7 th WORKSHOP ON LABOUR FORCE SURVEY METHODOLOGY

DATA PROCESSING AND DATA QUALITY

Madrid, Spain, 10 – 11 May 2012

C. Data processing: Methods for editing and imputation, weighting, non-response adjustment

C2– The Danish experience of revising the weight

Michael Frosch, Sammy Lauritsen– Denmark

The impact of the 2011 weighting method on the main figures of the Labour Force Survey

Abstract The Labour Force Survey (LFS) is a sample-based survey. Because of this, the results from the survey are weighted in order to correspond to the population. However, it is necessary to regularly review the basis of weighting for the LFS results. Particularly due to the circumstance that both the underlying register-based information and the explanatory power in a given weighting model change over time. Consequently, Statistics Denmark's LFS section has in cooperation with the division for methodology reviewed the basis of how the LFS results are weighted, and in this context it was concluded that there was room for improvement.

For example, it appeared from the review that the former definition of age was inconsistent and not sufficiently exact in its definition of the population. Consequently, a more exact age definition was introduced in connection with the new weighting method, which was implemented from the 3rd quarter of 2011.

The review of the weighting method also revealed that there was a possibility to compensate further with regard to the non-response by adjusting the auxiliary information. This was concluded on the basis of a review conducted by the division for methodology of about 40 potential variables, which could be used as auxiliary information. There is naturally a limitation as to the amount of auxiliary information that can be used in a given model, but the analysis of the division for methodology showed that it was possible to enhance the model's explanatory power. Therefore, the new model not only contains partly new auxiliary information, but also an adjustment of the auxiliary information that was already incorporated in the 2007 weighting method.

Introduction Statistics Denmark has a wide variety of high-quality register information about the Danish population. This enables the LFS unit to do in-depth analysis of non-respondents, and further try to amend this by creating detailed weighting schemes. This is also very important as the non-response is very high in Denmark – above 45 pct.

General changes – In the 2011 weighting method, a more exact age definition was introduced. The former age classification was inconsistent and not sufficiently precise with regard to the definition of the population. Furthermore, the 2011 weighting method was adjusted in such a manner that it could measure the age group 15-64 years, and with regard to StatBank Denmark from the 3rd quarter 2011 this age definition is used, instead of the 15-66 year-olds. The new age definition corresponds to that used by Eurostat.

– In the 2007 weighting method, auxiliary information was created by Statistics Denmark's division for methodology on the basis of data extracts from different registers. In the 2011 weighting method, register-based information is drawn directly from Statistics Denmark's database for personal statistics (PSD). It is thereby ensured that the most recent data are always being used. The five main data sources are the Population register, Personal income register, Register-based labour force statistics, Register of education and training statistics and the Registered unemployment.

	– The new 2011 weighting method replaces the 2007 weighting method. This implies that the figures in the LFS are revised on the basis of the new weighting method as from the 1st quarter of 2007.
New auxiliary information	New auxiliary information has been added to the 2011 weighting method.

– The number of children in the household is added as auxiliary information, as it has been proven that this is highly correlated with the non-response.

– The socio-economic status in the 2011 weighting method complies with the classification used in the register-based labour force statistics as this definition is more closely related to the definitions used by the International Labour Organization (ILO). The former classification was created by Statistics Denmark's Division for Welfare Statistics.

– In the new model, age is crossed by level of education. A more precise estimation of the educational level analysed by age groups is thereby achieved.

New grouping The age classification in the 2011 weighting method differs slightly from that used in the 2007 weighting method. The purpose of the previous 12-grouping was to fix a bias among the 15-year olds. This is no longer necessary, as the selection of the sample has been improved. The selection is now conducted, using an age definition based on the reference quarter. In this way, the classification can be limited to 11 groups. Furthermore, the grouping is created in such a way that there is consistency between the age grouping used in the weighting method and the grouping used in publishing the statistics.

The classification of auxiliary information concerning citizenship has been refined compared to the former model, and at the same time the categorisation of education has been adapted to the international ISCED classification.

Effects on the results For the purpose of looking at the impact of the changes described above, data are in the following weighted on the basis of the 2007 and 2011 weighting methods in order to enable a comparison of the main estimates.

Labour market attachment The general picture is that the new weighting method only has a limited impact on the labour market attachment at aggregated level for the 15-64 year-olds. It appears that this applies to the entire period 2007-2011.

In the latest quarter, both employment and unemployment is slightly lower when the new weighting method is applied. Consequently there is also a fall in the level of the labour force.

Employment in the 1st, 2nd, 3rd and 4th quarters of 2010 and the 2nd and 3rd quarters of 2011, are slightly lower with the new weights. The employment in the 1st quarter 2011 is a bit higher with the 2011-weights and the difference is somewhat higher in the 2nd quarter 2010, which is probably due to the problems with the 2007 weighting method in this quarter.

Applying the new weighting method, the level of unemployment is slightly lower – except in the 2nd quarter 2010 and the 2nd quarter 2011. The level of the group outside the labour force shows the greatest fluctuations, for three quarters the level is higher and for four quarters the level is lower when the new weighting method is applied.

	2010				2011		
	Q1	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**
Employed (2007 weights)	2,643,384	2,686,084	2,672,583	2,649,503	2,623,777	2,659,210	2,664,656
Employed (2011 weights)	2,642,842	2,664,033	2,669,180	2,646,928	2,625,164	2,652,090	2,664,207
Difference	**-542**	**-22,051**	**-3,403**	**-2,575**	**1,387**	**-7,120**	**-449**
Unemployed (2007 weights)	229,265	215,620	212,545	213,988	239,594	212,826	223,774
Unemployed (2011 weights)	228,403	219,559	212,180	211,239	235,005	213,741	219,296
Difference	**-862**	**3,940**	**-365**	**-2,750**	**-4,589**	**915**	**-4,478**
Outside labour force (2007 weights)	752,861	730,150	736,788	770,533	769,295	746,133	717,741
Outside labour force (2011 weights)	748,815	736,544	733,960	767,663	757,038	750,127	727,657
Difference	**-4,046**	**6,394**	**-2,829**	**-2,869**	**-12,257**	**3,994**	**9,916**

Impact from the more precise age classification

The difference in the labour market attachment between the 2007 and 2011 weighting methods is more pronounced when the attachment is split in age groups. It is likely that the age impact, among other factors, is a result of the more precise age classification, which creates a result that is more closely linked to the population recorded in the Population register. The clarification of the age definition will re-group persons between the different age groups. For example, the total number of persons in the age group 15-24 years has decreased by 13,000 persons in the 3rd quarter 2011, which results in a lower level of employed, unemployed and persons outside the labour force within this age delimitation.

Furthermore, the 2011 weighting method focuses, unlike the former method, on a target population aged 15-64 years. This may give rise to further differences, when comparison is made with the previous weighting method, as the 2007 weighting method focused on the age group 15-66 years.

Sub-classification of unemployment

At the general level, the unemployment level is only slightly changed by the new weighting method. The difference between unemployment according to the two weighting methods continues to be limited, when the unemployment is sub-classified into registered unemployed, persons participating in activation programmes, students and other unemployed persons.

Table 2. Difference in sub-classified unemployment between the 2011 and the 2007 weighting method (2011 minus 2007)

	2010				2011		
	Q1	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**
Registered unemployed	-3,054	1,612	-866	-5,505	-2,085	3,263	-3,504
Pers. in act. prog.	1,116	297	-320	797	-2,644	-402	-821
Students	976	13	-2,432	882	-1,024	-1,522	-1,068
Other unemployed	100	2,018	3,253	1,077	1,985	-423	915

Full- and part-time employment

The impacts on employment are somewhat more obvious, when employment is sub-classified by full- and part-time employed. Applying the new weighting method, part-time employment is generally somewhat lower, with three quarters in 2008 and four quarters in 2009, as a remarkable exception. Full-time employment on the other hand is generally increased.

Within recent years, there is thus, except for the 2nd quarter 2010 and the 2nd quarter 2011, a transfer in employment from part-time to full-time when the new

weighting method is applied. There is a clear relationship between the number of children and full-time employed, for which adjustments are made by the new weighting method.

Table 3. Difference in full- and part-time employment between the 2011 and 2007 weighting method (2011 minus 2007)

	2010				2011		
	Q1	Q2	Q3	Q4	Q1	Q2	Q3
Full-time	9,236	-13,586	2,805	5,960	15,158	-250	10,553
Part-time	-9,779	-8,472	-6,208	-8,535	-13,771	-6,871	-11,002

More self-employed

Furthermore, the level for the number of self-employed increases in the Labour Force Survey after the 2nd quarter 2008 when the new weighting method is applied. The level for employees is slightly lowered in most quarters. It is only in the 2nd quarter 2010 that the difference is pronounced, which is probably due to the problems with the 2007 weighting method in this quarter.

Impact changes in the weighting of education

In the 2011 weighting method, age is crossed by level of education. This appears, among other things, to have an impact on employment divided by age, so that employment is lowered among younger people, while employment is raised among older people.

The overall impact of the changes in the weighting of level of education is pronounced in terms of absolute figures, which is primarily due to changes in the delimitation of the population. However, the impact is limited, when shares are analysed. The group of tertiary education is slightly smaller for all quarters when the new weighting method is applied, whereas the groups of primary and secondary education as the highest completed level of education are slightly higher.

Table 4. Difference in shares, highest completed level of education between the 2011 and 2007 weighting methods (2011 minus 2007)

	2010				2011		
	Q1	Q2	Q3	Q4	Q1	Q2	Q3
Primary, incl. unknown	0.19	0.12	0.28	0.53	0.35	0.26	0.17
Secondary	0.39	0.51	0.40	0.05	0.19	0.48	0.43
Tertiary	-0.59	-0.64	-0.67	-0.58	-0.54	-0.74	-0.61

The impact is more obvious when the highest completed level of education is crossed by age. The share with tertiary education is lowered for young people, whereas it increases for older people.

The 30-34 year-olds are of special interest, as they are incorporated in the Lisbon 2020 target of the EU. This objective of the EU target is that at least 40 pct. in this age grops must have tertiary education. On the basis of the former weighting method, Denmark was well above this target for all 4 quarters in 2010 and in the first three quarters in 2011.

Table 5. Share of 30-34 year-olds with tertiary education, 2011 and 2007 weighting methods, persons with an unknown level of educaton are excluded in order to align the method to the one used in OECD's Education at a Glance

30-34 year-olds		2010				2011		
		Q1	Q2	Q3	Q4	Q1	Q2	Q3
2011 weighting method	Share with tertiary education	41.17	42.07	41.54	40.08	40.24	40.34	42.06
2007 weighting method	Share with tertiary education	45.52	48.25	48.29	45.27	43.60	43.36	46.90

When the new weighting method is applied, the share of persons with tertiary education aged 30-34 years is, on average, lowered by 5 percentage points for all quarters in 2010 and 2011. This implies that in the 3rd quarter of 2011, e.g., the share in Denmark declines from 46.9 pct. of persons with tertiary education to 42.1 pct. in this age group.

The reason for this is again that the educational level is split by age in the new weighting method. Thereby the level of education in the LFS, analysed by age, approaches the lower level in the educational register.

Conclusions and lessons learned

Concerning the results the conclusion seems to be that the overall effects were limited, as was also expected. The main effect is on the educational level – which is also the place where the main changes to the model were implemented.

Another important lesson learned was the communication with the users. During the entire process internal and external users were kept informed of the prospects and developments. This pro-active involvement from the side of Statistics Denmark made sure there were very few misunderstandings and quite little work informing users after the implementation.