Tutorial letter 103/2/2018

Forecasting

STA2604

Semester 2

Department of Statistics

TRIAL EXAMINATION PAPER

1. Trial Examination Papar
2. Trial Examination Paper Solutions

1 Trial Examination Paper

STA2604
FORECASTING II
TRIAL EXAM PAPER: SEMESTER 2
October 2018

Read the following instructions carefully:

1) Attempt all questions.
2) Show all relevant computations and steps.
3) You may use a non-programmable calculator.

QUESTION 1 **[17]**

(1.1) Consider the following data: $15 \quad 59 \quad 34 \quad 80 \quad 100 \quad 129 \quad 189 \quad 167$.
An observer stated that there is no evidence that these data can be classified as time series data. What is the reason of his/her statement? (2)

(1.2) Consider the monthly mean temperatures (in degrees Celsius) for a given city during the twelve months of a certain year: $25 \quad 26 \quad 24 \quad 20 \quad 10 \quad 14 \quad 15 \quad 18 \quad 19 \quad 22 \quad 24 \quad 21$.
Why these data can be classified as time series data? (2)

(1.3) A market researcher stated that the price of food has been depending on the time of the year during the last 15 years. Give the time series concept being referred to in such a statement. (2)

(1.4) Give the main difference between (a) seasonal variations and (b) cyclical variations in time series. (6)

(1.5) Define irregular variations in time series and give two examples. (5)

QUESTION 2 **[20]**

The following table presents actual and predicted mean number of visitors at a certain place during the last eight months.

Month	Actual number of visitors, y_t	Predicted number of visitors, $\widehat{y}_t$
Jul	160	170
Aug	125	135
Sep	125	130
Oct	130	125
Nov	140	150
Dec	130	125
Jan	110	120
Feb	106	100

Calculate:

(2.1) the forecast error for each month (3.5)

(2.2) the MAD (5.5)

(2.3) the MSE (5.5)

(2.4) the MAPE (5.5)

QUESTION 3 **[39]**

The data in the following table give quarterly sales of mountain bike for two consecutive years by a bicycle shop in Switzerland (in $\$100$). Assume a multiplicative decomposition time series model.

Year	Quarter	bike Sales
1	1	10
	2	31
	3	43
	4	16
2	1	11
	2	33
	3	45
	4	17

(3.1) Compute the appropriate four-point moving averages (4-MA) for the data. (2)

(3.2) Compute the centered moving averages (CMA) for the data. (2)

(3.3) Compute the adjusted seasonal indices sn_t (not in percentages) for these data. (11)

(3.4) Compute the deseasonalized observations $d_t = y_t/sn_t$ for these data. Use two-decimal places for the final results. (4)

(3.5) Assume that a linear trend $TR_t = \beta_0 + \beta_1 t$ describes the deseasonalized observations where $t = 1$ for quarter 1 of year 1, ..., 8 for quarter 4 of year 2. Calculate the least squares estimates of β_0 and β_1. (8)

(3.6) Use the results in part (3.5) to compute the estimated trend values tr_t for these data. (4)

(3.7) Compute the estimated values (point forecasts) $\widehat{y}_t = tr_t \times sn_t$. Use two-decimal places for the final results. (4)

(3.8) Compute the point sales forecasts for the four quarters of Year 3. Use two-decimal places for the final results. (4)

QUESTION 4 **[24]**

Consider the bike sales data presented in Question 3.

(4.1) Using the results in part (3.7), compute the eight residuals. (8)

(4.2) Explain why we cannot use the Durbin-Watson test to determine whether or not there are first-order positive or negative autocorrelations for these data. (2)

(4.3) Using the simple exponential smoothing with $\alpha = 0.2$, determine the smoothed levels for the first three observations. (6)

(4.4) Now assume that the Holt's trend corrected exponential smoothing, with $\alpha = 0.2$ and $\gamma = 0.1$, is more appropriate than the simple exponential smoothing.

(a) Give the initial estimates of the level l_0 and the growth rate b_0. (2)

(b) Calculate the value of l_1 for $y = 10$ (observation for the first quarter of Year 1). (2)

(c) Estimate the growth rate for $y = 10$. (2)

(4.5) Based on the results for l_1 in parts (4.3) and (4.4), would you recommend the Holt's trend corrected exponential smoothing or the simple exponential smoothing? Briefly explain. (2)

Formulae

(1) $e_t = y_t - \widehat{y_t}$

(2) $MAD = \dfrac{\sum_{t=1}^{n} |e_t|}{n}$

(3) $MSE = \dfrac{\sum_{t=1}^{n} (e_t)^2}{n}$

(4) $APE = \dfrac{100|e_t|}{y_t}$

(5) $MAPE = \dfrac{\sum_{t=1}^{n} APE_t}{n}$

(6) $\widehat{\beta}_1 = \dfrac{n\sum_{t=1}^{n} ty_t - \sum_{t=1}^{n} t \sum_{t=1}^{n} y_t}{n\sum_{t=1}^{n} t^2 - (\sum_{t=1}^{n} t)^2}$

(7) $\widehat{\beta}_0 = \dfrac{\sum_{t=1}^{n} y_t - \widehat{\beta}_1 \sum_{t=1}^{n} t}{n}$

(8) $l_T = \alpha y_T + (1-\alpha) l_{T-1}$

(9) $l_T = \alpha y_T + (1-\alpha)[l_{T-1} + b_{T-1}]$

(10) $b_T = \gamma (l_T - l_{T-1}) + (1-\gamma) b_{T-1}$

2 Trial Examination Solutions

QUESTION 1

(1.1) The chronological pattern of the times at which the data were collected is not given.

(1.2) The mean temperatures were collected monthly in a certain year. Therefore, the chronological pattern of the times at which the data were collected is given.

(1.3) The time series concept referred to in this case is the seasonal variation.

(1.4) The main difference between seasonal and cyclical variations is that seasonal variations are regular, generally during each year, while cyclical variations may appear after a long period, generally after more than one year, and are not necessarily regular.

(1.5) Irregular variations are fluctuations that do not follow a recognisable or a regular pattern, and thus are unpredictable. Two examples are: floods, droughts.

QUESTION 2

Let y_t and $\widehat{y_t}$ denote the actual and predicted number of visitors, respectively, in the table below:

Month	y_t	$\widehat{y_t}$	e_t	$\lvert e_t\rvert$	e_t^2	$APE_t = \frac{\lvert e_t\rvert}{y_t} \times 100$
Jul	160	170	-10	10	100	6.25
Aug	125	135	-10	10	100	8
Sep	125	130	-5	5	25	4
Oct	130	125	5	5	25	3.8462
Nov	140	150	-10	10	100	7.1429
Dec	130	125	5	5	25	3.8462
Jan	110	120	-10	10	100	9.0909
Feb	106	100	6	6	36	5.6604
Tot				51	511	47.8366

(2.1) The forecast error e_t for each month is given in the fourth column of the table.

(2.2) MAD $= \dfrac{\sum_{t=1}^{8} |e_t|}{8} = \dfrac{51}{8} = 6.375.$

(2.3) MSE $= \dfrac{\sum_{t=1}^{8} e_t^2}{8} = \dfrac{511}{8} = 63.875.$

(2.4) MAPE $= \dfrac{\sum_{t=1}^{8} APE_t}{8} = \dfrac{47.8366}{8} = 5.9796.$

QUESTION 3

(3.1) The four-point moving averages are in given in the fifth column of Table 1.

Table 1: Analysis of bike sales data using the multiplicative decomposition of a time series.

Year	Quarter	t	y_t	4MA	CMA_t	$sn_t \times ir_t$	sn_t	d_t	tr_t	$\widehat{y}_t = tr_t \times sn_t$
1	1	1	10				0.42	23.81	24.24	10.18
1	2	2	31				1.25	24.80	24.60	30.75
1	3	3	43	25	25.125	1.7114	1.71	25.15	24.96	42.68
1	4	4	16	25.25	25.5	0.6275	0.63	25.40	25.32	15.95
2	1	5	11	25.75	26	0.4231	0.42	26.19	25.68	10.79
2	2	6	33	26.25	26.375	1.2512	1.25	26.40	26.04	32.55
2	3	7	45	26.5			1.71	26.32	26.40	45.14
2	4	8	17				0.63	26.98	26.76	16.86

(3.2) The centered moving averages, $CMA_t = tr_t \times cl_t$, are in given in the sixth column of Table 1.

(3.3) First calculate the $sn_t \times ir_t = y_t/CMA_t$ values. These are reported in the seventh column of Table 1. Steps for the calculation of seasonal indices are given in the following tables:

Year	Quarter			
	1	2	3	4
1	-	-	1.7114	0.6275
2	0.4231	1.2512	-	-
Total	0.4231	1.2512	1.7114	0.6275
Mean (unadjusted seasonal indices)	0.4231	1.2512	1.7114	0.6275

The total index is: $0.4231 + 1.2512 + 1.7114 + 0.6275 = 4.0132$.

Adjusted seasonal indices (must sum to 4, the number of seasons; quarters here):

Quarter	Value
1	$\frac{0.4231 \times 4}{4.0132} = 0.4217$
2	$\frac{1.2512 \times 4}{4.0132} = 1.2471$
3	$\frac{1.7114 \times 4}{4.0132} = 1.7058$
4	$\frac{0.6275 \times 4}{4.0132} = 0.6254$

These adjusted seasonal indices, rounded to the nearest two-decimal place numbers, are copied in the eighth column of Table 1.

(3.4) The deseasonalised observations $d_t = y_t/sn_t$ are given in the ninth column of Table 1.

(3.5) The following table has necessary information for estimating β_1 and β_0.

t	y_t	td_t	t^2
1	24	24	1
2	25	50	4
3	25	75	9
4	25	100	16
5	26	130	25
6	26	156	36
7	26	182	49
8	27	216	64
36	204	933	204

$$\widehat{\beta}_1 = \frac{8\sum_{t=1}^{8} td_t - \sum_{t=1}^{8} t \sum_{t=1}^{8} d_t}{8\sum_{t=1}^{8} t^2 - \left(\sum_{t=1}^{8} t\right)^2} = \frac{8 \times 933 - 36 \times 204}{8 \times 204 - (36)^2} = \frac{120}{336} = 0.36$$

and

$$\widehat{\beta}_0 = \bar{d} - \widehat{\beta}_1 \bar{t} = \frac{\sum_{t=1}^{8} d_t - \widehat{\beta}_1 \sum_{t=1}^{8} t}{8} = \sum_{t=1}^{8} t8 = \frac{204 - 0.36 \times 36}{8} = \frac{191.04}{8} = 23.88.$$

Thus the fitted model for the trend is: $tr_t = 23.88 + 0.36t$.

(3.6) The trend values tr_t, obtained using $tr_t = 23.88 + 0.36t$, are given in the tenth column of Table 1.

(3.7) The point estimates $\widehat{y}_t = tr_t \times sn_t$ are given in the eleventh column of Table 1.

(3.8) The four quarters of year 4 correspond to $t = 9$, $t = 10$, $t = 11$ and $t = 12$, respectively.
We use $\widehat{y}_t = (23.88 + 0.36t) \times sn_t$.
For $t = 9$, we have $\widehat{y}_9 = (23.88 + 0.36 \times 9) \times 0.42 = 11.39$.
For $t = 10$, we have $\widehat{y}_{10} = (23.88 + 0.36 \times 10) \times 1.25 = 34.35$.
For $t = 11$, we have $\widehat{y}_{11} = (23.88 + 0.36 \times 11) \times 1.71 = 47.61$.
For $t = 12$, we have $\widehat{y}_{12} = (23.88 + 0.36 \times 12) \times 0.63 = 17.77$.

QUESTION 4

(4.1) The eight residuals are given in the third column of the following table:

y_t	$\widehat{y}_t$	$e_t = y_t - \widehat{y}_t$
10	10.18	-0.18
31	30.75	0.25
43	42.68	0.32
16	15.95	0.05
11	10.79	0.21
33	32.55	0.45
45	45.14	-0.14
17	16.86	0.14

(4.2) We cannot use the Durbin-Watson test since the number of residuals is less than 15.

(4.3) $l_0 = \frac{1}{8}\sum_{t=1}^{8} y_t = \frac{206}{8} = 25.75$.
We use $l_T = \alpha t_T + (1-\alpha)l_{T-1}$ with $\alpha = 0.2$ and $l_0 = 25.75$.
We obtain:
$l_1 = 0.2 \times 10 + 0.8 \times 25.75 = 22.6$.
$l_2 = 0.2 \times 31 + 0.8 \times 22.26 = 24.01$.
$l_3 = 0.2 \times 43 + 0.8 \times 24.01 = 27.81$.

(4.4) Use of the Holt's trend corrected exponential smoothing with $\alpha = 0.2$ and $\gamma = 0.1$. The Holt's trend corrected exponential smoothing model assumes a trend in the data, but no seasonality. Therefore, we have to fit the model

$$y_t = \beta_0 + \beta_1 t + \epsilon_t$$

to the observed (not deseasonalised) data. The following table gives the required values for estimating β_1 and β_0.

t	y_t	ty_t	t^2
1	10	10	1
2	31	62	4
3	43	129	9
4	16	64	16
5	11	55	25
6	33	198	36
7	45	315	49
8	17	136	64
36	206	969	204

The values of $\widehat{\beta}_1$ and $\widehat{\beta}_0$ are the following:

$$\widehat{\beta}_1 = \frac{8\sum_{t=1}^{8} ty_t - \sum_{t=1}^{8} t \sum_{t=1}^{8} y_t}{8\sum_{t=1}^{8} t^2 - \left(\sum_{t=1}^{8} t\right)^2} = \frac{8 \times 969 - 36 \times 206}{8 \times 204 - (36)^2} = \frac{336}{336} = 1.$$

$$\widehat{\beta}_0 = \bar{y}_t - \widehat{\beta}_1 \bar{t} = \frac{\sum_{t=1}^{8} y_t - \widehat{\beta}_1 \sum_{t=1}^{8} t}{8} = \frac{206 - 1 \times 36}{8} = \frac{170}{8} = 21.25.$$

Thus, the fitted model is: $\widehat{y_t} = 21.25 + t$.

(a) $l_0 = \widehat{\beta}_0 = 21.25$ and $b_0 = \widehat{\beta}_1 = 1$.

(b) We use $l_T = \alpha y_T + (1-\alpha)[l_{T-1} + b_{T-1}]$ where $T = 1$, $l_0 = 21.25$ and $\alpha = 0.2$.
We obtain:

$$\begin{aligned} l_1 &= \alpha y_1 + (1-\alpha)[l_0 + b_0] \\ &= 0.2 \times 10 + 0.8 \times (21.25 + 1) \\ &= 19.8 \end{aligned}$$

(c) We use $b_T = \gamma[l_T - l_{T-1} + (1-\gamma)b_{T-1}]$ where $T = 1$, $\gamma = 0.1$ and $b_0 = 1$.
We obtain:

$$\begin{aligned} b_1 &= \gamma[l_1 - l_0 + (1-\gamma)b_0] \\ &= 0.1(19.8 - 21.25) + 0.9 \times 1 \\ &= 0.76. \end{aligned}$$

(4.5) The forecast error using the simple exponential smoothing is: $10 - 22.6 = -12.6$ and the forecast error using the Holt's trend corrected exponential smoothing is $10 - 19.8 = -9.8$. Therefore, based on this single observation, the Holt's trend corrected exponential smoothing should be recommended since it gives the smallest forecast error.