LABOUR FORCE SURVEY STEERING GROUP

15 October 2010

Implementation of SIC 2007 in the Labour Force Survey

Note by Nick Palmer

Introduction

This paper presents some of the main findings from the extensive analysis the ONS has carried out into the effects of the implementation of SIC 2007 and a new coding tool in the LFS. The findings produce some important messages for users of time series of LFS estimates by industry and some decisions to be made regarding the LFS industry coding in the future.

Background

The industry class to which people in employment are coded in the LFS switched to SIC 2007 in January 2009. From that point onwards, all cases were coded to a new code, even when the respondent's situation had not changed, due to the introduction of SIC 2007. It was decided not to code each case in the survey to both SIC 1992 and SIC 2007 due to potential confusion amongst interviewers, i.e. no dual coding.

The transition to the new classification was accompanied by the implementation of a new automatic coding tool in the LFS questionnaire. A similar tool had been in place for the coding of occupations to the Standard Occupational Classification (SOC) for several years, but prior to 2009, industry had been coded manually, using a paper-based SIC volume. The new tool is seen by interviewers as a significant improvement in the coding of industry in the LFS. Its introduction was designed to bring increased accuracy and consistency to the industrial coding. Greater consistency was envisaged because cases with the same description are more likely to be allocated the same code with the coding tool than with the previous approach, which was subject to individual interviewers' own interpretations.

Mapping between SIC 2007 and SIC 1992

Quarterly LFS microdata for Q1 2009 onwards have been coded to SIC 2007 and contain conversion variables that map SIC 2007 classes (4 or 5 digit categories) to SIC 1992 divisions (2 digit categories). This is continuing indefinitely for all quarterly LFS datasets. A similar approach was taken for the Longitudinal datasets and for the Annual Population Survey (APS).

Historical LFS data sets back to Q1 2006 contain conversion variables that map the SIC 1992 classes (as per the collected data) to SIC 2007 divisions using ONS guidance about the assumed relationship between the two classifications. This ONS guidance is broadly consistent with the mappings assumed for ONS business surveys, although there are some simplifications.

For the historical microdata, a mapping was constructed from the lowest level SIC 1992 codes (industry classes) to SIC 2007 at the two digit level (industry divisions). The mapping was set up by assigning each of the SIC 1992 classes to a SIC 2007 division, depending on where it was assumed the people employed in that class would be coded under SIC 2007. This was relatively straightforward for most classes. However there were a number of SIC 1992 classes which were separated into several more detailed classes under SIC 2007 which aggregated to different SIC 2007 divisions. Where this occurred, the mapping had to be simplified since it was not practical to do any kind of apportionment. The simplification chosen was to map every person in that class to the SIC 2007

division to which the majority of people were estimated to belong, according to estimates derived from the Inter Departmental Business Register (IDBR).

An example of such classifications is as follows:

> Under SIC92, class 01.41 (Agricultural service activities including landscape gardening) mapped into 4 classes under SIC 2007, i.e.
>
> - 01.61 – Support activities for crop production
> - 01.63 – Post-harvest crop activities
> - 01.64 – Seed processing for propagation
> - 81.30 – Landscape service activities
>
> According to IDBR estimates, 30% of this class mapped into 01.61 under SIC 2007 and 70% to 81.30. Consequently, under the simplified approach adopted for the LFS, all of the class was mapped to 81.30.

The mapping from Q1 2009 onwards which maps the SIC 2007 classes to SIC 1992 divisions contains fewer simplifications because there are 26 more divisions under the newer classification. In other words, there are not so many classes under SIC 2007 which map to more than one division under SIC 1992.

Spreadsheets detailing the mappings are available to users, along with the syntax to carry out the mapping in SPSS datasets.

Impact of SIC 2007 implementation on LFS estimates by industry

Extensive analysis has been carried out, aiming to assess:

- the strengths and weaknesses of the mappings;
- the impact of the new classification itself;
- the impact of the change from the manually coded method to the automatic coding tool;
- the extent of any step-changes in the time series of aggregates on an industry basis.

The impact of the changes were much greater than originally envisaged and have required a significant amount of additional analysis to understand them. This paper just presents some of the key aspects of the analysis and the related findings.

The analysis was based primarily on comparisons of the recorded and mapped SIC codes for people surveyed in both Q4 2008 and Q1 2009 who reported that they had remained in the same job (referred to as “linked records”). The analysis exploits the panel nature of the LFS which meant that four-fifths of the total sample interviewed in Q1 2009 (when SIC 2007 was first used) had been previously interviewed in Q4 2008 (the last occasion when SIC 1992 was used). For those respondents who had not changed their job between the two interviews, the SIC 1992 codes from first quarter could be compared with the SIC 2007 code from the second. Two main sets of analysis were worked on:

- The mapped SIC 07 divisional codes for Q4 2008 compared with the recorded SIC 2007 divisional codes for Q1 2009); and
- The recorded SIC92 divisions for Q4 2008 compared with the mapped SIC92 divisions for Q1 2009

The analysis was carried out with unweighted data as well as with weighted estimates. It involved detailed investigation of the reasons behind the transitions that emerged and identification of whether they were caused by the change in classification, weaknesses in the mapping or effects of the new coding tool. The investigative work included obtaining the written descriptions of workplace activity recorded during the LFS interviews and re-running them through the coding tool, comparing the outcomes with those obtained by the interviewers.

Table 1 in the Annex to this paper compares the number of (unweighted) cases mapped to each SIC07 Section in Q4 2008 with the number coded to that Section in Q1 2009. The net change is given

along with those in the same Section in each quarter and also the gross “inflows” and “outflows”. (The “inflow” is the number who were mapped to a different Section in Q4 2008 and the “outflow” is the number that what was mapped to the Section in Q4 but is no longer in it in Q1.)

Similar analysis was done at Division level and on a SIC92 basis. The results will also be made available to users.

Chart 1 in the Annex shows the weighted estimates of total employment in selected SIC 2007 Sections derived from the LFS microdata, consistent with estimates published in Table 9 of the LFS Historical Quarterly Supplement. The Sections selected for the chart are those that exhibit the most significant changes over the transition period.

Findings: the mappings

The analysis revealed a few small errors in the mapping from SIC 2007 to SIC 92, as follows:

	Original mapping	**Revised mapping**
SIC2007 class/subclass	**Division**	**Division**
38210 waste treatment and disposal	14 Other mining and quarrying	90 Sanitation and sewerage
62090 other information technology and computer activities	30 Office machinery and computer manufacrture	72 Computer related activities
71200 Technical testing and analysis	75 Public admin	74 Other business activities
91011 Library and archive activities	75 Public admin	92 Recreation and cultural activities

The conversion variables in post-2008 datasets were subsequently corrected. The impacts were not significant.

Findings: the estimates

For the employment time series on a SIC 2007 basis there are a number of step changes at Section level between Q4 2008 and Q1 2009. These have been mainly caused by the introduction of the new coding tool. The limitations in the mapping also contributed, but to a much lesser extent. Step changes are also evident when producing time series on a SIC92 basis and are also apparent for many Divisions on both bases.

In addition, some of the assumed effects of the new classification, did not materialise in the way expected, partly because of the change in coding tool.

Some of the main step changes on a SIC 2007 basis can be explained as follows:

- Sections M and N (Professional, Scientific and Technical activities and Administrative & Support Services) gained significantly from people formerly coded to: Manufacturing; Construction; Information and Communication (J); Recreational (R); and Activities of households as employers.
- Manufacturing (Section C) has decreased in size, in particular losing people to various services within Sections M and N: and also to Distribution (G).
- Education (Section P) has gained mostly from people formerly coded to: Public Admin (O); Arts, Entertainment & Recreation (R) and Health & Social Work (Q).
- Section T (Households as Employers) reduced by 55%, mostly losing people to services in Section N.

Findings: the automatic coding tool

The comparisons of the industry descriptions given during the interview with the SIC07 codes recommended by the automatic coding tool in Q1 2009 and with the SIC92 codes assigned in the previous quarter revealed some unexpected results. For example, under the new coding tool, anybody who reports their industry as "Telecommunications" (or even "B.T.") is now coded to the Construction sector. Prior to 2009, the majority were coded to the "Telecommunications" sector. The key determinant here, and always has been, is that those whose dominant activity is the installation of telecommunications systems should be in Construction and those that maintain such systems should

be in Telecommunications. In theory, the knowledge base of the automatic coding tool could be enhanced to help differentiate correctly between the two sectors. Other cases were identified where the coding tool introduced an unrealistic or incorrect bias into the coding.

Any future enhancements to the coding tool should of course be managed carefully with all major changes documented along with an assessment of their impact on LFS time series by industry.

Conclusions and future actions

The implementation of the new coding tool into the LFS introduced many discontinuities at Q1 2009 for estimates by industry at both Section level and Division level. The majority of these discontinuities represent an improvement in the accuracy and consistency of the coding, correcting some weaknesses associated the previous coding method. Some simplifications in the mapping between the two SICs have also contributed to the discontinuities, but these effects are quite small.

Some weaknesses are evident in the new automatic coding tool and the knowledge base it uses needs to be enhanced.

The industry descriptions recorded during the interviews often contain in sufficient detail to code accurately and consistently. This has always been a difficulty but has been exacerbated by the increased detail required by the new classification.

Users need to be aware of step changes in LFS estimates by industry. A simple method to adjust out the step changes can be applied by using the figures in Table 1 to rescale the historical estimates (on a SIC 2007 basis). The results of this are shown in Chart 2 in the Annex. This could also be done at the Division level of course and a similar approach can be taken to rescale the post-2008 estimates when looking at time series on a SIC92 basis.

<u>Action for LFS Steering Group Members</u>

Members are requested to comment on the findings from the analysis and their views are sought on:

- **how to publicise the findings among users**
- **whether the time series published in the HQS and elsewhere should be adjusted to eliminate the step changes**
- **how to take forward the future enhancements to the knowledge base for the coding tool.**

ANNEX

TABLE 1

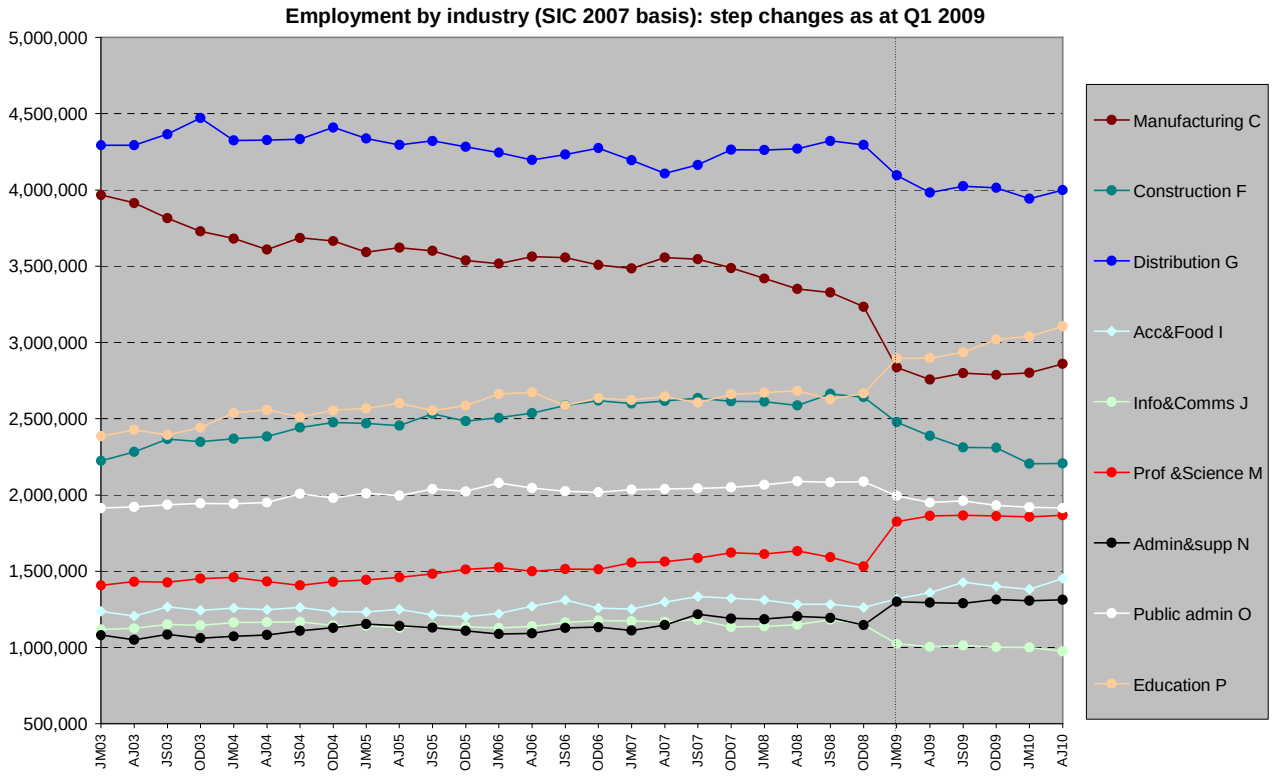
TRANSITION TO SIC 2007
SIC2007 Section in Q1 2009 compared with mapped SIC 2007 Section in Q4 2008

LFS JM09OD08_common

Filter: (inecac05_OD08 >=1 & inecac05_OD08 <=3) & (iout_OD08 = 1 or iout_OD08 = 2)

	SIC 2007 Section	Total Q4 08 (mapped)	Total Q1 09 (recorded)	Change Q4 08 to Q1 09	Change as %	In the Section in both periods	In the Section in both periods as % of Q1 09 total	Inflow	Outflow
A	Agriculture, forestry and fishing	402	392	-10	-2%	321	82%	71	81
B	Mining & quarrying	152	127	-25	-16%	87	69%	40	65
C	Manufacturing	3,899	3,549	-350	-9%	3,011	85%	538	888
D	Electricity, gas & air con supply	203	191	-12	-6%	142	74%	49	61
E	Water supply, sewerage & waste	282	275	-7	-2%	199	72%	76	83
F	Construction	3,020	2,894	-126	-4%	2,404	83%	490	616
G	Distribution	5,028	4,921	-107	-2%	4,347	88%	574	681
H	Transport & storage	1,920	1,881	-39	-2%	1,666	89%	215	254
I	Accommodation & food services	1,340	1,416	76	6%	1,243	88%	173	97
J	Information & communication	1,308	1,164	-144	-11%	942	81%	222	366
K	Financial & insurance services	1,372	1,336	-36	-3%	1,206	90%	130	166
L	Real estate services	288	328	40	14%	196	60%	132	92
M	Prof, scientific & technical activities	1,779	2,128	349	20%	1,288	61%	840	491
N	Admin & suppport services	1,291	1,503	212	16%	913	61%	590	378
O	Public admin & defence	2,658	2,569	-89	-3%	1,877	73%	692	781
P	Education	3,448	3,677	229	7%	3,232	88%	445	216
Q	Health & social work	4,683	4,741	58	1%	4,213	89%	528	470
R	Arts, entertainment & recreation	889	831	-58	-7%	686	83%	145	203
S	Other service activities	833	919	86	10%	606	66%	313	227
T	Households as employers	161	72	-89	-55%	48	67%	24	113
U	Extra territorial	10	52	42	420%	6	12%	46	4
	Total	34,966	34,966	0	0%	28,633	82%	6,333	6,333

Chart 1



ANNEX (cont)

Chart 2