EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY

HIGHER CERTIFICATE IN STATISTICS, 2015

MODULE 7 : Time series and index numbers

Time allowed: One and a half hours

Candidates should answer **THREE** *questions.*

Each question carries 20 marks.
The number of marks allotted for each part-question is shown in brackets.

Graph paper and Official tables are provided.

Candidates may use calculators in accordance with the regulations published in the Society's "Guide to Examinations" (document Ex1).

The notation log *denotes logarithm to base* ***e***.
Logarithms to any other base are explicitly identified, e.g. $\log_{10}$.

Note also that $\binom{n}{r}$ *is the same as* ${}^nC_r$.

This examination paper consists of 8 printed pages.
This front cover is page 1.
Question 1 starts on page 2.

There are 4 questions altogether in the paper.

1. (i) For a time series y_t write down

(a) a formula for smoothing the series using a symmetric $(2k+1)$-point moving average with weights a_r $(r = -k, \ldots, k)$.

(2)

(b) the equation, in simple exponential smoothing with smoothing parameter α, for calculating the smoothed estimate m_t of the series in terms of y_t and m_{t-1}.

(2)

(ii) The table below shows the half-yearly sales figures y_t of a certain product for the years 2011 to 2014.

Date		*Sales (£10K)*	*Smoothed A*	*Smoothed B*
2011	June	51		
	December	25		
2012	June	49		
	December	31		
2013	June	53		
	December	29		
2014	June	55		
	December	25		

Copy the table and complete the extra columns, giving (correct to 2 decimal places)

- Smoothed A: the 3-point simple symmetric moving average of y_t
- Smoothed B: the exponentially weighted moving average of y_t with smoothing parameter 0.3 and initial smoothed value set equal to the actual value for June 2011.

(6)

Describe the strengths and weaknesses of each of these methods from the point of view of

(a) forecasting,

(b) smoothing seasonal time series.

(6)

(iii) Explain briefly how moving averages can be used to extract the seasonal and trend components from a time series.

(4)

2. From a time series x_t of 156 daily values of electricity consumption, an analyst has calculated the series of first differences $y_t = x_t - x_{t-1}$.

(i) Suggest a reason why the analyst, having examined a time series plot of the data, decided to analyse the first differences.

(1)

The analyst fitted the following model to the series y_t:

$$y_t = 0.4y_{t-1} + \varepsilon_t + 0.7\varepsilon_{t-1}.$$

(ii) What assumptions are usually made about the errors, ε_t?

(3)

(iii) Write out the model equation for x_t in terms of past x_t and past and present ε_t only. The model for x_t can be classified as a member of the ARIMA(p, d, q) family having $p = 1$, $d = 1$ and $q = 1$. Explain these values of p, d and q in the light of your model equation.

(5)

(iv) Describe why and how the analyst could use a plot of the residual autocorrelation function (i.e. the correlogram of the residuals) in order to assess the quality of the model.

(4)

(v) Why might the analyst also examine

(a) a time series plot of the residuals,

(b) a Q-Q plot of the residuals?

(4)

As part of the computer output from the fitting routine, the analyst received the following information on portmanteau lack-of-fit tests for the residuals.

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	18.7	40.2	49.5	56.9
DF	10	22	34	46
P-Value	0.045	0.010	0.041	0.130

(vi) Briefly explain what the chi-square statistic here is measuring. What conclusion should the analyst reach about the adequacy of the model that has been fitted?

(3)

3. Standardised mortality indices are index numbers constructed in a way analogous to price indices in economics. Here the population of type i in region t is denoted by N_{it} and is analogous to quantity in the price index. The number of deaths amongst this population is D_{it} and is analogous to value. The ratio $R_{it} = \frac{D_{it}}{N_{it}}$ is the death rate for that specific type i in region t and plays a role analogous to price.

(i) For assessing the overall death rate in Health Board region A we can compare its data, indexed with subscript A, with those from a standard or baseline population indexed with subscript 0. For example the Laspeyres index is

$$P_L(0, \text{A}) = \frac{\sum_i N_{i0} R_{i\text{A}}}{\sum_i N_{i0} R_{i0}}$$

where the sums are over all types, i. Indicate any simplification that you can make in this formula. How would you interpret $P_L(0, \text{A})$ in a way that a non-expert could understand?

(4)

(ii) Write down the corresponding formula for the Paasche index, $P_P(0, \text{A})$, simplifying where appropriate. How would you interpret $P_P(0, \text{A})$ in a way that a non-expert could understand?

(4)

(iii) The table below gives data on population numbers and deaths for a Standard Population and for Health Board region A.

	Standard Population		*Health Board region A*		
Age Groups, i	*Population size (×1000)*	*Death rates per 1000*	*Population size (×1000)*	*Actual deaths*	*Death rates per 1000*
64 or under	1322	2.0	128	342	
65 – 74	131	21.8	14	373	
Over 75	121	87.2	9	825	
Total	1574		151	1540	

For Health Board region A, calculate

(a) the death rates per 1000 population (to 2 decimal places) for each age group i,

(b) the Laspeyres index, $P_L(0, \text{A})$.

(5)

Question continued on the next page

(iv) The following table contains similar data for Health Board region B.

Age Groups, i	*Population size (×1000)*	*Actual deaths*
64 or under	78	210
65 – 74	9	249
Over 75	7	651
Total	94	1110

Calculate

(a) the death rates per 1000 population for each age group *i*,

(b) the Laspeyres index, $P_L(0, \mathrm{B})$.

In comparing B with A, is the message from the Laspeyres indices consistent with that from the age-specific death rates?

(4)

(v) Show that

$$P_L(0, \mathrm{B}) - P_L(0, \mathrm{A}) = \frac{\sum_i N_{i0}(R_{i\mathrm{B}} - R_{i\mathrm{A}})}{\sum_i N_{i0}R_{i0}}.$$

Deduce that the Laspeyres index for a region B will exceed that for region A in any situation where each age-specific death rate in B exceeds the corresponding rate in A.

(3)

4. (a) (i) When used as a consumer price index the Paasche index is usually expected to be lower than the Laspeyres index. Without doing any detailed calculations, explain why this might be so.

(4)

(ii) Explain why the Fisher index is sometimes described as 'ideal'. Are there any reasons why the Laspeyres index might still be used in preference?

(3)

(b) (i) Suppose you have data available for three periods (0, 1 and 2) from which you have calculated Laspeyres and Paasche indices. Prove that the chain-linked Fisher index (referenced to period 0 and linked at period 1) is the geometric mean of the chain-linked Laspeyres index and the chain-linked Paasche index (both being referenced to period 0 and linked at period 1).

(6)

(ii) Suppose a country's economic output can be classified into three sectors according to the following table.

Economic sector	*2011 value ($ billions)*	*2012 value ($ billions)*	*2012 Laspeyres volume index (base period 2011)*	*2013 Laspeyres volume index (base period 2012)*
Agriculture	20	21	101.2	100.7
Manufacturing	40	40	97.3	100.2
Services	70	75	103.9	105.4

If the Paasche output indices for 2012 (based on 2011) and 2013 (based on 2012) are 100.7 and 102.4 respectively, calculate the chain-linked Fisher index for 2013 (referenced to 2011 and linked at 2012).

(7)

BLANK PAGE

BLANK PAGE