

## **Construction Output Sampling, Imputation And Grossing Methodology**

### **1. Production of construction output**

Construction output is a quarterly series of the output of the construction industry, in both the private and public sectors. Press Notices are published on the first Friday of March, June, September and December. The actual series is a combination of results from 2 surveys, the Quarterly Inquiry of Construction Activity and the Building and Civil Engineering Employment and Output Enquiry, plus an estimate for unrecorded output.

#### **The Quarterly Inquiry of Construction Activity**

The Quarterly Inquiry of Construction Activity (CA) measures the output (and employment) of private contractors, covering new construction, improvements and alterations, and repair and maintenance.

Output figures include:

- the value of building, civil engineering and associated work, excluding VAT, done by the contractor's directly employed staff which is chargeable to customers
- the value of materials used, labour costs, overheads and profits
- the value of work done on the contractors own initiative on buildings such as dwellings for eventual sale or lease
- the value of work done on demolition and site preparation
- the value of work done by the contractor on the construction or maintenance of its own premises
- the value of articles made by the contractor and used in construction work
- the value of any materials supplied by the contractor free of charge to subcontractors.

The following are excluded:

- the value of work done overseas
- the value of work done by sub-contractors of any type
- the value of any payments made to labour-only subcontractors
- the value of articles made for sale or materials sold
- the value of materials supplied to the contractors free of charge by firms in the industry
- the value of land
- any architects or consultants fees.

#### **1.1. Sampling and collection of Quarterly Construction Activity**

Information for Great Britain is based on quarterly returns from a sample of around 12,000 firms taken from the CISTATS universe of around 170,000 construction contractors. This universe is kept up-to-date through quarterly exchanges of information with the Inter-Departmental Business Register as well as information

from data suppliers. It contains details of VAT-registered companies classified to construction (Section F/Division 45 of the [Standard Industrial Classification 2003](#)) in Great Britain.

Construction firms on CISTATS are assigned to size groups (SGs) based on their employment levels. Initially the employment data supplied by ONS at the company's "birth" is used, this is then updated when the company responds to a DTI survey. The sample is selected using a rotational sampling methodology with an eight quarter rotation - rotational sampling is explained in section 2. The table below shows the sampling ratios used – SGs 0 and 1 have a variable sampling ratio to ensure that the sample size is always 12,000:

Size Group	Employment	Sampling Ratio
0	1	Variable
1	2-3	Variable
2	4-7	1 in 8
3	8-13	1 in 4
4	14-24	1 in 3
5	25-34	1 in 2
6	35-59	1 in 1
7	60-79	1 in 1
8	80-144	1 in 1
9	115-299	1 in 1
10	300-599	1 in 1
11	600-1199	1 in 1
12	1200-6499	1 in 1
13	6500+	1 in 1

The output information collected is broken down into New Work, Public Housing Repair and Maintenance, Private Housing Repair and Maintenance, Public Non-housing Repair and Maintenance and Private Non-housing Repair and Maintenance. The employment information is broken down into Working Proprietors, APTCs (administrative, professional, technical and clerical employees) and Operatives (manual workers).

## **1.2. Results Processing - Imputation and Grossing**

The following paragraphs describe the processing carried out to produce construction output from the information collected by DTI surveys. DTI uses the agreed common methodology used for all business surveys.

Pre-imputation and Imputation – CA firms which are in the sample but have not responded have a value imputed for them. For each Size Group, the system calculates an imputation link by taking the average movement between the previous quarter and the current quarter for those firms which have responded. This is then applied to those firms which responded in the previous quarter but not in the current one. Where a firm did not respond in the last quarter an average figure is imputed, which is the same for each affected firm in the Size Group. During pre-imputation, the statistician examines the imputation link for each question in each Size Group, and adjusts it if necessary.

Construct – a small number of the largest firms have been marked as key responders, and cannot be imputed for. For each of these firms that has not responded, the statistician must construct a response, usually based on the imputed value for the stratum, unless there is better information available.

Weights – The first stage of grossing is the calculation of the grossing weights, which will be used to multiply up the sample total to the universe total. CA uses 3 different grossing weights: model, design and outlier.

The model weight (g) adjusts the grossing according to known independent properties of the companies sampled. The independent variable used is the latest employment value held on the IDBR and collected through an ONS survey. This information is updated on CISTATS annually, prior to selection of the Q1 sample. The formula is:

$$g = \frac{X}{\left(\frac{N}{n}\right)^* x}$$

where

X = Size Group population total for ONS employment

x = Size Group sample total for ONS employment

N = population size at selection

n = sample size

For companies in Size Groups 6 and above, where there is 1 in 1 sampling, the model weight is 1 because X = x and N = n

The design weight (a) is the main grossing weight. The formula is:

$$a = \left(\frac{N}{n}\right)^* \left(1 + \frac{hd}{(n-d)}\right)$$

where

h = factor for births and deaths (1 for companies in strata 0-5, 0 for companies in strata 6-13)

d = number of deaths (closedowns) recorded

N = population size at selection

n = sample size

Again for Size Groups where there is 1 in 1 sampling (ie SG 6-13) the design weight is equal to one.

The outlier weight (w) ensures that responses which are deemed to be outliers (see below) are not grossed as they are not representative.

If the output per person is greater than the outlier limit set then

$$w = \frac{n}{N}$$

where

N = population size at selection

n = sample size

otherwise  $w = 1$

Outlier limits, that is the cut-off point whereby everything above that level is deemed to be an outlier, are set manually for each stratum every quarter, based on the output per person values for each return. For example, if an outlier limit of £60,000 is set, then all returns where the output per person is greater than this are deemed to be outliers, and will be grossed accordingly.

So, grossed value =  $a * g * w * \text{ungrossed value}$

where a, g and w are the grossing weights described above. Because non-responders are imputed for, returns for firms in Size Groups 6 and above, where there is 1 in 1 sampling, are not grossed.

### **1.3. The Building and Civil Engineering Employment and Output Enquiry (DLO)**

The Building and Civil Engineering Employment and Output Enquiry is usually known as DLO, for Direct Labour Organisation. DLOs are public sector organisations which employ their own construction workers rather than contracting out construction work to the private sector. In the past this included the public utilities such as the water boards however now these have largely been privatised it covers local authorities.

#### **Sampling and collection of DLO information**

As there are only about 215 DLOs, this quarterly enquiry is a census, using 1 in 1 sampling. The questions asked are the same as for the CA Inquiry.

#### **Results Processing - Imputation and Grossing**

DLO results are processed in a similar way to CA, going through Pre-imputation and Imputation. No grossing is necessary because of the 1 in 1 sampling.

### **1.4. Estimate of Unrecorded Output**

The Builders Address File, which DTI uses as a universe for all its contractor based surveys is aligned quarterly with the IDBR but only includes those firms over the VAT threshold. It is clear from comparing the ONS surveys of employment with the figures for construction employment collected in the CA and DLO surveys that a significant proportion of construction employment is either self-employment or with

firms not registered for VAT. The work done by these people is therefore not captured by either the CA or DLO surveys.

In order to compensate for this missing “unrecorded” output, we make an estimate of it. This is done by taking the difference between total construction employment (including the self-employed) as measured in the Labour Force Survey, and total construction employment reported to CISTATS. This gives us a measure of unrecorded employment, which tends to be around 600,000. This is then combined with wage information and the average output per person for Size Groups 0-2 to give an estimate of the output produced by the construction workers missed.

## **2. Rotational Sampling**

Construction Activity sampling is carried out using a rotational sampling method. Each firm is allocated to a Size Group based on employment. The sample required from each cell is determined by the pre-set sampling fraction appropriate to the Size Group.

Each firm is allocated a permanent random number, or PRN, when it is added to the Inter-Departmental Business Register (IDBR). This is used to produce a list of firms in each cell in random order. The first sample is the first  $n$  firms in the list, where  $n$  is the sample size.

To work out the second sample, we need a further value, the “period to rotate” or “ $r$ ”. This indicates the number of periods for which we want a firm to stay within the sample, and is pre-set. For Construction Activity the period to rotate is 8, which is 8 quarters or 2 years. We then work out  $n/r$ , which gives us the number of firms which are dropped from the start of sample 1. Starting from the first firm left in the list, we take the next  $n$  firms and this makes up sample 2.

New firms (“births”) are allocated a PRN and slotted in where appropriate – the only proviso is that a firm which has started its turn in the sample cannot fall out because of a birth, and in these cases the inclusion of such a firm will be forced. Firms which have closed down (“deaths”) leave the list, and the sample moves on as if they had never existed.

This concept is best illustrated using an example. Imagine a survey with a population ( $N$ ) of 100 firms, with a sampling fraction of 1 in 10 (so a sample size ( $n$ ) of 10 firms) and a period to rotate ( $r$ ) set to 5. We list all 100 firms in random order, and then start to sample. The first sample is the first 10 firms on the list. For the second sample, we drop the first 2 firms (sample/period to rotate =  $10/5 = 2$ ), and then, starting from firm 3, we take the next 10 firms. For the third sample, we again drop 2 firms, and take 10 firms starting from firm 5, and so on. After 5 samples, all the firms in the first sample have completed their turn, and the sixth sample contains none of the firms which were included in the first sample. These firms then have a survey holiday, which lasts in this example for approximately 45 periods – the exact length of a holiday is affected by births and deaths but the formula is:

$$\text{Holiday length} = \left( \frac{(N-n)}{n} \right) r$$

### Example of rotational sampling

Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6	No of periods in sample
Firm 1						1
Firm 2						1
Firm 3	Firm 3					2
Firm 4	Firm 4					2
Firm 5	Firm 5	Firm 5				3
Firm 6	Firm 6	Firm 6				3
Firm 7	Firm 7	Firm 7	Firm 7			4
Firm 8	Firm 8	Firm 8	Firm 8			4
Firm 9	Firm 9	Firm 9	Firm 9	Firm 9		5
Firm 10	Firm 10	Firm 10	Firm 10	Firm 10		5
	Firm 11	Firm 11	Firm 11	Firm 11	Firm 11	5
	Firm 12	Firm 12	Firm 12	Firm 12	Firm 12	5
		Firm 13	Firm 13	Firm 13	Firm 13	5
		Firm 14	Firm 14	Firm 14	Firm 14	5
			Firm 15	Firm 15	Firm 15	5
			Firm 16	Firm 16	Firm 16	5
				Firm 17	Firm 17	5
				Firm 18	Firm 18	5
					Firm 19	5
					Firm 20	5

### 3. Deflation

Construction output is deflated using a range of different price indices as appropriate to individual sectors and whether the work is carried out by contractors or public sector direct labour. The indices used and their sources are given below.

#### *Contractors output*

Sector	Deflator
Public Housing New Work - England and Wales	Tender Price Index for Social Housing (TPISH) (DTI)
Public Housing New Work - Scotland	Public Housing – Tender Price Index (Scottish Executive)
Private Housing New Work	0.5 * Index of House Prices at Mortgage Approval Stage (New Dwellings) (ODPM) + 0.5 * TPISH
Public Non-housing New Work	Tender Price Index of Public Sector Building Non-Housing (PUBSEC) (DTI)

Private Industrial New Work	Private Industrial Index (BCISPI) (Building Cost Information Service)
Private Commercial New Work	Private Commercial Index (Building Cost Information Service)
Infrastructure New Work - roads	Road Construction Price Index (RCPI) (DTI)
Infrastructure New Work - other	$0.7 * RCPI + 0.2 * PUBSEC + 0.1 * BCISPI$
Repair and Maintenance	$0.5 * \text{Materials R\&M (MATRM) (DTI/ONS)} + 0.5 * \text{Contractors Labour (LABC) (DTI)}$

### *Direct Labour output*

<b>Sector</b>	<b>Deflator</b>
Housing New Work	$0.5 * \text{Material New Housing (MATHO) (DTI/ONS)} + 0.5 * \text{DLO Labour (LABD) (DTI)}$
Non-housing New Work	$0.5 * \text{Material Non Housing (MATNH) (DTI/ONS)} + 0.5 * \text{DLO Labour (LABD) (DTI)}$
Housing R&M	$0.25 * \text{MATRM (DTI/ONS)} + 0.75 * \text{DLO Labour (LABD) (DTI)}$
Non-housing R&M	$0.33 * \text{MATRM (DTI/ONS)} + 0.67 * \text{DLO Labour (LABD) (DTI)}$

The materials indices (MATRM, MATHO and MATNH) are weighted averages of various materials Producer Price Indices using weights provided to ONS by DTI. These are published by DTI in the *Monthly Statistics of Building Materials and Components*. The price indices TPISH, PUBSEC and RCPI are published by DTI in *Quarterly Building Price and Cost Indices*.

LABC and LABD are combinations of labour cost indices for skilled labour (LABS), unskilled labour (LABU), heating and ventilation labour (HAV) and electrical installation labour (ELEC) using the following formulae:

$$LABC = (0.56 * LABS) + (0.29 * LABU) + (0.08 * HAV) + (0.07 * ELEC)$$

$$LABD = (0.72 * LABS) + (0.28 * LABU)$$

The four individual labour cost indices are published by DTI in the *Monthly Bulletin of Indices*.

## **4. Seasonal adjustment**

Seasonal factors are calculated annually for each sector, for contractors and DLO output separately, using X11ARIMA.