



Construction of Full Time Equivalents for the Register-based Swiss Structural Business Statistics

Desislava Nedyalkova and Daniel Assoulin

Swiss Federal Statistical Office, Statistical Methods Unit

European Establishment Statistics Workshop
EESW13, 9-11 September 2013, Nürnberg, Germany



Outline

Introduction

Method

Model

Model improvement

Evaluation

Discussion





Introduction

The Swiss business census will be replaced by the use of registers and complementary surveys. Main sources:

1. The Business register (BR)
 - ▶ records the new businesses,
 - ▶ updates the economic activity (NACE Rev. 2).
2. The Social security register (SR) provides
 - ▶ total employment (BETOT_SR) and wages by gender at the enterprise level.
3. The Quarterly survey of employment (JobStat) provides:
 - ▶ full-time equivalents (FTE_JS) for approx. 31 000 enterprises,
 - ▶ total employment (BETOT_JS) by gender, economic activity and geographical location (NUTS2).



Variable of interest

- ▶ Variable of interest (by gender):

$$FTE_SR := \frac{FTE_JS}{BETOT_JS} \times BETOT_SR. \quad (1)$$

- ▶ FTE_SR equals FTE_JS, but is adapted for differences in the variables BETOT coming from the two sources (due to measurement errors, definition of employment).
- ▶ This variable can be calculated in the case where all the information is available (survey and register).
- ▶ For the other enterprises in the SR - construct FTE by model estimation.



Linear model

- ▶ A prediction model based on enterprises contained in both JobStat and the register.
- ▶ A separate model is estimated by gender and economic sector (second, third).
- ▶ Explanatory variables: number of employees in four salary classes.



Linear model II

$$y_i = \sum_{j=1}^4 \beta_{jkl} V_{ij} + \epsilon_i, \quad (2)$$

where:

- ▶ y_i = FTE_SR of an enterprise i according to (1),
- ▶ V_{ij} , number of employees of enterprise i in salary class j ($j = 1, \dots, 4$) ($\sum V_{ij} = \text{BETOT_SR}$)
- ▶ β_{jkl} , regression coefficient for V_{ij} in NUTS2 k ($k = 1, \dots, 7$) and NACE (rev.2) section ℓ ,
- ▶ ϵ_i , residual with $E(\epsilon_i) = 0$ and $\text{Var}(\epsilon_i) = \sigma^2 \text{BETOT_SR}_i$.



Estimation

Estimation of the regression parameters by Weighted Least Squares (WLS).

- ▶ The weight for enterprise i : $w_{1,i} = w_i / \text{BETOT_SR}_i$, where w_i is the extrapolation weight according to JobStat.
- ▶ The factor $1 / \text{BETOT_SR}_i$ takes into account the assumed heteroscedasticity.



Construction of the salary classes

- ▶ Based on JobStat, for each activity sector OFS50 ¹, estimate the proportion of employees working on:
 - ▶ part time III (degree of employment less than 15%),
 - ▶ part time II (15%-49%),
 - ▶ part time I (50%-89%),
 - ▶ fulltime (> 90%).
- ▶ Calculate quantiles of the annual salaries' distribution from SR that correspond to the cumulative proportions.
- ▶ Based on these quantiles, define in each OFS50 four salary classes.

¹In Switzerland there is a standard of aggregations of the economic divisions called OFS50





Model selection

- ▶ Inconvenience of the model : a large number of parameters to be estimated.
- ▶ Proposed solution: a model selection procedure in order to reduce the number of parameters of the model.
- ▶ Initial model for model selection:

$$y_i = \sum_{j=1}^4 \alpha_j V_{ij} + \sum_{j=1}^4 \beta_{jkl} V_{ij} + \epsilon_i, \quad (3)$$





Model selection II

Starting from model (3):

- ▶ Using the GLMSELECT procedure in SAS (selection criteria = stepwise), choose the best model based on Akaike Information Criterion (AIC).
- ▶ For each gender and economic sector the term $\sum_{j=1}^4 \alpha_j V_{ij}$ is selected for the model.
- ▶ For certain NUTS2, k , and NACE sections, ℓ , a specific parameter β_{jkl} may be added in order to adjust the slope for the salary class j .



Robustification

- ▶ Aim: reduce the effect of extreme values on model selection and parameter estimation.
- ▶ Solution: a robustification procedure, which adjust the weights of the extreme observations.
 - ▶ Apply the SAS ROBUSTREG procedure on an initial model using only $\sum_{j=1}^4 \alpha_j V_{ij}$ as predictions.
 - ▶ Used technique: *LTS estimation*.
 - ▶ Reduce weight of observations with big standardized residuals.



Quality assessment

- ▶ Weighted R^2 : varies between 0.94 and 0.99 (better prediction for Industry than for Services).
- ▶ Possible quality measures:
 - ▶ weighted mean of absolute prediction errors,
 - ▶ *reldif*, measure based on the ratio between:
 - ▶ the extrapolated prediction errors (residuals) when estimating FTE_SR by our model,
 - ▶ and the extrapolated FTE_SR.
 - ▶ extrapolation is performed at the desired level of aggregation using appropriate weights.



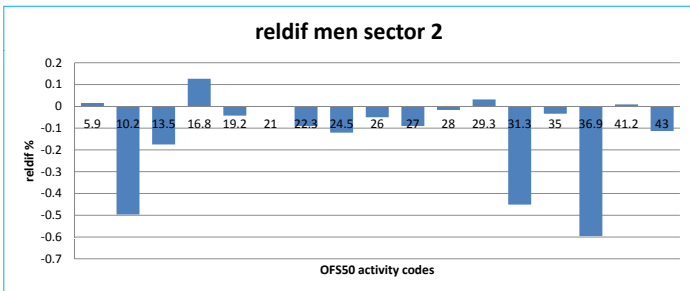
reldif by NACE OFS50 for economic sector 2

| NACE OFS 50 | extr. prediction errors | | extr. FTE | | reldif in % | |
|----------------|-------------------------|---------|------------|-----------|-------------|-------|
| | Men | Women | Men | Women | Men | Women |
| 5.9 | 0.58 | -1.84 | 3'798.63 | 383.05 | 0.02 | -0.48 |
| 10.2 | -206.96 | -158.68 | 41'678.96 | 23'301.49 | -0.50 | -0.68 |
| 13.5 | -11.57 | 100.58 | 6'613.57 | 9'168.60 | -0.17 | 1.10 |
| 16.8 | 67.85 | -219.64 | 53'586.01 | 13'883.82 | 0.13 | -1.58 |
| 19.2 | -9.98 | -13.05 | 23'443.97 | 7'119.84 | -0.04 | -0.18 |
| 21 | 0.06 | 0.06 | 22'065.98 | 14'028.48 | 0.00 | 0.00 |
| 22.3 | -28.66 | -27.23 | 31'705.99 | 8'126.38 | -0.09 | -0.34 |
| 24.5 | -94.28 | -215.34 | 78'052.09 | 15'113.61 | -0.12 | -1.42 |
| 26 | -33.00 | 25.26 | 65'804.96 | 37'872.16 | -0.05 | 0.07 |
| 27 | -24.14 | -37.11 | 26'678.77 | 9'220.19 | -0.09 | -0.40 |
| 28 | -11.09 | -108.49 | 67'072.31 | 11'080.71 | -0.02 | -0.98 |
| 29.3 | 4.31 | -8.93 | 13'608.50 | 1'743.78 | 0.03 | -0.51 |
| 31.3 | -165.12 | -168.63 | 36'601.51 | 12'146.29 | -0.45 | -1.39 |
| 35 | -7.22 | -5.98 | 21'304.91 | 4'116.98 | -0.03 | -0.15 |
| 36.9 | -69.25 | 0.38 | 11'628.31 | 1'429.56 | -0.60 | 0.03 |
| 41.2 | 8.56 | -88.15 | 96'753.14 | 5'454.28 | 0.01 | -1.62 |
| 43 | -208.15 | -137.49 | 183'552.95 | 18'564.15 | -0.11 | -0.74 |



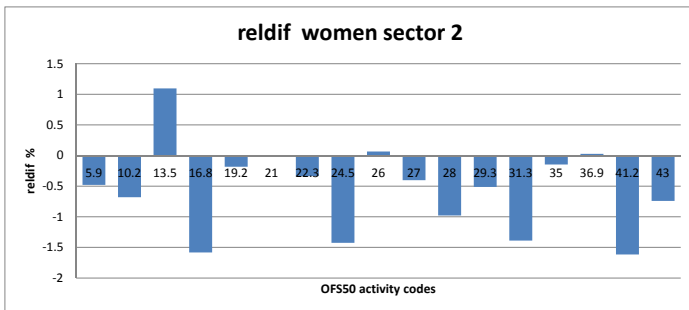


Men/Sector 2: reldif in % by NACE OFS50





Women/Sector 2: reldif in % by NACE OFS50





Discussion

With regard to the variable of interest BETOT_SR:

- ▶ FTE is not contained as such in SR.
 - ▶ Construction of FTE_SR using information about FTE's on enterprise level from survey data (JobStat).
 - ▶ The definition of FTE takes into account differences between employment information coming from JobStat and SR, respectively.
- ▶ Further investigations regarding these differences are planned.



Discussion II

With regard to enterprises not contained in JobStat:

- ▶ FTE can be estimated based on a linear model, which:
 - ▶ is based on carefully chosen salary classes which reflect different degrees of employment,
 - ▶ should be relatively stable over time (model selection, robustification),
 - ▶ leads to satisfactory prediction accuracy.
- ▶ Quality measures: an important tool for assessing and comparing the performance of different models for predicting FTE.