Tutorial set #1 Solution

Question 1:

1- Define a time series? mention some examples of time series data.

A time series is a collection of observations of some phenomenon collected sequentially over a period of time. This means that data have chronological order. Some examples are, monthly rain amounts, value of quarterly foreign remittances, daily car accidents, …

2- Mention four goals of time series analysis.

There are several goals for the analysis of time series, some of which are:

a- Description

which indicate describing and portraying the available information that shows how the studied phenomenon evolve over time. That is, describe the main features of the time series, which will help in determining the best mathematical model that can be appropriate to achieve the other goals of the analysis, and get to know the upward and downward movements in the time series and to identify the major components such as trend and seasonal changes.

b- Interpretation

Interpretation means explaining the changes occurring in the phenomenon using other time series that are related to it, or by using environmental factors affecting the phenomenon, for example, one can study how the sea level is affected by temperature, or how sales are affected by advertising.

c- Control

In production lines (in the factories), one may get time series that designate the product quality in the manufacturing process, and the goal here might be to control product quality so that it does not go below a specified level.

d- Forecasting

Forecasting is considered one of the most important goals of time series analysis. As one might want to know or expect the future values of a time series.

Analysis of time series usually starts by identifying an appropriate model that explains the evolution pattern of the series, and then uses the model to extrapolate this pattern into the future.

3- Explain briefly the components of a time series.

Time series data have four components:

a- trend component

If there exist a long term increase (or decrease) in the level of the series, then we say there exist a trend component in the series, for example, in time series of the number of births, or the number of pilgrims, or prices of goods annually.

b- seasonal component

A seasonal pattern exists when a series is influenced by seasonal factors (e.g., the quarter of the year, the month, or day of the week). Seasonality is always of a fixed and known period. for example, the electric power consumption reaches its peak in summer and fall in winter. Seasonal changes occur at periods less than a year, such as hour, day, week, month, quarter, etc.

c- cyclical component

These changes are similar to seasonal variation, but they appear in long periods of time (more than one year), and to discover the cyclical variation one need a very long annual series, for example, climate changes needs data of fifty years or more to discover its cycle. Also, economic cycles need a long periods of time, for example five or ten Years, to appear.

d- Random component

After getting rid of seasonal, trend, or cyclical components from the data, we are left with a residual series, which represent the irregular changes. These changes differ from the other components, as they can’t be predicted, and they do not occur according to any law or system.

4- Explain the difference between time series models and causal models.

a- causal models:

This approach is based on identifying the variable(s) that may have a causal relationship with the variable under study that we want to predict, this variable is called the dependent variable, then determine the appropriate statistical model or appropriate functional relationship which explains how the dependent variable is associated to the independent or explanatory variables. Using this model, we can predict the dependent variable under study. The main disadvantages of this approach are:

1-Difficulty of identifying all the explanatory variables that are related to the dependent variable.

2-Requires the availability of detailed historical information about all the explanatory variables, and the ability of knowing these variables or predicting them.

b- time series models:

This approach relies on analyzing historical data of the variable under study in order to determine the pattern it follows. Assuming that this pattern will continue in the future, we use it to predict future values of the variable.

The approach assumes that there exists a stochastic process able to produce the time series under study, and by imposing the stationarity condition, we then will be able to use the mathematical model that describes the correlation structure of the data to obtain reliable forecasts.

5- When measuring forecast accuracy (or error size), explain why don’t we just use the sum of the errors.

Because by definition, the errors can be positive and negative, and using the sum will not provide us with an accurate measure of their size, thus we resort to squaring or taking the absolute value of the errors to get a good estimate of their size.

6- Is the best forecasting technique always the most accurate one? explain.

The best forecasting method is not necessarily the method that achieves the highest accuracy or the smallest forecasting errors, but one method may be used because of type of the required forecast (point or confidence interval forecast), another because of only small number of observations are available, a third because it has a low cost, and a fourth because its theoretical assumptions comply with the data set in hand.

7- Give examples from the real life of a time series that have:
   a- seasonal component of length a year.
      for example, yearly amount of rain, or average riyal exchange rate against dollar for a period of 50 years.
   b- seasonal component of length a month.
      number of monthly car accidents for a period of 50 years.
   c- seasonal component of length a week.
      number of patients visiting a clinic weekly for a period of a year.

Question 2:

The following table monthly sales (in thousands of riyals) of some item, and estimated values of the sales calculated from some fitted model:

Year	1995	1996	1997	1998	1999	Total
Sales y_t	240	251	265	250	260	
estimated sales $\hat{y}_t$	235	258	260	260	255	
estimated errors $\varepsilon_t = \hat{y}_t - y_t$	-5	7	-5	10	-5	2
absolute errors $\lvert\varepsilon_t\rvert$	5	7	5	10	5	32
Absolute relative errors $\left\lvert\frac{y_t - \hat{y}_t}{y_t}\right\rvert$	0.020833	0.027888	0.018868	0.04	0.019231	0.12682
Squared errors $(\varepsilon_t)^2$	25	49	25	100	25	224

a- calculate the estimated errors.
   The estimated errors are:

$$\varepsilon_t = \hat{y}_t - y_t, \quad t = 1,2, \dots, n$$

$$\sum \varepsilon_t = 2$$

b- calculate mean squared errors MSD, the mean absolute deviances MAD, and mean absolute percentage errors MAPE.

$$\mathrm{MSD} = \frac{1}{k}\sum_{i=1}^{k}(y_t - \hat{y}_t)^2 = \frac{224}{5} = 44.8$$

$$MAD = \frac{1}{k}\sum_{i=1}^{k}|\varepsilon_t| = \frac{32}{5} = 6.4$$

$$MAPE = \frac{100}{k}\sum_{i=1}^{k}\left|\frac{y_t - \hat{y}_t}{y_t}\right| = \frac{100}{5}\,(0.12682) = 2.5364$$

Question 3:

The following table shows the loans financed by a bank (in millions of dollars) in the period 1995 to 2001:

Year	1995	1996	1997	1998	1999	2000	2001	2002
loans financed	12	13	11	13	12	14	11	
K=3	-	-	-	12	12.33	12	13	12.33
K=4	-	-	-	-	12.25	12.25	12.5	12.5

a- Use the method of simple moving averages method to find all the possible forecasts using k=3, and k=4. Also find the mean absolute deviances in each case.

simple moving average of order k=3:

$$ma_1(3) = \frac{y_3 + y_2 + y_1}{3} = \frac{12+13+11}{3} = 12$$

$$ma_2(3) = \frac{y_4 + y_3 + y_2}{3} = \frac{13+11+13}{3} = 12.33$$

$$ma_3(3) = \frac{y_5 + y_4 + y_3}{3} = \frac{12+13+11}{3} = 12$$

$$ma_4(3) = \frac{y_6 + y_5 + y_4}{3} = \frac{14+12+13}{3} = 13$$

$$ma_5(3) = \frac{y_7 + y_6 + y_5}{3} = \frac{11 + 14 + 12}{3} = 12.33$$

$$MAD = \frac{1}{k}\sum_{i=1}^{k}|\varepsilon_t| = \frac{1}{k}\sum_{i=1}^{k}|\hat{y}_t - y_t| = \frac{1}{4}(1 + 0.33 + 2 + 2) = 1.3325$$

simple moving average of order k=4:

$$ma_1(4) = \frac{y_4 + y_3 + y_2 + y_1}{4} = \frac{13 + 11 + 13 + 12}{4} = 12.25$$

$$ma_2(4) = \frac{y_5 + y_4 + y_3 + y_2}{4} = \frac{12 + 13 + 11 + 13}{4} = 12.25$$

$$ma_3(4) = \frac{y_6 + y_5 + y_4 + y_3}{4} = \frac{14 + 12 + 13 + 11}{4} = 12.5$$

$$ma_4(4) = \frac{y_7 + y_6 + y_5 + y_4 + y_3}{4} = \frac{11 + 14 + 12 + 13}{4} = 12.5$$

$$MAD = \frac{1}{k}\sum_{i=1}^{k}|\varepsilon_t| = \frac{1}{3}(0.25 + 1.75 + 1.5) = 1.1667$$

b- Forecast the amount of loans that the bank will finance in the year 2002 using the simple moving average method.
The forecast for the sales of 2002 using a simple moving average forecast of order k=3 is 12.33 .

c- Estimate the initial value $\hat{y}_0(1)$ using the mean of the series, then use the single exponential smoothing method to find all the forecasts, use $\alpha = 0.75$, and then $\alpha = 0.95$. Which one gives better forecasts? explain.

$$S_t = \alpha y_t + (1-\alpha)S_{t-1} \qquad , t = 1, \dots, n\ ; \quad S_0 = \bar{y}\ ; \quad 0 < \alpha < 1$$

$$\hat{y}_0(1) = S_0 = \bar{y} = 86/7 = 12.2857$$

now, simple exponential smoothing using $\alpha = 0.75$:

$$\hat{y}_t(1) = S_1 = 0.75\, y_1 + (1 - .75)\, S_0$$

$$= (0.75)(12) + (1 - .75)(12.2857) = 12.0714$$

$$\hat{y}_t(2) = S_2 = 0.75\, y_2 + (1 - .75)\, S_{t-1}$$

$$= (0.75)(13) + (1 - .75)(12.0714) = 12.7679$$

$$\hat{y}_t(3) = S_3 = 0.75\, y_3 + (1-.75)S_{t-1}$$

$$= (0.75)(11) + (1-.75)(12.7679) = 11.442$$

$$\hat{y}_t(4) = S_4 = 0.75\, y_4 + (1-.75)S_{t-1}$$

$$= (0.75)(13) + (1-.75)(11.442) = 12.6105$$

$$\hat{y}_t(5) = S_5 = 0.75\, y_5 + (1-.75)S_{t-1}$$

$$= (0.75)(12) + (1-.75)(12.6105) = 12.1526$$

$$\hat{y}_t(6) = S_6 = 0.75\, y_6 + (1-.75)S_{t-1}$$

$$= (0.75)(14) + (1-.75)(12.1526) = 13.5382$$

$$\hat{y}_t(7) = S_7 = 0.75\, y_7 + (1-.75)S_{t-1}$$

$$= (0.75)(11) + (1-.75)(13.5382) = 11.6346$$

In the same manner, for SES using $\alpha = 0.95$, we get:

$$\hat{y}_t(1) = 0.95y_1 + (1-.95)\hat{y}_0(1)$$

$$= (0.95)(12) + (1-.95)(12.2857) = 12.0143$$

$$\hat{y}_t(2) = 0.95y_2 + (1-.95)S_{t-1}$$

$$= (0.95)(13) + (1-.95)(12.0143) = 12.9507$$

$$\hat{y}_t(3) = 0.95y_3 + (1-.95)S_{t-1}$$

$$= (0.95)(11) + (1-.95)(12.9507) = 11.0975$$

$$\hat{y}_t(4) = 0.95y_4 + (1-.95)S_{t-1}$$

$$= (0.95)(13) + (1-.95)(11.0975) = 12.9049$$

$$\hat{y}_t(5) = 0.95y_5 + (1-.95)S_{t-1}$$

$$= (0.95)(12) + (1-.95)(12.9049) = 12.0452$$

$$\hat{y}_t(6) = 0.95y_6 + (1 - .95)S_{t-1}$$

$$= (0.95)(14) + (1 - .95)(12.0452) = 13.9023$$

$$\hat{y}_t(7) = 0.95y_7 + (1 - .95)S_{t-1}$$

$$= (0.95)(11) + (1 - .95)(13.9023) = 11.1451$$

To see which one gives better forecasts, we can check the sum of squared errors:

Actual value y_t	Forecast ($\hat{y}_t$)		squared errors $(\varepsilon_t)^2 = (y_t - \hat{y}_t)^2$	
	$\alpha = 0.75$	$\alpha = 0.95$	$\alpha = 0.75$	$\alpha = 0.95$
12	12.0714	12.0143	0.005098	0.0002045
13	12.7679	12.9507	0.053870	0.0024305
11	11.442	11.0975	0.195364	0.0095063
13	12.6105	12.9049	0.151710	0.0090440
12	12.1526	12.0452	0.023287	0.0020430
14	13.5382	13.9023	0.213259	0.0095453
11	11.6346	11.1451	0.402717	0.0210540
Total			1.04531	0.0538276

So we see that the SES with smoothing parameter $\alpha = 0.95$ gives forecasts the smallest squared errors, thus, we use it for forecasting the value of the loans the bank will finance for the year 2002.