THE ROYAL STATISTICAL SOCIETY

2007 EXAMINATIONS – SOLUTIONS

HIGHER CERTIFICATE

(MODULAR FORMAT)

MODULE 7

TIME SERIES AND INDEX NUMBERS

The Society provides these solutions to assist candidates preparing for the examinations in future years and for the information of any other persons using the examinations.

The solutions should NOT be seen as "model answers". Rather, they have been written out in considerable detail and are intended as learning aids.

Users of the solutions should always be aware that in many cases there are valid alternative methods. Also, in the many cases where discussion is called for, there may be other valid points that could be made.

While every care has been taken with the preparation of these solutions, the Society will not be responsible for any errors or omissions.

The Society will not enter into any correspondence in respect of these solutions.

Note. In accordance with the convention used in the Society's examination papers, the notation log denotes logarithm to base *e*. Logarithms to any other base are explicitly identified, e.g. $\log_{10}$.

Higher Certificate, Module 7, 2007. Question 1

(i) A time series is often explained as $Y = C + S + I$ or $Y = C \times S \times I$, where Y is an observed value, C is a regular component such as a regression or moving average, S is systematic (often seasonal) variation from C and I is irregular or random variation.

Seasonally adjusted estimates have the systematic (usually calendar-related) variation estimated and removed. The irregular component will still be present, but it is then easier to compare estimates across years. A widely used method for carrying this out is to use an iterative filter-based approach, such as the X-12-ARIMA package.

Trend estimates reduce the impact of the irregular component, and help to clarify the underlying direction of the time series. [In the process, it is possible that information useful to some users can also be removed.] The iterative filter-based approach also handles this.

(ii) The following points should be mentioned.

(a) Seasonal adjustment enables annual comparisons to be made.

(b) Corrections for special days, such as religious festivals (Easter, Ramadan, Diwali, etc), non-trading days (bank and public holidays), international holidays (US, China, etc), can be incorporated in the equation used to model data.

(c) Seasonal adjustment is a complex procedure, but leads to improved interpretability by users.

(d) Forecasting may improve estimation of seasonal factors near the end of a series.

(e) A filter-based approach to seasonal adjustment is likely to have significant revisions near the end of a series.

Solution continued on next page

(iii) The following points should be mentioned.

(a) The use of symmetric or asymmetric filters.

(b) The revision to the seasonally adjusted and trend measurements resulting from the use of different filters, and the parameter changes in the seasonal adjustment process.

(c) Because smoothing the seasonally adjusted estimates reduces the impact of the irregular component, the underlying direction of the time series can become easier to observe.

(iv) Modelling can be used to improve time series estimates in various ways. One is the specific modelling of the time series decomposition, explicitly determining the relationship between the seasonal components and trend (taking trend at time t = trend at $(t - 1)$ plus noise).

Another way is extending the original time series to attempt to minimise the impact of revisions on the seasonally adjusted and trend estimates. For example, ARIMA modelling uses past observations to give a model for using to project the missing points at the end of a series. This can reduce the effect of using asymmetric filter weights and reduce the revisions in seasonally adjusted and trend estimates. It is important to choose an appropriate ARIMA model – the wrong model could lead to reduced quality in the estimates.

Using modelling for particular effects is important, such as regression techniques to examine the effect of holidays etc mentioned in (ii)(b) above, and also shifts in the general level of the observations, large extreme values, changes in seasonal pattern. Such modelling improves the quality of estimates, but such effects should perhaps be modelled in the data, as spurious results can cause revisions when more information becomes available.

<u>Higher Certificate, Module 7, 2007. Question 2</u>

(i) Denote the original data by O_i and the trend by T_i. Note that the seasonal factors S_i are centred on i. Hence the decomposition must be a multiplicative model, so $O_t = T_t S_t I_t$.

(ii) According to the model, NOV05 should be $133.32 \times 1.1219 \times 0.9977 = 149.2277$, which is 149.2 to one decimal place, as is O_t.

Some reasons for agreement not being absolutely precise are (a) rounding, (b) the process of weighting for ensuring that entries add up over the year may not be fully accurate, (c) there may have been a holiday/non-trading day (etc.) for which an estimate was used but not shown in the table.

(iii) The change DEC04 to JAN05 for the original values is –42%; for the seasonally adjusted estimates is –1.9%; for the trend estimates is +1.3%.

Each estimate is useful in understanding the nature of movement in the series. The original values drop from a very high value to the lowest value in the entire series. The seasonal adjustment shows the short term movement, but not that due to the time of year, holidays/festivals, etc. The trend suggests a general, slow increase.

(iv) This is because the seasonal factor for DEC05 is 1.4521, considerably reducing the level of the seasonally adjusted series compared with the original value (from 195.77 to 195.77/1.4521), whereas the seasonal factor for AUG05 is 0.9499, marginally increasing the level of the seasonally adjusted series compared with the original value. The trend and irregular components for AUG05 and DEC05 are each roughly the same.

(v) In a multiplicative model, the sum of the seasonal factors should be 12 over the year. From JAN05 to DEC05 it is $0.8568 + \ldots + 1.4521 = 12.0082$.

Higher Certificate, Module 7, 2007. Question 3

Part (i)

$$P_L(0,t) = \frac{\sum_{i=1}^{N} p_{ti} q_{0i}}{\sum_{i=1}^{N} p_{0i} q_{0i}}$$

$$= \frac{\sum_{i=1}^{N} p_{ti} \frac{p_{0i}}{p_{0i}} q_{0i}}{\sum_{i=1}^{N} v_{0i}} \quad \text{where } v_{0i} = p_{0i} q_{0i} \text{, the value of commodity } i \text{ at time } 0$$

$$= \frac{\sum_{i=1}^{N} R_{0ti} v_{0i}}{\sum_{i=1}^{N} v_{0i}} \quad \text{where } R_{0ti} = \frac{p_{ti}}{p_{0i}} \text{, the price relative for commodity } i$$

$$P_P(0,t) = \frac{\sum_{i=1}^{N} p_{ti} q_{ti}}{\sum_{i=1}^{N} p_{0i} q_{ti}}$$

$$= \frac{\sum_{i=1}^{N} v_{ti}}{\sum_{i=1}^{N} p_{0i} \frac{p_{ti}}{p_{ti}} q_{ti}} \quad \text{where } v_{ti} = p_{ti} q_{ti} \text{, the value of commodity } i \text{ at time } t$$

$$= \frac{\sum_{i=1}^{N} v_{ti}}{\sum_{i=1}^{N} \frac{v_{ti}}{R_{0ti}}}$$

Part (ii)

A Laspeyres price index is typically more convenient to calculate than its Paasche equivalent because it requires value data (such as turnover) from the base period instead of from the current period. Current period value data can be difficult to obtain on a timely enough basis for the publication of price indices.

Solution continued on next page

Part (iii)

Laspeyres:

$$\frac{(20\times1.10)+(20\times1.15)+(20\times1.24)+(15\times1.07)}{20+20+20+15} \;=\; \frac{85.85}{75} \;=\; 1.145\,.$$

Paasche:

$$\frac{25+20+15+10}{\frac{25}{1.10}+\frac{20}{1.15}+\frac{15}{1.24}+\frac{10}{1.07}} \;=\; \frac{70}{61.56} \;=\; 1.137\,.$$

Part (iv)

Laspeyres price index numbers use base period quantities, whereas Paasche ones use current period quantities. Because quantity movements are generally negatively correlated with price movements, Laspeyres indices put more weight than Paasche indices on those commodities whose prices are rising the most (or falling the least). Furthermore, Laspeyres indices are arithmetic means of price relatives, whereas Paasche indices are harmonic means of price relatives. An arithmetic mean is greater than a harmonic mean calculated from the same (strictly positive) data.

Higher Certificate, Module 7, 2007. Question 4

Part (a)

$$\frac{P_P(0,t)}{P_L(0,t)} = \frac{\sum_{i=1}^{N} p_{ti}q_{ti} \sum_{i=1}^{N} p_{0i}q_{0i}}{\sum_{i=1}^{N} p_{0i}q_{ti} \sum_{i=1}^{N} p_{ti}q_{0i}} = \frac{\sum_{i=1}^{N} p_{ti}q_{ti}}{\sum_{i=1}^{N} p_{ti}q_{0i}} \frac{\sum_{i=1}^{N} p_{0i}q_{0i}}{\sum_{i=1}^{N} p_{0i}q_{ti}} = \frac{Q_P(0,t)}{Q_L(0,t)}.$$

Part (b)(i)

$$P_L(\text{Central}, \text{South}) = \frac{(55\times100)+(30\times100)+(38\times50)}{(70\times100)+(40\times100)+(45\times50)} = \frac{10400}{13250} = 0.785.$$

$$P_P(\text{Central}, \text{South}) = \frac{(55\times20)+(30\times40)+(38\times10)}{(70\times20)+(40\times40)+(45\times10)} = \frac{2680}{3450} = 0.777.$$

$$P_F(\text{Central}, \text{South}) = \sqrt{P_L(\text{Central}, \text{South})\times P_P(\text{Central}, \text{South})}$$

$$= \sqrt{0.785\times0.777} = 0.781.$$

Part (b)(ii)

$$Q_L(\text{Central}, \text{South}) = \frac{(70\times20)+(40\times40)+(45\times10)}{(70\times100)+(40\times100)+(45\times50)} = \frac{3450}{13250} = 0.260.$$

$$Q_P(\text{Central}, \text{South}) = \frac{(55\times20)+(30\times40)+(38\times10)}{(55\times100)+(30\times100)+(38\times50)} = \frac{2680}{10400} = 0.258.$$

$$Q_F(\text{Central}, \text{South}) = \sqrt{Q_L(\text{Central}, \text{South})\times Q_P(\text{Central}, \text{South})}$$

$$= \sqrt{0.260\times0.258} = 0.259.$$