THE ROYAL STATISTICAL SOCIETY

2010 EXAMINATIONS – SOLUTIONS

HIGHER CERTIFICATE

MODULE 7

TIME SERIES AND INDEX NUMBERS

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, 2010. Question 1
[Solution continues on next page]

(i) For an additive decomposition, $Y = C + S + I$. No transformation should be applied to the data before seasonal adjustment.

For a multiplicative decomposition, $Y = C \times S \times I$. A logarithmic transformation should be applied to the data before seasonal adjustment.

The decomposition could be chosen as follows. There may be prior knowledge of the series: an additive decomposition might be appropriate if the series is itself obtained by subtraction or contains zeros or negative values; a multiplicative decomposition might be appropriate if the series is an index or an economic series. We may have knowledge of related series (such as components or aggregates of the series under investigation, or the present series in an earlier period). Or we could fit a seasonally adjusted model using each decomposition and see which performs best (graphically by eye, or quantitatively by reviewing diagnostics).

The seasonally adjusted series is either $Y = C + I$ or $Y = C \times I$. It is achieved by estimating the seasonal component S (by first estimating and removing C) and then removing it from Y (by subtraction if additive or division if multiplicative).

(ii) Movement of Easter between March and April can cause a problem if there is an Easter effect on the series – i.e. if it is higher or lower specifically due to Easter. This is often the case; for example, some series may be lower as there are fewer normal trading days, whereas others may be higher if they are concerned with some form of trading that is enhanced at Easter. In such cases, movements are likely to be distorted simply due to the calendar, not due to any change in the nature of the series. For example, growth between February and March in one year may be smaller than in another simply because in the one year Easter was in March and in the other it was in April. This may lead to a misleading picture of growth/decline in the series – which can cause problems for users basing decisions on movements in the series.

The Easter effect can be tested for by seasonally adjusting with and without an adjustment for Easter and assessing the relative performances, by eye or by considering diagnostics. The diagnostics would be likely to include the significance of the parameter estimate of a dummy variable for Easter itself (as regressed against the irregular component).

If found, the Easter effect is adjusted for by estimating it and removing it from the series prior to seasonal adjustment – for instance by removing half the difference Easter makes from the "Easter month" and adding half the difference to the "non-Easter month". Note that the Easter adjustment is permanent, i.e. the adjustment is not reversed after seasonal adjustment.

Other calendar related effects might include trading day, leap year, any other moving holiday such as Chinese New Year or May Bank Holiday, or Ramadan. A case might also be made for a Christmas effect in terms of relating the series to the day of the week on which Christmas falls.

(iii) Two additive outliers have been identified – one in October 2004 and one in August 2005. These are effects that are not explained by normal seasonal variation, and are not are seen as indicative of permanent change but are one-off non-repeating events. Both outliers are significant at the 5% significance level – which means there is real evidence that they affect the seasonal adjustment of this series. Easter has been tested for, i.e. whether the movement of Easter between March and April has an effect on the series, but the parameter is not significantly different from zero so it appears that Easter does not have an effect on this series.

Both outliers could be adjusted for, depending on policy for dealing with such phenomena. It might be policy that all outliers are adjusted for, or it might be that adjustment is made only for outliers which can be seen to coincide with a real-world event of some kind. Easter should not be adjusted for.

The difference is that adjustment for outliers is temporary – it is reversed after seasonal adjustment has taken place – whereas (as mentioned in the solution to part (ii)) adjustment for Easter is permanent – it is not reversed.

Higher Certificate, Module 7, 2010. Question 2
[Solution continues on next page]

(i) A moving average of the unadjusted time series can be used to smooth the data and obtain an initial estimate of the trend component. This estimate of the trend can then be removed from the unadjusted time series data to leave the combined seasonal and irregular components. The resulting time series can then be smoothed, a month or quarter at a time, by another moving average to obtain an estimate of the seasonal component. This seasonal pattern can in turn be removed from the unadjusted time series to leave an estimate of the seasonally adjusted time series which can then be smoothed to get an improved estimate of the trend component.

(ii) The 5-point simple symmetric moving average for Q3 2005 is given by

$$\frac{50.0+36.5+43.0+44.5+38.9}{5}=42.58$$

and similarly for the others. This calculation cannot be performed for the first two quarters or the last two quarters. Thus we get the following.

Production of commodity, *y*

Period	2005 Q1	2005 Q2	2005 Q3	2005 Q4	2006 Q1	2006 Q2	2006 Q3	2006 Q4	2007 Q1	2007 Q2	2007 Q3	2007 Q4
Production	50.0	36.5	43.0	44.5	38.9	38.1	32.6	38.7	41.7	41.1	40.5	33.8
Moving av.			42.58	40.20	39.42	38.56	38.00	38.44	38.92	39.16		

A *symmetric* moving average (with equal or unequal weights) gives the same degree of importance to values of the series that are the same distance before and after the point in question, which is often intuitively sensible. But, as noted above for this particular series, no such moving average can be calculated at the beginning or end of a series. For the example of this particular series, it contains 12 data items but the smoothed series only has 8 terms. Symmetric moving averages simply cannot be used at the ends of series.

To obtain smoothed estimates in the regions at the ends of series, asymmetric moving averages may be used. A smoothed value is calculated "off centre", with the average being determined using more data from one side of the point than the other, according to what data are available. Alternatively, modelling techniques may be used to extrapolate the series, after which a symmetric moving average is used with the extended series.

(iii) By definition, equally-weighted moving averages give equal weight to all points that are included in the calculation. Often this means that too much weight is given to the points further away from the point in question, relative to the "nearby" points. The "nearby" points are more likely to be "influenced" by the point in question and, relatively, less so by natural variability in the series. Unequal weights, with more weight given to "nearby" points, are therefore often preferred. Another feature of unequal weights is that they can be chosen so as to give approximations to polynomials that efficiently capture trend turning points and points of inflexion.

(iv) In general, the longer the moving average, the smoother the resulting trend, as it takes information from a greater number of data points.

There is however a penalty in making the average longer. A linear trend will seldom continue for any long time, although it may be an adequate approximation over a short period. If the trend has some curvature, using too long an average will distort it. Thus the choice of length for a moving average will be a matter of balancing the objectives of smoothing and following the trend.

Higher Certificate, Module 7, 2010. Question 3

Part (i)

$$\frac{P_L(Jan, Mar)}{P_L(Jan, Feb)} = \frac{\dfrac{\sum_i p_{Mar,i}q_{Jan,i}}{\sum_i p_{Jan,i}q_{Jan,i}}}{\dfrac{\sum_i p_{Feb,i}q_{Jan,i}}{\sum_i p_{Jan,i}q_{Jan,i}}} = \frac{\sum_i p_{Mar,i}q_{Jan,i}}{\sum_i p_{Feb,i}q_{Jan,i}} ,$$

where *p* refers to price, *q* refers to quantity and *i* identifies individual commodities.

Part (ii)

$$P_{Lo}(Feb, Mar; Jan) = \frac{\sum_i p_{Mar,i}q_{Jan,i}}{\sum_i p_{Feb,i}q_{Jan,i}}$$

$$= \frac{\sum_i p_{Mar,i}\dfrac{p_{Feb,i}}{p_{Feb,i}}\dfrac{p_{Jan,i}}{p_{Jan,i}}q_{Jan,i}}{\sum_i p_{Feb,i}\dfrac{p_{Jan,i}}{p_{Jan,i}}q_{Jan,i}} = \frac{\sum_i R_{Feb,Mar,i}R_{Jan,Feb,i}v_{Jan,i}}{\sum_i R_{Jan,Feb,i}v_{Jan,i}} ,$$

where *R* denotes a price relative and *v* denotes a monetary value.

The Lowe price index weights are $R_{Jan,Feb,i}v_{Jan,i}$. These are January values inflated by price movements from January to February.

Solution continued on next page

Part (iii)

$$\frac{P_P(Jan, Mar)}{P_P(Jan, Feb)} = \frac{\dfrac{\sum_i p_{Mar,i} q_{Mar,i}}{\sum_i p_{Jan,i} q_{Mar,i}}}{\dfrac{\sum_i p_{Feb,i} q_{Feb,i}}{\sum_i p_{Jan,i} q_{Feb,i}}} .$$

Part (iv)

$$\frac{P_P(Jan, Mar)}{P_P(Jan, Feb)} = \frac{\sum_i v_{Mar,i}}{\sum_i v_{Feb,i}} \frac{\sum_i p_{Jan,i} q_{Feb,i}}{\sum_i p_{Jan,i} q_{Mar,i}} = \frac{\dfrac{\sum_i v_{Mar,i}}{\sum_i v_{Feb,i}}}{Q_{Lo}(Feb, Mar; Jan)} .$$

Thus the growth in value is divided by a Lowe volume index.

Part (v)

$$\frac{P_F(Jan, Mar)}{P_F(Jan, Feb)} = \frac{\sqrt{P_L(Jan, Mar) P_P(Jan, Mar)}}{\sqrt{P_L(Jan, Feb) P_P(Jan, Feb)}}$$

$$= \sqrt{\frac{P_L(Jan, Mar)}{P_L(Jan, Feb)} \frac{P_P(Jan, Mar)}{P_P(Jan, Feb)}}$$

$$= \sqrt{\frac{\sum_i p_{Mar,i} q_{Jan,i}}{\sum_i p_{Feb,i} q_{Jan,i}} \times \frac{\dfrac{\sum_i v_{Mar,i}}{\sum_i v_{Feb,i}}}{Q_{Lo}(Feb, Mar; Jan)}}$$

$$= \sqrt{\frac{\sum_i v_{Mar,i}}{\sum_i v_{Feb,i}} \frac{P_{Lo}(Feb, Mar; Jan)}{Q_{Lo}(Feb, Mar; Jan)}}$$

from parts (i), (ii) and (iv).

Higher Certificate, Module 7, 2010. Question 4

Part (i)

The formula for the Laspeyres birth rate index is

$$\frac{\sum_i r_{ti} w_{0i}}{\sum_i r_{0i} w_{0i}}$$

where r denotes birth rate, w denotes number of women, i denotes age group, 0 is the base period and t is the current period. Also let b denote the number of live births.

Then the formula can be written alternatively as $\dfrac{\sum_i r_{ti} w_{0i}}{\sum_i b_{0i}}$ or $\dfrac{\sum_i \frac{r_{ti}}{r_{0i}} b_{0i}}{\sum_i b_{0i}}$.

The formula for the Paasche birth rate index is

$$\frac{\sum_i r_{ti} w_{ti}}{\sum_i r_{0i} w_{ti}}$$

using the same notation.

This can be written alternatively as $\dfrac{\sum_i b_{ti}}{\sum_i r_{0i} w_{ti}}$ or $\dfrac{\sum_i b_{ti}}{\sum_i \frac{r_{0i}}{r_{ti}} b_{ti}}$.

Solution continued on next page

Part (ii)

Age group	*Number of women in 2008*	*Number of live births in 2008*	*Number of women in 2009*	*Number of live births in 2009*
15–17	3095	30	2985	28
18–21	4056	439	4027	401
22–29	7483	984	7514	975
30–39	10247	426	10473	453
40–49	9835	37	9626	52

The Laspeyres birth rate index is (as calculated using the basic formula above)

$$\frac{\frac{28}{2985}\times 3095 \;+\; \frac{401}{4027}\times 4056 \;+\; \frac{975}{7514}\times 7483 \;+\; \frac{453}{10473}\times 10247 \;+\; \frac{52}{9626}\times 9835}{\frac{30}{3095}\times 3095 \;+\; \frac{439}{4056}\times 4056 \;+\; \frac{984}{7483}\times 7483 \;+\; \frac{426}{10247}\times 10247 \;+\; \frac{37}{9835}\times 9835}$$

$$= \frac{1900.25}{1916} = 0.9918,$$ or 99.18 if expressed as a percentage.

The Paasche birth rate index is (as calculated using the basic formula above)

$$\frac{\frac{28}{2985}\times 2985 \;+\; \frac{401}{4027}\times 4027 \;+\; \frac{975}{7514}\times 7514 \;+\; \frac{453}{10473}\times 10473 \;+\; \frac{52}{9626}\times 9626}{\frac{30}{3095}\times 2985 \;+\; \frac{439}{4056}\times 4027 \;+\; \frac{984}{7483}\times 7514 \;+\; \frac{426}{10247}\times 10473 \;+\; \frac{37}{9835}\times 9626}$$

$$= \frac{1909}{1924.48} = 0.9920,$$ or 99.20 if expressed as a percentage.

The Fisher birth rate index is $\sqrt{(0.9918)(0.9920)} = 0.9919$, or 99.20 as a percentage.

Solution continued on next page

Part (iii)

An index for number of women would be a weighted average of the changes in the numbers of women in each age group, weighted by the corresponding birth rates. It could be regarded as trying to give an indication of how the overall change in number of women has contributed to the change in the number of live births.

Thus the index is calculated as

$$\frac{\sum_i w_{ti} r_{0i}}{\sum_i w_{0i} r_{0i}}$$

(where w denotes number of women, r denotes birth rate, i denotes age group, 0 is the base period and t is the current period)

$$= \frac{2985\times\frac{30}{3095} + 4027\times\frac{439}{4056} + 7514\times\frac{984}{7483} + 10473\times\frac{426}{10247} + 9626\times\frac{37}{9835}}{3095\times\frac{30}{3095} + 4056\times\frac{439}{4056} + 7483\times\frac{984}{7483} + 10247\times\frac{426}{10247} + 9835\times\frac{37}{9835}}$$

$$= \frac{1924.4807}{1916} = 1.0044,$$ or 100.27 if expressed as a percentage.