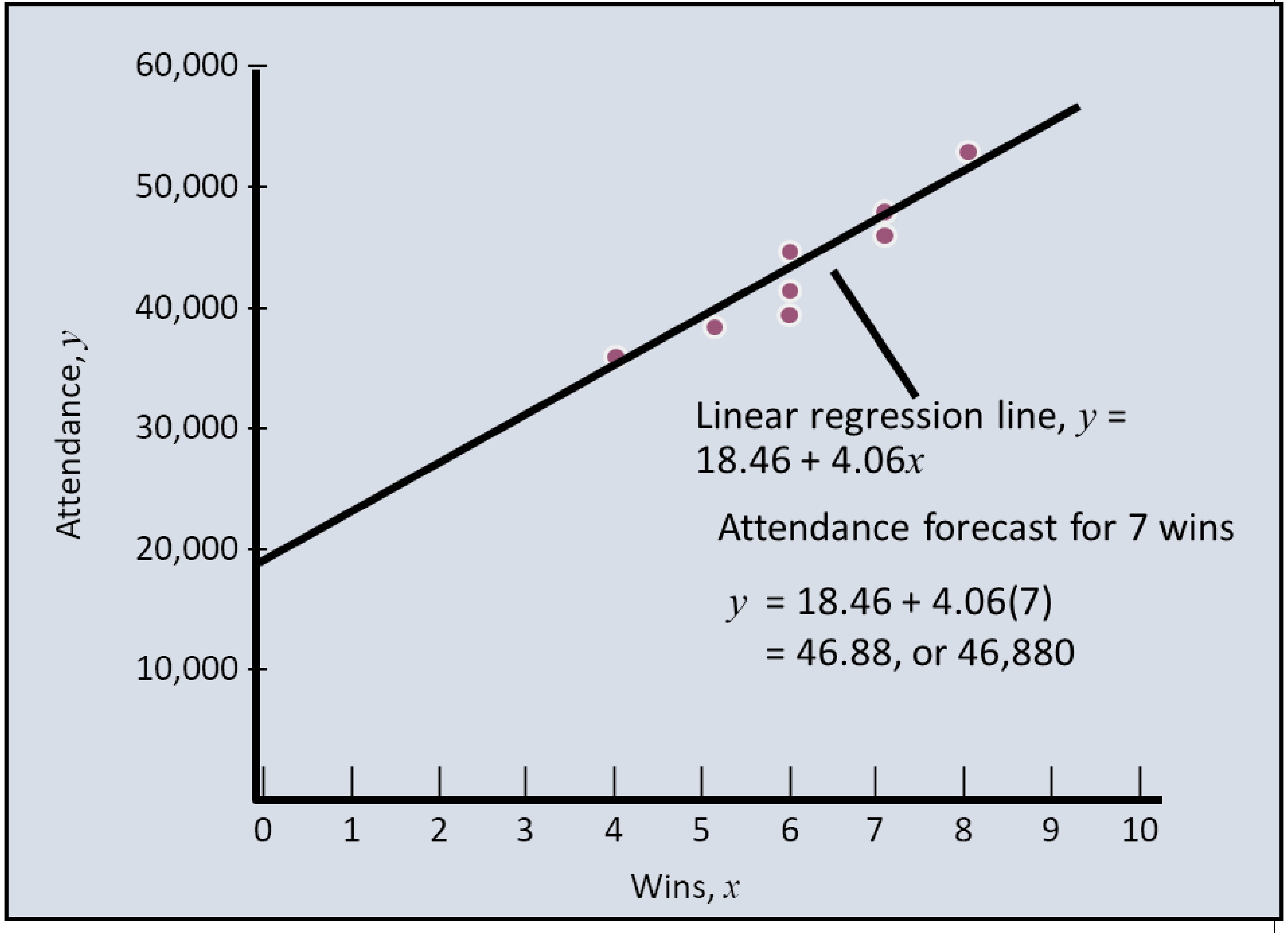
$$= 4.06$$

$$a = \bar{y} - b\bar{x}$$

$$= 43.36 - (4.06)(6.125)$$

$$= 18.46$$

Linear Regression Example



Correlation and Coefficient of Determination

- Correlation, r
  - Measure of strength of relationship
  - Varies between -1.00 and +1.00
- Coefficient of determination, r^2
  - Percentage of variation in dependent variable resulting from changes in the independent variable

ARIMA Models and Forecasting

- If we can describe the way the points in the series are related to each other (the autocorrelations), then we can describe the series using the relationships that we've found
- AutoRegressive Integrated Moving Average Models (ARIMA) are mathematical models of the autocorrelation in a time series
- One way to describe time series

ARIMA METHOD OF FORECASTING

- The Autoregressive (AR) Model
  - The following is an **AR(p)** model:

$$Y_t = B_0 + B_1 Y_{t-1} + B_2 Y_{t-2} +, \ldots, B_p Y_{t-p} + u_t$$

- where u_t is a white noise error term.

- The Moving Average (MA) Model
  - We can also model Y_t as an the **MA(q)** model, a weighted, or moving, average of the current and past white noise error terms:

- $$Y_t = C_0 + C_1 u_t + C_2 u_{t-1} +, \ldots, C_j u_{t-q}$$

ARIMA METHOD OF FORECASTING (CONT.)

- The Autoregressive Moving Average (ARMA) Model
  - The ARMA (p,q) model is a combination of AR (autoregressive) and MA (moving average) terms.

- The Autoregressive Integrated Moving Average (ARIMA) Model
  - The BJ methodology is based on the assumption that the underlying time series is stationary or can be made stationary by differencing it one or more times.
  - This is known as the ARIMA (p, d, q) model, where d denotes the number of times a time series has to be differenced to make it stationary.

Autocorrelation

- The major statistical tool for ARIMA models is the sample autocorrelation coefficient

$$r_k = \frac{\sum_{t=k+1}^{n} (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{n} (Y_t - \bar{Y})^2}$$

Autocorrelations

- r_1 indicates how successive values of Y relate to each other,
- r_2 indicates how Y values two periods apart relate to each other, and so on.

ACF

- Together, the autocorrelations at lags 1, 2, 3, etc. make up the autocorrelation function or ACF and then we plot the autocorrelations by the lags
- The ACF values reflect how strongly the series is related to its past values over time.

Autocorrelation Function Plot

- Autocorrelation refers to how correlated a time series is with its past values.

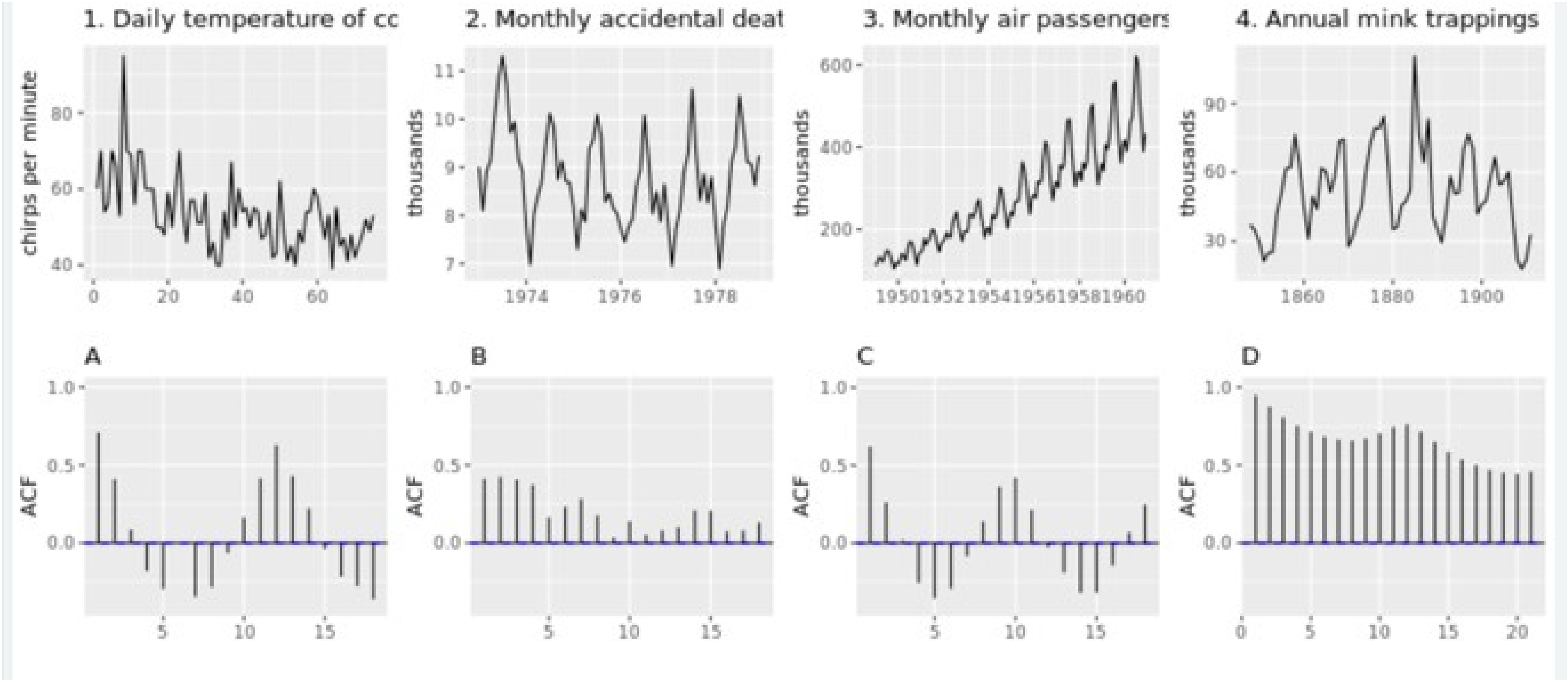
- When data have a trend, the autocorrelations for small lags tend to be large and positive.

- When data are seasonal, the autocorrelations will be larger at the seasonal lags

- When data are trended and seasonal, you see a combination of these effects

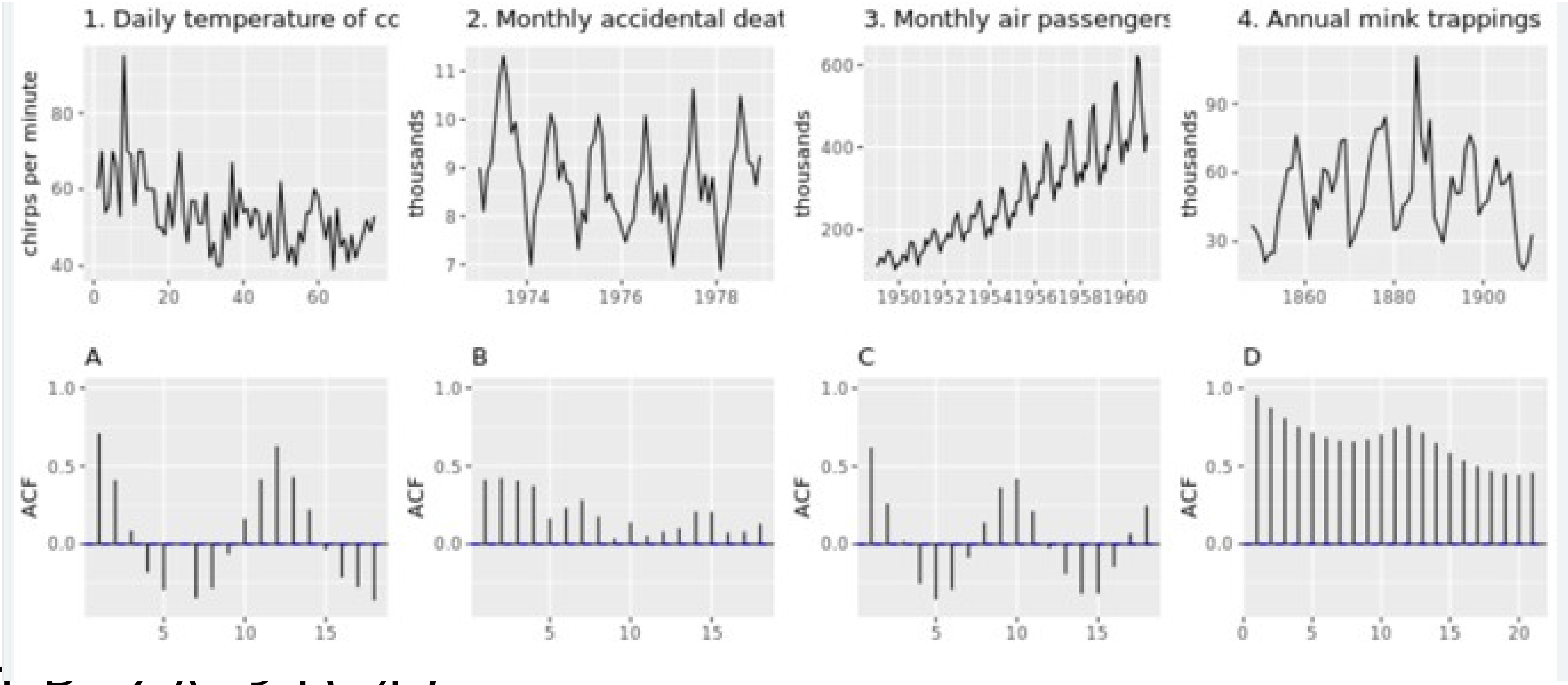
Quiz

> Match the ACF to the time series



Quiz

> Match the ACF to the time series



> 1-B, 2-A, 3-D, 4-C

Time Series Forecasting Process

Look at the data (Scatter Plot) → Forecast using one or more techniques → Evaluate the technique and pick the best one.

Observations from the scatter Plot	Techniques to try	Ways to evaluate
Data is reasonably **stationary** (no trend or seasonality)	**Heuristics - Averaging methods** • Naive • Moving Averages • Simple Exponential Smoothing	• MAD • MAPE • Standard Error • BIAS
Data shows a consistent **trend**	**Regression** • Linear • Non-linear Regressions (not covered in this course)	• MAD • MAPE • Standard Error • BIAS • R-Squared
Data shows both a **trend** and a **seasonal pattern**	**Classical decomposition** • Find Seasonal Index • Use regression analyses to find the trend component	• MAD • MAPE • Standard Error • BIAS • R-Squared

Evaluation of Forecasting Model

- **BIAS**
  - The arithmetic mean of the errors
  - n is the number of forecast errors
- **Mean Absolute Deviation - MAD**
  - Average of the absolute errors

Evaluation of Forecasting Model

- **Mean Square Error - MSE**

  -

- **Standard error**
  - Square Root of MSE
- **Mean Absolute Percentage Error - MAPE**
  - Calculate the % error using the absolute error, then average the results

Controlling the quality of forecast

- Necessary to monitor forecast to ensure that the forecast is performing adequately
- This is accomplished by comparing forecast errors to predetermined values
- Errors that fall within the limits are considered acceptable
- Errors outside either limit indicates that corrective action is needed
- Tracking signal values are compared to predetermined limits (+4,-4) based on judgment and experience

Choosing a Forecasting Technique

- Moving Averages and Exponential Smoothing are short range techniques. They produce forecast for the next period
- Trend equations are used for much longer time horizons.
- More than one forecasting techniques might be used to increase confidence
- The forecast horizon
- Forecasting frequency
  - Forecasting is not free
  - Consider cost and accuracy
    - Weigh cost-accuracy trade-offs carefully