THE ROYAL STATISTICAL SOCIETY

2008 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, 2008. Question 1

(i) Easter is an example of a moving holiday. (There are several similar holidays in various parts of the world.) The Easter holiday may fall in either March or April, between 22 March and 25 April in the Western Christian calendar, more commonly in April. For seasonal adjustment of a monthly time series, and especially for a quarterly series, this must be allowed for. Many, though not all, retail sales outlets are closed for extra days at Easter. Before the holiday, there may be extra activity such as sales of special items traditionally associated with Easter. Even when Easter is in April, this extra activity could happen in March. A term (or terms) should be included in the regression model used to study the series to allow for these variations in activity as far as possible.

Christmas Day is always on 25 December (in the Western calendar, though this is different in some other calendars), but obviously not always on the same day of the week. There will be at least one day (often two days) when retail outlets are completely closed. An increase in activity may be expected, at least for some types of commodity – clothing, food and household items, for example. This increase may be for quite a long period in December. For some commodities, it may begin earlier than December, although for others it is likely to be concentrated in mid-December. Thus series for different items or with different survey periods are likely to need different treatment. Again, a term (or terms) in the regression model can be used.

(ii) Controlling for identifiable effects, usually called prior correction, aims to ensure that the seasonal factors can be reliably identified and estimated to a high quality. Otherwise, these factor estimates may not be satisfactory.

The key steps in choosing a correction for Easter are that the effects of March and April should compensate one another and there should be an "impact window".

As an example, the following might be satisfactory. Let w be the number of days in the impact window (i.e. on or before Easter), and n_e the number of these days that fall in March; then an Easter correction would be $E(w, t) = n_e/w$ in March, $-n_e/w$ in April and 0 otherwise.

(iii) Seasonally adjusted estimates have the systematic calendar-related component removed, thus containing only the trend and irregular components. Hence they provide useful in-yearly comparisons. However, if there is a large irregular component, the seasonally adjusted estimates may provide misleading information. Trend estimates represent the underlying direction of the time series, and this is a useful component if the series has a large irregular component. However, trend estimates have the disadvantage that they are normally revised at the current end of the time series.

In analysis, the most commonly used estimates are the seasonally adjusted. These include the impact of the irregular component and (notwithstanding the point made above) this can provide useful additional information that is useful in explaining any real-world impacts.

Higher Certificate, Module 7, 2008. Question 2

(i) Since we are using a regression framework, a possible approach is to model the original series by a regression-ARIMA method. In this, the original series Y is modelled as the sum of an ARIMA component Z and dummy variables for trading day (TD), Easter (E) and the two outliers ($AO1$ and $AO2$), all components being assumed additive, and with TD being a vector of six components. Thus the model is

$$Y_t = Z_t + \beta_1 (TD)_t + \beta_2 (E)_t + \beta_3 (AO1)_t + \beta_4 (AO2)_t \; .$$

This means that the impacts of Easter, Trading Day and Outliers have themselves neither been impacted on nor distorted by the seasonal adjustment process. If adjusted values were modelled, some of the potential impact for the variable of interest might have already been removed as part of the seasonal adjustment.

An alternative approach would model the irregular component. However, this would not be optimal if a filter-based approach had been used to estimate the seasonal component (it would cause distortions). In such a case, modelling the original series would be preferable.

(ii) No p-values or degrees of freedom are given for the t values in the output. Assuming that there are adequate residual (error) degrees of freedom for a 5% test to require t not much greater than 2 for significance, the Easter effect and the 2002.8 outlier effect need to be retained in the model. The p-value quoted for "chi-square" at the foot of the output indicates that TD is not needed in the model.

Note that 2002.4 is an April outlier and this may be related to the significant Easter effect. Perhaps this could be removed from the model and a new model fitted to examine the significance of TD.

Both the Easter effect and the 2002.8 effect have $t < 0$, i.e. a fall. Practical explanations of each of these should be sought to see if that might have been expected.

(iii) A stock time series is unlikely to have trading day identified, or an Easter effect, because a long series would be needed to obtain a robust estimate. A user-defined regressor would need to be constructed for an Easter effect; the impact of this will depend on the actual time series being used.

Solution continued on next page

(iv) Temporary prior adjustments remove real world impacts prior to identifying (and removing) the seasonal component. Temporary priors are added back into the seasonally adjusted series. Permanent priors remove systematic calendar effects, but are not added back in.

A temporary correction for outliers would be used to keep the economic effect in the seasonally adjusted series without distorting the seasonal adjustment analysis. But it is wise to have a good reason to make such a correction (see the discussion above regarding outliers).

A permanent correction would be used for fixed systematic calendar effects; such effects should not be included in the seasonally adjusted data. In the present case, Easter and Trading Day are candidates for such permanent corrections, though the analysis of the present data set suggests that Easter appears to need retention but Trading day not (assuming this is not post-hoc inclusion, ie there was a theoretical reason to test for both in the first place).

Higher Certificate, Module 7, 2008. Question 3

(i) $Q_L\left(2004, 2005\right) = \dfrac{(20\times 101.2)+(40\times 97.3)+(70\times 103.9)}{20+40+70} = \dfrac{13189}{130} = 101.5$

(ii) $Q_L\left(2005, 2006\right) = \dfrac{(21\times 100.7)+(40\times 100.2)+(75\times 105.4)}{21+40+75} = \dfrac{14027.7}{136} = 103.1$

(iii) $Q_L{*}\left(2006; 2004\right) = \dfrac{Q_L\left(2004, 2005\right)\times Q_L\left(2005, 2006\right)}{100} = \dfrac{101.5\times 103.1}{100} = 104.6$

(iv) $M_L\left(0, t; r\right) =$ Laspeyres volume index $\times$ value in reference period r

$$= Q_L\left(0, t\right)\times \sum_{i\in U} p_{ri}q_{ri} = \frac{\sum_{i\in U} p_{0i}q_{ti}}{\sum_{i\in U} p_{0i}q_{0i}} \sum_{i\in U} p_{ri}q_{ri} .$$

When r = 0, this gives

$$M_L\left(0, t; 0\right) = \frac{\sum_{i\in U} p_{0i}q_{ti}}{\sum_{i\in U} p_{0i}q_{0i}} \sum_{i\in U} p_{0i}q_{0i} = \sum_{i\in U} p_{0i}q_{ti} .$$

(v) Value in 2004 = 20 + 40 + 70 = 130 ($ billion).

$M_L\left(2004, 2005; 2004\right) = 1.015\times 130 = 132.0$ ($ billion).

$M_L\left(2005, 2006; 2004\right) = 1.031\times 130 = 134.0$ ($ billion).

$M_L{*}\left(2006; 2004\right) = 1.046\times 130 = 136.0$ ($ billion).

<u>Higher Certificate, Module 7, 2008. Question 4</u>

<u>Part (i)</u>

(a) For each commodity group g we have

base period income $\sum_{i \in g} p_{0i} q_{0i}$

current period income $\sum_{i \in g} p_{ti} q_{ti}$.

We wish to construct a Laspeyres volume index $Q_L(0,t) = \dfrac{\sum_{i \in g} p_{0i} q_{ti}}{\sum_{i \in g} p_{0i} q_{0i}}$.

We have the denominator for this, but not the numerator. To proceed, we first multiply and divide the numerator by the current period income:

$$\text{numerator} = \sum_{i \in g} p_{0i} q_{ti} \frac{\sum_{i \in g} p_{ti} q_{ti}}{\sum_{i \in g} p_{ti} q_{ti}}$$

$$= \frac{\sum_{i \in g} p_{0i} q_{ti}}{\sum_{i \in g} p_{ti} q_{ti}} \sum_{i \in g} p_{ti} q_{ti} = \frac{\sum_{i \in g} p_{ti} q_{ti}}{P_P(0,t)} ,$$

where $P_P(0, t)$ is the Paasche price index. Thus the Paasche price index is the appropriate deflator for current period incomes when constructing a Laspeyres volume index.

(b) The Laspeyres price index is the appropriate deflator for current period incomes when constructing a Paasche volume index.

(c) The Fisher price index is the appropriate deflator for current period incomes when constructing a Fisher volume index.

Solution continued on next page

Part (ii)

Deflating the 2007 sales data as indicated in the answer to part (i)(a) gives the following contributions to the numerator.

Cheeses:	£520 million / 1.004 = £518 million
Butters:	£210 million / 0.997 = £211 million
Creams:	£150 million / 0.982 = £153 million.

For milk, the base period price is £690 million / 720 million litres, i.e. £0.958 per litre. This gives a contribution to the numerator of 750 million litres × £0.958 per litre, i.e. £719 million.

Thus the Laspeyres volume index is

$$Q_L\left(0,t\right)=\frac{518+211+153+719}{500+220+120+690}\times 100$$

$$=\frac{1601}{1530}\times 100$$

$$= 104.6.$$