

00

Statistical Basis and Overviews

338-0077

Construction of full-time equivalents for the Swiss structural business statistics 2011

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Home Affairs FDHA
Federal Statistical Office FSO

Neuchâtel 2017

Methodology Reports by the Statistical Methods Section

The methodology reports describe the mathematical and statistical methodology which forms the basis of the results and analyses of public statistics. They also present the evaluation and development of new methods with regard to their use in the future. These publications serve to document the methodology used or planned in order to ensure transparency and scientific integrity. They also encourage collaboration with the universities and world of science.

Although the numerical results presented in the methodology reports are used to illustrate the mathematical concepts described, they are not the official results of the surveys concerned. Similarly, the methodology actually used may differ slightly from that described in these reports.

The methodology reports are available in electronic form at the FSO internet site.

Construction of full-time equivalents for the Swiss structural business statistics 2011

Editors Daniel Assoulin, FSO; Desislava Nedyalkova, FSO
Published by Federal Statistical Office (FSO)

Neuchâtel 2017

Published by: Federal Statistical Office (FSO)

Information: Desislava Nedyalkova, FSO, tel. +41 58 463 68 07, desislava.nedyalkova@bfs.admin.ch; Daniel Assoulin, FSO, tel. +41 58 463 67 51, daniel.assoulin@bfs.admin.ch

Editors: Desislava Nedyalkova, FSO; Daniel Assoulin, FSO

Contents: Statistical Methods Section, FSO

Series: Swiss Statistics

Topic: 00 Statistical basis and overviews

Original text: English

Layout: DIAM Section, Prepress/Print
This document has been automatically created from data banks and does not conform with the typographic norms for FSO publications.

Front page: FSO; Concept: Netthoevel & Gaberthüel, Biel; Photograph: © NorthShoreSurfPhotos – Fotolia.com

Printed: In Switzerland

Copyright: FSO, Neuchâtel 2017
Reproduction with mention of source authorised (except for commercial purposes).

Print format orders: Federal Statistical Office, CH-2010 Neuchâtel, tel. +41 58 463 60 60, fax +41 58 463 60 61, order@bfs.admin.ch

Price: free of charge

Downloads: www.statistics.admin.ch (free of charge)

FSO number: 338-0077

ISBN: 978-3-303-00569-9

Contents

Preface 5

Summary 5

1 Introduction 7

2 Description of the data 8
2.1 Data used for estimation of the FTE model . . . 8

3 The FTE model and its validation 9
3.1 Development of the FTE model . . . 11
3.2 Construction of the salary classes . . . 13
3.3 Model selection . . . 16
3.4 Weights used for the estimation of the model . . . 17
3.5 Measures used for the evaluation of the model . . . 18

4 Harmonization of employment variables 19
4.1 Prorating . . . 19
4.2 A harmonization approach that accounts for the type of employment . . . 19
4.2.1 Treatment of case 1 . . . 20
4.2.2 Treatment of case 2 . . . 22
4.3 Variables used to estimate the FTE model after data harmonization . . . 23
4.4 Consequences of data harmonization for constructing FTE in STATENT . . . 24
4.5 Effects of the harmonization . . . 24

5 Estimation of the FTE model and special cases 25
5.1 Estimated parameters . . . 26
5.2 Quality measures . . . 26

6 Conclusions 26

Appendix 30

References 35

Methodology reports published by the FSO's Statistical Methods Section 37

Preface

On demand of the Business Registers Data section (URD) of the Swiss Federal Statistical Office (SFSO), the Statistical Methods section (METH) provided the methodological support for the construction of the variable "number of full-time equivalents" for the enterprises included in the new Swiss structural business statistics (STATENT). This document gives a description of the main methodological challenges and chosen solutions. The described methodology reflects the state of the developed techniques for the definitive results of STATENT 2011 which were published in August 2014. The methodological practices were developed by Desislava Nedyalkova and Daniel Assoulin. We would like to thank Nicolas Tissot, Nadège Jacot and Mathieu Gunzinger from URD for the fruitful collaboration. We would also like to thank Beatriz Andrade, Jann Potterat and Philippe Eichenberger from METH who were also implied in the project for their valuable contribution. Finally, we would like to thank Lionel Qualité from METH for reviewing the document and for his valuable remarks.

Summary

From 2011 onwards, the Swiss business census is replaced by the new Swiss structural business statistics (STATENT). STATENT is mainly based on register data and complemented by several business surveys. Full-time equivalents (FTE) of employment are not available in the register data and should be constructed. This report presents the methodology used for the construction of FTE for STATENT. Two important topics are treated. First, we describe the model used for estimating FTE based on survey data which are matched to register data. The different phases in the development of the model such as model selection, robustification and quality assessment are presented. Then, we discuss how data inconsistency due to the integration of different sources is treated. The method which we apply to treat the inconsistencies takes into account the fact that they may be related to the type of employment. Finally, we explain how this data harmonization affects the estimation of the model and the construction of FTE for STATENT.

1 Introduction

In the past, a periodical business census (BC) played an important role for producing various statistics on the structure of the Swiss economy and for updating the business register (BR). The BC was held for the last time in 2008. For the reference year 2011 it has been replaced by an integrated system called STATENT (Statistiques des entreprises in french or Swiss structural business statistics). STATENT is mainly based on the BR, the register of the compensation offices of the Old-Age and Survivors Insurance (OASI) [1] and complementary surveys like the quarterly survey of employment (JobStat).
JobStat is the survey which is the closest in structure to the BC. It gives total employment and full-time equivalents (FTE) by gender for approximately 30'000 enterprises. The OASI register provides information about gender, monthly employment and wages (annual salary) at the employee level as well as total employment calculated at the enterprise level. The OASI register is linked to the BR at the enterprise level. It is the main source of information about employment at the enterprise level while the business register provides information like classification of the economic activity, legal form or enterprise structure for the multi-establishment enterprises.
FTE of employment is an important target variable which measures the employment capacity of an enterprise. In the OASI register, FTE are not available. For this reason, the construction of such a variable plays an important role for STATENT. For enterprises not included in a complementary survey, FTE for STATENT are constructed using a linear prediction model based on explanatory variables coming from the OASI register and the BR. This model is fitted on matched data coming from the OASI register, the BR and some complementary surveys. In the case where an enterprise has FTE collected from a complementary survey, these FTE will be, usually, used in STATENT.
The integration of data coming from different sources reveals the existence of inconsistencies regarding the number of employees. These different datasets need to be reconciled before the FTE prediction model is fitted.
This document is organized as follows. Section 2 gives a description of the data and the different motivations for constructing FTE (by gender) for STATENT. Next, in Section 3, we describe the linear model used for predicting FTE. We present also some quality measures to assess the quality of the chosen model. In Section 4, we discuss the treatment of inconsistencies regarding the number of employees according to the different sources when combining administrative data with survey data. We show also how data harmonization affects the estimation of the FTE model. In Section 5, we present some results and applications of the FTE model.

[1] "The old-age and survivors" insurance, better known as OASI, is the main pillar of the Swiss social security system." Source: http://www.zas.admin.ch/org/00723/00725/index.html?lang=en

2 Description of the data

The main sources for STATENT are the BR and the OASI register data. These administrative sources are complemented by different surveys. The BR is the main source of information about economic activity and the structure of enterprises and establishments included in STATENT. The OASI register is the main source for STATENT with regard to employment. It provides information about the number of employees and wages (by gender) without consideration for their occupational level. Different additional sources that provide information about employment - FTE, number of employees, structure of employment (occupational levels) or that are used to keep up-to-date the information in the BR - are taken into account in STATENT. We have, for instance

- Profiling and Profiling-Light - these are direct surveys which cover large enterprises, multi-establishment enterprises and groups (including public administration).
- Quarterly survey of employment (JobStat).
- The quarterly survey on new enterprises (ERST).
- NOGA-QS survey - a survey used to update the economic classification code.
- Surveys covering parts of the first sector of economic activities which is Agriculture, forestry and fishing (see Appendix).

The transition from the BC to STATENT induces several changes in definition and methodology. The principal differences concern the covered units, definition of employment and periodicity. The BC took into account only establishments and enterprises in which a certain activity is carried out for a minimum of 20 hours per week, whereas, by definition, STATENT takes into account enterprises (including independent employees) that are paying any social contribution for persons with an annual wage of at least CHF 2 300 (criteria for reference year 2011). The BC was conducted every 3-4 years, whereas STATENT is published each year. For the two universes the criteria for defining which employees are counted are also different. In the BC an employee is a person who has worked at least 6 hours per week in an enterprise or an establishment. In STATENT all persons working in an enterprise (as wage-earner or as independent) and paying their mandatory contributions to the OASI for a minimum annual wage of CHF 2 300 are counted. Another main difference between BC and STATENT is the way FTE are calculated. In the past, FTE were derived from the information in the BC (occupational levels) which is not available in the STATENT.

2.1 Data used for estimation of the FTE model

JobStat (Renaud, A., 2008; Renaud, A., Panchard, C. and Potterat, J., 2008) is the survey which is closest to the BC in its structure (definition of employment, variables). The survey frame for quarter 4, year 2011, is based, to a large extent, on the BC 2008. JobStat takes into account the second (Industry) and third (Services) sectors of economic activity. The number of establishments participating in the survey is around 60 000. Data from JobStat survey were matched to OASI register data at the enterprise level. This matched dataset is the basis for the development of the FTE prediction model. For the fourth quarter of 2011, in the matched dataset, we have: 20 834 single-establishment enterprises (EUNT), 5 016 multi-establishment enterprises (MUNT) and 2 551 enterprises coming from the Profiling which account for approximately half of the establishments participating to the survey. The basic employment variables

from JobStat are the number of employees working on part-time and on full-time. For an enterprise i, we have the total number of employees working on part-time III (occupational level of less than 15%) denoted by T_{i1}, the total number of employees working on part-time II (15%-49%) denoted by T_{i2}, the total number of employees working on part-time I (50%-89%) denoted by T_{i3} and the total number of employees working on full-time ($\geq 90\%$) denoted by T_{i4}. These variables are also available separately for the populations of men and women.

Due to the difference in the definition of the units covered by the BC and by STATENT, the number of entreprises counted in STATENT is larger than those counted in the BC and this concerns mainly the small enterprises. This gave the motivation to complement our dataset (matched data JobStat-OASI register) with the "small" enterprises coming from the survey collecting employment information from new enterprises, ERST (quarter 4, 2011). This complementary dataset is denoted as ERST-MINI and is considered as a census of very small enterprises entering the BR. The criterium "small" is defined as follows: the enterprise does not have employees working on full-time or part-time I and has no more than two employees working on part-time II. This can be explained by the fact that, if we have at least three employees working on part-time II, the enterprise will be covered by JobStat. For the fourth quarter of 2011, the number of small enterprises in the ERST survey is equal to 24 261 (see also Table 1). In general, FTE for MUNT are based on survey information (Profiling, Profiling-Light). So, the FTE model was established for predicting FTE for EUNT and also the estimation of the model is based on the EUNT. Table 2 shows how EUNT from the JobStat and ERST-MINI surveys are distributed by NACE Rev.2 [2] section (EUROSTAT, 2008).

Sector	JobStat	ERST-MINI	Total
2	7'366	1'228	8'594
3	12'071	23'033	35'104
Total	19'437	24'261	

Table 1 Number of EUNT in the matched dataset by economic activity sector

3 The FTE model and its validation

We have already described the dataset used for the construction of the model, based on the JobStat and ERST surveys (quarter 4, 2011) matched to OASI register data (2011, reference month December). These are EUNT belonging to the second and third sectors of economic activity, for which we know the number of FTE, wages and the economic activity according to the statistical classification of economic activities NACE Rev. 2. Another variable which will be used for modeling FTE is the main geographical region or NUTS-2 [3]. This variable, denoted by GREG, is given in Table 3.
Employment and wages represent the most intuitive auxiliary information which can be used for predicting FTE. The OASI register delivers an annual wage and we have also information on duration of employment (each working month is known). As STATENT has its reference date in December, only employees that worked in December are taken into account for calculating the

[2] The Swiss standard which is based on the NACE Rev.2 is called NOGA 2008. NOGA is available for different levels of aggregation (section, division, OFS50) and plays an important role for our model. In the Appendix, tables describing the NOGA at different aggregation levels are given.

[3] The NUTS classification (Nomenclature of territorial units for statistics) is a hierarchical system for dividing up the geographic territory of the EU. As a member of the EFTA, Switzerland is included.

Section	JobStat	ERST-MINI
B	121	1
C	4’977	751
D	150	17
E	212	14
F	1’902	445
G	2’265	1’389
H	1’079	241
I	1’026	412
J	1’001	766
K	717	212
L	280	739
M	2’349	4’111
N	572	1’289
O	10	2’876
P	255	4’718
Q	1’481	2’639
R	292	4’303
S	557	3’611

Table 2 Number of EUNT in the matched dataset by NOGA section

GREG /NUTS-2	REGION	NUTS-2 code
1	Lake Geneva region	CH01
2	Espace Mittelland	CH02
3	Northwestern Switzerland	CH03
4	Zurich	CH04
5	Eastern Switzerland	CH05
6	Central Switzerland	CH06
7	Ticino	CH07

Table 3 Main geographical regions of Switzerland

number of employees. In order to correct for different duration of employment, the calculation of an employee’s standardized wage in the register was necessary. The construction of the standardized wage will be illustrated by an example. Suppose that an employee has worked only during the last 5 months of the year (including December). The OASI register reports an annual wage bill of 25 000 CHF. We suppose that his average monthly salary is 5 000 CHF (the same each working month of the year), which gives an estimated standardized annual wage bill of 60 000 CHF. This procedure is applied for all employees.
It is clear that, in this way, employees which have not worked during the month of December will not be included in the variable "Total employment" of the enterprise according to the OASI register. The following variables by gender will also be used in the model:

- EMPTOT_R: total employment (number of employees) according to the OASI register.
- EMPTOT_S: total employment according to survey information.
- FTE_S: the number of FTE per enterprise according to survey information.

These variables form the basis of our prediction model and data treatment procedures. The variable of interest is the number of FTE of the enterprise. This variable should be consistent with OASI register data which is the reference source for STATENT.
Before presenting the model which was retained for the definitive STATENT 2011 (published in 2014), we will briefly discuss the different steps in the development of the FTE model. It should be noted that a separate model is estimated by gender and economic activity sector (second, third). Hence, four models are used for the estimation of FTE. As the same approach was used for all four models, for the sake of simplicity, we will not discuss all these models separately.

3.1 Development of the FTE model

The first developed model was a model on the dependent variable $\log$ of the mean occupational level (MOL) of the enterprise which is equal to the number of FTE divided by the total employment (EMPTOT):

$$\log_{10} \text{MOL} = \log_{10} (\text{FTE}/\text{EMPTOT}) = \log_{10} \text{FTE} - \log_{10} \text{EMPTOT}. \tag{1}$$

For an enterprise i we define the following model:

$$y_i = \log_{10}(\text{MOL}_i) = \alpha_0 + \sum_j \alpha_j X_{ij} + \epsilon_i, \tag{2}$$

where the explanatory variables of the model, X_{ij}, are a mix of categorical variables (NOGA, NUTS-2, gender) and quantitative variables (characteristics of the wage distribution, like the mean of the log wages ($\log_{10}$ (WAGES)), where $\log_{10}$ (WAGES) is a logarithmic transformation of the standardized annual wages. In general, $\log_{10}$ MOL should be positively correlated with the mean of $\log_{10}$ (WAGES), that is why this variable is a rather natural predictor. A feasibility study was performed on the Swiss earnings structure survey (SESS) which is a biennial business survey that provides data on occupational levels and wages. The advantage of testing a model on this survey is that information comes from the same source (Graf, M. and Potterat, J.). Model (2) has the drawback that predicting FTE from the logarithm of the MOL is not straightforward in the sense that, in order to interpret prediction plots and plots of residuals, a certain number of transformations need to be done. This led us to the following idea. Knowing (1), we have decided to set as dependent variable $\log_{10}$ FTE and let $\log_{10}$ EMPTOT as a predictor. The new model has then the following form:

$$y_i = \log_{10}(\text{FTE}_i) = \alpha_0 + \sum_j \alpha_j \tilde{X}_{ij} + \epsilon_i, \tag{3}$$

where the explanatory variables of the model, $\tilde{X}_{ij}$, are regional information (GREG), economic activity (NOGA section), $\log_{10}$(EMPTOT) and $\log_{10}$ (WAGES). This model still has the drawback that predicted FTE are biased due to the logarithmic transformation. Different measures for correcting the bias (Newman, 1993) were considered and applied to predicted FTE but our analyses showed that the bias was never completely removed. However, there was another problem with this model - for small values of FTE we observed very large residuals. Figure 1 shows the structure of the residuals in the log scale, for the model fitted on the population of men in the second economic activity sector. We can see that residuals are large for the really small values of FTE, and decreasing with the size of log(FTE). Another challenge when working with this model was the treatment of large standardized annual wages.
These reasons led us to develop a model, which estimates directly the number of FTE of an enterprise and is robust against extreme salaries. For that purpose, the distribution of wages is used to construct salary classes and the variable EMPTOT_R is then decomposed in

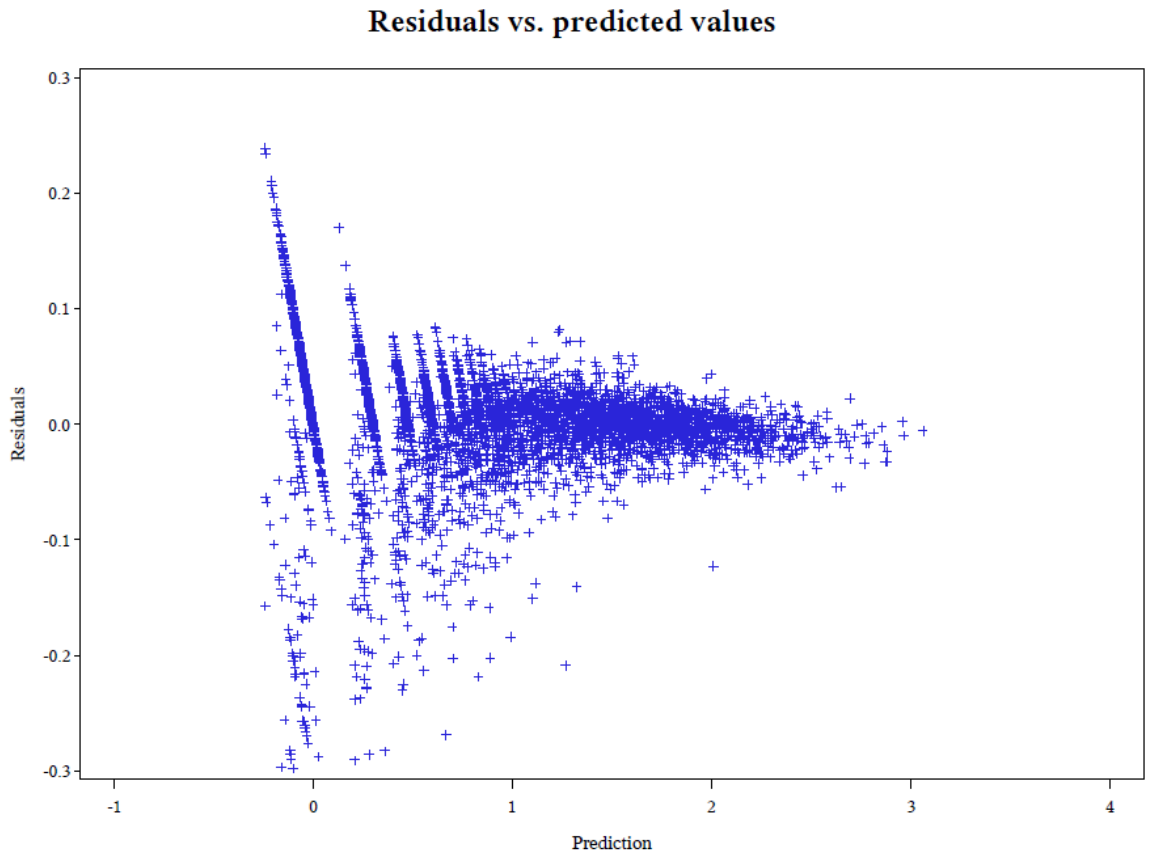


Figure 1 Residual plot for the log model

number of employees by salary class. We have studied and compared different variants for the construction of the salary classes. The first sketch was to have a limit for the first salary class set at CHF 10 000 and then by brackets of CHF 20 000. In this way, we have chosen to construct 9 salary classes. The advantages over Model (3) are that there is no obvious technical bias and the model and related quality diagnostics are easy to interpret. The major disadvantage of this model is that the number of parameters which should be estimated can become very large. For instance, for the population of men in the second economic activity sector we have 310 parameters (7 NUTS-2, 9 salary classes and 5 NOGA sections, see Table 15). In comparison, with Model (3) we have only 13 parameters to estimate.

Thus we have decided, to further develop this model by reducing the number of salary classes to four. We will explain in the following section how the new limits for the salary classes are computed and how the variables of the model are constructed. So, the explanatory variables for the final FTE model are the number of employees in four salary classes with coefficients by GREG and NOGA section adjusted for the three largest salary classes. We have decided not to adjust the first salary class by GREG and NOGA section due to the small number of observations in this salary class.

For an enterprise i, the initial model used for estimating FTE is:

$$y_i = \alpha_1 V_{i1} + \sum_{j=2}^{4} \alpha_{jkl} V_{ij} + \epsilon_i, \tag{4}$$

where:

- y_i = FTE of enterprise i in NUTS-2 k and NOGA section ℓ representing employment according to the OASI register,
- $V_{ij},\ j = 1, ..., 4$, number of employees of enterprise i in salary class j with

$$\sum_j V_{ij} = \text{EMPTOT_R}_i,$$

  the number of employees of enterprise i,
- α_1, regression coefficient for V_{i1},
- $\alpha_{jk\ell}$, regression coefficient for $V_{ij},\ (j = 2, ..., 4)$ in NUTS-2 k $(k = 1, ..., 7)$ and NOGA section ℓ,
- ϵ_i, residual with $\mathbf{E}(\epsilon_i) = 0$ and Var$(\epsilon_i) = \sigma^2$EMPTOT_R$_i$.

For some practical reasons, we have chosen a model without intercept (for instance, predicted FTE is zero in case of no employees). A separate model according to Equation (4) is used for the men and women in each of the two economic activity sectors (Industry and Services). This leads to 4 models in total.

3.2 Construction of the salary classes

As we have seen in Equation (4), the explanatory variables of the model are the number of employees in four different salary classes. Suppose we have already calculated the limits for the salary classes, by NOGA and by gender, denoted by limit 1, limit 2 and limit 3. If I_j denotes the set of employees in the salary class j, $j = 1, ..., 4$, then the variable number of employees per salary class for enterprise i is defined as follows: $V_{ij} = \sum_{k \in E_i} 1(k \in I_j)$, where E_i is the set of employees of enterprise i and $j = 1, ..., 4$ is the subscript for the class. Note that the total number of employees of enterprise i, EMPTOT_R$_i$, is equal to the sum of the V_{ij}. In what follows we explain how the limits for the salary classes are calculated.
In the JobStat survey, as well as in the ERST survey, for each enterprise, we know the number of part-time III, part-time II, part-time I and full-time employees. This auxiliary information can be used for the estimation of the proportion of employees working part-time. Based on these proportions, we calculate cumulative proportions corresponding to increasing occupational levels. Then we use the corresponding percentiles in the OASI standardized wage distribution as class limits for our salary classes.
This construction is based on the assumption that the order of employees according to their occupational levels would be similar to the one obtained from their standardized wages. Thus, for example, the proportion of employees working on part-time III (smallest occupational level) will correspond, in the distribution of the annual wages, to employees receiving the smallest salaries. This proportion of part-time III workers will define the salary class limit for the first salary class.

Data and weights used for the calculations

The whole dataset (JobStat and complementary surveys) comprises, in total, 52 235 enterprises. We know that survey data and OASI register data have some inconsistencies between the variables EMPTOT_S and EMPTOT_R. Thus, we have decided to exclude from this dataset

enterprises that have differences exceeding certain limits. The excluded enterprises represent approximately 24% of the data. Let denote by

$$\textit{absdifemp} = |\text{EMPTOT_S} - \text{EMPTOT_R}|$$

the absolute difference between total employment according to survey and OASI register and by

$$\textit{absreldif} = 100 \times |\text{EMPTOT_S} - \text{EMPTOT_R}| / \min(\text{EMPTOT_S}, \text{EMPTOT_R})$$

the absolute relative difference between the two variables. These two differences are calculated separately for men and for women. Then we construct a rule which determines if the enterprise will be included in the dataset for the calculation of the proportion of part-time workers. Thus, if *absreldif* < 10 or *absdifemp* ≤ 1, for men or for women, the enterprise will be included in the final dataset and excluded otherwise.
For the estimation of proportions, a special set of weights was created. The sampling weights of the enterprises who are not excluded from the total dataset were adapted by calibration ensuring coherence between extrapolated EMPTOT_S values and the known totals by NOGA section and gender. For ERST-mini the sampling weights were set to 1.This procedure reduces the effect of enterprises with large employment differencies (register vs. survey) on the proportion estimates. The proportions are thus estimated using these extrapolation weights (by gender), denoted by $w_i^{(1)}$, and the restricted dataset (39 493 enterprises for the year 2011).

Procedure for the calculation of the salary class limits

- Based on JobStat and complementary surveys, we estimate for each NOGA section the proportions of part-time III, part-time II, part-time I and full-time employees.
- Let EXTRTOT_S $= \sum w_i^{(1)}$EMPTOT_S$_i$ denote the extrapolated sum of employees of enterprise i according to the survey. The cumulative proportions of part-time workers (by gender and NOGA section) using the extrapolation weights $w_i^{(1)}$ are defined as follows:

$$\begin{aligned}
\text{CUMPR_PT3} &= \sum w_i^{(1)} T_{i1} / \text{EXTRTOT_S} \\
\text{CUMPR_PT32} &= \text{CUMPR_PT3} + \sum w_i^{(1)} T_{i2} / \text{EXTRTOT_S} \\
\text{CUMPR_PT321} &= \text{CUMPR_PT32} + \sum w_i^{(1)} T_{i3} / \text{EXTRTOT_S}
\end{aligned}$$

We have noticed that for certain NOGA sections, especially for the men, certain proportions of part-time workers can be really small. For example, $0 \leq$ CUMPR_PT3 ≤ 2 and $2 <$ CUMPR_PT32 ≤ 6. In order to ensure a reasonable number of observations per category, we have modified the obtained cumulative proportions in the following way:

$$\begin{aligned}
\text{CUMPRC_PT3} &= \text{CUMPR_PT3} \\
\text{CUMPRC_PT32} &= \text{CUMPRC_PT3} + \max(\text{CUMPR_PT32} - \text{CUMPRC_PT3}; 0.10) \\
\text{CUMPRC_PT321} &= \text{CUMPRC_PT32} + \max(\text{CUMPR_PT321} - \text{CUMPRC_PT32}; 0.10),
\end{aligned}$$

where CUMPRC_PT3 denotes the corrected cumulative proportion of employees working on part-time III in the survey, CUMPRC_PT32 denotes the corrected cumulative proportion of employees working on part-time III and part-time II and CUMPRC_PT321 - the corrected cumulative proportion of employees working on part-time III, part-time II and part-time I.

- In order to ensure the stability between years 2011 and 2012, we have calculated the arithmetic mean of the corrected cumulative proportions by NOGA section for years 2011 and 2012. These average proportions are denoted by acpm3, acpm32, acpm321 for the population of men and acpw3, acpw32 and acpw321 for the population of women, respectively. For instance, acpm3 is calculated as the mean of the corrected cumulative proportion of part-time III workers (men) in the data for 2011 and the corrected cumulative proportion of part-time III workers (men) in the data for 2012. The average proportions for men and women by NOGA section are given in Table 4.

Section	acpm3	acpm32	acpm321	acpw3	acpw32	acpw321
Total	1.8	11.8	21.8	6.2	25.3	56.2
A	9.3	19.3	29.3	26.1	74.4	84.4
B	1.2	11.2	21.2	14.8	32.9	57.8
C	0.6	10.6	20.6	2.8	13.9	37.2
D	2.0	12.0	22.0	7.9	20.3	47.9
E	2.6	12.6	22.6	4.4	18.0	57.5
F	0.5	10.5	20.5	10.1	37.2	64.1
G	1.7	11.7	21.7	5.0	25.0	55.6
H	2.0	12.0	22.0	4.3	21.0	54.7
I	2.2	12.2	22.2	8.2	27.3	49.9
J	1.4	11.4	21.4	5.4	23.1	52.3
K	1.2	11.2	21.2	2.7	12.7	38.4
L	5.7	16.5	26.5	5.7	27.4	57.3
M	1.9	11.9	21.9	6.5	24.0	50.0
N	2.5	14.0	30.7	8.6	40.7	67.5
O	0.1	10.1	20.1	0.2	10.2	52.7
P	7.5	26.3	48.7	13.7	41.8	71.6
Q	3.9	13.9	31.5	6.2	26.5	67.4
R	7.4	21.2	37.1	14.0	46.4	72.9
S	1.5	11.5	23.0	5.9	22.3	50.1

Table 4 Average proportions for 2011 and 2012 (as percentages) by NOGA section

- Within OASI register data, for each year, we have at our disposal a file containing information on apprentices - social security number, gender, birth year, duration of the apprenticeship contract, etc. For the calculation of the salary class limits described below, we have removed apprentices from the population of employees because apprentices usually receive smaller annual salaries than regular employees, but work mostly full-time. Thus, the salary class limits were calculated without the apprentices. After defining the salary class limits, and for prediction purposes, we have added all apprentices to the employees in the fourth salary class, which is associated to the largest FTE.

  The limits for the salary classes are calculated as follows. In the distribution of standardized annual wages according to OASI register (without apprentices), we calculate the percentiles corresponding to the average proportions above. For example, for 2011, the limit for the first salary class (men) is given by the acpm3-percentile, which is the value of standardized wages below which acpm3 percents of the wages can be found. These limits are reported in Table 5, which gives also the number of employees (men, respectively women) in each NOGA section according to the dataset used for calculating the class limits. Note that for the first sector of economic activity (section A), due to the small

number of enterprises and employees in this sector, the limits were calculated without separation by gender that is why the same limit is reported for men and women.

Section	Number of men	Limit 1	Limit 2	Limit 3	Number of women	Limit 1	Limit 2	Limit 3
A	984	4'264	8'000	13'242	197	4'264	8'000	13'242
B	4'029	4'000	41'230	62'023	610	7'700	24'180	53'320
C	470'386	4'900	46'317	58'981	201'627	5'166	18'353	41'600
D	19'849	3'926	43'515	71'449	4'981	4'708	14'000	55'394
E	10'862	3'544	16'000	43'800	2'021	3'715	9'910	48'030
F	274'407	4'262	36'552	53'960	37'333	6'403	21'720	47'100
G	316'868	5'126	24'600	46'122	324'480	6'077	21'000	43'965
H	159'430	4'810	21'450	43'550	60'067	6'126	19'942	50'050
I	103'391	4'329	14'633	26'765	126'757	6'412	17'028	31'805
J	106'040	3'715	21'340	54'000	43'873	4'835	21'299	59'187
K	150'426	3'200	25'000	64'800	100'139	5'618	27'447	62'300
L	28'523	3'531	6'000	10'000	30'437	3'537	8'960	31'512
M	199'862	4'267	19'000	41'900	143'622	4'913	17'400	48'200
N	149'890	4'350	13'737	33'957	108'330	5'725	16'803	37'658
O	2'296	2'320	4'200	9'967	2'336	2'400	9'374	61'818
P	69'460	4'205	12'930	41'560	78'723	6'514	23'273	57'286
Q	116'300	4'284	10'046	44'865	378'031	5'918	23'540	57'477
R	36'915	3'646	7'200	13'500	31'530	4'303	13'641	38'910
S	53'200	2'500	4'225	8'046	97'959	3'600	9'030	24'381

Table 5 Limits of the classes by NOGA Section, 2011

This procedure is also applied to the 2012 dataset.

3.3 Model selection

The inconvenience of the model given in Equation (4) is that there is still a large number of parameters to be estimated. For example, for the third economic activity sector, our model will have a parameter for each combination of salary class (4) with NUTS-2 (7) and NOGA section (12, see Table 15, omitting sections O, T and U) or in total more than 300 parameters. In order to reduce the number of parameters and select a more parsimonious model, we perform a stepwise selection (see, e.g., Hastie, T., Tibshirani, R. and Friedman, J., 2009) based on the Akaike Information Criterion (AIC) (Akaike, 1973) on the initial model given in Equation (4). For programming issues, we reparametrize the parameters $\alpha_{jk\ell}$ as follows:

$$\alpha_{jk\ell} = \alpha_j + \alpha^*_{jk\ell},\ j = 2, 3, 4,$$

The initial model in Equation (4) can then be rewritten as follows:

$$\begin{aligned} y_i &= \alpha_1 V_{i1} + \sum_{j=2}^{4} (\alpha_j + \alpha^*_{jk\ell}) V_{ij} + \epsilon_i \\ &= \sum_{j=1}^{4} \alpha_j V_{ij} + \sum_{j=2}^{4} \alpha^*_{jk\ell} V_{ij} + \epsilon_i. \end{aligned} \tag{5}$$

The prediction model obtained by means of the GLMSELECT procedure in SAS (Cohen, 2009) with selection criterion "stepwise" is given by:

$$\widehat{y}_i = \sum_{j=1}^{4} \widehat{\alpha}_j V_{ij} + \sum_{j=2}^{4} 1_{jk\ell} \widehat{\alpha}^*_{jk\ell} V_{ij}. \tag{6}$$

It is important to note that, as a result, for each gender and economic activity sector, the term $\sum_{j=1}^{4} \widehat{\alpha}_j V_{ij}$ is always selected in the model. For certain combinations of NUTS-2, k, and NOGA sections, ℓ, a specific parameter $\widehat{\alpha}^*_{jk\ell}$ may be added in order to adjust the parameter $\widehat{\alpha}_j$ for the salary class j depending on the parameter selection. This is reflected by the variable $1_{jk\ell}$ which takes the value 1 if $\widehat{\alpha}^*_{jk\ell}$ is selected and 0 otherwise. This selection procedure is applied for the principal model selection for STATENT 2011 and STATENT 2012.

In order to obtain stable results throughout the consecutive years, the salary class limits used for constructing the auxiliary variables of the models for 2011 and 2012 are obtained by calculating the percentiles of the distribution of wages in function of the estimated average proportion of part-time workers between year 2011 and year 2012 as described in Section 3.2. In order to ensure stability in model estimation through the years we have also applied a special procedure for model selection. The procedure is based on the model selection for year 2011 and the model selection for year 2012 and uses as a criterion the t-values of the estimated parameters for each year. The effects (model parameters) which appear in both models (years 2011 and 2012) after model selection are kept in the final selection. To these common number of effects we add specific effects appearing either in the model selection for 2011 or in the model selection for 2012 based on the following criterion:

- for Sector 2: $|tval11| > 4$ or $|tval12| > 4$,
- for Sector 3: $|tval11| > 6$ or $|tval12| > 6$,

where $tval11$ denotes the t-value of the parameter estimates for the model 2011 and $tval12$, respectively the t-value for year 2012. The difference of the cutoff for sector 2 and 3 is explained by the wish to keep the number of variables that enter the model reasonable. For instance, for the model "men/sector 3", if we set the cutoff at 4, as for sector 2, we would obtain a selection of 121 effects. With the value of 6, we obtain 81 effects. Using this specific selection, for each year, we refit the model given in Equation (5) by means of the GLM procedure in SAS.

3.4 Weights used for the estimation of the model

Throughout the development and estimation of the model different sets of weights are used. This section describes and justifies the use of these specific weights. The matched data (survey-register) contains the survey extrapolation weights, w_i. This is our initial set of weights. We have seen that for the calculation of the proportion of part-time workers a special set of extrapolation weights, $w_i^{(1)}$ was used. If we assume that for each person in enterprise i the FTE is predicted, then FTE of the enterprise is equal to the sum of EMPTOT individual FTE. The residual for enterprise i is then the sum of EMPTOT residuals. Assuming independence, this leads to $\sigma^2_\epsilon \propto$ EMPTOT, i.e. we can assume that the variance of the residuals in (4) is proportional to the variable number of employees. Thus, another set of weights is produced by adjusting the survey extrapolation weight of an enterprise, w_i, in order to account for the heteroscedasticity present in the model. It is given by $w_i^{(h)} = w_i/\text{EMPTOT}_i$. This set of weights is used for the initial fit of the model.

Another set of weights stems from the desire to take into account and reduce the effect of extreme values on model selection and parameter estimation. A robustification procedure was applied to the FTE model. We have chosen not to eliminate observations from the data used for model estimation, but rather to adjust the weights. We have applied the SAS ROBUSTREG procedure (Chen, 2002) on a simple model using only $\sum_{j=1}^{4} \widehat{\alpha}_j V_{ij}$ as predictions (initial fit). The chosen technique for estimating the regression coefficients is least trimmed squares (LTS) estimation (see Rousseeuw, 1984). In this way we obtain estimates for the scale parameter σ and residuals ϵ_i^*. By choosing an appropriate tuning constant c, we define a correction factor for the weights $w_i^{(h)}$ as follows:

$$u_i = \begin{cases} \frac{c\widehat{\sigma}\sqrt{\text{EMPTOT_R}_i}}{|\epsilon_i^*|} & \text{if } |\epsilon_i^*| > c\widehat{\sigma}\sqrt{\text{EMPTOT_R}_i} \\ 1 & \text{otherwise} \end{cases}$$

In order to account for the number of robustified enterprises and keep this number relatively small, the tuning constant was set to $c = 3$, respectively to $c = 5$, for the men of sector 3. The new weights $w_i^* = u_i \times w_i^{(h)}$ are used to select and fit the final model.

3.5 Measures used for the evaluation of the model

The evaluation of the model is based on different criteria. Besides the coefficient of determination R^2, different measures were used in the process of comparing models throughout the development phase and assessing their performance on different aggregation levels. Among these measures are the weighted mean of absolute prediction errors, denoted by *meanabsdif*, and a quality measure based on estimated relative prediction errors for totals, denoted by *reldif*. More precisely, *meanabsdif* is defined as follows :

$$\textit{meanabsdif} = \frac{\sum_{EUNT} w_i |y_i - \widehat{y}_i|}{\sum_{EUNT} w_i},$$

where w_i are the survey extrapolation weights and y_i are the FTE. The last measure, *reldif*, is defined as the ratio between

- the extrapolated prediction errors (residuals) when estimating FTE for EUNT by our model:
- and the extrapolated FTE,

where extrapolation is performed at the desired level of aggregation using appropriate weights. In our case:

$$\textit{reldif} = \frac{\sum_{EUNT}(w_i^{(2)} - 1)(y_i - \widehat{y}_i)}{\sum_{EUNT} w_i^{(2)} y_i + \sum_{MUNT} w_i^{(2)} y_i + \sum_{PROF} y_i}, \tag{7}$$

where $w_i^{(2)}$ are extrapolation weights [1], y_i are FTE and PROF denotes the enterprises coming from the Profiling. In Equation (7), the differences $y_i - \widehat{y}_i$ are weighted with $w_i^{(2)} - 1$ and not with $w_i^{(2)}$. This is due to the fact that for the final calculation of STATENT totals the use of the model is restricted to enterprises without survey data. Thus, this quality measure takes into account the values of the extrapolated FTE not only for the data used in the model, but globally for all data from the JobStat survey including MUNT. These measures are calculated at different aggregation levels.

[1] The weights $w_i^{(2)}$ are constructed like $w_i^{(1)}$ presented in Section 3.2. The only difference is that the strict filter in Section 3.2 is replaced by a much weaker criterion described in Section 4.2, which excludes only enterprises with very large discrepancies between EMPTOT_S and EMPTOT_R. The justification is that the quality measures are calculated after the data harmonization described in the next section. The exclusion of data with inconsistencies is then to a wide extent replaced by a treatment of inconsistencies.

4 Harmonization of employment variables

The FTE model is based on matched data (survey-register). This dataset shows differences between the variables number of employees reported in the survey (EMPTOT_S) and number of employees reported in the OASI register (EMPTOT_R). These differences may for instance be due to measurement errors or to discrepancies regarding the definition of employment and its interpretation between survey and OASI register data. In what follows we present the different treatments that we have done in order to overcome inconsistency.

4.1 Prorating

In the literature, we can find different methods for variable adjustment and treating inconsistent data like, for example, *prorating* and *generalized ratio adjustment* (Pannekoek, J. (2011); Pannekoek (2014)). For instance, the *prorating* method is a simple multiplicative adjustment applied on variables involved in editing rules. The first approach we present is based on these adjustment methods. It was used for the temporary results STATENT 2011 which appeared at the end of the year 2013. The basic idea is that, for each enterprise i, we define a new variable FTE_R (by gender) as follows:

$$\text{FTE_R}_i = \eta_i \text{FTE_S}_i, \tag{8}$$

where $\eta_i = \text{EMPTOT_R}_i/\text{EMPTOT_S}_i$, i.e. we use the MOL of the survey to impute the MOL of the OASI register.
This new variable which is now *harmonized with the OASI register data* can be used as the dependent variable in the FTE model (4). In this way predicted FTE will be consistent with the values of EMPTOT_R. If the inconsistencies are treated according to Equation (8), we can rewrite the equation of Model (4) as follows:

$$\eta_i \text{FTE_S}_i = \alpha_1 V_{i1} + \sum_{j=2}^{4} \alpha_{jkl} V_{ij} + \epsilon_i, \tag{9}$$

or

$$\text{FTE_S}_i = \alpha_1 \frac{V_{i1}}{\eta_i} + \sum_{j=2}^{4} \alpha_{jkl} \frac{V_{ij}}{\eta_i} + \frac{\epsilon_i}{\eta_i}, \tag{10}$$

In the case $\eta_i > 1$, we can interpret Equation (10) as follows: the adjustment between EMPTOT_R and EMPTOT_S is done uniformly in the four salary classes by reducing the number of employees with the factor η_i. If the inconsistencies are independent of the salary classes, then this approach could be justified. However, analysis of records on a case to case basis put some doubt on the general validity of such an assumption.

4.2 A harmonization approach that accounts for the type of employment

The prorating approach was not satisfactory in the sense that we could not take into account information about the type of employment (part-time or full-time) or the distribution of employees in the salary classes. So, we looked for an alternative method in which the inconsistencies in the number of employees are no longer treated in an uniform manner. We consider separately the following three cases:

- Case 1: EMPTOT_R $>$ EMPTOT_S.

- Case 2: EMPTOT_R $<$ EMPTOT_S.
- Case 3: EMPTOT_S $=$ EMPTOT_R.

In case 3 no inconsistency is detected and therefore no harmonization is needed. Therefore we may concentrate ourselves on case 1 and case 2. Let us define

$$\text{diff_ab} = \text{EMPTOT_R} - \text{EMPTOT_S}$$

and

$$\text{diff_ba} = \text{EMPTOT_S} - \text{EMPTOT_R},$$

the differences (by gender) in the number of employees coming from the two sources. If, for an enterprise, the difference between these two variables is very large, we think that the two data sources are irreconcilable and we remove it from the dataset used for model fitting. More precisely, if any of the following conditions is fulfilled (either for men or women), then the enterprise is treated as if only OASI register data is available. These conditions are based on the assumption that the variance of observed differences is proportional to the total number of employees (EMPTOT_R in case 1 and EMPTOT_S in case 2) and are defined as follows:

$$\text{diff_ab}/\sqrt{\text{EMPTOT_R}} > 5$$
$$\text{diff_ba}/\sqrt{\text{EMPTOT_S}} > 5$$

For the year 2011, among all enterprises presenting differences in the variables number of employees (case 1 and case 2), 2.12% of the enterprises are excluded using this filter. If we look at what part this represents among all enterprises used for modelling, then this percentage becomes smaller (only 1.48% of the total number of enterprises). We have examined different cut-off points. The choice of the value 5 was deliberate and was done in order to, firstly, account for really large differences and, secondly, not exclude a too large number of enterprises.

4.2.1 Treatment of case 1

As EMPTOT_R $>$ EMPTOT_S, our aim is to reduce EMPTOT_R and the V_{ij} so that they are coherent with EMPTOT_S. Knowing the number of employees in each salary class, we estimate the coefficients of the following model (by gender and economic activity sector):

$$\text{diff_ab}_i = \sum_{j=1}^{4} \beta_j V_{ij} + \epsilon_i,$$

under the hypothesis $\text{Var}(\epsilon_i) = \sigma^2 \text{EMPTOT_R}_i$. This is not done with the aim to predict the difference between EMPTOT_R and EMPTOT_S. We are rather interested in the estimated coefficients, $\widehat{\beta}_j$, which can be interpreted as an estimation of the proportion of persons in the salary class j which are in the OASI register but not in the survey.

Table 6 contains the estimated coefficients, $\widehat{\beta}_j$, obtained with the ROBUSTREG procedure in SAS with weights proportional to $1/\text{EMPTOT_R}_i$. It indicates, for example, that the coefficients for the class of employees having the smallest wages (salary class 1) are larger than the coefficients for the salary class 2. Thus, at least in this subpopulation (case 1), a uniform treatment using Equation (8) seems not to be appropriate.

	$\widehat{\beta}_1$		$\widehat{\beta}_2$		$\widehat{\beta}_3$		$\widehat{\beta}_4$	
	Estimator	StdErr	Estimator	StdErr	Estimator	StdErr	Estimator	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 6 Values of $\widehat{\beta}_j$ and the estimated standard errors by gender (m/w) and economic activity sector (s2/s3)

Initial idea - probability proportional to size (PPS) sampling of fixed size

For each enterprise i, we suppose that the persons which are not in the survey come from a sample s_i of fixed size $n_i = \text{diff_ab}_i$ of probabilities proportional to $\widehat{\beta}_j$ (see Table 6). The probability that a person d belonging to the set of wage-earners in the salary class j is not counted in the survey is therefore given by:

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

where $\text{mos}_d = \widehat{\beta}_j$ and $\text{mos}_i = \sum_{d \in i} \text{mos}_d = \sum_{j=1}^{4} \widehat{\beta}_j V_{ij} = \widehat{n_i}$.

If $\widehat{\beta}^*_{ij} \geq 1$, then the person will be automatically removed (see Särndal, C.E., Swensson, B. and Wretman, J., 1992, p.89).

Calculation of the average number of persons which should be removed in each salary class

The inconveniences of using PPS sampling of persons are its random aspect with a potential impact on comparability over years and the difficulty of implementation in production (matched data is on enterprise and not on personal level). This has led us to develop an adapted procedure in order to replace the random PPS sample. Instead of drawing a sample, we calculate the expected number of persons in salary class j to be selected in the sample s_i. This number is given by:

$$E\left[\sum_{d \in \mathcal{V}_{ij}} 1(d \in s_i)\right] = V_{ij} P(d \in s_i) = V_{ij} \widehat{\beta}^*_{ij}, \tag{11}$$

where $\mathcal{V}_{ij}$ denotes the set of employees of enterprise i in the salary class j, and $1(d \in s_i)$ is an indicator variable denoting if a person d is selected in the sample s_i.
As in the case of PPS sampling, our procedure first removes all persons for which $\widehat{\beta}^*_{ij} \geq 1$. Next, we calculate the mean number of persons which has to be eliminated according to Equation (11). At the end of this procedure, we obtain new variables

$$\widetilde{V}_{ij} = V_{ij} - V_{ij} \widehat{\beta}^*_{ij},$$

such that $\sum_{j=1}^{4} \widetilde{V}_{ij} = \text{EMPTOT_S}_i$. These new variables will replace the variables V_{ij} in the estimation of model (4), where y_i will be given by FTE_S_i.

4.2.2 Treatment of case 2

If EMPTOT_S $>$ EMPTOT_R, we want to reduce EMPTOT_S and T_{ij} so that they are coherent with EMPTOT_R. Knowing the number of employees working at part-time III (T_{i1}) , part-time II (T_{i2}) , part-time I (T_{i3}) and full-time (T_{i4}) from the survey data, we estimate the following model (by gender and economic activity sector):

$$\text{diff_ba}_i = \sum_{j=1}^{4} \gamma_j T_{ij} + \epsilon_i,$$

The used procedure is PROC ROBUSTREG in SAS with weights proportional to $1/\text{EMPTOT_S}_i$. The estimated coefficients as well as the corresponding standard errors are given in Table 7. For example, the estimated coefficients $\widehat{\gamma}_1$ can be seen as estimation of the proportion of persons working at part-time III that are in the survey but not in the register. It can be seen from Table 7 that the coefficients for the persons working at part-time III are larger than the coefficients for the persons working full-time.

	$\widehat{\gamma}_1$		$\widehat{\gamma}_2$		$\widehat{\gamma}_3$		$\widehat{\gamma}_4$	
	Estimator	StdErr	Estimator	StdErr	Estimator	StdErr	Estimator	StdErr
m/s2	0.469	0.039	0.657	0.041	0.256	0.032	0.071	0.002
m/s3	0.448	0.014	0.237	0.011	0.203	0.014	0.123	0.003
w/s2	0.548	0.025	0.329	0.019	0.063	0.018	0.065	0.007
w/s3	0.442	0.010	0.251	0.008	0.002	0.007	0.140	0.005

Table 7 Values of $\widehat{\gamma}_j$ and the estimated standard errors by gender (m/w) and economic activity sector (s2/s3)

Adaptation of the harmonization procedure to case 2

We know that, in order to eliminate the differences between the employment variables EMPTOT_S and EMPTOT_R and make them consistent, we have to eliminate, for each enterprise, a fixed number of persons, $n_i^* = \text{diff_ba}_i$. Using the coefficients $\widehat{\gamma}_j$, we apply the same procedure as for case 1, with the required modifications. In this way we obtain the new variables $\widetilde{T}_{ij}$ such that $\sum_{j=1}^{4} \widetilde{T}_{ij} = \text{EMPTOT_R}_i$.

Elimination procedure for case 2 based on the average number of persons that should be eliminated in each occupational level class

Again, we calculate the average number of persons of the class j which are selected in the sample s_i^* of persons which are in the OASI register but not in the survey. This number is given by:

$$E\left[\sum_{d \in T_{ij}} 1(d \in s_i^*)\right] = T_{ij} P(d \in s_i^*) = T_{ij}\widehat{\gamma}_{ij}^*. \tag{12}$$

As in the case of PPS sampling, our procedure first removes all persons for which $\widehat{\gamma}_{ij}^* \geq 1$. Next, we calculate the mean number of persons which has to be eliminated according to Equation (12). At the end of the procedure we obtain new variables denoted by

$$\widetilde{T}_{ij} = T_{ij} - T_{ij}\widehat{\gamma}_{ij}^*,$$

such that $\sum_{j=1}^{4} \widetilde{T}_{ij} = \text{EMPTOT_R}_i$.

Harmonization of the variable FTE_S

Let us suppose that FTE_S can be modeled as follows:

$$\text{FTE_S}_i = \sum_{j=1}^{4} \delta_j T_{ij} + \epsilon_i, \tag{13}$$

where $\text{Var}(\epsilon_i) = \sigma^2 \text{EMPTOT_S}_i$. The estimated model, by gender and economic activity sector, is adjusted via PROC GLM from SAS with weights proportional to $1/\text{EMPTOT_S}$. The estimated coefficients and their standard errors are given in Table 8.

	$\widehat{\delta}_1$		$\widehat{\delta}_2$		$\widehat{\delta}_3$		$\widehat{\delta}_4$		
	Est.	StdErr	Est.	StdErr	Est.	StdErr	Est.	StdErr	R^2
m/s2	0.110	0.006	0.300	0.007	0.632	0.005	0.998	0.000	0.999
m/s3	0.078	0.004	0.278	0.003	0.645	0.004	0.998	0.001	0.997
w/s2	0.095	0.008	0.293	0.006	0.654	0.006	0.994	0.002	0.997
w/s3	0.091	0.004	0.281	0.003	0.660	0.003	0.988	0.002	0.992

Table 8 Values of $\widehat{\delta}_j$ and the estimated standard errors

Using the estimated coefficients of Model (13), we calculate a new variable, adjusted from FTE_S, denoted by FTE_R, that is coherent with the variable EMPTOT_R:

$$\text{FTE_R}_i = \min\left(\text{FTE_S}_i \frac{\sum \widehat{\delta}_j \widetilde{T}_{ij}}{\sum \widehat{\delta}_j T_{ij}}, \text{EMPTOT_R}_i\right). \tag{14}$$

We can explain the minimum in Equation (14) by the fact that MOL of harmonized data, FTE_R/EMPTOT_R, should be bounded by 1.

4.3 Variables used to estimate the FTE model after data harmonization

Table 9 presents the variables used in the estimation of the model of Equation (4) in cases 1, 2 and 3, respectively. From the table, it can be seen that in case 1 it is the explanatory variables that are adjusted in order to correspond to FTE recorded in the survey. In case 2, where the total number of employees is larger for the survey than for the OASI register, the explanatory variables remain unchanged but the dependent variable for the model is adjusted. The three datasets were set together and unique dependent and independent variables were created.

Case	Variable of interest	Explanatory variables
1	FTE_S	$\widetilde{V}_{ij}$
2	FTE_R	V_{ij}
3	FTE_S	V_{ij}

Table 9 Variables used in Model (4)

The fit of the model as well as the calculation of the quality measures described in Section 3.5 are based on these harmonized data.

4.4 Consequences of data harmonization for constructing FTE in STATENT

In this section, we show how we calculate FTE for enterprises for which survey data is available. For all the other enterprises, the prediction model (4) and OASI register data are used.

- Case 1: Let us denote

$$\text{FTE_R}_{i,model} = \sum_{j=1}^{4} \widehat{\alpha}_j V_{ij}$$

the FTE calculated using the estimated FTE model parameters and the variables V_{ij}, and

$$\text{FTE_S}_{i,model} = \sum_{j=1}^{4} \widehat{\alpha}_j \widetilde{V}_{ij}$$

the FTE calculated using the FTE model parameters and the variables $\widetilde{V}_{ij}$.
Note that, for simplicity of notation, $\widehat{\alpha}_j$ is written here without possible parameter adjustments, see Equation (6). Then, the ratio of these variables is applied to FTE_S in order to produce the harmonized FTE, denoted by FTE_R. Thus,

$$\text{FTE_R}_i = \min\left(\text{FTE_S}_i \frac{\text{FTE_R}_{i,model}}{\text{FTE_S}_{i,model}}, \text{EMPTOT_R}_i\right). \tag{15}$$

The minimum is explained by the fact that mean occupational level of harmonized data should be bounded by 1.

- Case 2: the harmonized FTE_R is calculated by applying a multiplicative or enhanced ratio adjustment as defined in Equation (14).

- Case 3: in this case we have that FTE_R $=$ FTE_S.

To summarize: the simple ratio adjustment based on the number of employees is replaced by an enhanced ratio adjustment using full-time equivalents based on the FTE model.

4.5 Effects of the harmonization

We illustrate the effects of harmonization on the data by a few examples taken from real data. Figure 2 shows a comparison, for case 1, between the MOL of survey and OASI register data. From the graph it can be seen that in general MOL based on the enhanced ratio adjustment (mol_harm) are smaller than MOL based on the simple ratio adjustment (mol_ratio) defined in Section 4.1. This can be explained by the fact that the surplus of employment in the OASI register corresponds rather to small FTE while the simple ratio adjustment is based on the hypothesis that MOL in the OASI register are the same as in the survey (the values on the diagonal of the graph).
Next, Table 10 shows the values of $\widetilde{V}_{ij}$ for the female employees of a given enterprise according to the OASI register and to the survey. For the second economic activity sector, model for women, the estimated FTE model coefficients are, respectively, $\widehat{\alpha}_1 = 0.223$, $\widehat{\alpha}_2 = 0.364$, $\widehat{\alpha}_3 = 0.601$ and $\widehat{\alpha}_4 = 0.940$. For this particular example it can be seen that the enterprise has 9 employees according to the OASI register (variable EMPTOT_R) and 5 employees according to the survey (variable EMPTOT_S) and FTE_S is equal to 4. This enterprise falls in the case 1 for data treatment. As for this enterprise $\widehat{\beta}_1 = 0.724$, we will proportionally eliminate most of the employees in the first salary class. The values of the harmonized FTE for this enterprise, using the simple and enhanced ratio adjustments, are given in Table 11. We note

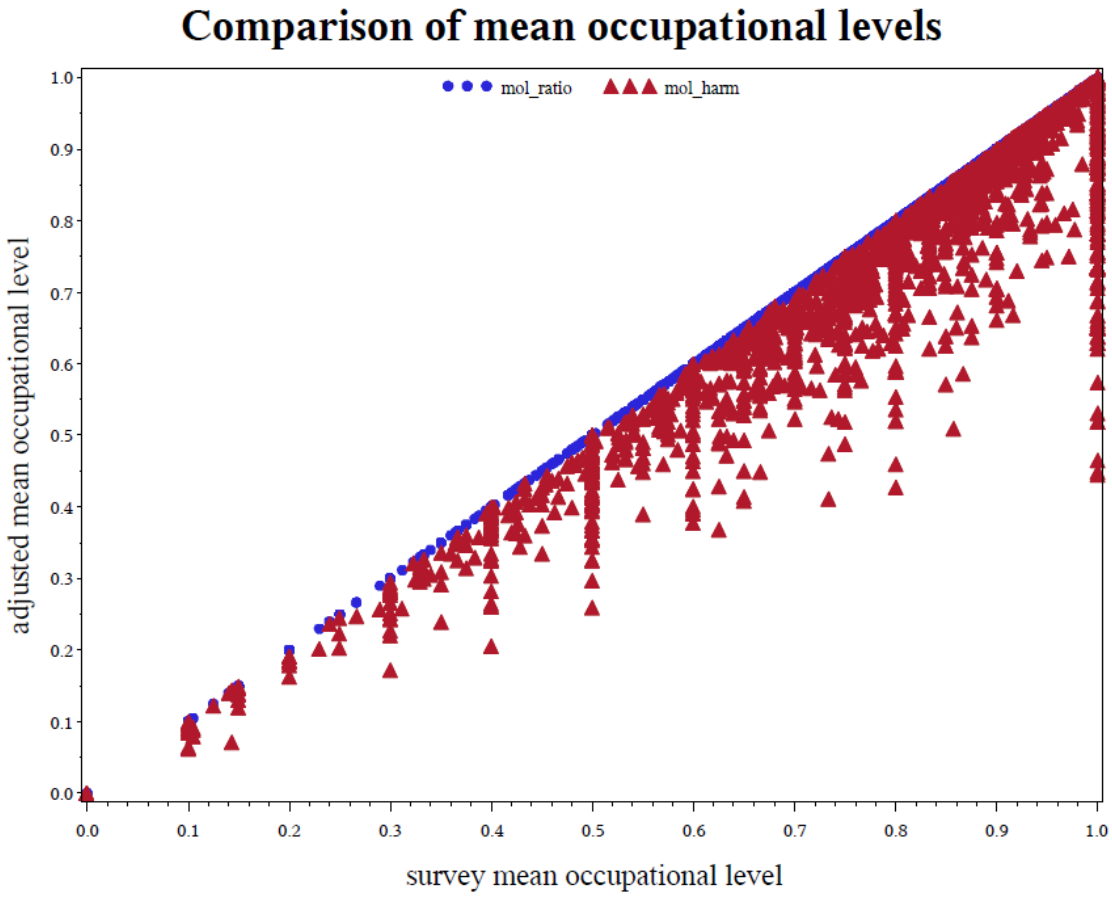


Figure 2 Mean occupational levels, women /sector 2

that, FTE_R$_{i,model} = 4.915$ and FTE_S$_{i,model} = 3.692$. The ratio of these two values applied to FTE_S gives the value of the harmonized FTE using the enhanced adjustment as given in Equation (15). We can see from Table 11 that the harmonized FTE (enhanced adjustment) is much more in accordance to the value of FTE_S than the one using the simple ratio adjustment (according to Equation (8)) and that the MOL is smaller after harmonization. Thus, this result is consistent with the results of Figure 2.

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

Table 10 Number of employees per salary class

5 Estimation of the FTE model and special cases

We have seen that, at the end of the data treatment process, we have three different cases, with explanatory and dependent variables given in Table 9. In the following, we discuss the output of

	Source		Adjustment	
	Survey (S)	OASI (R)	simple	enhanced
EMPTOT	5	9		
FTE	4		**7.2**=$4 \times \frac{9}{5}$	**5.325**=4 $\times \frac{4.915}{3.692}$
FTE model	3.692	4.915		
MOL			**0.8**	**0.59**

Table 11 Effects of harmonization on FTE

estimation of the FTE model presented in Section 3.3 which is fitted on a unique dataset which comprises the three different cases. The procedures used for selecting and fitting the model are described in Section 3.

5.1 Estimated parameters

Table 12 gives the values of $\widehat{\alpha}_j$, the estimated main parameters of Model (6) and their corresponding standard errors. Note that for the model "men/sector 2" (denoted as m/s2 in the table) the number of classes was reduced to three (V_{i3} and V_{i4} were put together) as fitted model parameters showed no relevant differences between these two salary classes. What we can see from the table is that the estimated main coefficients are gradually increasing. The standard errors are small, indicating stability of the model.

	$\widehat{\alpha}_1$		$\widehat{\alpha}_2$		$\widehat{\alpha}_3$		$\widehat{\alpha}_4$		
Model	Est.	StdErr	Est.	StdErr	Est.	StdErr	Est.	StdErr	R^2
m/s2	0.371	0.024	0.807	0.006	0.990	0.001			0.996
m/s3	0.279	0.016	0.449	0.008	0.692	0.006	0.973	0.001	0.979
w/s2	0.223	0.014	0.364	0.007	0.601	0.006	0.940	0.003	0.968
w/s3	0.203	0.007	0.312	0.004	0.594	0.003	0.917	0.002	0.958

Table 12 Main model parameters for STATENT 2011

5.2 Quality measures

Tables 13 and 14 present the values (in percentages) of the quality measure *reldif*, defined in Equation (7), at NOGA OFS50 level for sectors 2 and 3, which correspond to the considered 46 levels of NOGA OFS50, and by GREG, respectively. What we can see from these tables is that the prediction quality of our model is quite good. The largest differences are observed for the population of women in sector 3 (min = -2.89%; max= 2.24%) for OFS50 in $(68, 73.5)$. The majority of the measures lie in the interval $(-1\%; 1\%)$. These results are also illustrated by the boxplots in Figure 3.

6 Conclusions

This report documents the data and the methodology used for constructing the variable FTE for the new enterprise statistics STATENT 2011. For all enterprises not contained in a survey,

NOGA	*reldif* in %		NOGA	*reldif* in %	
OFS 50	Men	Women	OFS 50	Men	Women
5.9	-0.04	-0.25	53	-0.05	0.03
10.2	-0.23	-0.05	55	0.24	0.65
13.5	0.12	0.89	56	-0.55	0.34
16.8	0.06	-1.20	58.6	0.85	2.07
19.2	-0.01	-0.08	61	0.06	-0.08
21	-0.00	-0.00	62.3	0.55	0.48
22.3	0.02	-0.10	64	-0.01	0.02
24.5	-0.08	-1.23	65	-0.56	0.06
26	-0.02	0.07	66	0.74	1.14
27	-0.03	0.01	68	1.67	-2.89
28	-0.02	-0.72	69	-0.20	0.06
29.3	0.10	-0.38	70	0.63	0.10
31.3	-0.19	-1.91	71	0.02	-1.16
35	-0.02	-0.09	72	-0.02	-0.02
36.9	-0.93	-1.52	73.5	1.04	2.24
41.2	-0.00	-0.39	79.2	-0.06	-0.82
43	0.09	-0.02	84	-0.09	0.05
45	0.85	-0.43	85	-0.01	0.05
46	-0.04	-0.68	86	-0.02	-0.40
47	-0.17	0.70	87	-0.24	0.12
49	-0.03	0.23	88	-0.25	0.80
50.1	0.16	0.67	90.3	-0.11	-1.16
52	0.08	0.27	94.6	-0.55	0.77

Table 13 Quality measure *reldif* by NOGA OFS50

sector 2	*reldif* in %		sector 3	*reldif* in %	
GREG	Men	Women	GREG	Men	Women
1	-0.07	0.00	1	0.15	0.21
2	-0.10	-0.24	2	-0.07	-0.26
3	-0.00	-0.03	3	0.21	-0.06
4	-0.00	-0.45	4	-0.04	-0.21
5	0.00	-0.24	5	0.29	0.66
6	-0.15	-0.23	6	0.39	0.12
7	0.10	-0.10	7	0.02	1.33
total	-0.02	-0.17	total	0.09	0.06

Table 14 Quality measure *reldif* by GREG

FTE must be estimated using an appropriate technique. The finally retained approach based on linear regression, predicting the FTE of an enterprise by means of the variables number of employees in four different salary classes, is relatively easy to interpret and to understand. The model (4 different models adjusted by sex and sector) is based on carefully chosen salary classes which reflect different occupational levels and allows for separate slopes for NUTS-2 and NOGA sections, where this improves the model in terms of AIC. We have seen also that

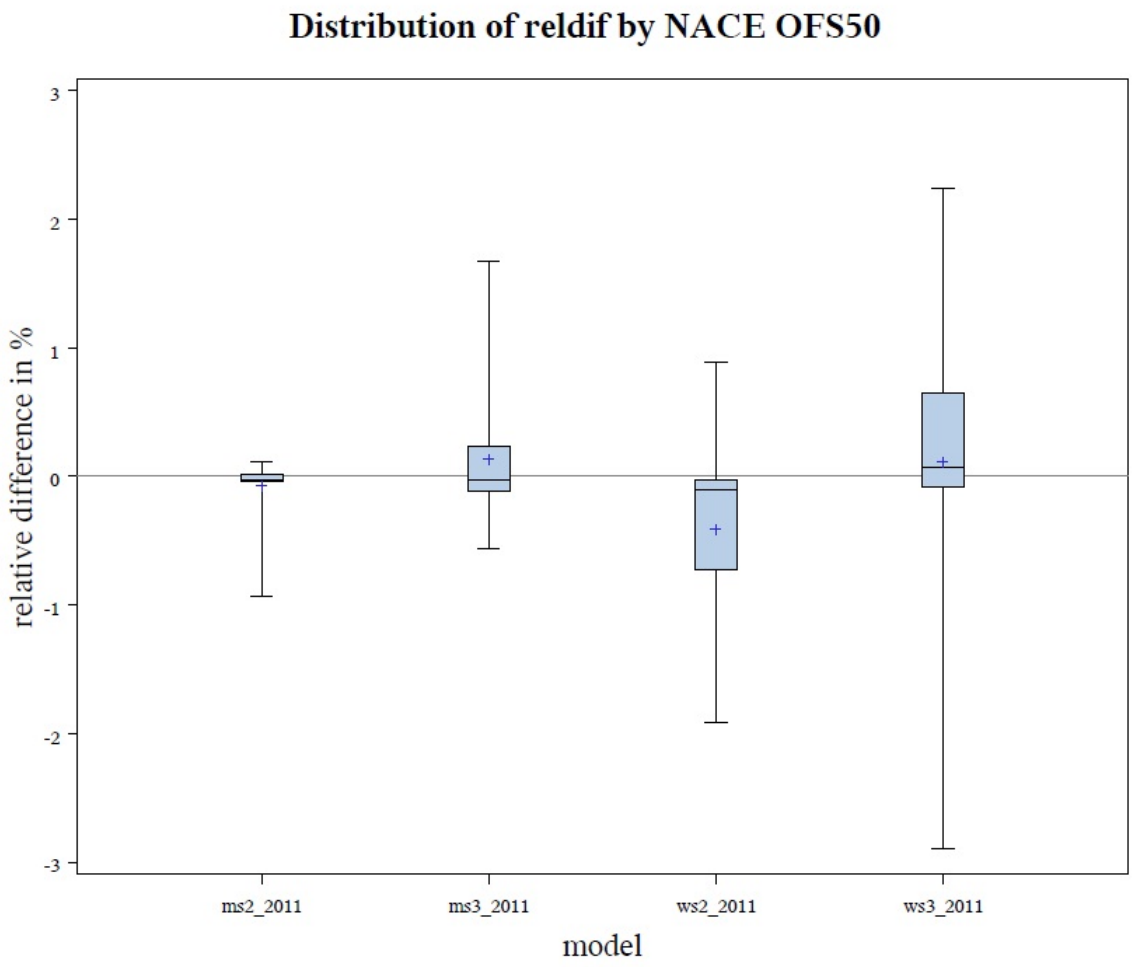


Figure 3 Quality measure *reldif* by NOGA OFS50

the estimates of the model parameters are relatively stable over time (parsimonious model by model selection procedure, robustification by reducing the weights of potential leverage points) and leads to satisfactory prediction accuracy assessed by the considered quality measures. Quality measures proved to be a very important tool when assessing and comparing the performance of different models for predicting FTE. For the majority of NOGA OFS50, the estimated relative prediction error for men is in the range $(-1\%; 1\%)$. For women, the estimated relative prediction errors are larger, but for 36 out of the 46 considered NOGA OFS50, they are still in the range $(-1\%; 1\%)$.

The integrated data revealed inconsistences concerning the variable number of employees coming from different data sources. As this variable is closely related to the variable number of FTE recorded from the survey, data inconsistency has an important impact on modelling. The procedure we have developed for treating the inconsistencies is based on probability-proportional to size sampling of fixed size and takes into account information about type of employment. However, it has the inconvenience to be random and difficult to apply in practice. By using expected sample sizes instead of random sampling for adjusting the number of employees in the different salary classes we have succeeded to overcome this inconvenience for treating the inconsistencies.

In our methodology, data harmonization has two main applications. The first is to treat the in-

consistencies between the independent and dependent variables of the FTE estimation model. The second concerns the cases where both information from the registers and from the surveys is available for constructing FTE for STATENT. In that case, the value of FTE coming from the survey is adjusted in such a way that available survey information becomes coherent with register data.
This kind of modelling, based on integrated data, has been a pioneer in the domain of business statistics at FSO and showed its feasability. However, modelling in this context highlited several unpredicted obstacles which had to be tackled.

Appendix

Economic activity sector	NOGA section	Title
1	A	Agriculture, forestry and fishing
2	B	Mining and quarrying
2	C	Manufacturing
2	D	Electricity, gas, steam and air-conditioning supply
2	E	Water supply, waste management
2	F	Construction
3	G	Wholesale and retail trade; repair of motor vehicles and motorcycles
3	H	Transportation and storage
3	I	Accommodation and food service activities
3	J	Information and communication
3	K	Financial and insurance activities
3	L	Real estate activities
3	M	Professional, scientific and technical activities
3	N	Employment activities
3	O	Public administration
3	P	Education
3	Q	Human health and social work activities
3	R	Arts, entertainment and recreation
3	S	Other service activities
3	T	Activities of households as employers and producers
3	U	Activities of extraterritorial organisations

Table 15 NOGA Sections

Section	Code	Title
A	01	Crop and animal production, hunting and related service activities
A	02	Forestry and logging
A	03	Fishing and aquaculture
B	05	Mining of coal and lignite
B	06	Extraction of crude petroleum and natural gas
B	07	Mining of metal ores
B	08	Other mining and quarrying
B	09	Mining support service activities
C	10	Manufacture of food products
C	11	Manufacture of beverages
C	12	Manufacture of tobacco products
C	13	Manufacture of textiles
C	14	Manufacture of wearing apparel
C	15	Manufacture of leather and related products
C	16	Manufacture of wood and of products of wood and cork, except furniture
C	17	Manufacture of paper and paper products
C	18	Printing and reproduction of recorded media
C	19	Manufacture of coke and refined petroleum products
C	20	Manufacture of chemicals and chemical products
C	21	Manufacture of basic pharmaceutical products and pharmaceutical preparations
C	22	Manufacture of rubber and plastic products
C	23	Manufacture of other non-metallic mineral products
C	24	Manufacture of basic metals
C	25	Manufacture of fabricated metal products, except machinery and equipment
C	26	Manufacture of computer, electronic and optical products
C	27	Manufacture of electrical equipment
C	28	Manufacture of machinery and equipment n.e.c.
C	29	Manufacture of motor vehicles, trailers and semi-trailers
C	30	Manufacture of other transport equipment
C	31	Manufacture of furniture
C	32	Other manufacturing
C	33	Repair and installation of machinery and equipment
D	35	Electricity, gas, steam and air-conditioning supply
E	36	Water collection, treatment and supply
E	37	Sewerage
E	38	Waste collection, treatment and disposal activities; materials recovery
E	39	Remediation activities and other waste management services
F	41	Construction of buildings
F	42	Civil engineering
F	43	Specialized construction activities

Table 16 NOGA Divisions 01 to 43

Section	Code	Title
G	45	Wholesale and retail trade and repair of motor vehicles and motorcycles
G	46	Wholesale trade, except of motor vehicles and motorcycles
G	47	Retail trade, except of motor vehicles and motorcycles
H	49	Land transport and transport via pipelines
H	50	Water transport
H	51	Air transport
H	52	Warehousing and support activities for transportation
H	53	Postal and courier activities
I	55	Accommodation
I	56	Food and beverage service activities
J	58	Publishing activities
J	59	Motion picture, video and television programme production
J	60	Programming and broadcasting activities
J	61	Telecommunications
J	62	Computer programming, consultancy and related activities
J	63	Information service activities
K	64	Financial service activities, except insurance and pension funding
K	65	Insurance, reinsurance and pension funding, except compulsory social security
K	66	Activities auxiliary to financial services and insurance activities
L	68	Real estate activities
M	69	Legal and accounting activities
M	70	Activities of head offices; management consultancy activities
M	71	Architectural and engineering activities; technical testing and analysis
M	72	Scientific research and development
M	73	Advertising and market research
M	74	Other professional, scientific and technical activities
M	75	Veterinary activities
N	77	Rental and leasing activities
N	78	Employment activities
N	79	Travel agency, tour operator reservation service and related activities
N	80	Security and investigation activities
N	81	Services to buildings and landscape activities
N	82	Office administrative, office support and other business support activities
O	84	Public administration and defence; compulsory social security
P	85	Education
Q	86	Human health activities
Q	87	Residential care activities
Q	88	Social work activities without accommodation
R	90	Creative, arts and entertainment activities
R	91	Libraries, archives, museums and other cultural activities
R	92	Gambling and betting activities
R	93	Sports activities and amusement and recreation activities
S	94	Activities of membership organisations
S	95	Repair of computers and personal and household goods
S	96	Other personal service activities
T	97	Activities of households as employers of domestic personnel
T	98	Undifferentiated goods- and services-producing activities
		of private households for own use
U	99	Activities of extraterritorial organisations and bodies

Table 17 NOGA Divisions 45 to 99

Position	Code	Divisions	Title
1	A	01 to 03	Agriculture, forestry and fishing
2	B	05 to 09	Mining and quarrying
3	CA	10 to 12	Manufacture of food and tobacco products
4	CB	13 to 15	Manufacture of textiles and apparel
5	CC	16 to 18	Manufacture of wood and paper products, and printing
6	CD + CE	19 + 20	Manufacture of coke, chemicals and chemical products
7	CF	21	Manufacture of pharmaceutical products
8	CG	22+23	Manufacture of rubber and plastics products
9	CH	24+25	Manufacture of metal products
10	CI	26	Manufacture of computer, electronic and optical products;
			watches and clocks
11	CJ	27	Manufacture of electrical equipment
12	CK	28	Manufacture of machinery and equipment n.e.c.
13	CL	29+30	Manufacture of transport equipment
14	CM	31 to 33	Other manufacturing, repair and installation
15	D	35	Electricity, gas, steam and air-conditioning supply
16	E	36 to 39	Water supply, waste management
17	F	41 to 42	Construction of buildings and Civil engineering
18	F	43	Specialized construction activities
19	G	45	Trade and repair of motor vehicles and motorcycles
20	G	46	Wholesale trade
21	G	47	Retail trade
22	H	49	Land transport and transport via pipelines
23	H	50 to 51	Water transport and Air transport
24	H	52	Warehousing and support activities for transportation
25	H	53	Postal and courier activities
26	I	55	Accommodation
27	I	56	Food and beverage service activities
28	JA	58 to 60	Publishing, audiovisual and broadcasting activities
29	JB	61	Telecommunications
30	JC	62 + 63	IT and other information services
31	K	64	Financial service activities
32	K	65	Insurance
33	K	66	Activities auxiliary to financial services and insurance activities
34	L	68	Real estate activities
35	M	69	Legal and accounting activities
36	M	70	Activities of head offices; management consultancy activities
37	M	71	Architectural and engineering activities
38	M	72	Scientific research and development
39	MC	73 to 75	Other professional, scientific and technical activities
40	N	77 + 79 to 82	Administrative and support service activities
41	N	78	Employment activities
42	O	84	Public administration
43	P	85	Education
44	QA	86	Human health activities
45	QB	87	Residential care activities
46		88	Social work activities without accommodation
47	R	90 to 93	Arts, entertainment and recreation
48	S	94 to 96	Other service activities
49	T	97 + 98	Activities of households as employers and producers
50	U	99	Activities of extraterritorial organisations

Table 18 NOGA OFS50

BC	Business census
BR	Business register
EMPTOT	Total employment
EMPTOT_S	Total employment according to survey
EMPTOT_R	Total employment according to register
ERST	Survey on new enterprises
EUNT	Single-establishment enterprise
FTE	Full-time equivalents
GREG	Main geographical region
JobStat	Quarterly survey of employment
MOL	Mean occupational level
MUNT	Multi-establishment enterprise
NACE / NOGA	Statistical classification of economic activities
NUTS-2	Main geographical region
absdif	Absolute difference
reldif	Relative difference
STATENT	Swiss structural business statistics
OASI	Old-age and survivors' insurance

Table 19 Table of abbreviations

References

[Akaike 1973] AKAIKE, H.: Information theory and an extension of the maximum likelihood principle. In: *Second International Symposium on Information Theory*, 1973, S. 267–281

[Chen 2002] CHEN, C.: Robust Regression and Outlier Detection with the ROBUSTREG procedure. In: *Proceedings of the SUGI conference*, URL http://www2.sas.com/proceedings/sugi27/p265-27.pdf, 2002

[Cohen 2009] COHEN, R. A.: Introducing the GLMSELCT Procedure for Model Selection. In: *Proceedings of the SUGI conference*, URL http://www2.sas.com/proceedings/sugi31/207-31.pdf, 2009

[EUROSTAT 2008] EUROSTAT: NACE Rev. 2 - Statistical classification of economic activities in the European Community / EUROSTAT. URL http://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF, 2008. – Technical report.

[Graf, M. and Potterat, J.] GRAF, M. AND POTTERAT, J.: Construction of Ful Time Equivalent for the Swiss Business Frame. In: *The 2nd European Establishment Statistics Workshop, Neuchâtel, Switzerland, 12-14 September 2011*, URL http://www1.unece.org/stat/platform/display/ENBES/Programme+EESW11#ProgrammeEESW11-Session2

[Hastie, T., Tibshirani, R. and Friedman, J. 2009] HASTIE, T., TIBSHIRANI, R. AND FRIEDMAN, J.: *The elements of statistical learning: data mining, inference and prediction*. Second. Springer, 2009

[Newman 1993] NEWMAN, Michael C.: Regression analysis of log-transformed data: statistical bias and its correction. In: *Environmental Toxicology and Chemistry* 12 (1993), S. 1129–1133

[Pannekoek 2014] PANNEKOEK, J.: Method: Reconciling Conflicting Microdata. In: *Memobust Handbook on Methodology of Modern Business Statistics*. URL http://www.cros-portal.eu/content/reconciling-conflicting-microdata-method, 2014

[Pannekoek, J. 2011] PANNEKOEK, J.: Models and algorithms for micro-integration. URL http://www.cros-portal.eu/content/wp2-development-methods, 2011. – Final Report on WP2: Methodological developments, ESSNET on Data Integration

[Renaud, A. 2008] RENAUD, A.: Statistique de l'emploi. Révision 2007 : cadre de sondage et échantillonage. / Swiss Federal Statistical Office. 2008. – Methodological report.

[Renaud, A., Panchard, C. and Potterat, J. 2008] RENAUD, A., PANCHARD, C. AND POTTERAT, J.: Statistique de l'emploi. Révision 2007 : méthodes d'estimation. / Swiss Federal Statistical Office. 2008. – Methodological report.

[Rousseeuw 1984] ROUSSEEUW, P.J: Least Median of Squares Regression. In: *Journal of the American Statistical Association* 79 (1984), S. 871–880

[Särndal, C.E., Swensson, B. and Wretman, J. 1992] SÄRNDAL, C.E., SWENSSON, B. AND WRETMAN, J.: *Model Assisted Survey Sampling*. Springer, 1992

Methodenberichte der Sektion Statistische Methoden des BFS
Rapports de méthodes de la section méthodes statistiques de l'OFS
Methodology reports published by the FSO's Statistical Methods Section

Nedyalkova, D., Assoulin, D. (2017). Construction of full-time equivalents for the Swiss structural business statistics 2011. Order number: 338-0077

Ferster, M. (2016). Energieverbrauchsstatistik EVS2014 - Stichprobe, Hochrechnung und Vergleichbarkeit mit der EVS2013. Bestellnummer: 338-0076

Panchard, C. (2014). Enquête sur la participation aux activités culturelles 2008. Plan de sondage et estimations. Bestellnummer: 338-0073

Assoulin, D. (2014). Wertschöpfungsstatistik. Revision 2009: Statistische Datenaufbereitung und Hochrechnung. Bestellnummer: 338-0072

Wilhelm, M. (2014). Echantillonnage boule de neige. La méthode de sondage déterminé par les répondants. Numéro de commande: 338-0071

Assoulin, D. (2013). Wertschöpfungsstatistik. Revision 2009: Stichprobenrahmen und Stichprobenplan. Bestellnummer: 338-0070

Ferster, M. (2013). EVS I - Energieverbrauchsstatistik 2002 bis 2007: Stichprobenplan und Hochrechnung. Bestellnummer: 338-0069

Assoulin, D. (2013). Zusatzerhebung für die landwirtschaftliche Betriebszählung 2010: Stichprobenplan und Hochrechnung. Bestellnummer: 338-0068

Potterat, J. (2012). Use of conversion keys for NOGA2002-2008. Order number: 338-0067-05

Andrade, B., Salamin P.-A. (2012). Enquête sur la situation sociale et économique des étudiant-e-s des hautes écoles suisses 2009. Cadre de sondage, plan d'échantillonnage et méthodes d'estimation. Numéro de commande: 338-0066

Potterat, J., Panchard, C., Kilchmann, D. (2012). Umweltschutzausgaben der Unternehmen 2009 (UWSA2009). Stichprobenplan, Einsetzungen, Gewichtung und Schätzverfahren. Bestellnummer: 338-0065

Potterat, J. (2012). Benutzung der Umsteigeschlüssel NOGA 2002-2008. Bestellnummer: 338-0064

Potterat, J. (2011). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe im Jahr 2009 (KNBB09). Stichprobenplan, Gewichtung und Schätzverfahren. Bestellnummer: 338-0063

Kilchmann, D., Potterat, J., Genoud, S. (2011). Gütertransporterhebung 2008. Stichprobenplan, Datenaufbereitung, Gewichtung und Schätzverfahren. Bestellnummer: 338-0062

Graf, E. (2010). Enquête suisse sur la santé 2007. Plan d'échantillonnage, pondérations et analyses pondérées des données. Numéro de commande: 338-0061

Eichenberger P., Hulliger B., Potterat J. (2010). Describing the Anticipated Accuracy of the Swiss Population Survey. Order number: 338-0060

Graf, E. (2010). Étude empirique de l'attrition du Panel Suisse de Ménages : vers une caractérisation du profil du non-répondant. Numéro de commande: 338-0059

Graf, E. (2009). Weightings of the Swiss Household Panel: SHP_I wave 9, SHP_II wave 4, SHP_I et SHP_II combined. Order number: 338-0058

Graf, E. (2009). Pondérations du Panel Suisse de Ménages: PSM_I vague 9, PSM_II vague 4, PSM_I et PSM_II combinés. Numéro de commande: 338-0057-05

Qualité, L., Tillé, Y. (2009). Estimation de la précision d'évolutions dans l'enquête sur la valeur ajoutée. Numéro de commande: 338-0056

Renaud, A., Panchard, C. et Potterat, J. (2008). Statistique de l'emploi. Révision 2007 : méthodes d'estimation. Numéro de commande: 338-0055

Graf, E. (2008). Pondérations du PSM. PSM_I vague 8, PSM_II vague 3, PSM_I et PSM_II combinés. Numéro de commande: 338-0054

Andrade, B., Graf, M. (2008). Enquête suisse sur la structure des salaires 2006. Aspects méthodologiques du modèle des salaires "Salarium". Numéro de commande: 338-0053

Renaud, A. (2008). Statistique de l'emploi. Révision 2007 : cadre de sondage et échantillonage. Numéro de commande: 338-0052

Graf, E. (2008). Pondérations du SILC pilote. SILC_I vague 2, SILC_II vague 1, SILC_I et SILC_II combinés. Numéro de commande: 338-0051

Kilchmann, D. (2008). Statistik der sozialmedizinischen Institutionen 1999-2004 und Krankenhausstatistik 1999-2002. Einsetzungen für fehlende Daten. Bestellnummer: 338-0050

Renaud, A. (2008). Technologies de l'information et de la communication. Estimations sur la base de la statistique de la valeur ajoutée. Numéro de commande: 338-0049

Assoulin, D. (2007). Wertschöpfungsstatistik. Einsetzungsversuche für fehlende Antworten grosser Unternehmen. Bestellnummer: 338-0048

Kilchmann, D. (2007). Beherbergungsstatistik Campingplätze. Stichprobenrahmen und Schätzverfahren 2005/06. Bestellnummer: 338-0047

Gabler, S., Häder, S. (2007). Haushalts- und Personenerhebungen. Machbarkeit von Random Digit Dialing in der Schweiz. Bestellnummer: 338-0046

Ferrez, J., Graf, M. (2007). Enquête suisse sur la structure des salaires. Programmes R pour l'intervalle de confiance de la médiane. Numéro de commande: 338-0045

Renaud, A. (2007). Harmonisation de la scolarité obligatoire en Suisse (HarmoS). Design général de l'enquête et échantillon des écoles. Numéro de commande: 338-0044

Potterat, J. (2007). Betriebszählung 2005. Statistische Methoden zur Schätzung der provisorischen Ergebnisse. Bestellnummer: 338-0043

Hulliger, B. (2006). Umweltschutzausgaben der Unternehmen 2003, Stichprobenplan, Datenaufbereitung und Schätzverfahren. Bestellnummer: 338-0042

Renfer, J.-P. (2006). Enquête sur les chiffres d'affaires du commerce de détail. Plan d'échantillonnage et méthodes d'estimation. Numéro de commande: 338-0041

Salamin, P.-A. (2006). Statistique de l'aide sociale dans le domaine de l'asile. Plan de sondage et extrapolations pour l'enquête pilote 2005. Numéro de commande: 338-0040

Renaud, A. (2006). Statistique suisse des bénéficiaires de l'aide sociale. Pondération des communes 2004. Numéro de commande: 338-0039

Graf, M. (2006). Swiss Earnings Structure Survey 2002-2004. Compositional data in a stratified two-stage sample: Analysis and precision assessment of wage components. Order number: 338-0038

Potterat, J. (2006). Pensionskassenstatistik 2004. Statistische Methoden zur Schätzung der provisorischen Ergebnisse. Bestellnummer: 338-0037

Potterat, J. (2006). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe im Jahr 2004. Stichprobenplan, Gewichtung und Schätzverfahren. Bestellnummer: 338-0036

Kilchmann, D. (2006). Vierteljährliche Wohnbaustatistik. Stichprobenplan, statistische Datenaufarbeitung und Schätzverfahren 2005. Bestellnummer: 338-0035

Kilchmann, D. (2006). Erhebung über Forschung und Entwicklung in der schweizerischen Privatwirtschaft 2004. Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren. Bestellnummer: 338-0034

Kilchmann, D., Eichenberger, P., Potterat, J. (2005). Volkszählung 2000. Statistische Einsetzungsverfahren Band 2. Bestellnummer: 338-0033

Kilchmann, D., Eichenberger, P., Potterat, J. (2005). Volkszählung 2000. Statistische Einsetzungsverfahren Band 1. Bestellnummer: 338-0032

Graf, M., Matei, A. (2005). Enquête suisse sur la structure des salaires 2002. La précision du salaire brut standardisé médian. Numéro de commande: 338-0031

Graf, E., Renfer, J.-P. (2005). Enquête suisse sur la santé 2002. Plan d'échantillonnage, pondération et estimation de la précision. Numéro de commande: 338-0030

Potterat, J. (2005). Mietpreis-Strukturerhebung 2003. Gewichtung und Schätzverfahren. Bestellnummer: 338-0029

Potterat, J. (2005). Landwirtschaftliche Betriebszählung 2003. Schätzverfahren für die Zusatzerhebung. Bestellnummer: 338-0028

Renaud, A. (2004). Coverage estimation for the Swiss population census 2000. Estimation methodology and results. Order number: 338-0027

Kilchmann, D. (2004). Revision des Schweizerischen Lohnindex. Schätzmethoden der Lohnindices und deren Varianzschätzer. Bestellnummer: 338-0026

Graf, M. (2004). Enquête suisse sur la structure des salaires 2002. Plan d'échantillonnage et extrapolation pour le secteur privé. Numéro de commande: 338-0025

Renaud, A. (2004). Analyse de données d'enquêtes. Quelques méthodes et illustration avec des données de l'OFS. Numéro de commande 338-0024

Renaud, A., Potterat, J. (2004). Estimation de la couverture du recensement de la population de l'an 2000. Echantillon pour l'estimation de la sous-couverture (P-sample) et qualité du cadre de sondage des bâtiments. Numéro de commande: 338-0023

Graf, M. (2004). Fusion de données. Etude de faisabilité. Numéro de commande: 338-0022

Potterat, J. (2003). Mietpreis-Strukturerhebung 2003. Entwicklung des Stichprobenplans und Ziehung der Stichprobe. Bestellnummer: 338-0021

Potterat, J. (2003). Landwirtschaftliche Betriebszählung 2003. Stichprobenplan der Zusatzerhebung. Bestellnummer: 338-0020.

Renaud, A. (2003). Estimation de la couverture du recensement de la population de l'an 2000. Echantillon pour l'estimation de la sur-couverture (E-sample). Numéro de commande: 338-0019

Hulliger, B. (2003). Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren. Erhebung über F+E in der schweizerischen Privatwirtschaft 2000. Bestellnummer: 338-0018

Renfer, J.-P. (2003). Enquête 2000 sur la recherche et le développement dans l'économie privée en Suisse. Plan d'échantillonnage. Numéro de commande: 338-0017

Potterat, J. (2003). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe. Schätzverfahren. Bestellnummer: 338-0016

Graf, M., Matei, A. (2003). Stratégie de choix des modèles de désaisonnalisation. Application aux séries de l'emploi total. Numéro de commande: 338-0015

Potterat, J., Salamin, P.A. (2002). Betriebszählung 2001. Methoden für die Datenbereinigung. Bestellnummer: 338-0014

Renaud, A. (2002). Programme international pour le suivi des acquis des é lèves (PISA). Plans d'échantillonnage pour PISA 2000 en Suisse. Numéro de commande: 338-0013

Renfer, J.-P. (2002). Enquête 2001 sur les coûts et l'utilité de la formation des apprentis du point de vue des établissements. Plan d' échantillonnage. Numéro de commande: 338-0012

Potterat, J., Salamin, P.A. (2002). Betriebszählung 2001. Stichprobenplan und Schätzverfahren für die provisorischen Ergebnisse. Bestellnummer: 338-0011

Graf, M. (2002). Enquête suisse sur la structure des salaires 2000. Plan d'échantillonnage, pondération et méthode d'estimation pour le secteur privé. Numéro de commande: 338-0010

Renaud, A., Eichenberger P. (2002). Estimation de la couverture du recensement de la population de l'an 2000. Procédure d'enquête et plan d'échantillonnage de l'enquête de couverture. Numéro de commande: 338-0009

Kilchmann, D., Hulliger, B. (2002). Stichprobenplan für die Obstbaumzählung 2001. Bestellnummer: 338-0008

Graf, M. (2002). Passage du concept établissement au concept entreprise. Numéro de commande: 338-0007

Salamin, P.A. (2001). La technique de la double enquête pour la statistique du transport routier de marchandise. Numéro de commande: 338-0006

Peters, R., Renfer, J.-P. et Hulliger, B. (2001). Statistique de la valeur ajoutée 1997-1998. Procédure d'extrapolation des données. Numéro de commande: 338-0005

Potterat, J., Hulliger, B. (2001). Schätzung der Sägereiproduktion mit der Sägerei-Erhebung PAUL. Bestellnummer: 338-0004

Graf, M. (2001). Désaisonnalisation. Aspects méthodologiques et application à la statistique de l'emploi. Numéro de commande: 338-0003

Hüsler, J., Müller, S. (2001). Schlussbericht Betriebszählung 1995 (BZ 95), Mehrfach imputierte Umsatzzahlen. Bestellnummer: 338-0002

Renaud, A. (2001). Statistique suisse des bénéficiaires de l'aide sociale. Plan d'échantillonnage des communes. Numéro de commande: 338-0001

Hulliger, B., Eichenberger, P. (2000). Stichprobenregister für Haushalterhebungen: Umstellung auf Telefonnummern ohne Namen und Adressen, Abläufe für Erstellung und Stichprobenziehung. Bestellnummer: 338-0000

The FSO's publications

As the central statistical agency of the Confederation, the Federal Statistical Office (FSO) has the task of providing Swiss statistical information to a wide range of users. Dissemination is done by topic with different information media via several channels.

The statistical topics

00 Statistical basis and overviews
01 Population
02 Territory and environment
03 Work and income
04 National economy
05 Prices
06 Industry and services
07 Agriculture and forestry
08 Energy
09 Construction and housing
10 Tourism
11 Mobility and transport
12 Money, banks and insurance
13 Social security
14 Health
15 Education and science
16 Culture, media, information society, sports
17 Politics
18 General Government and finance
19 Crime and criminal justice
20 Economic and social situation of the population
21 Sustainable development, regional and international disparities

The key publications

Statistical Yearbook of Switzerland



The "Statistical Yearbook of Switzerland" (German/French) published by the Federal Statistical Office has been the standard reference book for Swiss statistics since 1891. It contains the most important statistical findings regarding the Swiss population, society, government, economy and environment.

Statistical Data on Switzerland



Statistical Data on Switzerland is an appealing and entertaining summary of the year's most important figures. With 52 pages in a practical A6/5 format, the publication is free of charge and available in five languages (German, French, Italian, Romansch and English).

The FSO online – www.statistics.admin.ch

The Swiss Statistics website offers you a modern, attractive and up-to-date gateway to all statistical information. We would like to draw your attention to the following popular offerings:

Publication database – publications offering further information

Almost all publications published by the FSO are available in electronic form on the website free of charge. Print publications can be ordered by telephone on 058 463 60 60 or by emailing order@bfs.admin.ch.
www.statistics.admin.ch → Look for statistics → Catalogues and Databases → Publications

NewsMail – always up to date



Email subscriptions by topic with details and information on the latest findings and activities.
www.news-stat.admin.ch

STAT-TAB – the interactive statistical database



The interactive statistical database offers simple and customisable access to statistical results as well as the option of downloads in various formats.
www.stattab.bfs.admin.ch

Statatlas Switzerland – regional database and interactive maps



With more than 3 000 interactive thematic maps, the Statistical Atlas of Switzerland gives you an up-to-date and permanently available overview of captivating regional issues covering all FSO topics. Available in German and French.
www.statatlas-switzerland.admin.ch

Individual inquires

The FSO's statistical information service
058 463 60 11, info@bfs.admin.ch

From 2011 onwards, the Swiss business census is replaced by the new Swiss structural business statistics (STATENT). STATENT is mainly based on register data and complemented by several business surveys. Full-time equivalents (FTE) of employment are not available in the register data and should be constructed. This report presents the methodology used for the construction of FTE for STATENT. Two important topics are treated. First, we describe the model used for estimating FTE based on survey data which are matched to register data. The different phases in the development of the model such as model selection, robustification and quality assessment are presented. Then, we discuss how data inconsistency due to the integration of different sources is treated. The method that we apply to treat the inconsistencies takes into account the fact that they may be related to the type of employment. Finally, we explain how this data harmonization affects the estimation of the model and the construction of FTE for STATENT.

Print format orders
Tel. 058 463 60 60
Fax 058 463 60 61
order@bfs.admin.ch

Price
Free of charge

Downloads
www.statistics.admin.ch
(free of charge)

FSO number
338-0077

ISBN
978-3-303-00569-9

Statistics
counts for you.

www.statistics-counts.ch