THE ROYAL STATISTICAL SOCIETY

2009 EXAMINATIONS – SOLUTIONS

HIGHER CERTIFICATE

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, 2009. Question 1

(i) Many components could be mentioned. Among them are the following.

The trend, representing the long term change in the mean level.

The seasonal, representing movements around the trend which have a fixed period.

Cycles, representing oscillations around the long term variation which are not characterised by fixed periods (as the seasonal variation) but are still predictable to some extent.

The irregular, consisting of the remaining variation in the time series left after other components (e.g. trend, seasonal, cycles) have been removed.

Holiday effects, for example representing public holidays such as Christmas and Easter.

Trading day effects – the day of the trading week can be important, for example trade on Monday or Friday might be different from that on other days.

Outliers, such as special or "one-off" events that have large effects.

Level shifts, where the measurement (level) suddenly takes a step change and does not revert to its previous value.

Seasonal breaks, where the seasonal pattern changes permanently.

(ii) With additive seasonality, seasonal fluctuations exhibit constant amplitude with respect to the trend. With multiplicative seasonality, the amplitude of the seasonal fluctuations is a function of the trend, e.g. the amplitude of the seasonal fluctuations increases as the trend increases.

The model transformation becomes of crucial importance as regression in this context can only be additive. This implies that if the underlying pattern of seasonality is multiplicative, a logarithmic transformation of the model becomes necessary in order to validate the use of regression.

(iii) In the case of additive seasonality, the decomposition model for series X_t could take the form $X_t = TC_t + S_t + \varepsilon_t$ where the trend and cycle have been collapsed into a single component TC_t.

One way to identify and remove the seasonal component is as follows.

(1) Identify the component TC_t using moving averages.

(2) Subtract the component TC_t from the time series X_t.

(3) Estimate the seasonal component from the de-trended series $X_t - TC_t$ and then remove it.

Higher Certificate, Module 7, 2009. Question 2

(i) The ARIMA(1,0,1) model is

$$X_t = \phi_1 X_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

where ϕ_1 is the autoregressive parameter, θ_1 is the moving average parameter and there is no differencing

The fitted values are $\hat{X}_t = \hat{\phi}_1 X_{t-1} + \hat{\theta}_1 \varepsilon_{t-1}$ where $\hat{\phi}_1$ and $\hat{\theta}_1$ are estimates of the parameters. The residuals are then simply given by residual = observation – fitted value, i.e. residual = $X_t - \hat{X}_t$.

(ii) The best way to assess the degree and the type of any autocorrelation is to look at the correlogram. If we had fitted the correct model, the residuals would be Normally distributed and, accordingly, the autocorrelation at any lag would be Normally distributed with mean zero and variance $1/N$. However, not knowing the correct model, we have to rely on confidence interval estimation. In general, $1/\sqrt{N}$ continues to be an upper bound for the standard errors of the autocorrelations of the residuals, implying that autocorrelations lying outside $\pm 2/\sqrt{N}$ (where 2 is used as an approximation to 1.96, the double-tailed 5% point of the N(0, 1) distribution) are significantly different from zero and hence provide evidence that the wrong form of the model has been fitted. This evidence has to be weighed against the fact that the choice of $\pm 2/\sqrt{N}$ as giving the confidence interval means that, if the model is in fact correct, 5% of autocorrelations will be outside by definition.

(iii) The simplest and most intuitive way to analyse the residuals is probably to plot them as a time plot. This procedure will allow a visual check for outliers and for cyclical patterns in the residuals. An alternative is the *portmanteau lack-of-fit test* on the residuals. This test is again based on the correlation of the residuals, but this time considering jointly the correlation of the residuals at all the lags. The statistic for this test is

$$Q = N \sum_{k=1}^{K} r_{z,k}^2$$

where N is the number of term in the series (if differences have been applied to the series than it represents the number of terms in the differenced series) and K is the number of lags which is usually chosen between 15 and 30.

[Other uses of residuals might also be mentioned.]

Higher Certificate, Module 7, 2009. Question 3

Part (i)

Chain-linked Laspeyres: $P_L^*(2;0) = P_L(0,1)P_L(1,2) \div 100.$

Chain-linked Paasche: $P_P^*(2;0) = P_P(0,1)P_P(1,2) \div 100.$

Chain-linked Fisher: $P_F^*(2;0) = P_F(0,1)P_F(1,2) \div 100$

$$= \sqrt{P_L(0,1)P_P(0,1)}\sqrt{P_L(1,2)P_P(1,2)} \div 100$$

$$= \sqrt{P_L(0,1)P_L(1,2) \div 100}\sqrt{P_P(0,1)P_P(1,2) \div 100}$$

$$= \sqrt{P_L^*(2;0)P_P^*(2;0)}.$$

Part (ii)

Notation for this part of the question:

$R^*_{2007,2008,i}$: 2008 volume relative for commodity i, using 2007 as the base period

$s_{2007,i}$: 2007 sales for commodity i

$s_{2008,i}$: 2008 sales for commodity i

(a) Laspeyres

$$Q_L(2007,2008) = \frac{\sum_i R^*_{2007,2008,i}\, s_{2007,i}}{\sum_i s_{2007,i}}$$

$$= \frac{(98.2 \times 20) + (101.3 \times 5) + (107.5 \times 12)}{20 + 5 + 12} = \frac{3760.5}{37} = 101.6.$$

(b) Paasche

$$Q_P(2007,2008) = \frac{\sum_i s_{2008,i}}{\sum_i \frac{s_{2008,i}}{R^*_{2007,2008,i}}}$$

$$= \frac{21 + 6 + 15}{\frac{21}{98.2} + \frac{6}{101.3} + \frac{15}{107.5}} = \frac{42}{0.4126} = 101.8\,.$$

Solution continued on next page

(c) Fisher

$$Q_F\left(2007,2008\right)=\sqrt{Q_L\left(2007,2008\right)Q_P\left(2007,2008\right)}$$

$$=\sqrt{101.6\times 101.8}=101.7\,.$$

(d) Geometric Laspeyres

$$Q_{GL}\left(2007,2008\right)=\prod_i\left(R^*_{2007,2008,i}\right)^{y_i}\qquad\text{where } y_i=\frac{s_{2007,i}}{\Sigma\, s_{2007,i}}$$

$$=98.2^{\frac{20}{37}}\times 101.3^{\frac{5}{37}}\times 107.5^{\frac{12}{37}}=101.6\,.$$

(e) Törnqvist

$$Q_T\left(2007,2008\right)=\prod_i\left(R^*_{2007,2008,i}\right)^{y_i}\qquad\text{where } y_i=\frac{1}{2}\left(\frac{s_{2007,i}}{\Sigma\, s_{2007,i}}+\frac{s_{2008,i}}{\Sigma\, s_{2008,i}}\right)$$

$$=98.2^{\frac{1}{2}\left(\frac{20}{37}+\frac{21}{42}\right)}\times 101.3^{\frac{1}{2}\left(\frac{5}{37}+\frac{6}{42}\right)}\times 107.5^{\frac{1}{2}\left(\frac{12}{37}+\frac{15}{42}\right)}=101.7\,.$$

Higher Certificate, Module 7, 2009. Question 4

[Solution continues on next page]

Part (a)

(i) It is most likely that A is the Paasche index and B the Laspeyres index (because under normal economic conditions Laspeyres indices tend to be larger numbers than the corresponding Paasche indices).

(ii) August 2008 Fisher price index = 100.0

November 2008 Fisher price index = $\sqrt{102.7 \times 111.1}$ = 106.8

February 2009 Fisher price index = $\sqrt{96.2 \times 114.8}$ = 105.1

May 2009 Fisher price index = $\sqrt{93.6 \times 107.4}$ = 100.3

(iii) If there is no seasonality then chain-linking the Laspeyres index should produce numbers which lie between the Laspeyres price index and the Fisher price index.

Part (b)

Notation for this part of the question:

$R_{2007,2008,i}$: 2008 price relative for service i, using 2007 as the base period

$s_{2007,i}$: 2007 sales for service i

$s_{2008,i}$: 2008 sales for service i

(i) Paasche price index

$$P_P\left(2007, 2008\right) = \frac{\sum_i s_{2008,i}}{\sum_i \frac{s_{2008,i}}{R_{2007,2008,i}}} = \frac{1.2 + 0.7 + 0.5}{\frac{1.2}{105.2} + \frac{0.7}{102.7} + (0.5 \times 0)} = \frac{2.4}{0.0182} = 131.9 .$$

[Note. The final index number value is found as 131.7 if the denominator 0.0182 is used to full accuracy as calculated from the preceding fractions.]

[Note that the undefined price relative for the super wash is not a problem for Paasche price index calculation because it appears in the formula as a reciprocal and the term is set equal to zero.]

(ii) Recognising that the product of the price relative and the base period sales in the numerator of the Laspeyres price index formula is a replacement for the product of the current period price and the base period quantity, then the contribution from the super wash is zero because its base period quantity is zero.

(iii) Laspeyres

$$P_L\left(2007,2008\right)=\frac{\sum_i R_{2007,2008,i}s_{2007,i}}{\sum_i s_{2007,i}}$$

$$=\frac{(105.2\times 1.0)+(102.7\times 0.8)+0}{1.0+0.8+0.0}=\frac{187.36}{1.8}=104.1.$$

(iv) Paasche volume index

We first need volume relatives, denoted here by R^*. Those for the quick and full washes can be calculated by recognising that the growth in sales is the product of the growth in prices and the growth in quantities (all growths being from 2007 to 2008). Thus

Quick wash: $$R^*_{2007,2008,1} = 100\,\frac{\dfrac{s_{2008,1}}{s_{2007,1}}}{\dfrac{R_{2007,2008,1}}{100}} = 100\,\frac{\dfrac{1.2}{1.0}}{\dfrac{105.2}{100}} = 114.1.$$

Full wash: $$R^*_{2007,2008,2} = 100\,\frac{\dfrac{s_{2008,2}}{s_{2007,1}}}{\dfrac{R_{2007,2008,2}}{100}} = 100\,\frac{\dfrac{0.7}{0.8}}{\dfrac{102.7}{100}} = 85.2.$$

The volume relative for the super wash cannot be calculated. However, recognising that the ratio of the current period sales to the volume relative in the denominator of the Paasche volume index formula is a replacement for the product of the current period price and the base period quantity, the contribution from the super wash is zero because its base period quantity is zero. Thus

$$Q_P\left(2007,2008\right)=\frac{\sum_i s_{2008,i}}{\sum_i \dfrac{s_{2008,i}}{R^*_{2007,2008,i}}}=\frac{1.2+0.7+0.5}{\dfrac{1.2}{114.1}+\dfrac{0.7}{85.2}+(0.5\times 0)}=\frac{2.4}{0.0187}=128.3.$$