

QUARTERLY SURVEY OF EMPLOYMENT (JOBSTAT) - HOW TAKING INTO ACCOUNT DEMOGRAPHICAL CHANGES IN THE ENTERPRISE POPULATION

Prepared by Daniel Assoulin (Daniel.Assoulin@bfs.admin.ch) and Jann Potterat (Jann.Potterat@bfs.admin.ch), Swiss Federal Statistical Office

1 Introduction

The Swiss Survey of Employment (JobStat) provides quarterly estimates for job variables like number of jobs in total, by gender, by degree of employment or also full time equivalents. JobStat is currently under revision. One major change consists in including JobStat into FSO's Sample Coordination System for enterprise surveys. The sampling units in the Coordination System are enterprises while extrapolation in JobStat is done on establishment level. Therefore, the new sampling plan is based on a cluster sample, where enterprises are the primary and establishments the second stage sampling units (see f.e. [4]), while so far sampling and extrapolation has been performed on establishment level (methodological details see [5] and [6]). Table 1 gives an overview of population¹ and sample sizes by sectors (2=industry, 3=services) in terms of enterprises (ent) and establishments (est) for the JobStat survey of 2nd Quarter 2015. The variable Profiling in Table 1 displays the number of enterprises and establishments belonging to Profiling² Groups. As for these enterprises JobStat information is based on Profiling, there is no additional data collection and they can be considered as an exhaustively surveyed strata.

Sector	Primary sampling units (ent)			Establishments (est)		
	Population size	Sample size	Prof	Population size	Sample size	Prof
2	95 887	5 776	283	102 648	8 276	1 620
3	479 673	13 004	1 278	566 156	57 750	38 818
Total	575 560	18 780	1 561	668 804	66 026	40 438

Table 1: Population and sample sizes

It is planned that each year a new sample is drawn. One of the main goals of JobStat is to produce reliable quarterly evolutions, which should not be distorted by administrative processes. Under this aspect the concept described below was set up to handle demographical and classification (f.e. NACE code) changes for the quarterly estimates within a given year. In contrast to conjunctural changes we interpret and summarize demographical and classification changes as structural changes.

For non-Profiling units:

- notification of structural changes for non-Profiling units in the business register may lag behind reality. Relying on the actualisation date in the business register could result in undesirable effects in the estimates due to administrative processes.

¹Source: sampling frame, based on the FSO business register

²Profiling provides quarterly updates on employment and structural information for a selection of big enterprises, enterprise groups respectively.

- structural changes in Non-Profiling units (enterprises and establishments) are to a wide extent ignored for the quarterly estimates. Estimates are produced by extrapolation based on the original frame, slightly adapted following the rules discussed in the next section (extrapolation frame).

It is planned, that for the first survey (quarter) based on a new sample estimates will also be produced based on the old sample (double sample), that the coordination of the old and the new sample is positive and that the samples will be based on a similar sampling design. So, there is a big overlap between the two samples, keeping the additional data collection effort / response burden relatively small and allowing to interpret differences in the estimates (\hat{y}_{new} vs \hat{y}_{old}) as reflecting structural changes. The quarterly estimates based on the old sample are then revised in order to reflect these changes. The revision methodology used in the past is described in [7]: A characteristic of the method is, that the three increase factors within the considered year are affected uniformly by the factor³ $(\hat{y}_{new}/\hat{y}_{old})^{\frac{1}{3}}$.

For enterprises belonging to a Profiling group (Profiling units):

- notification of structural changes in Profiling units in the Business Register are considered to be timely.
- structural changes (enterprise and establishment level) are taken into account for the quarterly estimates

The different treatment of Profiling and non-Profiling units leads to the question on how to deal with enterprises that go in or out of Profiling within the year. As we don't want such administrative changes to influence our quarterly estimates the initial status (Profiling / Non-Profiling) is relevant for the treatment of the enterprise throughout the year. In the following section we give an overview of identified demographical changes within the JobStat enterprise population between T0 (December 2014) and T1 (July 2015) based on business register data. We then discuss ideas on a categorization of these changes in connection with their possible treatment regarding maintenance of the sample and extrapolation.

2 Categorisation of changes and possible treatments

First in Table 2 we represent frequencies of identified demographical events in the union of units for the enterprise populations in frames T0 and T1 (603 434 enterprises). The source for change information is the business register reporting, documenting business register changes. The indicator *prof* distinguishes Profiling and non-Profiling units. As next, in Table 3, we map the eventcodes of Table 2 into categories according to the planned treatment for maintaining the JobStat sample and for extrapolation. Treatment codes D (one to one) and E (complex) are both represented by events 2, 6 in Table 2. They can be distinguished by the variable *predecessors*, reflecting the number of predecessors: The 2528+7 enterprises with event 6 and one predecessor reflect treatment category D (one to one) while the 24 cases with 2 or 6 predecessors belong to treatment category E (complex). At the other hand all enterprises with eventcode 2 (death with successor) had only one identified successor, therefore treatment code E restricts to 24 cases. From Table 2 we see, that around 91 % of cases are covered by the treatment category „no change“ (event 0). So, we discuss in detail the treatment of this „simple“ situation. The aim is to illustrate the treatment strategy, which is then also used to solve the other cases, just briefly discussed at the end of the section.

³Explicitly: If $\hat{y}_{old,t}$ denotes the original estimation for quarter t , the revised value is $\hat{y}_{rev,t} = \hat{y}_{old,t} \cdot (\hat{y}_{new,4}/\hat{y}_{old,4})^{\frac{t-1}{3}}$; $t = 1, \dots, 4$, where $t = 4$ is supposed to be the quarter with the double sample. In fact this adaptation is applied on different aggregation levels.

event	predecessors	prof	frequency	percent	interpretation
0	.	0	548 827	90.95	no change
0	.	1	1 486	0.25	
1	.	1	52	0.01	death without successor
1	.	0	21 346	3.54	
2	.	0	2 578	0.43	death with successor
2	.	1	13	0	
5	.	1	10	0	emigration out of population
5	.	0	1 248	0.21	
6	1	0	2 528	0.42	new with predecessor
6	1	1	7	0	
6	2	0	22	0	
6	6	0	1	0	
6	6	1	1	0	
7	.	0	8 171	1.35	new without predecessor
7	.	1	15	0	
8	.	0	17 023	2.82	new due to reactivation
8	.	1	13	0	
9	.	0	90	0.01	immigration into population
9	.	1	3	0	

Table 2: Events between T0 and T1

treatment	event	description
A	0	no change
B	1,5	disappear
C	7,8,9	appear
D	2,6	change in appearance, one to one
E	2,6	change in appearance, complex (f.e. many to one)

Table 3: Treatment categories

Category A - no change

Discussing category A (event 0) we take also into account whether the enterprise belongs to Profiling or not at time points T0, T1. This leads to 4 cases for which the treatment is summarized in Table 4. For explaining the notation we concentrate on subcategory 0.1, an enterprise belonging to the frames T0 and T1, which is not in Profiling in T0 ($\text{Prof}(T0)=0$) and not in Profiling in T1 ($\text{Prof}(T1)=0$). T0 in column

- *unit* means that in the sample (data collection) and for extrapolation (updated Frame) we use the same unit as in T0.
- $1(s)$ means that the unit keeps its sample indicator: if the unit is drawn into the sample at T0 it remains in the sample.
- π^{-1} means that the unit keeps its initial sampling weight, which provides the basis for constructing the estimator.
- x means that the auxiliary information used for extrapolation (calibration with the updated frame) is based on T0.

Subcategory	Prof(T0)	Prof(T1)	Information for extrapolation / sample at T1				
			unit	1(s)	π^{-1}	x	d
0.1	0	0	T0	T0	T0	T0	T0
0.2	0	1	T0	T0	T0	T0	T0
0.3	1	0	T0	T0	T0 (1)	.	T0
0.4	1	1	T0	T0	T0 (1)	.	T1

Table 4: Treatment of category A, no change

- d means that domain information (f.e. NACE Code, NUT) is based on T0. For non-Profiling multi-establishment enterprises this implies, that the list of establishments belonging to the enterprise at T0 is kept stable.

Remarks:

- The treatment of subcases 0.2 and 0.3 ignores immigration into / emigration out of Profiling. For extrapolation and sample purposes, the relevant status is the one at T0. Hence, the fact that an enterprise moves into a Profiling group between T0 and T1 does not imply, that it is automatically included in the sample, although the information on the target variables are available. The motivation for this treatment is, that we don't want administrative changes to influence our estimates.
- For non-Profiling multi-establishment enterprises, data collection is performed on enterprise level. The employment data on enterprise level is then distributed to the local units by an allocation key. This key is enterprise specific and based on Profiling-Light information. Profiling-Light covers multi-establishment enterprises outside Profiling and provides the same information as Profiling, but just on a yearly frequency. As the distribution to establishment is used to provide estimates for geographical and activity aggregates, the approach of keeping the domain information as in T0 can be interpreted as working with the establishment structure of T0.
- With exception of one small change (the unit in the updated frame is based on T1), Table 4 holds also for treatment category D (change in appearance, one to one).

The treatment of the other cases are also based on the idea of stability regarding sample and estimates for non-Profiling units and therefore follow the same principles. We briefly describe some aspects.

Categories B (disappear) and C (appear): For non-Profiling units both cases are ignored for construction of the extrapolation frame, meaning that these changes are not reflected in population totals and its estimation. Based on Table 2 we notice that 25 284 (=8 171 + 17 023 + 90) appearances face 22 594 (=21 346+1 248) disappearances. So, according to the Business Register we have a population increase of 2 690 enterprises that is not reflected in our estimates (before revision). In case of Profiling units: new enterprises within a Profiling group are included for estimation and disappeared enterprises are excluded for extrapolation.

Category E: For these more complicated but relatively rare events, the treatment for non-Profiling units is currently based on the following principles.

- an enterprise existing at T1 is taken into the sample if it has a relationship to at least one enterprise in T0 belonging to the sample.

- an enterprise existing at T1 is taken into the extrapolation frame if it has a relationship to at least one enterprise in T0.
- the weights for the new firms are constructed on the basis of the Generalized Weight Share Method described in [1] and [2] and for which procedures are already in place.

According to [1], page 14, or alternatively [2], Result 3, the use of the weight share method provides unbiased estimates for the population of enterprises at T1 which have at least one connection/relationship to an enterprise in T0. So it seems to be a bit in contradiction with the idea of keeping the impact of demographic changes on the estimates for non-Profiling units as small as possible. An alternative could be the Pro-Rating approach described in [3] (Solution 2), imputing the values for the variables of interest collected for enterprises in T1 back into the original enterprises in the sample of T0 and allowing extrapolation in the original frame. However, according to the discussion regarding Table 3, the number of concerned and identified cases is low (23 cases for the considered period). The situation simplifies if all involved enterprises in T0 and T1 are Profiling enterprises: extrapolation and sample is based on the enterprises present at T1.

3 Sampling Frame vs Extrapolation Frame

The sampling frame represents the population at T0 and serves as basis for drawing the sample. The extrapolation frame results from an adaptation of the sampling frame based on information about demographic changes between T0 and T1 and the treatment rules discussed in the previous section. The final weights for non-Profiling enterprises are obtained by calibration on employment totals (full-time equivalents) of the extrapolation frame. Therefore, comparing totals between the extrapolation and the sampling frame provides a good idea of the potential impact of demographic changes on our estimates. From Table 5 it follows that in case of non-Profiling enterprises, population totals (at least by sector) for variables N (number of enterprises), fte (full-time equivalents) remain very stable with absolute differences below 0.02%. So, we do not expect big effects of structural changes on our estimates at the considered levels. For Profiling units differences are much larger (up to 2.7 %), as the information in the extrapolation frame reflects T1.

prof	sector	N(T0)	fte(T0)	N(T1)	fte(T1)	rd_N (%)	rd_fte (%)
0	2	95'604	890'250	95'615	890'313	0.012	0.007
0	3	478'395	1'974'607	478'353	1'974'544	-0.009	-0.003
0	Total	573'999	2'864'857	573'968	2'864'857	-0.005	0.000
1	2	283	119'758	277	116'531	-2.120	-2.694
1	3	1'278	871'119	1'248	864'141	-2.347	-0.801
1	Total	1'561	990'877	1'525	980'672	-2.306	-1.030
Total	.	575'560	3'855'734	575'493	3'845'529	-0.012	-0.265

Table 5: Population total sampling (T0, December 14) vs extrapolation frame (T1, July 15)

4 Conclusions

An important goal of JobStat is to produce reliable estimates for evolutions which are not affected by administrative processes to update the structural information (enterprises, their establishments, codification) contained in the business register. We distinguish the situations for non-Profiling and for Profiling units:

- the information about structural changes for Profiling units is considered to be timely and is therefore taken into account for quarterly estimation.
- the information about structural changes for Non-Profiling units may lag behind reality. The time point these changes becomes evident in the Business Register rather reflects administrative processes. This motivates for the construction of an estimation procedure in two phases:
 - A first estimation based on the information / population reflected in the original frame. Update information is mainly used to maintain the sample but to a large extent ignored for estimation.
 - For the first quarter based on the new sample drawn in the new frame, estimates are produced also with the old sample allowing a model based revision of the quarterly estimates to reflect also structural changes.

For data collection and continuity reasons the sample is adapted to demographical changes. An analysis of recorded demographical changes in the Business Register between December 2014 and July 2015 indicated, that the big majority of enterprises existing at least in one of the two frames belong to categories for which treatment remains simple for data collection (keep the enterprise, replacing the enterprise by a successor, excluding enterprise) as also for extrapolation (using information from original frame).

References

- [1] Lavallée, P. (2002), *Le sondage indirect ou la méthode généralisée du partage des poids*. Editions de l'université des Bruxelles et Ellipses.
- [2] Lavallée, P., Deville, J-C. (2006), *Indirect sampling: The Foundations of the Generalized Weight Share Method*. Survey Methodology, Vol 32.
- [3] Lavallée, P., Struijs, P. (1998): Treatment of enterprise restructuring in business panels. *Assemblée annuelle de la SSC 1998*, http://www.ssc.ca/survey/documents/SSC1998_P_Lavallee.pdf
- [4] Särndal, C.E., Swensson, B. et Wretman, J. (1992), *Model Assisted Survey Sampling*, Springer.
- [5] Renaud, A., Panchard, C. and Potterat, J. (2008), *Statistique de l'emploi. Révision 2007: méthodes d'estimation*. Numéro de commande: 338-0055, Swiss Federal Statistical Office.
- [6] Renaud, A. (2008), *Statistique de l'emploi. Révision 2007: cadre de sondage et échantillonnage*. Numéro de commande: 338-0052, Swiss Federal Statistical Office.
- [7] Graf, M. (2001), *Désaisonnalisation: Aspects méthodologiques et application à la statistique de l'emploi*. Numéro de commande: 338-0003, Swiss Federal Statistical Office.