

# Quarterly Employment Survey – Summary Methodology Guide

## Contents

1.1 Sample design and methodology .....	2
1.2 Sampling Methodology.....	2
2.1 Data Capture and Processing .....	4
2.2 Data Validation .....	4
2.3 Data Congruency checks.....	4
2.4 Imputation for missing or invalid responses.....	5
3.1 Grossing /Estimation.....	6
3.2 Confidence Intervals and precision of estimates .....	6
4.1 Seasonal Adjustment.....	7
5.1 Revisions to QES Estimates.....	8
Further Information.....	8

## 1.1 Sample design and methodology

The sample for inclusion in the survey is drawn biennially from the Inter Departmental Business Register (IDBR) (following completion of the Census of Employment and incorporation of these results on the register system). The selection is actually taken in two parts so that there is an element which is effectively full coverage and also an element which is a statistical sample. The full coverage includes all public sector units, private sector units with 25 or more employees and private sector units with more than one industry activity. The sample is a stratified random sample of small businesses with less than 25 employees.

The IDBR is a dynamic register, which is continuously updated to reflect the most recent information (i.e. it takes account of company births, deaths, mergers, etc.). The QES sample is 'topped up' every quarter by new 25+ openings identified from IDBR and the IDBR is also used to estimate the number of employees in small businesses that have recently opened.

### QES Sampling Strata

<b>Strata by each 2 Digit SIC (07)</b>	<b>Employment Size / Sector</b>	<b>Coverage</b>
01 – 96	0-9 emps	Sample
01 – 96	10-24 emps	Sample
01 – 96	MULTI*	Full Count
01 – 96	25+	Full Count
01 – 96	Public Sector	Full Count

\*Multiple Industry Activities.

## 1.2 Sampling Methodology

The QES sample is selected using a Neyman Allocation which is an optimal form of allocation where the sample is allocated to achieve the smallest possible standard error of estimate of the population mean for a given size of sample, i.e. the highest precision of the estimate. This methodology ensures that consideration is given to the variability of the observations within the strata as well as the number of cases in the strata. This is important since strata with relatively large variation in the observations require a larger sample size to obtain a good estimate than do strata with less variability.

The QES sample uses the Neyman allocation to provide an accuracy of (+/-) 5% at the Industry Sector level except for Manufacturing which has an accuracy of (+/-) 2%. Results are published at the 2 digit Standard Industrial Classification (SIC07) level and are split into the categories and then aggregated to form the corresponding sections.

The following formulae show the methodology that is applied

**Neyman Allocation selection Formulae:**

$n$  = the number of sampling units for required accuracy

$$n = \frac{\left( \sum_{i=1}^L N_i \sigma_i \right)^2}{\frac{B^2}{4} + \sum_{i=1}^L N_i \sigma_i^2}$$

$n_i$  = the number of sampling units allocated in each stratum

$$n_i = n \left( \frac{N_i \sigma_i}{\sum_{i=1}^L N_i \sigma_i} \right)$$

Where:

$L$  = Number of strata

$N_i$  = Number of units in the population Stratum  $i$

$N$  = Number of units in the population

$n_i$  = number of sampling units in stratum  $i$

$n$  = number of sampling units

$\sigma_i$  = population Standard deviation stratum  $i$

$B$  = Bound error in terms of % of total

## **2.1 Data Capture and Processing**

The QES runs on a 14-15 week cycle. The forms are despatched by second-class post to arrive by the survey date. Contributors are asked to return the completed form within two weeks of receipt. Non-respondents receive a written reminder at the beginning of week three. At the end of week four non-respondents with a valid Fax number / Email address receive a Fax / Email reminder. Finally, during weeks 4-8 all those who have still not responded will be contacted by telephone to obtain employment data. The speed with which forms are returned can be affected by holiday periods but where possible QES forms and telephone calls are addressed to named contacts to ensure as high a response rate as possible.

For all data collected, best estimates are accepted where precise figures are not available. The final response rate to the QES is typically around 90% of all businesses on the survey. Data is received by post, fax and telephone and is keyed onto the Quarterly computer system via the data prep team at Equiniti / ICS Computing Ltd.

## **2.2 Data Validation**

There are a number of hard edit and logic checks designed to ensure that values cannot be keyed in error (e.g. components must sum to total). Further soft edit checks takes place to identify results that are extremely implausible but not impossible. This will take the form of validation against previous returns to identify any unusual increases and decreases. Contributors whose quarterly returns show an increase or decrease above a certain threshold are contacted and asked for an explanation (if they have not already provided one on their original returned QES form) to ensure that the employment return is correct. Where a return does not make it through the “validation gate” and no explanation is forthcoming, these returns will be imputed until an explanation or correction is provided.

## **2.3 Data Congruency checks**

Further validation of the QES returns is carried out by cross checking against other survey and admin sources, such as redundancy notices, PAYE information from the IDBR, previous QES returns, press releases and information from other surveys to ensure their accuracy. Detailed congruence checks take place using responses from the NI Census of Employment which is carried out biennially. This takes place as part of the biennial benchmarking of QES employee jobs estimates to NI Census of Employment Returns.

## **2.4 Imputation for missing or invalid responses**

Imputation is defined as a procedure for entering a value for a specific data item where the response is missing or unusable. This could be because it is missing due to non-response or unusable due to errors or inconsistent responses. Simplistically, imputation can improve the accuracy of the estimates by reducing non-response bias. Imputed values for the QES are calculated based on the contributor's previous returns and/or the PAYE data held on the Inter-Departmental Business Register (IDBR).

### 3.1 Grossing /Estimation

Estimation is the process of approximating some characteristic of a population (e.g. total employees for the Manufacturing industry) when information on that characteristic is only available for a sample of the population. The QES samples around 5,500 of all the VAT and PAYE registered businesses on the IDBR and uses a “Ratio Estimation” methodology to calculate total employee jobs estimates by 2 digit SIC.

The IDBR holds information about every unit in the population, for example the number of employees in every business. We call this information auxiliary data. If we can assume some kind of relationship between this information and the characteristic which we are trying to estimate, then we can incorporate the auxiliary data into our estimate. For instance the auxiliary variable displays the relationship between the sample and population employees at the time of sample selection. This is used to calculate a “grossing factor” which can be applied to the returned sample value for the current quarter.

The following formula describes how the ratio estimator is used on the QES:

$$\hat{Y} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} \sum_{i=1}^N x_i$$

Where:

$\hat{Y}$  is the ratio expansion employee job estimate of  $Y$

$x_i$  is the value of the auxiliary variable for the  $i$ th business (IDBR employee value at time of selection)

$y_i$  is (sample employee value at reference period)

$n$  is sample size

$N$  is number of units in the population

### 3.2 Confidence Intervals and precision of estimates

Although the QES sample is selected to provide a level of accuracy at the broad sector level, there are currently no confidence intervals published around the employee jobs estimates. A work programme for developing confidence intervals around employee jobs estimates at the 2 digit SIC level is underway and the first set of QES confidence intervals are scheduled to be published in 2012.

## 4.1 Seasonal Adjustment

Data that are collected over time form a time series. Those analysing time series typically seek to establish the general pattern of the data, the long term movements, and whether any unusual occurrences have had major effects on the series. This type of analysis is not straightforward when one is reliant on raw time series data, because there will normally be short-term effects, associated with the time of the year, which obscure or confound other movements. For example, retail sales rise each December due to Christmas. The purpose of seasonal adjustment is to remove systematic calendar related variation associated with the time of the year, i.e. seasonal effects. This facilitates comparisons between consecutive time periods.

The Northern Ireland employee jobs estimates from the QES are an important short term economic indicator. The estimates are therefore seasonally adjusted using the X12 ARIMA software to enable timely interpretation of time series data. This seasonally adjusted employment series provides a better indication of underlying trends in employment.

The X12 ARIMA method, developed by Statistics Canada in 1980, uses Box Jenkins AutoRegressive Integrated Moving Average (ARIMA) models to extend a time series and is used extensively in statistical surveys. The program splits the series into a trend, seasonal, and irregular components. If the series is modelled additively, summing the three parts gives the unadjusted data. If it is modelled multiplicatively, the raw data is the product of the three components. The seasonal component cannot be found without knowing the trend component, yet the trend component cannot be found without knowing the seasonal component. Thus the programs perform a series of iterations, obtaining a better estimate for the trend and seasonality with each one.

The program fits an autoregressive integrated moving average model to the data, using forecasts for one year to improve the estimation of the seasonal factors at each end of the series.

Published estimates for the QES are unadjusted at the 2 digit SIC level with seasonal adjustment being carried out at the higher Industry level (Manufacturing, Construction, Services and “Other” industries) for males and females where seasonality is present. There are also published seasonally adjusted estimates for the Public and Private sector.

At the latest review, seasonality was not found in the construction sector or Other industries.

## **5.1 Revisions to QES Estimates**

It is normal practice for Quarterly Employment Survey estimates to be revised. Estimates for each quarter back to the latest published Census of Employment reference period are provisional and subject to revision in each subsequent first release. These revisions will occur primarily in light of (a) late responses to surveys and administrative sources, (b) where cross survey congruence checks provide more accurate data and (c) where there are revisions to seasonal adjustment factors.

A more extensive revisions exercise also takes place biennially as part of benchmarking QES estimates to the NI Census of Employment and the reallocation of the QES sample.

Further detail about [QES Revisions](#) including the QES revisions policy, and the extent of the revisions can be accessed on the QES page of the DETINI website.

## **Further Information**

Further details about the methodology employed in the Quarterly Employment Survey can be obtained by contacting:

Email: [darren.hetherington@dfpni.gov.uk](mailto:darren.hetherington@dfpni.gov.uk)

Tel: (028) 9052 9655