Comparing Subsample Approaches

Ole Villund, Statistics Norway

Presentation to the 9th Workshop on Labour Force Survey Methodology Rome, 15th – 16th May 2014

TOPIC

Survey design and weighting – Effective data collection designs

Sub-sampling of annual modules using a wave approach

In a modular EU LFS the wave approach is used for sub-sampling. Sub-sampling reduces the respondent burden and increases flexibility. However, data collection and processing become more complicated and demanding. We are looking for positive experiences with this approach and successful ways of dealing with the obstacles or processing complications.

ABSTRACT

The purpose of this paper is to discuss some practical and methodological issues regarding subsampling. The Norwegian Labour Force Survey has used three different subsampling approaches, and gained some practical and methodological results from that experience.

Practical considerations include IT management, interviewer workload and respondent burden. In a CATI-only setting, with flexible IT-systems and a capable interviewer staff, the wave approach seems like the best solution. With paper forms or more arbitrary IT-solutions, subsampling by a period in time may be more practical. A well-organized wave approach makes the interviewer workload more even. Perhaps it also improves measurement quality by balancing the experience with multiple questionnaires.

Any subsampling can introduce selection- or nonresponse effects. For instance, seasonal variation in response could affect the representativity of a period-based subsample. Generally, a wave approach evens out seasonal effects. However, time-approach could be better for some topics regarding specific groups. Another argument is that we observe lowest response rate in first wave, which might affect a wave-based subsample.

In modules that went from one approach to another, we saw no obvious break in the time series.

The Norwegian LFS estimation procedure, both for full-sample and subsample, uses post-stratification weighting. Auxiliary variables are register-based employment and demographic categories. The subsample weights seem to have little effect on overall figures; this applies to both structural variables and ad-hoc topics.

Outline

First we state some general information about the survey, emphasizing facts directly related to the subsampling approach. The following discussion will focus on practical considerations, though some possible selection- and nonresponse effects are also discussed.

We then include a description of our relatively simple weighting procedure, and suggest an even simpler.

Finally some results are presented, comparing subsample approaches and estimates on subsample topics.

Sample Design

Regular Sample

The Norwegian LFS uses CATI only, sampling families and allowing proxy interviews. 3000 people are sampled every quarter, adding to a rotating panel that lasts 2 years. That means a "quarterly sample" has 24000 people. Yearly averages are based on 96000 records, which effectively represent 33000 independent units.

Each sampled person is scheduled for a serial of 8 interviews, 13 weeks apart. About 1/13 of people are distributed to each week in each "quarter" (13-week period). This scheduling is done in advance and fixed whether the interview takes place or not.

A wave is the group of people scheduled for the same serial number. For instance, "wave 3" are the people scheduled for their third interview. As a result of the even allocation, any of the regular data collections (week, month, and quarter) consists of approximately 1/8 of each wave.

Subsample

Subsamples are used when some questions are asked only to a part of the regular sample:

- Structural variables – questions that need not be asked very frequently.
- Ad-hoc modules – sets of extra questions on specific topics.

There are two main subsample approaches discussed here, selecting a period in time vs picking one or more waves. A more drastic alternative, which we only mention here, is to select a subsample completely randomly.

In addition to Eurostat modules the Norwegian LFS also includes national ad-hoc modules. Before 2006 a common approach was used, different from both of the current. That means historically, that LFS data have three different subsampling types, specifically:

2nd Quarter only	All modules before 2006
1st and 8th Wave every Quarter	Eurostat modules after 2006
2nd through 7th Wave in 2nd Quarter only	National modules after 2006

Due to the different topics, there are no common variables of interest at the same time. Consequently, most estimates have no directly comparable figure, which of course would have been be most interesting in order to compare approaches.

Discussion

Both time approach and wave approach can have practical pros and cons, and methodological pros and cons. As the LFS is a large and continuous production, practical considerations can often take precedence. The point of subsamples is to save resources, but it doesn't make things easier to have more complicated work on top of the regular LFS.

Practical arguments include data management, interviewer workload and respondent burden. These will all affect economical- and HR considerations. Important factors may be the flexibility of IT systems and interviewer staff management. Some kind of modularity is beneficial, that means forms or procedures that are more or less reusable for different modules. We also appreciate a stable and in-house interviewer staff, which are motivated through and continuous education and dialogue.

Methodological arguments include selection- and nonresponse effects, and estimation issues. For instance seasonal variation or panel attrition can systematically affect the composition of a subsample. The choice of an estimation method must take into account several demands, for instance consistency and precision.

Our 2nd-quarter-only approach was devised before today's CATI-only mode. This approach may be more sensible when you use paper forms and face-to-face mode. Then you would use different forms in different quarters, but only one form at a time. A wave approach in this setting would mean that the interviewer must bring different forms at the same time, or that you allocate interviewer by wave. Both solutions would be less practical.

In the CATI environment, different questionnaire forms can be available to every interviewer at all times. On the other hand, that means careful planning and management of IT systems in order for this to work smoothly. If the IT systems include a lot of "manual" procedures, then a quarter-approach can be more efficient than a wave approach.

Generally, a wave approach evens out seasonal patterns which is desirable for many purposes. However, a period-approach could be better than wave-approach for specific topics. For instance when you ask young people about education, work or transitions between, it could be an advantage to avoid typical holiday periods. A subsample pattern tailored specifically to this end (for instance selecting certain months) could be optimal with respect to data quality. A more complicated approach would of course require more work in planning and IT management.

A strong argument for wave approach is that it evens out the interviewer workload through the year. That means easier management in terms of hiring and allocating man-hours. In addition it should increase performance as interviewers get better practice with the changing questionnaires.

We know now that when 2nd-quarter-only approach was used in the period 1996 – 2005, the response rate was markedly lower in 2nd quarter. Extra questions mean a heavier burden on both the interviewer and the interviewee. It is plausible that the lower response rate in 2nd quarter was caused by the module in one way or the other, since we from 2006 onward do not observe this pattern. When you are trying to estimate from a subsample instead of full sample, nonresponse is extra unwelcome. From that perspective, subsampling in a way that increases nonresponse seems like a bad idea.

Another matter is nonresponse bias. We have found that young people have lower response rate, as well as different seasonal pattern for labour market status, than older people. It would be no surprise if the combined effect was a biased subsample.

The wave-1-and-8 approach subsamples around the same size during a year as one quarterly sample. The first and last wave was chosen in order to reduce nonresponse due to the extra burden. The idea was that the first interview is longer anyway, and we try to motivate the

respondent by telling that subsequent interviews are much shorter. And the last wave was chosen simply because it would not affect subsequent interviews.

However, we know that response rates are not uniform across waves. Studies of panel data often mention attrition, i.e. lower response rate as the panel wears out. Contrary to this, in our data we find increasing response rate with waves. First wave has clearly the lowest response rate, waves two to eight see increasing rate but less clear pattern. At the same time, we know that nonresponse is not random. We must consider it possible that the wave response pattern can affect the representativity of the response subsample. It could well be that for instance choosing 3rd wave would be more advantageous. That wave has a more stable response rate and the interviewer relation is more established. On the other hand, this would reduce the effective sample size, due to our rotation design.

Estimation

In this section we describe and discuss a little about the current practice. A few results are included in the last chapter. Ideally, we would like to study more closely how the different approaches affect the representativity of the total subsample and of the response subsample. For the time being, we have a rather limited assessment of the subsample estimation.

Description

Currently, we use the same estimation method for all topics, variables and subsample types.

The subsample weights are calculated by multiplying the full-sample weight with an inflation factor:

w_i Full-sample Weight

$$x_h = \frac{\sum_{i=1}^{n_h} w_{i,h}}{\sum_{i=1}^{s_h} w_{i,h,s}}$$ Expansion Factor

$$w^s_{i,h} = w_i \cdot x_h$$ Subsample Weight

N = population size
n = sample size
s = subsample size
h = stratum
i = unit

The full-sample weighting uses post-stratification by register-based employment status in addition to demographics. The subsample weighting uses a more aggregate stratification along the same dimensions.

The regular procedure includes calibration with an iterative method resulting in a unique number for each unit. This ensures consistency with several population figures, but also introduces some extra variance. Users of the ad-hoc modules are often interested in cross-tabulation by other attributes than the calibration marginals. The extra variance may result in less precision of those tabulations.

Alternative

A simpler method would be:

$$w_h^s = \frac{\sum_{i=1}^{n_h} w_{i,h}}{s_h}$$ Subsample Weight

These estimates is consistent under the partition $H = \{1,2,3, \dots, h\}$. The weights would be equal within each stratum, and some variance reduction could be achieved. This method will be efficient as far as you have an efficient stratification in the first place. The strongest argument is perhaps that it is simple to implement.

For both methods we assume that the subsample stratification is a more aggregated version of the regular stratification; that there are no extra variables and no crossing of existing categories. In this way, the more detailed (full sample) stratification can be consistent with the more aggregated (subsample) stratification. Generally, you would not make the stratification too fine, as strata become too small or empty. This problem will be extra noticeable in a subsample.

Results

What effect has the subsample methods on the published figures? As mentioned we have limited evaluation of the estimates for the subsample topics. Comparing weighted and unweighted estimates for selected subsample questions shows that the current method has little effect on the overall results. Since demographic variables are used in post-stratification, it could improve cross-tabulations by the same variables.

The next section presents some figures comparing subsample approaches. This is intended as a basis for further discussion, not deciding which approach is best.

The last two sections present some results of the subsample weighting. Again, the intention is to start a discussion, not finish it.

Comparing Subsamples

In order to say something about representativity, we compared two simulated subsamples. Table 1 show the distribution of age, gender and labour market status for: yearly average, wave approach and time-approach. We find that both subsamples are rather representative, when compared to the yearly average. Remember that we compare respondents only. Nonresponse is substantial and its effects are hidden from this table.

Of the two, the second-quarter subsample was largest and had a distribution closest to that of the yearly average.

Table 1: Comparing subsamples. Norwegian LFS 2013.

	Sample size			Percent		
	Year	Wave 1 and 8	Second Quarter	Year	Wave 1 and 8	Second Quarter
Total	75821	18722	19019	100.0	100.0	100.0
Men	39145	9611	9818	51.6	51.3	51.6
Women	36676	9111	9201	48.4	48.7	48.4
15 - 24 years	13077	3248	3295	17.2	17.3	17.3
25 - 29 years	5574	1381	1406	7.4	7.4	7.4
30 - 34 years	5925	1404	1504	7.8	7.5	7.9
35 - 39 years	6411	1577	1571	8.5	8.4	8.3
40 - 44 years	7610	1826	1933	10.0	9.8	10.2
45 - 49 years	7709	1931	1886	10.2	10.3	9.9
50 - 54 years	7096	1757	1778	9.4	9.4	9.3
55 - 59 years	6727	1751	1692	8.9	9.4	8.9
60 - 64 years	6228	1552	1579	8.2	8.3	8.3
65 - 74 years	9464	2295	2375	12.5	12.3	12.5
Employed	53209	13113	13366	70.2	70.0	70.3
Unemployed	1592	448	415	2.1	2.4	2.2
Inactive	21020	5161	5238	27.7	27.6	27.5
Employed	48604	11930	12139	64.1	63.7	63.8
Student	9365	2341	2438	12.4	12.5	12.8
Retired, old age	6301	1518	1575	8.3	8.1	8.3
Retired, early	1366	335	343	1.8	1.8	1.8
Retired, disabled	6282	1582	1567	8.3	8.4	8.2
Homemaker	830	215	217	1.1	1.1	1.1
Unemployed	2074	529	503	2.7	2.8	2.6
Conscripts	157	50	32	0.2	0.3	0.2
Other	842	222	205	1.1	1.2	1.1

We also compared time series for employment and unemployment, for the same subsample types. It seems to us that wave approach result in a slight but consistent overestimation of employment (and underestimation of unemployment). However, we do not know whether this would affect estimates for questions that are conditional on employment status (e.g. shift work, work related injuries).

Ad-Hoc Variables

Table 2 compares three estimates of the proportion of disabled people. There are little differences between estimates at the same time. However, the proportion varies noticeably[1] over time, which is at odds with our assumptions about the population. We think this variation must be due to measurement error or similar factors other than population changes.

We do not observe an obvious break in the time series at the time of change in subsample approach. It is possible that the variation overshadows a minor break.

[1] The change from one year to another is up to 2.6 percentage points. In comparison the margin of error is below 0.2.

Table 2: comparing estimates of the proportion of disabled people. LFS 2002–2013.

	Unweighted	Regular Weight	Subsample Weight
2002	16.5	16.4	16.6
2003	15.2	15.2	15.7
2004	17.8	18.1	18.3
2005	17.3	17.6	17.7
2006	16.8	17.0	17.1
2007	16.7	16.7	16.7
2008	16.9	16.7	17.0
2009	16.0	15.8	16.0
2010	17.4	17.1	17.2
2011	15.7	16.0	16.1
2012	14.5	14.7	15.0
2013	15.9	16.1	16.5

Structural variables

Table 3 compares three estimates of the proportion of shift work. We observe that neither weighting procedure have much influence on the overall figures[2].

From 2006 onwards wave approach was introduced. The results show no marked break in the time series.

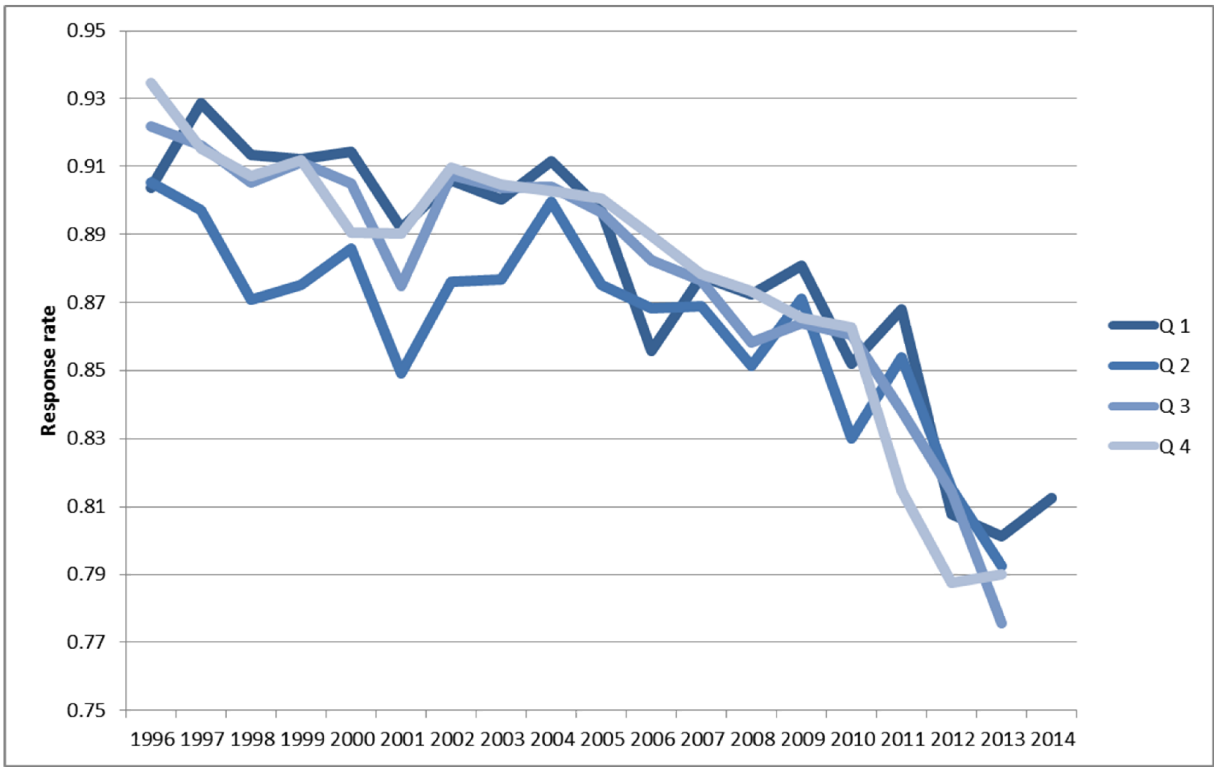
Table 3: comparing estimates of the proportion of shift work. LFS 2001–2013.

	Sample	Regular weights	Wave weights
2001	21.2	20.7	
2002	22.1	21.5	
2003	23.1	22.7	
2004	22.1	22.0	
2005	22.5	22.4	
2006	21.3	21.1	21.2
2007	21.9	21.6	21.7
2008	22.0	21.9	21.9
2009	21.9	21.7	21.7
2010	22.5	22.3	22.4
2011	21.9	21.7	21.7
2012	22.6	22.4	22.4
2013	23.0	23.1	23.1

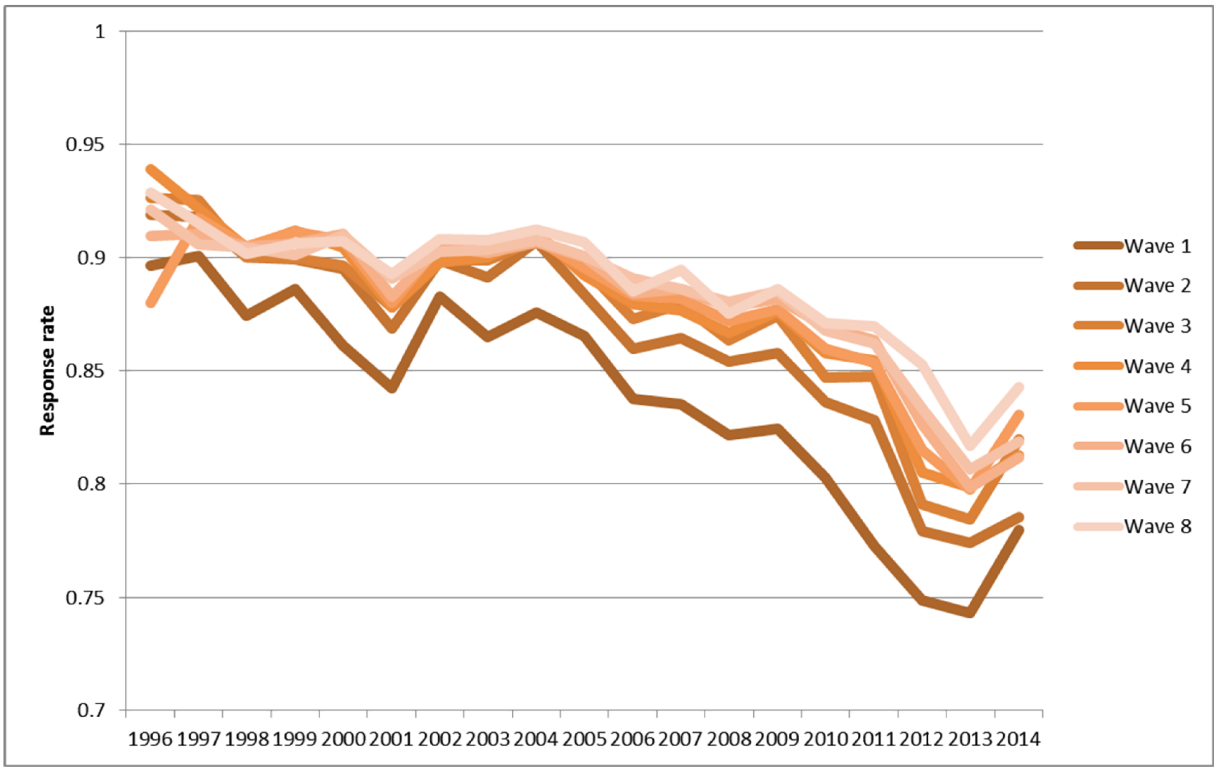
[2] The effect is slight, but it seems that both weights systematically adjust the figures down.

Revised May 22nd : Included diagrams from the presentation

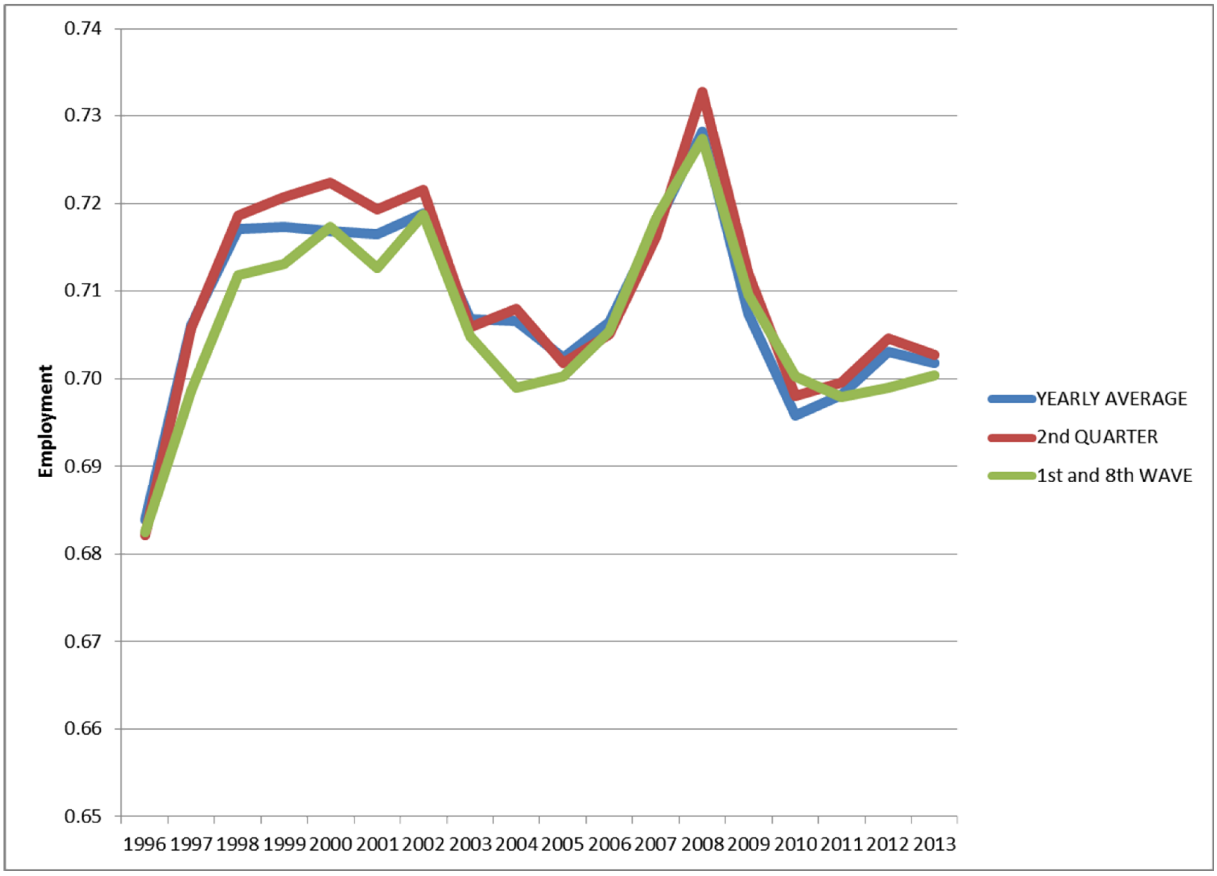
Response by QUARTER



Response by WAVE



Comparing subsamples (simulation)



Comparing weights

Other sources of error (questions etc.)?