BUS 173 Applied Statistics

Lecture 15-20

Prepared By: Mohammad Kamrul Arefin Lecturer, School of Business, North South University

Modeling for Time Series Forecasting

✓ Forecasting is a necessary input to planning, whether in business, or government. Often, forecasts are generated subjectively and at great cost by group discussion, even when relatively simple quantitative methods can perform just as well or, at very least; provide an informed input to such discussions.

✓ Statistical Forecasting: The selection and implementation of the proper forecast methodology has always been an important planning and control issue for most firms and agencies. Often, the financial well-being of the entire operation rely on the accuracy of the forecast since such information will likely be used to make interrelated budgetary and operative decisions in areas of personnel management, purchasing, marketing and advertising, capital financing, etc.

Modeling for Time Series Forecasting

✓ For example, any significant over-or-under sales forecast error may cause the firm to be overly burdened with excess inventory carrying costs or else create lost sales revenue through unanticipated item shortages. When demand is fairly stable, e.g., unchanging or else growing or declining at a known constant rate, making an accurate forecast is less difficult. If, on the other hand, the firm has historically experienced an up-and-down sales pattern, then the complexity of the forecasting task is compounded.

✓ There are two main approaches to forecasting. Either the estimate of future value is based on an analysis of factors which are believed to influence future values, i.e., the explanatory method, or else the prediction is based on an inferred study of past general data behaviour over time, i.e., the extrapolation method.

Modeling for Time Series Forecasting

✓ For example, the belief that the sale of doll clothing will increase from current levels because of a recent advertising blitz rather than proximity to Christmas illustrates the difference between the two philosophies. It is possible that both approaches will lead to the creation of accurate and useful forecasts, but it must be remembered that, even for a modest degree of desired accuracy, the former method is often more difficult to implement and validate than the latter approach.

Forecasting Performance Measures

✓ If the forecast error is stable, then the distribution of it is approximately normal. With this in mind, we can plot and then analyze forecast errors on the control charts to see if there might be a need to revise the forecasting method being used.

 \checkmark To do this, if we divide a normal distribution into zones, with each zone one standard deviation wide, then one obtains the approximate percentage we expect to find in each zone from a stable process.

✓ Control limits could be one-standard-error, or two-standard-error, and any point beyond these limits (i.e., outside of the error control limit) is an indication the need to revise the forecasting process.

✓ The plotted forecast errors on this chart, not only should remain with the control limits, they should not show any obvious pattern, collectively. We define the best forecast as the one which yields the forecast error with the minimum variance.

Forecasting Performance Measures and Control Chart for Examine Forecasting Errors



Forecasting Methods

✓ Regression Analysis

✓ Trend Analysis: Uses linear and nonlinear regression with time as the explanatory variable, it is used where pattern over time have a long-term trend.

✓ Modeling Seasonality and Trend: Seasonality is a pattern that repeats for each period. For example annual seasonal pattern has a cycle that is 12 periods long, if the periods are months, or 4 periods long if the periods are quarters. We need to get an estimate of the seasonal index for each month, or other periods, such as quarter, week, etc, depending on the data availability.

Forecasting Methods

✓ Modeling Seasonality and Trend:

✓ Seasonal Index: Seasonal index represents the extent of seasonal influence for a particular segment of the year. The calculation involves a comparison of the expected values of that period to the grand mean. A seasonal index is how much the average for that particular period tends to be above (or below) the grand average. A seasonal index of 1.00 for a particular month indicates that the expected value of that month is 1/12 of the overall average. The formula for computing seasonal factors is: S_i = D_i/D, where: S_i = the seasonal index for ith period,

 D_i = the average values of ith period,

D = grand average,

i = the ith seasonal period of the cycle.

Trend + Seasonal Forecasting Method

✓ Incorporating seasonality in a forecast is useful when the time series has both trend and seasonal components. The final step in the forecast is to use the seasonal index to adjust the trend projection. One simple way to forecast using a seasonal adjustment is to use a seasonal factor in combination with an appropriate underlying trend of total value of cycles.

✓ A Numerical Application: The table on next page provides monthly sales (\$1000) at a college bookstore. The sales show a seasonal pattern, with the greatest number when the college is in session and decrease during the summer months.

		Tr	end	+ Se	ason	al F	orec	astir	ng M	lethe	od		
M T	Jan	Feb	Mar	Apr	Мау	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
1	196	188	192	164	140	120	112	140	160	168	192	200	1972
2	200	188	192	164	140	122	132	144	176	168	196	194	2016
3	196	212	202	180	150	140	156	144	164	186	200	230	2160
4	242	240	196	220	200	192	176	184	204	228	250	260	2592
Mean:	208.6	207.0	192.6	182.0	157.6	143.6	144.0	153.0	177.6	187.6	209.6	221.0	2185
Index:	1.14	1.14	1.06	1.00	0.87	0.79	0.79	0.84	0.97	1.03	1.15	1.22	12

✓ Suppose we wish to calculate seasonal factors and a trend, then calculate the forecasted sales for July in year 5.

 \checkmark The first step in the seasonal forecast will be to compute monthly indices using the past four-year sales. For example, for January the index is:

S(Jan) = D(Jan)/D = 208.6/182.083 = 1.14

where D(Jan) is the mean of all four January months, and D is the grand mean of all past four-year sales.

Trend + Seasonal Forecasting Method

✓Next, a linear trend is calculated using the annual sales:

Y = 1684 + 200.4T,

✓ The main question is whether this equation represents the trend. If we construct a scatter-diagram, we may notice that a Parabola might be a better fit. Using the Polynomial Regression, the estimated quadratic trend is:

 $Y = 2169 - 284.6T + 97T^2$

✓ Predicted values using both the linear and the quadratic trends are presented in the table in the next page. Comparing the predicted values of the two models with the actual data indicates that the quadratic trend is a much superior fit than the linear one, as often expected.

Trend + Seasonal Forecasting Method

Determination of the Annual Trend for the Numerical Example

Year No:	Actual Sales	Linear Regression	Quadratic Regression
1	1972	1884	1981
2	2016	2085	1988
3	2160	2285	2188
4	2592	2486	2583

✓ We can now forecast the next annual sales; which, corresponds to year 5, or T = 5 in the above quadratic equation:

Y = 2169 - 284.6(5) + 97(5)2 = 3171 sales for the following year. The average monthly sales during next year is, therefore: 3171/12 = 264.25.

✓ Finally, the forecast for month of July is calculated by multiplying the average monthly sales forecast by the July seasonal index, which is 0.79; i.e., (264.25).(0.79) or 209.

Techniques for Averaging

- Moving average
- Weighted moving average
- Exponential smoothing

Simple Moving Average Formula

- The simple moving average model assumes an average is a good estimator of future behavior.
- The formula for the simple moving average is:

$$F_{t} = \frac{A_{t-1} + A_{t-2} + A_{t-3} + ... + A_{t-n}}{n}$$

$$F_{t} = \text{Forecast for the coming period}$$

$$N = \text{Number of periods to be averaged}$$

$$A_{t-1} = \text{Actual occurrence in the past period for}$$

$$up \text{ to "n" periods}$$

Simple Moving Average Problem (1) $F = A_{t-1} + A_{t-2} + A_{t-3} + \dots + A_{t-n}$

		Б —
Week	Demand	Γ_t –
1	650	
2	678	
3	720	
4	785	
5	859	
6	920	
7	850	
8	758	
9	892	
10	920	
11	789	
12	844	



- Question: What are the 3-week and 6-week moving average forecasts for demand?
- Assume you only have 3 weeks and 6 weeks of actual demand data for the respective forecasts



Plotting the moving averages and comparing them shows how the lines smooth out to reveal the overall upward trend in this example.



Simple Moving Average Problem (2) Data

Week	Demand
1	820
2	775
3	680
4	655
5	620
6	600
7	575

- Question: What is the 3 week moving average forecast for this data?
- Assume you only have 3 weeks and 5 weeks of actual demand data for the respective forecasts

Simple Moving Average

Problem (2) Solution

Week	Demand	3-Week	5-Week
1	820		
2	775		
3	680		
4	655	758.33	
5	620	703.33	
6	600	651.67	710.00
7	575	625.00	666.00



Weighted Moving Average Formula

While the moving average formula implies an equal weight being placed on each value that is being averaged, the weighted moving average permits an unequal weighting on prior time periods.

The formula for the moving average is:

$$F_{t} = W_{1}A_{t-1} + W_{2}A_{t-2} + W_{3}A_{t-3} + \dots + W_{n}A_{t-n}$$

w_t = weight given to time period "t" occurrence. (Weights must add to one.)

 $\sum_{i=1}^{n} w_i = 1$



Weighted Moving Average Problem (1) Solution

Week	Demand	Forecast
1	650	
2	678	
3	720	
4		693.4

$$F_4 = 0.5(720) + 0.3(678) + 0.2(650) = 693.4$$

Compute a three period moving average forecast given demand for shopping carts for the last five periods

Period	Age	Demand	
1	5	42	
2	4	40	
3	3	43	
4	2	40 } the 3 most recent dem	ands
5	1	41	
43 -	+ 40 + 41	- Augura -	
$F_6 =$	3	= 41.33	

If actual demand in period 6 turns out to be 39, the moving average forecast for period 7 would be





Period	Demand		
1	42		
. 2	40		
3	43		
4	40		
5 .	41		
E = 400	(41) + 30(40) + 20(4)	(3) + 10(40) = 41.0	
		5) 1 .10(40) = 41.0	
" Fy = .40((39) + .30(41) + .20(4)	(0) + .10(43) = 40.2	
Note that if t	four weights are used, o	only the four most recent demands are used to	pre
Has farmant			

Weighted Moving Average Problem (2) Data

Question: Given the weekly demand information and weights, what is the weighted moving average forecast of the 5th period or week?

Week	Demand
1	820
2	775
3	680
4	655

Weights: t-1 .7 t-2 .2 t-3 .1

Weighted Moving Average Problem (2) Solution

Week	Demand	Forecast
1	820	
2	775	
3	680	
4	655	
5		672



Exponential Smoothing $F_t = F_{t-1} + \alpha(A_{t-1} - F_{t-1})$

- Premise--The most recent observations might have the highest predictive value.
 - Therefore, we should give more weight to the more recent time periods when forecasting.

Example of Exponential Smoothing

Period	Actual	Alpha = 0.1	Error	Alpha = 0.4	Error
1	42				
2	40	42	-2.00	42	-2
3	43	41.8	1.20	41.2	1.8
4	40	41.92	-1.92	41.92	-1.92
5	41	41.73	-0.73	41.15	-0.15
6	39	41.66	-2.66	41.09	-2.09
7	46	41.39	4.61	40.25	5.75
8	44	41.85	2.15	42.55	1.45
9	45	42.07	2.93	43.13	1.87
10	38	42.36	-4.36	43.88	-5.88
11	40	41.92	-1.92	41.53	-1.53
12		41.73		40.92	



Exponential Smoothing Model

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

 α = smoothing constant

- Premise: The most recent observations might have the highest predictive value.
- Therefore, we should give more weight to the more recent time periods when forecasting.

Exponential Smoothing Problem (1) Data

Week	Demand
1	820
2	775
3	680
4	655
5	750
6	802
7	798
8	689
9	775
10	

- Question: Given the weekly demand data, what are the exponential smoothing forecasts for periods 2-10 using α=0.10 and α=0.60?
- Assume $F_1 = D_1$

Answei values.	r: The respect Note that yo	tive alphas col ou can only for	umns denote t ecast one time	the forecast period into the	Э			
future.	Week	Demand	0.1	0.6				
	1	820	820.00	820.00				
	2	775	820.00	820.00				
	3	680	815.50	820.00				
	4	655	801.95	817.30				
	5	750	787.26	808.09				
	6	802	783.53	795.59				
	7	798	785.38	788.35				
	8	689	786.64	786.57				
	9	775	776.88	786.61				
	10		776.69	780.77				
Mohammad Kamrul Arefin, Lecturer-SOB, NSU								



Exponential Smoothing Problem (2) Data

Week	Demand
1	820
2	775
3	680
4	655
5	

Question: What are the exponential smoothing forecasts for periods 2-5 using a =0.5?

Assume
$$F_1 = D_1$$



F	orecasts l	based of	averages.	Given	the f	following	data:

Period	Number of Complaints	
1	60	
2	65	
3	55	
4	58	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
5	64	

Prepare a forecast using each of these approaches:

- a. The appropriate naive approach.
- b. A three-period moving average.
- c. A weighted average using weights of .50 (most recent), .30, and .20.
- d. Exponential smoothing with a smoothing constant of .40.
- a. The values are stable. Therefore, the most recent value of the series becomes the next forecast: 64.

b.
$$MA_3 = \frac{55 + 58 + 64}{3} = 59$$

c.
$$F = .50(64) + .30(58) + .20(55) = 60.4$$

d.

	Number of		
Period	Complaints	Forecast	Calculations
1	60		The previous value of series is used
2	65	60	as the starting forecast.]
3	55	62	60 + .40(65 - 60) = 62
4	58	59.2	62 + .40(55 - 62) = 59.2
5	64	58.72	59.2 + .40(58 - 59.2) = 58.72
6		60.83	59.72 + .40(64 - 58.72) = 60.83

Associative Forecasting

- <u>Predictor variables</u> used to predict values of variable interest
- <u>Regression</u> technique for fitting a line to a set of points
- <u>Least squares line</u> minimizes sum of squared deviations around the line

Simple Linear Regression Model The simple linear regression

various data over time.

 $Y_t = a + bx$



0 1 2 3 4 5 x (Time)

Is the linear regression model.

Yt is the regressed forecast value or dependent variable in the model, **a** is the intercept value of the the regression line, and **b** is similar to the slope of the regression line. However, since it is calculated with the variability of the data in mind, its formulation is not as straight forward as our usual notion of slope.



Reasonable



4^c Regression analysis. The owner of a small hardware store has noted a sales pattern for window locks that seems to parallel the number of break-ins reported each week in the newspaper. The data are:

 Sales:	46	18	20	22	27	34	14	37	30
Break-ins:	9	3	3	5	4	7	2	6	4

- a. Plot the data to determine which type of equation, linear or nonlinear, is appropriate.
- b. Obtain a regression equation for the data.
- c. Estimate sales when the number of break-ins is five.



The graph supports a linear relationship. b. The computations for a straight line are:

×	Y	xy	x ²	y ²	Hence
9	46	414.	81	2,116	You
3	18	54	9	324	replace
3	20	60	9	400	
5	22	110	25	484	
4	27	108	16	729	c. For $x = \frac{1}{2}$
7	34	238	49	1,156	
2	14	28	4	196	1
6	37	222	36	1,369	
4	30	120	16	900	
43	248	1,354	245	7,674	
b =	$\frac{n(\Sigma xy) - (\Sigma xy)}{n(\Sigma x^2) - (\Sigma xy)}$	$\frac{x(\Sigma y)}{(\Sigma x)^2} = \frac{9}{9}$	(1,354) - 9(245) -	$\frac{-43(248)}{-43(43)} =$	4.275
a =	$\frac{\sum y - b(\sum x)}{n}$	$=\frac{248-4}{9}$.275(43)	= 7.129	

Hence, the equation is: $y_c = 7.129 + 4.275x$.

You can obtain the regression coefficients using the appropriate Excel template. Simply replace the existing data for x and y with your data. Note: be careful to enter the values for

For x = 5, $y_c = 7.129 + 4.275(5) = 28.50$.

Sides of 19-men color television sets and three-month tagged unemployment are shown in the following table. Determine if unemployment levels can be used to predict demand for 19-inch color TVs and, if so, derive a predictive equation.

Period	1	2	3	4	5	6	7	8	. 9.	10	11
Units sold	20	41	17	35	25	31	38	50	15	19	14
Unemployment %											
(three-month lag)	7.2	4.0	7.3	5.5	6.8	6.0	5.4	3.6	8.4	7.0	9.0

1. Plot the data to see if a linear model seems reasonable. In this case, a linear model seems appropriate for the range of the data.

Plot the data to see if a linear model seems reasonable. In this case, a linear model seems appropriate for the range of the data.



Comp	ite the co	orrelation coe	efficient to c	onfirm that it
×	Y	xy	x ²	y ²
7.2	20	144.0	51.8	400
40	41	164.0	16.0	1,681
7.3	17	124.1	53.3	289
5.5	35	192.5	30.3	1,225
6.8	25	170.0	46.2	625
6.0	31	186.0	36.0	961
5.4	38	205.2	29.2	1,444
3.6	50	180.0	13.0	2,500
8.4	15	126.0	70.6	225
7.0	19	133.0	49.0	361
9.0	14	126.0	81.0	196
70.2	305	1,750.8	476.4	9,907

$$r = \frac{11(1,750.8) - 70.2(305)}{\sqrt{11(476.4) - (70.2)^2} \cdot \sqrt{11(9.907) - (305)^2}} = -.966$$

This is a fairly high negative correlation.

3. Compute the regression line:

$$b = \frac{11(1.750.8) - 70.2(305)}{11(476.4) - 70.2(70.2)} = -6.91$$
$$a = \frac{305 - (-6.9145)(70.2)}{11} = 71.85$$

y = 71.85 - 6.89x

Note that the equation pertains only to unemployment levels in the range 3.6 to 9.0, because sample observations covered only that range. Healthy Hamburgers has a chain of 12 stores in northern Illinois. Sales figures and profits for the stores are given in the following table. Obtain a regression line for the data, and predict profit for a store assuming sales of \$10 million.

Sales, x	Profits, y
(in millions	s of dollars)
\$ 7	\$0.15
2	0.10
6	0.13
4	0.15
14	0.2:
15	0.27
16	0.24
12	0.20
14	0.27
20	0.44
15	0.34
7	0.17



4				
*	Y.	тy		37
7	0.15	1.05	49	0.0225
.2	0.10	0.20	4	0.0100
. 6	0.13	0.78	36	0.0169
4	0.15	0.60	16	0.0225
14	0.25	3.50	196	0.0625
1.5	0.27	4.05	225	0.0729
16	0.24	3.84	256	0.0576
12	0.20	2.40	144	0.0400
14	0.27	3.78	196	0.0729
20	0.44	8.80	400	0.1936
15	0.34	5.10	225	0.1156
7	0.17	1.19	49	0.0289
132	2.71	35.29	1.796	0.7159

Substituting into the equation, you find: .

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} = \frac{12(35.29) - 132(2.71)}{12(1,796) - 132(132)} = 0.01593$$
$$a = \frac{\sum y - b(\sum x)}{n} = \frac{2.71 - 0.01593(132)}{12} = 0.0506$$

Thus, the regression equation is: $y_c = 0.0506 + 0.01593x$. For sales of x = 10 (i.e., \$10 million), estimated profit is: $y_c = 0.0506 + 0.01593(10) = 0.2099$, or \$209,900. (It may







Linear Trend Problem Data

Question: Given the data below, what is the linear trend model that can be used to predict sales?

Week	Sales
1	150
2	157
3	162
4	166
5	177

Linear Trend Equation Example

t		У	
Week	t^2	Sales	ty
1	1	150	150
2	4	157	314
3	9	162	486
4	16	166	664
5	25	177	885
Σt = 15	$\Sigma t^2 = 55$	Σy = 812	Σ ty = 2499
$(\Sigma t)^2 = 225$			





Cell phone sales for a California-based firm over the last 10 weeks are shown in the for the sales for a California-based firm over the last 10 weeks are shown in the for the sales for the data, and visually check to see if a linear trend line would be appropriate. Then determine the equation of the trend line, and predict sales for weeks 11 and 12

Week	Unit Sales	
1	700	
2	724	
3	720	
4	728	
5	740	
6	742	
7	758	
2	750	
4	770	
10	775	

11)	y	ty
1	700	700
2	724	1,448
3	720	2,160
4	728	2,912
5	740	3,700
6	742	4,452
7	758	5,306
8	750	6,000
9	770	6,930
10	775	7,750
	7,407	41,358

From Table 3–1, for n = 10, $\Sigma t = 55$ and $\Sigma t^2 = 385$. Using Formulas 3–4 and 3–5, you can compute the coefficients of the trend line:

$$b = \frac{10(41.358) - 55(7.407)}{10(385) - 55(55)} = \frac{6.195}{825} = 7.51$$
$$a = \frac{7.407 - 7.51(55)}{10} = 699.40$$

Thus, the trend line is $y_t = 699.40 + 7.51t$, where t = 0 for period 0.

c. Substituting values of t into this equation, the forecasts for the next two periods (i.e., t = 11 and t = 12) are:

 $y_{11} = 699.40 + 7.51(11) = 782.01$

 $y_{12} = 699.40 + 7.51(12) = 789.52$



Linear trend line. Plot the data on a graph, and verify visually that a linear trend line is approprime. Develop a line trend equation for the following data. Then use the equation to predict the maxt two values of the series.



Period,	Demand,	***	
1	у	19	
 1	44	44	From Table $3-1$, with $n = 9$,
2	52	104	
3	50	150	$\Sigma t = 45$ and $\Sigma t^2 = 285$
4	54	216	
5	55	275	
6	55	330	
7	60	420	
8	56	448	
9	62	558	
2	488	2,545	

$$b = \frac{n\Sigma t y - \Sigma t \Sigma y}{n\Sigma t^2 - (\Sigma t)^2} = \frac{9(2.545) - 45(488)}{9(285) - 45(45)} = 1.75$$
$$a = \frac{\Sigma y - b\Sigma t}{n} = \frac{488 - 1.75(45)}{9} = 45.47$$

Thus, the trend equation is $y_t = 45.47 + 1.75t$. The next two forecasts are: $y_{10} = 45.47 + 1.75(10) = 62.97$

 $y_{11} = 45.47 + 1.75(11) = 64.72$

Trend Adjusted Exponential-Smoothing-Method Or Double Exponential Smoothing Method

✓ The trend-adjusted forecast (TAF) is composed of two elements: Exponential smoothed forecast and a trend factor.

 $TAF_{t+1} = S_{t+1} + T_{t+1},$

✓ Where, S_{t+1} = Exponential Smoothed Forecast

 T_{t+1} = Trend Factor

```
and S_{t+1} = TAF_t + \alpha(A_t - TAF_t)
```

$$T_{t+1} = T_t + \beta(TAF_t - TAF_{t-1} - T_t)$$

Vhere α and β are smoothing constants. In order to use this method, one must select values of α and β (usually through trial and error) and make a starting forecast and estimate of trend.

✓ Initial Trend Estimate = (728-700)/3 = 9.33

✓ Starting Forecast = 728 + 9.33 = 737.33

✓ Find trend-adjusted exp. smoothing to prepare forecast from period 5 using $\alpha = 0.4$ and $\beta = 0.3$

Wk	Unit Sales	St	Tt	TAFt	St+1	Tt+1
1	700					
2	724					
3	720					
4	728					
5	740	728	9.33	737.33	737.33+0.4(740- 737.33)= 738.40	9.33+0.3(0)=9.33
6	742	738.4	9.33	747.73	747.73+0.4(742- 747.73)= 745.44	9.33+0.3(747.73-737.33- 9.33)= 9.65
7	758	745.44	9.65	755.09	755.09+0.4(758- 755.09)= 756.25	9.65+0.3(755.09-747.73- 9.65)= 8.96
	Mohammad Kamrul Arefin, Lecturer-SOB, NSU 59					59