

Methodology Changes in Scottish Corporate Sector Statistics 2011

Scottish Corporate Sector Statistics provides information about the number of enterprises operating in Scotland broken down by various groupings including industry, company size, local authority area, and country of ownership.

The publication of [Scottish Corporate Sector Statistics \(SCSS\) 2011](#), released on 12 October 2011, contains several methodological changes. These are as follows:

(1) SCSS 2011 is the first time that all data is published using the 2007 Standard Industrial Classification only.

(2) Within the methodology used to estimate unregistered enterprises, we are introducing:

(a) a new method of obtaining more accurate estimates of the self-employed population from the Labour Force Survey.

(b) HMRC Self Assessment data, to replace the Survey of Personal Incomes to estimate the ratio between sole traders and partners.

The purpose of this note is to make users aware of the changes and the reasons for implementing them and to explain the impact they have on the estimates.

Note that the changes to the unregistered methodology are in line with recent changes made by the Department for Business, Innovation and Skills (BIS) in their new [Business Population Estimates for the UK and Regions](#) publication. However, there are still some differences between the SCSS and the BIS estimates, which are explained in the document '[UK Business Data Sources](#)'.

(1) Standard Industrial Classification 2007

Scottish Corporate Sector Statistics presents data about the industrial sectors businesses belong to. The industrial sectors are defined by the [UK Standard Industrial Classification](#) (SIC) which is used to classify business establishments and other statistical units by the type of economic activity in which they are engaged.

The Standard Industrial Classification recently underwent a major revision to enable it to more accurately reflect the structure of the modern economy and, in January 2008, SIC 2007 replaced SIC 2003. The revised classification reflects the growing importance of service activities in economies over the last fifteen years, mainly due to the developments in information and communications technologies.

Until now, SCSS has been published on a SIC 2003 basis. From 2008 onwards, businesses on the Inter-Departmental Business Register (IDBR), which is the source of the estimates of registered businesses within the SCSS, have been dual coded to SIC 2003 and SIC 2007. While the main publication continued to use SIC 2003, 2008-

2010 tables for **registered enterprises** were also released on a SIC 2007 basis so that users could familiarise themselves with the new classification. Data on unregistered enterprises continued to be available on a SIC 2003 basis only.

It was announced in Scottish Corporate Sector Statistics 2010 that it would be the final publication to be based on SIC 2003. The transition to SIC 2007 was postponed in 2010 because data on the IDBR was still being dual coded (thus data was still available on both classifications) and it was decided to implement several methodological changes (as outlined in this document) simultaneously so that there would only be one discontinuity in the series. It is necessary to adopt the SIC 2007 classification because the ongoing quality of the 2003 classification on the IDBR, so long as dual coding is in place, is becoming progressively less reliable over time as it is no longer being maintained as rigorously as SIC 2007. **SCSS 2011 is the first time that all data is published on a SIC 2007 basis only.**

Detailed information about the impacts of moving from SIC 2003 to SIC 2007 on **registered enterprises** can be found in the information note "[Standard Industrial Classification \(SIC\) 2007 – Assessing the Impact on Registered Enterprises](#)" on our website.

With regard to **unregistered enterprises**, the main impact of moving to SIC 2007 is in terms of the Labour Force Survey (LFS) data on self-employment. This is used to provide an estimate of unregistered enterprises by broad industrial sector. The LFS has been coded on a SIC 2007 basis since 2009. While SCSS still used SIC 2003 it was necessary to use LFS data that had been converted back from SIC 2007 to SIC 2003. For SCSS 2010, issues with this conversion process meant that it was not possible to provide a sectoral breakdown for total enterprises on a SIC 2003 basis. Publishing SCSS 2011 on a SIC 2007 basis ensures that the data is of greater quality and it is possible to provide a sectoral breakdown of total enterprises. SCSS 2011 includes a sectoral breakdown for total enterprises for 2010 and 2011 on a SIC 2007 basis.

Overview of Old Unregistered Enterprises Methodology

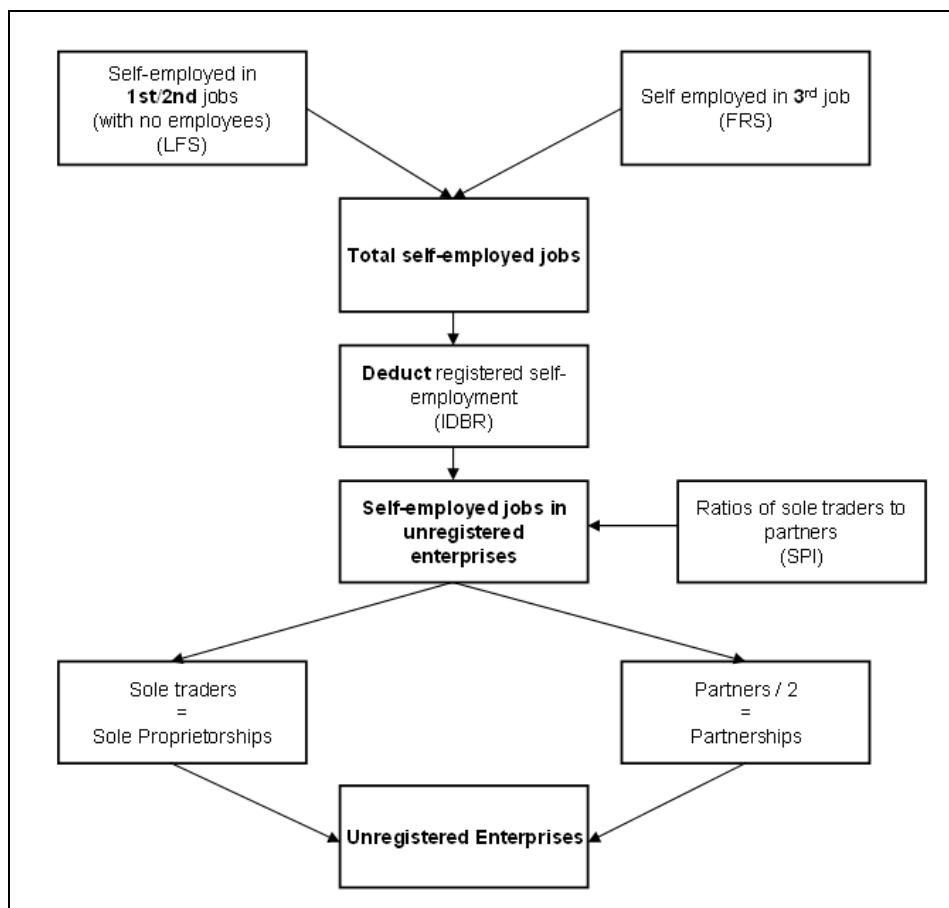
There are two main components to the data on enterprises contained in SCSS – registered and unregistered enterprises.

Data on registered enterprises is obtained via a Scottish extract of the IDBR, which is a database of all enterprises operating in the UK that are registered for VAT and/or PAYE. It is maintained by the Office for National Statistics and covers 99% of economic activity in the UK.

A large number of very small firms are excluded from the IDBR because they are small soletraders/partnerships with no employees and an annual turnover of less than the VAT threshold (£70,000 as at March 2011). The number of unregistered enterprises is estimated using several different data sources.

The [Labour Force Survey](#) (LFS) provides data on the number of people who are self-employed in their first or second jobs (with no employees). The figure for second jobs is augmented by data from the [Family Resources Survey](#) (FRS) on people who are self-employed in a third job, to obtain an estimate of the total number of self-employed jobs. This estimate is compared to the level of self-employment that is already registered on the IDBR in enterprises with no employees. As many self-employed people are not required to pay VAT or register for PAYE, the figure from the LFS/FRS is generally higher. The difference between these two figures (self-employed jobs in unregistered enterprises) is the starting point for estimating the number of additional sole traders and partners with no employees.

Data from the HMRC [Survey of Personal Incomes](#) (SPI) were used to obtain the ratio of self-employed people working alone to those working in partnership, which was then used to derive the number of unregistered sole traders and partnerships in each industry. It was assumed that employment is one in a sole proprietorship and two in a partnership.



The following sections describe the changes to this methodology that have been implemented in Scottish Corporate Sector Statistics 2011. Note that comparisons are based on the **original 2010 data** published on 28 October 2010. The 2010 data was based on SIC 2003 and the LFS data has since been reweighted. These comparisons are based on the old weights. **Therefore the figures obtained for 2010**

unregistered enterprises, under the new methodology, in this document are not the same as those published for 2010 in SCSS 2011.

(2) (a) Self-Employment Reclassification

It has been recognised that there is a degree of over-counting of the self-employed population obtained from the LFS^{1,2} because survey respondents self-report their employment status, which can lead to misclassification.

Reclassification of Main Job Employment Status

Since Q1 2007, the LFS has contained a question which asks both (i) people who are self-employed in their **main job** and (ii) employees **not** paid a wage or salary by their employer, who it was that paid their National Insurance and Income Tax. People who pay their own National Insurance and Income Tax are legally considered to be self-employed and should be registered as self-employed with HMRC.

The question on who pays NI and Tax can be combined with the question asking respondents who paid their wage or salary to obtain a better estimate of 'true' self-employment. If the following conditions are met, then an individual is considered to be self-employed (regardless of whether their self-reported status is self-employed or employee):

- (a) respondent pays their own National Insurance and/or Income Tax **and** is not a sole director of a limited company; **or**
- (b) respondent does not answer the question about National Insurance and Income Tax **and** does not answer the question about who pays their wage or salary. (In which case it is necessary to proceed with respondent's initial self-reported employment status – self-employed.)

Sole directors of limited companies should always be classed as employees rather than self-employed, as should those who respond that National Insurance and Income Tax are deducted by their organisation.

The combination of responses to these questions can be used to reclassify those who have mis-categorised themselves, to obtain a 'true' estimate of self-employed and employees. Diagram 1 illustrates the reclassification process. Due to small numbers in part (b) of the diagram (approximately 1,000 in total) the cells have been left blank, as the main purpose of the diagram is to illustrate the methodology.

Part (a) represents those whose self-reported employment status was self-employed (with no employees). The green cells represent those respondents who fall within

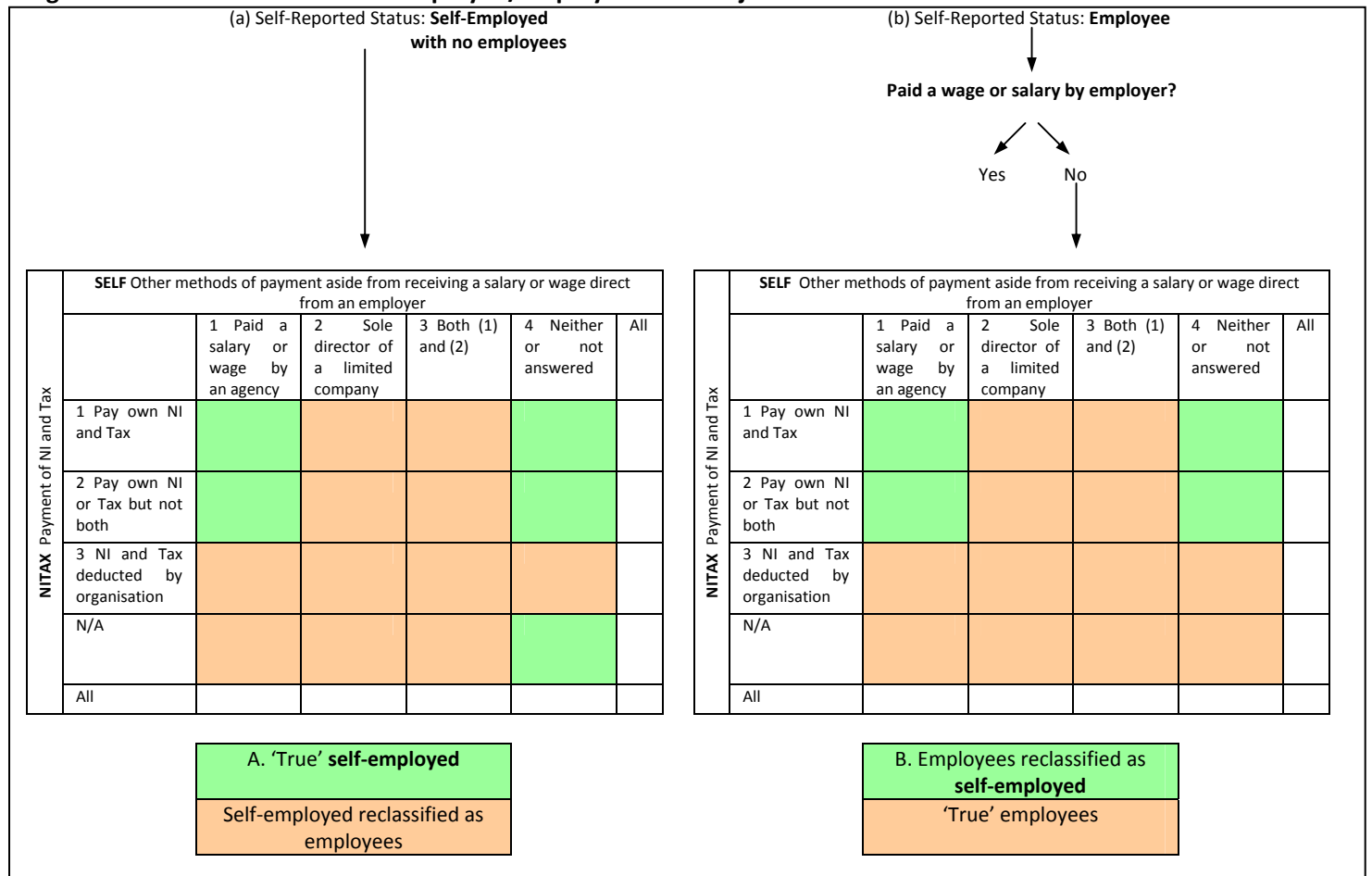
¹ <http://www.ons.gov.uk/ons/rel/elmr/economic-and-labour-market-review/april-2011/index.html>

² <http://www.ons.gov.uk/ons/rel/elmr/economic-and-labour-market-review/no--3--march-2008/economic---labour-market-review.pdf>

the definition of self-employment described above and are therefore ‘truly’ self-employed whilst the orange cells represent those who are actually employees and should be reclassified as such. Part (b) represents those whose self-reported employment status was employee. In this case, the orange cells represent respondents who have correctly classified themselves as employees whereas the green cells represent those who fall within the definition of self-employment above – and who should be reclassified as self-employed. Note that the assumption is being made that employees being reclassified as self-employed do not have employees. There is no way to verify the accuracy of this assumption but, given the small numbers involved, it is unlikely to have a major impact.

Under the new methodology, those who are ‘truly’ self-employed in part (a) are added to those employees in part (b) who should be reclassified as self-employed, to obtain a total estimate of ‘true’ self-employment in main job.

Diagram 1: Reclassification of self-employed/employees in main job



Reclassification of Second Job Employment Status

The LFS variables used to reclassify main job employment status are not available for those with second jobs so another method must be used to obtain a figure for ‘true’ self-employment in second jobs.

This involves calculating the ratio of ‘true’ self-employment to ‘self-reported’ self-employment in main job and applying this ratio to the self-reported data on self-employment in second jobs.

Impact of Improved LFS Methodology on Self Employed Jobs

Table 1: Impact of improved LFS methodology on self-employed jobs³

	Self-Employed in Main Job	Self-Employed in Second Job	Total Self-Employed
Old Method	195,600	27,600	223,200
New Method	169,600	24,300	193,900
Change	-26,000	-3,300	-29,300
% Change	-13.3%	-12.0%	-13.1%

Based on the original methodology, used in SCSS 2010, the total number of self-employed was 223,200. Using the new methodology, the total number of self-employed decreases by 13% to 193,900 (driven by sole directors of limited companies being reclassified as employees).

Impact of Improved LFS Methodology on Unregistered Enterprises

The number of unregistered enterprise jobs is derived by subtracting self-employed jobs in registered enterprises (from the IDBR) from the total self employed jobs from the LFS. Under the original methodology, the number of self-employed jobs in registered enterprise was defined as the employment of those enterprises on the IDBR with zero employees. However, this included directors of limited companies who, it has been established, should actually be classed as employees. Therefore, to coincide with the improved LFS data, self-employed jobs in registered enterprises are now defined as sole traders/partnerships with zero employees. The impact of this new definition is shown in Table 2 below:

Table 2: Impact of new definition of self-employed jobs in registered enterprises

	Registered Enterprise Jobs
Old Method All enterprises with zero employees	64,715
New Method Sole traders/partnerships with zero employees	36,405
Change	-28,310
% Change	-43.7%

Table 2 shows that there is a marked reduction in the number of registered enterprise jobs under the new definition. This means that a smaller number is being

³ Note that this comparison is based on using the new LFS methodology only - it does not take account of any of the other methodological changes outlined in this document.

subtracted from total self-employment from the LFS. The combined effect of these changes is shown below in Table 3.

Table 3: Impact of both changes on unregistered enterprises

	Unregistered Enterprises
Old Method	143,325
New Method	140,840
Change	-2,485
% Change	-1.7%

While the new LFS method of obtaining ‘true’ self-employment leads to a substantial reduction in total self-employment, the modified definition of registered enterprise jobs counteracts this to some extent, leading to a modest decrease (1.7%) in the number of unregistered enterprises.

(2) (b) HMRC Self Assessment Data

Previous versions of Scottish Corporate Sector Statistics used data from the HMRC Survey of Personal Incomes (SPI) to estimate the ratio between sole traders and partners by industrial sector. These ratios were then used in conjunction with self employment data from the LFS to estimate the number of unregistered sole traders and partnerships.

The 2007/08 SPI data used in SCSS 2010 contained more stringent statistical disclosure control measures which meant that it was not possible to obtain ratios for some industrial sectors and older data had to be used for these instead. Furthermore, delays were expected in terms of the 2008/09 SPI data.

Consequently, it was decided to make use of HMRC Self Assessment data. These data are based on a complete record of all self-employed people in the UK, unlike the SPI which was a sample survey. Also, the Self Assessment dataset provides more detail as it includes a marker indicating whether the turnover generated by each business is above/below the VAT threshold. This marker enables the data to be used as a proxy for non-VAT registered businesses. It is important to emphasise that this is a proxy, as there will be self employed people with turnover below the VAT threshold who are registered for VAT. The presence of the marker, however, enables ratios to be calculated using only those enterprises whose turnover is below the VAT threshold, whereas ratios calculated from the SPI were based on the whole sample.

Another advantage of the Self Assessment dataset is that it contains information about the number of partners in each partnership. Whereas previous versions of SCSS assumed that there were two partners per partnership, the Self Assessment data enables more accurate estimates to be made, based on the average number of partners per partnership for each sector.

The main drawback of the SA data is that, for approximately 15 per cent of the records, the industrial classification is missing. However, the benefits, as outlined above, outweigh this issue.

Impact of Self Assessment Data

Table 4 shows that the number of unregistered enterprises in SCSS 2010 (based on SPI data) was 143,325.

Table 4: Comparison of Results Using SA and SPI data⁴

Method	Unregistered Enterprises	% Change from SCSS 2010
SCSS 2010 (SPI data)	143,325	N/A
(i) Using SA data and 2 partners per partnership	148,865	+ 3.9%
(i) and (ii) Using SA data & improved average no. partners per partnership	147,015	+ 2.6%

Using Self Assessment data, the number of unregistered enterprises increases by 2.6% to 147,015. This increase is caused by changes in the methodology relating to the number of partners per partnership. The number of partners has changed in two ways:

(i) Using SA data makes it possible to approximately exclude people who are self-employed in registered enterprises. Partnerships with a greater number of partners are more likely to be registered enterprises, therefore by excluding these, the number of partners decreases. As the number of partners decreases, so the number of partnerships increases.

(ii) Using SA data makes it possible to improve the assumption that there are two partners per partnership, as data is available on the average number of partners per partnership. This means that for all SIC codes, the average number of partners has increased to more than two. This effect in isolation would result in the number of partnerships decreasing, as the number of partners is divided by a greater average number of partners.

The impact of each of these components is shown in Table 4. This shows that effect (i) (ability to exclude 'registered' enterprises) makes the greatest contribution to the overall change in the number of unregistered enterprises. Whilst improving the average number of partners per partnership alone would decrease the number of partnerships, when implemented in conjunction with (i), Table 4 shows that there is still an increase in unregistered enterprises of 2.6%.

⁴ Note that this comparison is based on using SA data only - it does not take account of any of the other methodological changes outlined in this document.

Overall Impact of Methodology Changes

Table 5 below shows the combined impact on the number of unregistered enterprises of all the methodological changes (except in relation to SIC 2007) outlined in this document.

Table 5: Summary of effects of methodological changes to unregistered methodology

	Unregistered Enterprises	Change	% Change
Data published in 2010	143,325	N/A	N/A
New LFS methodology incl IDBR registered changes (in isolation)	140,840	-2,485	-1.7%
HMRC SA data (in isolation)	147,015	+3,690	+2.6%
Combined impact of both changes	144,205	+880	+0.6%

When all changes are made, the number of unregistered enterprises in 2010 increases by 0.6% to 144,205. Note that this is **not** the same as the business enterprise count for 2010 published in SCSS 2011 under the new methodology, **since the 2010 unregistered count (contained in SCSS 2011) is based on the SIC 2007 classification and reweighted LFS data.**

Data on unregistered enterprises, based on the new methodology, are provided for 2010 and 2011 in SCSS 2011. Consequently, there is a break in the time series **for the number of unregistered enterprises** (and hence total enterprises) in Scotland. As these changes only affect unregistered enterprises, differences in the registered enterprise counts reflect only the impact of moving from SIC 2003 to SIC 2007 (as detailed in "[Standard Industrial Classification \(SIC\) 2007 – Assessing the Impact on Registered Enterprises](#)"). Registered enterprise counts on a SIC 2007 basis are available in SCSS 2011 from 2008 to 2011.