Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Home Affairs FDHA
Federal Statistical Office FSO

Swiss Structural Business Statistics: Data Harmonization for the Construction of Full-time Equivalents

Desislava Nedyalkova and Daniel Assoulin

Swiss Federal Statistical Office, Statistical Methods Unit

European Establishment Statistics Workshop
EESW15, 7-9 September 2015, Poznan, Poland

Outline

Introduction

From 2011 onwards, the Swiss Structural Business Statistics (STATENT) replaces the Business census. Main sources/variables:

1. The Business register (BR):
   - economic activity (NACE Rev. 2), geographical region (NUTS2).
2. The Social security register (SR):
   - total employment (BETOT_R) and wages by gender at the enterprise level.
3. Survey data, for instance the Quarterly survey of employment (JobStat):
   - full-time equivalents (FTE_S), total employment (BETOT_S) by gender.

FTE in the framework of STATENT

- FTE are not recorded in the SR.
- In general, survey FTE are used in STATENT.
- For enterprises without survey FTE:
  - FTE are constructed using a linear prediction model based on explanatory variables coming from the register.
  - The model is fitted on matched data coming from the register and complementary surveys.

Differences in the variables BETOT

- Matched data present certain differences between the variables number of employees coming from the different sources.
- These differences may, for instance, be due to measurement errors or to the definition of employment.
- For STATENT, the reference source for employment is the SR.
- Aim: harmonise survey FTE in order to make them consistent with the register.

The FTE model in brief

Prediction of FTE of an enterprise i by

$$\hat{y}_i = \sum_{j=1}^{4} \hat{\alpha}_j V_{ij}. \tag{1}$$

- Explanatory variables: V_{ij}, number of employees in four salary classes ($\sum_{j=1}^{4} V_{ij} = \text{BETOT\_R}$).
- V_{i1} contains the smallest, V_{i4} the biggest salaries.
- The regression coefficients $\hat{\alpha}_j$ vary by gender, economic sector of activity and NUTS2 region.

Harmonized FTE - a first approach

For each enterprise i, we define a new variable, FTE_R (by gender) as follows:

$$\text{FTE\_R}_i = \frac{\text{BETOT\_R}_i}{\text{BETOT\_S}_i}\text{FTE\_S}_i \tag{2}$$

Motivation for this choice comes from techniques like

- *prorating* - a simple multiplicative adjustment applied on control variables in data editing rules or
- *generalized ratio adjustment* (Panekoek, 2014).

A new approach

- Aim: treat the differences in the number of employees by taking into account information about employment.
- We examine separately each of the following three cases:
  - Case 1: BETOT_R $>$ BETOT_S.
  - Case 2: BETOT_R $=$ BETOT_S.
  - Case 3: BETOT_R $<$ BETOT_S.
- The focus will be put on Case 1, the treatment of Case 3 being similar.

Case 1

- For each enterprise i, we define (by gender):

$$\text{dif\_ab}_i = \text{BETOT\_R}_i - \text{BETOT\_S}_i.$$

- We fit the following model:

$$\text{dif\_ab}_i = \sum_{j=1}^{4} \beta_j V_{ij} + \epsilon_i, \tag{3}$$

  under the hypothesis $\text{Var}(\epsilon_i) = \sigma^2 \text{BETOT\_R}_i$.

- We are interested in the estimated coefficients $\widehat{\beta}_j$, and not in the prediction of dif_ab_i.
- $\widehat{\beta}_j$ may be seen as an estimation of the proportion of persons in the salary class j which are in the register but not in the survey.

The $\widehat{\beta}_j$

	$\widehat{\beta}_1$		$\widehat{\beta}_2$		$\widehat{\beta}_3$		$\widehat{\beta}_4$	
	Est.	StdErr	Est.	StdErr	Est.	StdErr	Est.	StdErr
m/s2	0.659	0.063	0.562	0.016	0.153	0.019	0.039	0.004
m/s3	0.808	0.025	0.624	0.012	0.329	0.012	0.088	0.003
w/s2	0.724	0.023	0.495	0.015	0.241	0.013	0.051	0.006
w/s3	0.693	0.012	0.390	0.009	0.125	0.006	0.129	0.004

Table : Values of $\widehat{\beta}_j$ and estimated standard errors by gender and sector

We can note that coefficients for the salary class having the smallest salaries (salary class 1) are larger than the coefficients for salary class 2.

Initial idea

- We suppose that, for an enterprise i, the observed difference $n_i = \text{dif\_ab}_i$ corresponds to a sample s_i
  - of fixed sample size (n_i),
  - and probabilities proportional to $\widehat{\beta}_j$

  of persons which are in the register but not in the survey.
- In this way, for a person d in salary class j we have:

$$P(d \in s_i) \quad = \frac{n_i \widehat{\beta}_j}{\sum_{j=1}^{4} \widehat{\beta}_j V_{ij}} = \widehat{\beta}_j \frac{n_i}{\widehat{n}_i} = \widehat{\beta}^*_{ij},$$

  where $\sum_{j=1}^{4} \widehat{\beta}_j V_{ij} = \widehat{n}_i$.
- if $\widehat{\beta}^*_{ij} \geq 1$ then the person is automatically removed (Särndal et al. (1992, p.89)).

Adaptation of the initial idea

- Inconveniences of a PPS sample of persons: random aspect, difficulty of implementation in production.
- Proposed alternative: calculation of an average number of persons which should be removed in each salary class:

$$E\Big(\sum_{d\in\mathcal{V}_{ij}} 1(d \in s_i)\Big) \quad = \quad V_{ij}P(d \in s_i) = V_{ij}\widehat{\beta}^*_{ij}, \qquad (4)$$

  where $\mathcal{V}_{ij}$ denotes the set of employees of the enterprise i in the salary class j.
- As for the initial idea, if $\widehat{\beta}^*_{ij} \geq 1$, then the person will be automatically removed.

Consequences on survey FTE for Case 1

- Use of the harmonization procedure in the framework of fitting the FTE model:
  - The variables V_{ij} are replaced by the variables $\tilde{V}_{ij} := V_{ij} - V_{ij}\widehat{\beta}^*_{ij}$ with $\sum_{j=1}^{4} \tilde{V}_{ij} = \text{BETOT\_S}_i$,
  - y_i is given by FTE_S_i.
- Consequences for STATENT:

$$\text{FTE\_R}_i = \text{FTE\_S}_i \frac{\sum_{j=1}^{4} \widehat{\alpha}_j V_{ij}}{\sum_{j=1}^{4} \widehat{\alpha}_j \widetilde{V}_{ij}},$$

where $\widehat{\alpha}_j$ are the estimated coefficients of the FTE model.

Example of an entreprise: case 1, women, second economic sector of activity

Table : Values of BETOT et EPT

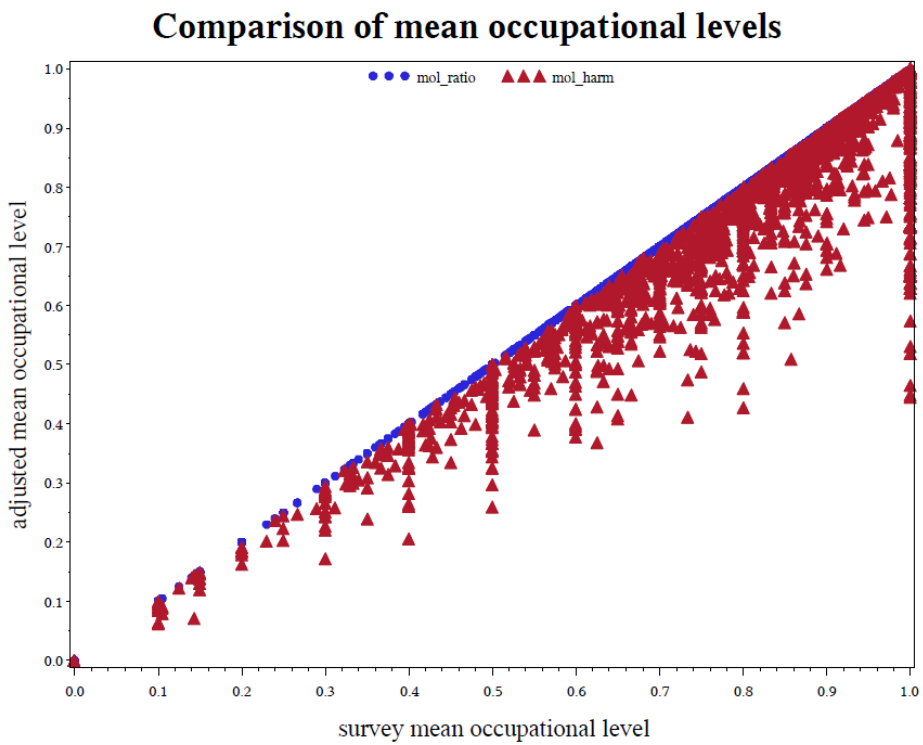
	Source		Adjustment	
	Survey	SR	simple	harmonized
BETOT	5	9		
EPT	4		**7.2**$=4 \times \frac{9}{5}$	**5.325**$=4 \times \frac{4.915}{3.692}$
EPTmodel	3.692	4.915		
mean occupational level			**0.8**	**0.59**

Table : Number of employees per salary class

V_1	V_2	V_3	V_4	Sum
4	0	2	3	9
$\widetilde{V}_1$	$\widetilde{V}_2$	$\widetilde{V}_3$	$\widetilde{V}_4$	Sum
0.718	0	1.455	2.827	5

Effects of harmonization: sector 2, case 1

The mean occupational levels after harmonization are in general smaller than before. This is due to the fact that the surplus of employment in the SR corresponds rather to small FTE.



Conclusion

The inconsistencies between survey and register data need to be treated.

- A first approach: a simple ratio adjustment of FTE by the ratio BETOT_R/BETOT_S.
  - Divergences due rather to small salaries or small occupational levels.
- A second approach: a *PPS* sample of fixed size allows for a treatment that takes into account information about employment.
  - Inconveniences: its random aspect, difficult to use in production.
- Improvement: use of expected sample sizes in order to adjust the number of employees in the different salary classes.

References

[1] Pannekoek, J. (2014) Method: Reconciling Conflicting Microdata, dans *Memobust Handbook on Methodology of Modern Business Statistics*, disponible sur http://www.cros-portal.eu/content/reconciling-conflicting-microdata-method

[2] Särndal, C.E., Swensson, B. et Wretman, J. (1992), *Model Assisted Survey Sampling*, Springer.