ANNEX 7

METHODS FOR DERIVATION OF STANDARD ERROR AND CONFIDENCE INTERVALS

METHODS FOR DERIVATION OF STANDARD ERROR AND CONFIDENCE INTERVALS

1. STANDARD ERROR OF GROSSED-UP EXPENDITURE FOR INDIVIDUAL CELLS

For grossing up, the individual cells are taken to be those partitioned by SIC code and the two employee size bands.

The standard error of grossed-up expenditure was derived as follows:

$$SE = M \times s \times \sqrt{n} \qquad \text{Equation 1}$$

where: $M = \frac{E}{e}$

SE = standard error of grossed-up expenditure

s = standard error of expenditure x_i, (from Equation 2)

n = number of valid responses of businesses in the grossing-up cell

E = population of employees in that cell for the UK

e = number of employees for the responding businesses in that cell

$$s = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{n-1}} \qquad \text{Equation 2}$$

Equation 2 includes valid zero expenditure x_i responses.

Note that Equation 1 is derived from the following more basic equation for standard errors of grossed-up totals:

$$SE = \frac{P}{\sqrt{n}} \times s \qquad \text{Equation 3}$$

Where:

P = population of businesses in the grossing-up cell

Equation 3 is obtained by assuming that $\frac{P}{n} = \frac{E}{e} = M$

Finite population correction

When the sampling from a finite population without replacement, the standard error calculated using Equation 1 can be reduced to allow for the knowledge gained about part of the population from the sample. Equation 4 shows this adjustment. The analysis only made a finite population adjustment if $M \leq 5$. This threshold was chosen because if M > 5 the correction would be less than 0.89 and therefore not considered worthwhile.

SE is reduced using Equation 2 if $M \leq 5$.

$$SE_F = SE \times \sqrt{1 - \frac{1}{M}} \qquad \text{Equation 4}$$

Where:

SE_F = standard error from Equation 3 adjusted for a finite population

2. STANDARD ERROR OF SUM OF TOTALS FOR EACH EMPLOYEE SIZE BAND

Responding companies in each SIC Sector are stratified according to company size defined by numbers of employees.

All SIC sectors will be analysed for two size bands, i.e.,

b_1 10 to 249 employees

b_2 250+ employees

Standard error, (from Karmel & Polasek, 1978)

$$SE_T = P\sqrt{\sum(\alpha_b^2 \frac{s_b^2}{n_b})} \qquad \text{Equation 5}$$

and

$$\alpha_b = \frac{P_b}{P} \qquad \text{Equation 6}$$

Where:

SE_T = standard error of Total

s_b = standard error of expenditure x_i, size band *b*

n_b = number of valid responses of businesses in the grossing-up cell size band *b*

P = overall population of businesses across all size bands

P_b = population of businesses in size band *b*.

Substituting Equation 3 into Equation 5 gives the following Equation for combining standard errors of each size band to calculate the standard error of the total. This simplification of Equation 5 occurs because for this survey P = $\sum P_b$

$$SE_T = \sqrt{\sum (SE_b)^2} \qquad \text{Equation 7}$$

where:

SE_b = standard error for size band b (calculated using Equation 1).

Equation 7 is used to combine the Standard errors of different cells of the grossing-up frame to estimate the standard errors of the overall totals.

3. STANDARD ERROR FOR THE AGGREGATION OF DIFFERENT TYPES OF ENVIRONMENTAL EXPENDITURES

Some results consist of an overall expenditure that is the sum of disaggregate types of expenditure, (see Annex 6 for method of grossing-up). For example, combining the different types of environmental expenditure to obtain a total environmental expenditure for comparison with UK Gross Output and GDP.

Opex, Capex, Total Net, and Total Gross expenditure are examples of results that are the sum of individual questions.

However, the standard error of the total needs to be calculated specifically using Equation 8. This method takes into account the likely correlation between the responses to the different questions. For example, high environmental expenditure companies often have expenditure for most questions whereas many companies have zero expenditure for all questions.

$$SE = M \times s_r \times \sqrt{n} \qquad \text{Equation 8}$$

Where:

SE = standard error of the grossed up expenditure of the sum of individual questions

And:

$$s_r = \sqrt{\frac{\sum_{i=1}^{N}(x_{ri} - \bar{x}_r)^2}{n-1}} \qquad \text{Equation 9}$$

Where

n = number of businesses *(i)* that have given a valid response to all relevant questions

x_{ri} = sum of valid expenditure responses of business i

$\bar{x}_r =$ mean of x_r for all businesses n in cell of grossing-up frame

Note that combining the individual standard errors of the component questions using the following equation used by the 1997 survey, would not take this into account and would underestimate the *standard error, because it does not take into account the correlation in response between different questions.*

$$SE = \sqrt{\sum (SE_q)^2} \qquad \text{Equation 10}$$

Where:

SE_q = standard errors of grossed-up expenditures for the individual questions q

4. CONFIDENCE INTERVALS

There is no good reason for replacing z by the critical t-value even for small samples because the distribution of business expenditures can currently be regarded as unknown and is not likely to be normal. The following is an approximate confidence interval:

$$T - zSE_T, T + zSE_T \qquad \text{Equation 11}$$

Where:
z = 1.96 for a 95% confidence interval

5. REFERENCE

Karmel P.H. & M. Polasek 1978 'Applied Statistics for Economists,' 4th Edition, Pitman International: (Chapter 6).