2015–
2021

00

Statistical basis
and overviews

Neuchâtel 2024

Swiss structural business statistics (STATENT) 2015–2021

Estimation and precision of the full-time equivalents

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Home Affairs FDHA
Federal Statistical Office FSO

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.

Swiss structural business statistics (STATENT) 2015–2021

Estimation and precision of the full-time equivalents

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

Neuchâtel 2024

Published by:	Federal Statistical Office (FSO)
Information:	Desislava Nedyalkova, tel. +41 58 463 68 07, desislava.nedyalkova@bfs.admin.ch; Daniel Assoulin, tel +41 58 463 67 51, daniel.assoulin@bfs.admin.ch
Editor(s):	Desislava Nedyalkova, FSO; Daniel Assoulin, FSO
Series:	Swiss Statistics
Topic:	00 Statistical base and overviews
Original text:	English
Layout:	Publishing and Dissemination PUB, FSO
Online:	www.statistics.admin.ch
Print:	www.statistics.admin.ch Federal Statistical Office, CH-2010 Neuchâtel, order@bfs.admin.ch, tel. +41 58 463 60 60 Printed in Switzerland
Copyright:	FSO, Neuchâtel 2024 Reproduction with mention of source authorised (except for commercial purposes).
FSO number:	338-0088
ISBN:	978-3-303-00737-2

Contents

Preface

Since 2011 the Swiss structural business statistics (STATENT) have replaced the business census. The initial methodology developed for the construction of full-time equivalents (FTE) for STATENT was described in the methodological report of Nedyalkova, D., and Assoulin, D. (2017). This document complements the first report by describing the modifications that have been applied to the original methodology for all STATENTs from 2015 onwards as well as some recent developments. These practices were developed by Desislava Nedyalkova and Daniel Assoulin. We would like to thank Nadège Jacot and Mathieu Gunzinger from the STATENT production team for the fruitful collaboration. We would also like to thank Djalel Meskaldji from the Statistical methods section (METH) for his valuable contribution on the bootstrap variance estimation part. Finally, we would like to thank Raphaël Jauslin, Daniel Kilchmann, Jann Potterat and Jean-Pierre Renfer from METH for reviewing the document and for their relevant remarks.

Summary

STATENT is published yearly starting with the reference year 2011. 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. The FTE model was developed on matched data coming from the register and two surveys - the quarterly survey of employment (JobStat) and the quarterly survey of new enterprises (ERST). In this report, we first describe the modifications in the methodology used for the construction of FTE for STATENT following the revision of the JobStat survey in 2015. Then, we describe subsequent changes in the model and in the data following the revision of STATENT in 2021. Finally, we describe a bootstrap method for calculating the variance of total FTE and show the results for STATENT 2020 and 2021.

1 Introduction

The FTE prediction model allows to construct FTEs by gender at the enterprise level for data coming from the register of the compensation offices of the Old-Age and Survivors Insurance (OASI), which is the main source of employment for STATENT. The methodology is described in Nedyalkova, D., and Assoulin, D. (2017).

The implementation of the FTE model for STATENT 2015 and the subsequent analyses showed an atypical trend in FTEs compared to the trend in employment (total employment increased but FTEs remained stable). This phenomenon was mainly observed in the tertiary sector, but also to some extent in the secondary sector. The reason for this was found to be in the integration of STATENT into the survey frame of the JobStat survey and the revision of its sampling design in the second quarter of 2015 (Potterat, J., Assoulin, D. and J.-M. Nicoletti (2019)). It should be noted that JobStat, in combination with OASI register data, is the main source used to model FTEs. Indeed, the new sampling frame of JobStat 2015 covered for the first time the STATENT universe which is larger than the one from the Business Census, the previous basis for the frame that did not cover enterprises with a very low level of employment. This change impacted the FTE model and the results derived from it. For this reason, the published STATENT 2015 results on FTE were first based on the model parameters estimated for STATENT 2014. This allowed us to continue the analyses with the STATENT 2015 data in order to adapt the FTE model to changes in the data on which the model is based. The methodological developments which we will describe were first carried out on the STATENT 2015 data and then applied to the STATENT 2016 data. They include four issues: a) changes in the definition of the dataset for modelling, b) changes to the model, c) changes in the weighting scheme and d) a procedure for the treatment of inversions in model parameters.

Further developments of the model were implemented in 2021 in the context of a revision of the whole series of STATENT 2011-2018. The revision was not conceptual but concerned mostly corrections in the source data and in the published results. However, we have used this as an opportunity to consolidate the FTE model and apply the same model to the revised STATENT 2015-2018 and to STATENT 2019. Some of these new developments were also motivated by the gradual reduction in the size of the data coming from the ERST survey. For STATENT 2011-2014, it was decided to apply the model for STATENT 2015 (same parameters). Following the revision, and with the results of five consecutive years of STATENT where FTE were calculated using the same FTE model, we have developed a new tool to monitor and analyse evolutions.

A further development concerned a precision measure of the estimator of total FTE of employment. FTE for STATENT are a mixture of FTE coming from the surveys, harmonised with the OASI register data, FTE predicted using the FTE model and imputed FTE for enterprises from the public sector which are not included in a survey. The FTE for STATENT is thus a complex estimator, which could have led us to consider the variance estimation by linearisation. Indeed, this type of variance estimation is often the method of choice for measuring the precision of complex estimators. However, this method requires an analytical form of the variance estimator, which was difficult to derive for STATENT. In order to take into account both the sampling design and the complexity of the estimator, we used the bootstrap technique. In practice, we adapted the bootstrap algorithm for unequal probability designs proposed by Chauvet (2007), which consists of generating a pseudo-population from the initial sample, in which we draw bootstrap samples that respect the original sampling design.

This document is organised as follows. Section 2 describes the developments made for STATENT 2015 and STATENT 2016 in order to take into account the new sample of JobSTAT in the FTE model. Next, in Section 3, we describe the additional adjustments in data and the FTE model following the revision of the whole series of STATENT. In Section 4, we show the bootstrap variance estimation procedure for estimating the precision of the estimator of total

FTE of employment for STATENT 2020 and 2021.

2 Methodological developments for STATENT 2015 and 2016

In this section, we present the main methodological developments following the revision of the JobStat survey design and the integration of STATENT in its survey frame.

2.1 Extending the dataset for modelling FTE

Prior to the adjustments, the data used to construct the FTE model for STATENT 2015 were those from the JobStat[1] 2015 survey, fourth quarter, matched with data from the OASI register (month of December 2015). These were single-establishment enterprises (EUNT) excluding public administration, Profiling and Profiling light, belonging to the secondary and tertiary economic sectors, for which we know the number of FTE, income and other variables such as the main geographical region or NUTS 2[2] (denoted by GREG) or variables for different statistical classifications of the economic activities (NACE Rev. 2 [3]) which we will denote by NOGA. In Appendix A.3, tables describing the NOGA at different aggregation levels are given. To the data mentioned above the small enterprises from the ERST 2015 survey (the quarterly survey of new enterprises that updates the business register), denoted by $ERST\_MINI$, were added. This was motivated by the fact that the JobStat reference population definition, prior to its revision in 2015, did not cover the very small enterprises. It should be noted that new enterprises, covered by the ERST 2015 survey, are not part of the JobStat 2015 survey frame, which dates from December 2014 updated by STATENT 2012 (published in 2014).
Following the changes in the JobStat survey, several options were considered for defining a new dataset for modelling. As the new JobStat also covers the population of very small enterprises, it made more sense to include all ERST enterprises rather than just the $ERST\_MINI$ population. We also considered a variant that only included EUNT from JobStat. However, this variant was discarded as it resulted in too few observations for modelling. Therefore, the variant retained is the one that includes all EUNT enterprises from the JobStat and ERST surveys. We denote the resulting dataset $BESTA\_EUNT + ERST\_EUNT$ as BE. This approach leads to a much larger dataset for modelling than in the past. To illustrate, $BESTA\_EUNT$ + $ERST\_MINI$ enterprises covered only $31\,485$ of the $69\,055$ $BESTA\_EUNT$ + $ERST\_EUNT$ enterprises in 2015 (data as of 31.05.2021). The difference between these two datasets, used to model FTEs before and after adjustments, comes only from the ERST survey as the number of $BESTA\_EUNT$ enterprises is the same. The new BE dataset requires an adjustment of the JobStat extrapolation weights, described below in Section 2.2. The enterprise weights in the ERST survey are all set to 1 as the survey is considered to be exhaustive.

[1] The German name for the JobStat survey is BESTA and this abbreviation will be used in part of the notation, for instance, when describing the different datasets.

[2] 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.

[3] 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 (50 groupings of NOGA sections used for the production of national statistics only) and plays an important role for our model. See, for details, https://www.kubb-tool.bfs.admin.ch/en.

2.2 Changes in the calibration procedure

In Nedyalkova, D., and Assoulin, D. (2017), we described the different sets of weights used at different stages: data processing (calibration weights), model estimation (weights accounting for heteroscedasticity, robustified weights) and model validation (calibration weights) stages. As in the first developments of the FTE model, the JobStat sample weights are adjusted through calibration to ensure consistency with STATENT in terms of employment. However, there are differences in the way these weights are calculated before and after the adjustments to the dataset described in Section 2.1. These new weights are described below.

2.2.1 Weights used for modelling

The first set of weights is obtained by a calibration technique that adjusts sample totals to known population totals (see Deville and Särndal, 1992), using the CALMAR2 macro in SAS (Sautory and Le Guennec, 2003). A unique weight is calculated for each enterprise. Let us denote by $AVS\_STATENT$ all enterprises included in the OASI register[4] data and which correspond to the definition of a STATENT eligible enterprise and for which we know total employment (EMPTOT). The primary sector, public sector enterprises and enterprises belonging to NOGA 97, 98 and 99 are not part of the considered STATENT universe. We have first excluded from this initial dataset those enterprises in the JobStat survey that are covered by Profiling and Profiling Light, which are in principle exhaustive surveys covering employment information for the concerned enterprises. We denote this dataset as $BESTA\_PROF$. In addition to the enterprises of $BESTA\_PROF$, we also exclude the enterprises of the ERST survey from $AVS\_STATENT$. The resulting dataset will be referred to as $AVS\_STATENT\_CAL$ and will constitute our calibration frame.
Let $BESTA\_STATENT$ denote the set of enterprises from the JobStat survey comprising the single-establishment, the multi-establishment (MUNT) and the enterprises coming from the Profiling and Profiling Light (a total of $14\ 170$ enterprises in JobStat 2015, quarter 4, as of 31.05.2021). The dataset that will be calibrated consists of the set $BESTA\_STATENT$, from which we exclude the enterprises belonging to $BESTA\_PROF$. It is denoted by $BESTA\_CAL$ (a total of $12\ 750$ enterprises, as of 31.05.2021).
The calibration is done with the variable EMPTOT_R (employment according to the OASI register) at the level of NOGA OFS50 (a special combination of economic activities used in Switzerland, see Appendix A.3). The totals of the dataset $BESTA\_CAL$ are calibrated to the totals of the dataset $AVS\_STATENT\_CAL$ (the population totals). Using the JobStat weights corrected for non-response as initial weights (see Potterat, J., Assoulin, D. and J.-M. Nicoletti, 2019, Section 6.5), we obtain new calibrated weights for each enterprise i, denoted as $w_i^{(1)}$. Enterprises of the ERST survey, which is considered exhaustive, have a weight of 1 ($w_i^{(1)} = 1$). These weights are then used for the data harmonisation procedure and for fitting the model. In order to validate the FTE model, we have constructed another set of weights, which we will describe below.

2.2.2 Weights used for calculating the quality measures

We use several measures to validate and compare our models. These measures are described in Section 3.5 of Nedyalkova, D., and Assoulin, D. (2017). As a reminder, these measures are *meanabsdif*, the weighted mean of the absolute prediction errors (based on the weights $w_i^{(1)}$), and *reldif*, the ratio between the extrapolated model residuals and the extrapolated FTEs. To

[4] AVS is the French acronym for this data source.

calculate *reldif*, we define an adjusted weight $w_i^{(2)}$, which is calculated on the basis of the weight $w_i^{(1)}$. These adjustments are made by NOGA OFS50. Thus, for each NOGA OFS50, two weighted sums of EMPTOT_R are calculated. The first is the weighted sum over all enterprises in the dataset BE used for modelling (see Section 2.1). Next, we discard from the BE dataset enterprises with very large differences between EMPTOT_R and EMPTOT_S (see Nedyalkova, D., and Assoulin, D., 2017, Section 4.2). Let BEF denote the resulting subset of BE. The second total is the weighted sum over the enterprises of the BEF dataset. Then, if for each NOGA OFS50,

$$\text{TOT\_ALL} = \sum_{i \in BE} w_i^{(1)} \text{EMPTOT\_R}_i,$$

and

$$\text{TOT\_ALL\_FILT} = \sum_{i \in BEF} w_i^{(1)} \text{EMPTOT\_R}_i,$$

we define for each enterprise i in BEF

$$w_i^{(2)} = w_i^{(1)} \frac{\text{TOT\_ALL}}{\text{TOT\_ALL\_FILT}}.$$

2.3 Development of the FTE model for STATENT 2015 and 2016 following data adjustments

In line with the methodology of the FTE model described in Nedyalkova, D., and Assoulin, D. (2017), we consider here four levels of modelling according to gender and sector of economic activity: men/sector 2, men/sector 3, women/sector 2, women/sector 3, denoted respectively by m/s2, m/s3, w/s2 and w/s3. This section describes the main changes and methodological developments made to the FTE model. It includes different model variants and results on the quality of the models developed within the domains of analysis corresponding to the four modelling levels.

2.3.1 Salary classes

Salary classes play an important role in the construction of the explanatory variables for the FTE-model. Salary class limits are based on estimated cumulative proportions corresponding to increasing occupational levels. The estimation is based on JobStat and ERST surveys from 2011 and 2012, and is described in detail in Section 3.2 of Nedyalkova, D., and Assoulin, D. (2017). We then use the corresponding percentiles in the standardised OASI income distribution for 2015 and 2016 as class limits for the salary classes. The main objective was to adapt the FTE model to changes in the data sources; therefore, the calculation of new cumulative proportions was considered of secondary importance. As a consequence, no further improvements have been made with regard to this issue.

2.3.2 A new parameterisation of the FTE model

We have first implemented on the extended dataset comprising the EUNT from the JobStat and ERST surveys the FTE model based on the methodology described in Nedyalkova, D., and Assoulin, D. (2017), which we consider as reference model. This model has as main variables the number of employees per salary class, with parameter adjustments allowing separate slopes for some combinations of GREG and NOGA sections. The separate slopes model is based on a specific parameter selection. For stability reasons this selection was kept constant since

STATENT 2012 (see Nedyalkova, D., and Assoulin, D., 2017, Section 3.3). This model is designated 15a/16a, for STATENT 2015 and 2016, respectively. The availability of a larger dataset allowed us to implement a new parameterisation of the model. This makes it possible to obtain a specific parameter for each combination of GREG and NOGA section.
This alternative model for estimating FTEs for STATENT 2015 and 2016, which we have commonly called "nested", has as main explanatory variables (main effects) the variables "number of employees per salary class" ($V_{ij}, j = 1, ..., 4$ for each enterprise i) described in Section 3.1 of Nedyalkova, D., and Assoulin, D. (2017). These effects, with the exception of the parameter for the first salary class, V_{i1}, which we leave fixed, are completed by their interactions with GREG and NOGA section, which are categorical variables. The variable GREG has 7 categories. For NOGA, we have 5 categories in economic activity sector 2 and 13 in economic activity sector 3. Thus, for this new "nested" model, we apply a finer adjustment than for the 15a/16a model described above. This new model will be referred to as 15an/16an, for STATENT 2015 and 2016, respectively.
In the model code, see example women/sector 2 (w/s2) below, the main variables are denoted slightly differently, e.g. `uV1_w` for V_{i1}, and correspond to the variables after the data harmonisation phase (see Nedyalkova, D., and Assoulin, D., 2017, Section 4.2) and subsequent processing. The variable `unique_y` refers to the harmonised FTE variable (see Nedyalkova, D., and Assoulin, D., 2017, Table 9, Section 4.3).

```
proc glm data=dout.good_rob2_fs2_&outdate.;
  class greg &noga_s.;
  weight adjweights;
  model unique_y = uV1_w uV2_w uV3_w uV4_w
  uV2_w*greg uV3_w*greg uV4_w*greg
  uV2_w*&noga_s. uV3_w*&noga_s. uV4_w*&noga_s. / noint solution;
  output out = dout.outglm_nested_fs2 residual = Reptavs predicted = Peptavs;
  ods output parameterestimates = dout.estim_nested_fs2_&outdate.;
run;
```

As intended and confirmed by the description of GLM models provided by SAS (see SAS Institute, 2018, Parameterization of PROC GLM models), our model uses an interaction of quantitative and qualitative variables. The regression parameter for each of the main variables (`uV2_w`, `uV3_w`, `uV4_w`) is adjusted for each GREG and each NOGA section except the last ones, which are used as a baseline. Table 1 shows the estimated parameter values for `uV2_w`, for the women/sector 2, model 16an, STATENT 2016, for which we have five NOGA sections (B, C, D, E and F) and seven GREG. In the first column, we have the value of the main parameter. A negative estimated parameter for a certain GREG (or NOGA), means that this value will be subtracted from the main parameter and implies a smaller estimated FTE for this GREG (or NOGA). On the contrary, a positive estimated parameter for a certain GREG (or NOGA) means that the value of this estimated parameter will be added to the main parameter and implies a larger value of FTE for this GREG (or NOGA). These parameter estimates reflect an intermediate state of our estimates and are given for illustrative purposes only.
Based on Table 1, we give the resulting parameter estimation of FTE for some combinations of GREG and NOGA.

- If the parameter adjustment for GREG 3 is 0.150 and the parameter adjustment for NOGA section B is -0.120, then the FTE parameter for `uV2_w` in the combination of GREG 3 and NOGA section B is calculated as follows:

$$0.379 + 0.150 - 0.120 = 0.409$$

Table 1 Model 16an, w/s2: parameters for `uV2_w`

	main parameter	GREG	adjustment	NOGA	adjustment
		1	0.023	B	-0.120
`uV2_w`	0.379	2	0.048	C	0.017
		3	0.150	D	-0.271
		4	-0.056	E	-0.338
		5	0.002	F	-
		6	-0.106		
		7	-		

- For NOGA section F and GREG 7, which are used as the baseline, the FTE parameter for `uV2_w` in GREG 7 and NOGA section F will actually be the same as the estimated parameter for the main variable `uV2_w`, which is 0.379.
- For NOGA section F and GREG 3, the parameter is calculated as follows:

$$0.379 + 0.150 = 0.529$$

To facilitate the implementation and comparison of parameters, we calculate estimated parameters for each combination of main variable (`uV2_w`, `uV3_w`, `uV4_w`) * GREG * NOGA, as shown in the examples above. This matrix form

- makes it easier to study the distribution of the parameters (minimum and maximum values; first, second and third quartiles),
- facilitates comparison between years.

The new parameterisation of the FTE model has the advantage of setting a specific parameter for each GREG and NOGA section. In terms of quality measures, models 15an and 16an performed better than models 15a and 16a. This is illustrated in the boxplots (shown in Appendix A.4, which show the distribution of the quality measure *reldif* described in Section 3.5 of Nedyalkova, D., and Assoulin, D. (2017) for the year 2016 (the results are similar for 2015). However, this parameterisation requires a sufficient number of observations in each NOGA section and each GREG. It has also accentuated the phenomenon of parameter inversions that we describe below (see Section 2.4), a phenomenon that was also present, but to a lesser extent, in the original FTE model applied before the revision.

2.3.3 Grouping of NOGA sections

Considering that for some combinations of GREG and NOGA section we have a small number of observations, some NOGA sections needed to be grouped to improve parameter stability. This led to the grouping of NOGA B with NOGA C in sector 2 due to the small size of NOGA B (61 enterprises for the m/s2 model and 31 enterprises for the w/s2 model). The resulting model is denoted by 15angr/16angr (**al**ternative, **n**ested, **gr**ouped) in the boxplots of the quality measures (see Figures 1, 2, 3 and 4).

2.4 Inversion of parameters

In principle, under the assumption that employees in a higher salary class are supposed to have a higher occupational level, the estimated FTE parameters should follow an ascending order.

The violation of this assumption is considered as a parameter inversion. This is illustrated in Table 2 for the w/s2 model, 2016, before grouping the NOGA sections and for GREG 6 and NOGA section B:

Table 2 Model 16an, w/s2: parameters for the combination of GREG 6 and NOGA B

uVj_w	GREG	NOGA	parameter
uV1_w	6	B	0.205
uV2_w	6	B	0.153
uV3_w	6	B	0.789
uV4_w	6	B	0.889

Table 2 shows that, in this combination of GREG and NOGA section, the parameter for the number of employees in the second salary class (uV2_w) is smaller than the parameter for the number of employees in the first salary class (uV1_w). The analyses showed that these inversions are only present in certain NOGA sections. In sector 2, these were mainly NOGA sections B, C, D and E. In sector 3, the largest number of inversions was observed in NOGA sections K, L, R and S. Table 3 shows the number of inversions for each of the models 15a/16a, 15an/16an (before grouping) and 15angr/16angr (after grouping of NOGA B and C). The difference before/after grouping is therefore only visible in sector 2. The total number of parameters is given in the "total" column (same number of parameters for men and women in sector 3 - column s3).

Table 3 Total number of inversed estimated parameters by model

	w/s2	w/s2	m/s2	m/s2	w/s3	m/s3	s3
model	inversions	total	inversions	total	inversions	inversions	total
15a	1	140	5	105	5	12	364
15an	11	140	14	105	21	27	364
15angr	6	140	13	105	21	27	364
16a	0	140	5	105	7	21	364
16an	16	140	13	105	21	40	364
16angr	13	140	12	105	21	40	364

We can see from Table 3 that the grouping of NOGAs B and C in sector 2, which we adopted to ensure the stability of the parameters, had the effect of eliminating some inversions, especially in NOGA B for w/s2, but the phenomenon was still present.

Treatment of the inversions

After discussions with the STATENT production team, it was decided that it would be advantageous to deal with inversions in order to guarantee non-decreasing coefficients for increasing salary classes. We considered two options. The first solution, called "maximum value replacement", consisted of iteratively checking for an inversion and, if present, correcting the inversion by adopting the value of the parameter of the lower salary class. In the example from Table 2, the parameter for uV2_w, which was estimated to be 0.153, is replaced by the value of the parameter for uV1_w, which is 0.205. This gives us the new values of the estimated parameters, which are uV1_w = uV2_w $= 0.205$. This version of the model has been denoted by 15angrc and 16angrc, respectively. The quality measures we use to validate our models show that the variant 15angrc/16angrc introduces a slight upward bias, which is also evident in the analyses of certain results.

In order to reduce the risk of bias in the treatment of inversions, we developed a second solution, called "mean-value replacement". It consists of iteratively calculating an average of the parameters for the salary classes involved in the inversions. Using the example of Table 2, we obtain the new, corrected for inversions, values of the estimated parameters as follows: `uV1_w` = `uV2_w` $= (0.205 + 0.153)/2 = 0.179$. We then check whether the parameter for `uV3_w` is less than the calculated average, and if so, we adjust the three parameters by calculating an average of the three parameters. It is this version of the model, called 15angrcm/16angrcm, that was finally retained.

2.5 Results of the implementation of the new FTE model for constructing FTE for STATENT 2015 and 2016

We carried out various analyses to assess the quality of each of the proposed models and the stability of the parameters. These results include:

- Statistical measures of the distribution of the parameters (quartiles, median), which make it possible to assess the variability of the parameters and their stability from one year to the next;
- Boxplots of the quality measures, which make it possible to compare the different models in terms of quality of adjustment;
- Graphs showing the changes in FTE and employment over the years.

Here we will only present the boxplots for the *reldif* quality measure for STATENT 2016 which are given in Appendix A.4 (see Figures 1, 2, 3 and 4). The different variants of the model are summarised in Appendix A.2. Note that the results for 2015 were similar.
The various analyses allowed us to draw the following conclusions. The reference for the comparisons will be the model 15a/16a consisting in applying the original FTE model, based on a fixed selection of parameters applied on the extended dataset of enterprises. For the "nested" variants (15/16an, 15angr/16angr, 15angrc/16angrc, 15angrcm/16angrcm), as there is a specific parameter for each combination of NOGA and GREG, we generally observe more dispersion between the parameters within one year, but also more dispersion in the evolution of the parameters between two years. On the other hand, we do not observe any parameters or year-to-year variations in the parameters that are as extreme as those observed in the 15a/16a model. Based on the comparisons of *reldif*, we can conclude that the different "nested" variants have a very similar quality, which is superior to that of variant 15a/16a. Variant 15angrcm/16angrcm only slightly impacts the quality of the estimation compared to the variant 15an/16an. Therefore it seems to be a good compromise between the aim for stability, the treatment of inversions and a quality close to variant 15an/16an (no grouping, no treatment of inversions). For completeness, the model denoted by 15fch/16fch, which is the reference FTE model that adjusts a selection of parameters on the $BESTA\_EUNT$ + $ERST\_MINI$ dataset, is also shown in the boxplots.The various analyses allowed us to draw the following conclusions. The reference for the comparisons will be the model 15a/16a consisting in applying the original FTE model, based on a fixed selection of parameters applied on the extended dataset of enterprises. For the "nested" variants (15/16an, 15angr/16angr, 15angrc/16angrc, 15angrcm/16angrcm), as there is a specific parameter for each combination of NOGA and GREG, we generally observe more dispersion between the parameters within one year, but also more dispersion in the evolution of the parameters between two years. On the other hand, we do not observe any parameters or year-to-year variations in the parameters that are as extreme as those observed in the 15a/16a model. Based on the comparisons of *reldif*, we

can conclude that the different "nested" variants have a very similar quality, which is superior to that of variant 15a/16a. Variant 15angrcm/16angrcm only slightly impacts the quality of the estimation compared to the variant 15an/16an. Therefore it seems to be a good compromise between the aim for stability, the treatment of inversions and a quality close to variant 15an/16an (no grouping, no treatment of inversions). For completeness, the model denoted by 15fch/16fch, which is the reference FTE model that adjusts a selection of parameters on the $BESTA\_EUNT$ + $ERST\_MINI$ dataset, is also shown in the boxplots.

2.6 Adaptations to the model for NOGA 78

The model for NOGA 78 (Employment activities) was fitted to a dataset that included the enterprises from the secondary and tertiary sector excluding those in NOGA 78. This approach takes into account the fact that the majority of employees in NOGA 78 are temporary workers, who in principle work in all sectors of economic activity. For the original model (a separate model for men and women, respectively, with no separation by economic activity), a parameter was estimated for each of the four main variables (number of employees per salary class) with slope adjustments for a selection of GREGs. We have adapted the model for NOGA 78 in line with the adaptations to the model for the other NOGAs (the principal FTE model). A nested model was therefore also fitted to the population of enterprises excluding NOGA 78. However, unlike the nested model for the other NOGAs described in Section 2.3, but in line with the original model for NOGA 78, only the variable GREG was used as a categorical variable for the separate slope model. The results obtained showed that the adaptation of the model for NOGA 78 with the new parameterisation and on the new dataset enhances the stability in the distribution of the parameters between different years.

2.7 Concluding remarks on the new FTE model

Due to the revision of JobStat for the reference year 2015, the original FTE model was based on a dataset that was no longer adequate. This is the main reason why it was not selected for the calculation of FTE for STATENT 2015. In this context, model 15a/16a could be seen as a minimal adjustment - estimating the original model on the new dataset by taking into account the changes in the JobStat survey. However, all our analyses showed that, regardless of the model, the FTEs calculated on the basis of the new JobStat data (for the years 2015 and 2016) are lower than those obtained with the original model estimated on the basis of the 2014 data used in the STATENT 2015 publication. This result reflects the fact that the sampling frame for the new JobStat sample includes smaller enterprises, which tend to have employees with lower occupational levels than those observed in the reference population for JobStat before 2015. In line with the quality measures and the treatment of inversions, we have chosen to use model 15angrcm/16angrcm for the construction of FTEs for STATENT 2016 and to implement the same model also for STATENT 2015 which was republished with STATENT 2016.

3 Methodological developments of the FTE model for STATENT 2015-2019 following the revision of the STATENT in 2021

3.1 Context

The main objectives of the revision of STATENT for the years 2011 to 2018 were to improve the quality of the data and the comparability of the results of STATENT and of the business demography statistics (UDEMO), by incorporating corrections transmitted after publication, e.g.

following new deliveries of OASI register and Profiling data. The new input data therefore resulted only in an update of the results under the same concepts. However, these revisions of the results gave us the opportunity to test enhancements of the FTE model in order to have a consistent model for the whole series as well as for STATENT 2019, which was produced in parallel with the revised series. Several variants of the modifications were analysed and tested on preliminary data in close collaboration with the STATENT production team. The variant finally selected for the STATENT revision is as follows:

- For STATENT 2015-2018, the nested model is re-estimated on the newly delivered data, ensuring consistency in robustification and making minimal changes to the model in order to ensure consistency in estimation over the years. All these developments and the resulting analyses, such as comparisons of coefficients with relative differences between years and boxplots of quality measures, are presented below.
- For STATENT 2011-2014, the FTE model is not re-estimated. In fact, it would not make sense to estimate the new model on 2011-2014 data because the JobStat sample has changed in 2015 and the populations are different. Therefore, the coefficients of the nested model for STATENT 2015 are used to calculate FTE in order to minimise the risk that model effects impact the comparability of the series.

The following sections present various analyses describing the quality and stability of the FTE model in the period 2015-2019.

3.2 Data used for modelling FTE for STATENT 2015-2019

The final datasets on which our developments are based were delivered on 31.05.2021. These data include all single-establishment enterprises from JobStat, quarter 4 and the ERST sample. Following the disappearance of the NOGA-QS subsample[5] from the ERST sample (progressive reduction and disappearance in 2019), we removed all enterprises belonging to this subsample for the 2015 to 2018 data. Table 4 gives an overview of the size of the dataset used for modelling FTEs for STATENT 2015-2019.

Table 4 Size of the data used for modelling as of 31.05.2021

STATENT	$BESTA\_EUNT$	$BESTA\_EUNT$ + $ERST\_EUNT$
2015	10 661	69 055
2016	10 330	59 346
2017	9 941	55 445
2018	10 933	52 486
2019	10 925	49 223

The table shows a gradual reduction in the data available for modelling. As the size of the JobStat sample remains relatively stable, the reduction is due to the ERST data. Compared to STATENT 2015, we have approximately 29% less data available for modelling FTE for STATENT 2019.

[5]The NOGA-QS survey covered enterprises not included in other surveys and aimed at updating information about the NOGA economic activities. From 2015 to 2018 it was entirely integrated in the ERST survey.

3.3 Effect of data on calibration margins

Following the reduction in the amount of data available, we carried out analyses to determine the number of enterprises per sector available for fitting the model and for the calibration of survey weights. The calibration approach for the revised STATENT 2015-2018 and STATENT 2019 is the same as the one described in Section 2.2. Before revision, calibration was done at the NOGA OFS50 level. In order to avoid calibration cells with less than 50 enterprises, we have decided to group the following NOGAs together:

- NOGA 19.2 with NOGA 21
- NOGA 49 with NOGA 50.1
- NOGA 52 with NOGA 53
- NOGA 61 with NOGA 62.3
- NOGA 64 with NOGA 65

We have analysed, for each year the ratios between the initial weights used for calibration and the calibration weights on the $BESTA\_CAL$ dataset. A summary of the distribution of these ratios is given in Table 5.

Table 5 Distribution of weight ratios

STATENT	2015	2016	2017	2018	2019
min	0.25	0.25	0.25	0.25	0.25
median	1.00	0.99	0.99	1.00	1.00
max	2.53	1.61	2.99	3.78	3.45

3.4 Analysis of the obtained results

This section presents the main results that allow us to assess the quality of the model, compare the results between STATENTs and monitor the evolution of total employment and FTE.

3.4.1 Robustification constants

The procedure for robustifying the enterprise weights used in the FTE model is described in Nedyalkova, D., and Assoulin, D. (2017). The robustification constants remain fixed in principle, but changes can be made from year to year if the percentage of enterprises for which weights are robustified differs significantly from the value in previous years. Table 6 shows the finally chosen values of the robustification constants for each of the four models (m/s2, m/s3, w/s2 and w/s3) for STATENT 2015-2019. Table 7 shows, for each model, the percentage of enterprises for which the weights have been robustified. We can see from the tables that both the constants and the percentages of robustified weights are relatively stable, with small variations between years.

3.4.2 Model parameters for the main FTE model

In order to assess the quality and the stability of the model and the evolution in parameters, we have first compared the distribution of the parameters by salary class, GREG and NOGA sector. We have also studied the differences in the parameters following two data deliveries

Table 6 Robustification constants, data as of 31.05.2021

STATENT	w/s2	w/s3	m/s2	m/s3
2015	3	3	5	5
2016	3	3	7	5
2017	3	3	7	5
2018	3	3	7	5
2019	3	3	7	7

Table 7 Percentage of enterprises with robustified weights, data as of 31.05.2021

STATENT	w/s2	w/s3	m/s2	m/s3
2015	0.58	0.71	0.50	0.29
2016	0.55	0.73	0.29	0.38
2017	0.84	0.70	0.27	0.38
2018	0.75	0.76	0.44	0.46
2019	0.71	0.77	0.78	0.48

or between two consecutive years. The median of the distribution of model parameters, by salary class, for each year and each model (m/s2, m/s3, w/s2 and w/s3) is given in Table 8. Note that for the m/s2 model the number of salary classes is only 3 because V3 and V4 have been combined (see Nedyalkova, D., and Assoulin, D., 2017, Section 5.1). Table 9 presents the median of the differences in parameters between two consecutive years.

Table 8 Median of the estimated parameters, data as of 31.05.2021

Variable	Year	m/s2	m/s3	w/s2	w/s3
V1	2015	0.370	0.211	0.209	0.207
V1	2016	0.279	0.282	0.200	0.228
V1	2017	0.338	0.220	0.238	0.244
V1	2018	0.310	0.246	0.263	0.198
V1	2019	0.143	0.244	0.244	0.211
V2	2015	0.693	0.380	0.343	0.349
V2	2016	0.617	0.383	0.331	0.299
V2	2017	0.652	0.376	0.341	0.312
V2	2018	0.663	0.369	0.318	0.324
V2	2019	0.630	0.366	0.331	0.358
V3	2015	0.989	0.641	0.568	0.602
V3	2016	0.990	0.661	0.582	0.618
V3	2017	0.992	0.691	0.583	0.607
V3	2018	0.991	0.713	0.575	0.636
V3	2019	0.991	0.716	0.610	0.626
V4	2015		0.945	0.947	0.893
V4	2016		0.940	0.950	0.880
V4	2017		0.940	0.942	0.879
V4	2018		0.939	0.948	0.881
V4	2019		0.935	0.936	0.879

Our analyses showed that, in general, the median of the parameters remains relatively stable from STATENT 2015 to 2019. What we observed for STATENT 2019 was the very low coeffi-

Table 9 Median of differences, data as of 31.05.2021

Variable	Difference	m/s2	m/s3	w/s2	w/s3
V1	16/15	-0.091	0.071	-0.008	0.021
V1	17/16	0.059	-0.062	0.037	0.016
V1	18/17	-0.028	0.026	0.025	-0.046
V1	19/18	-0.167	-0.002	-0.019	0.013
V2	16/15	-0.032	0.011	-0.042	-0.011
V2	17/16	0.014	-0.022	0.017	0.013
V2	18/17	0.013	0.012	-0.002	0.006
V2	19/18	-0.037	-0.003	0.000	0.019
V3	16/15	0.003	0.026	0.002	0.015
V3	17/16	0.000	0.004	0.004	-0.011
V3	18/17	-0.003	0.001	0.006	0.032
V3	19/18	0.000	0.012	0.028	-0.024
V4	16/15		-0.001	-0.007	-0.009
V4	17/16		-0.003	0.003	0.000
V4	18/17		0.003	0.001	-0.003
V4	19/18		-0.003	-0.008	-0.006

cient for the first salary class for m/s2, which decreased from 0.310 to 0.143. Various studies have been carried out to try to identify the cause of this change, including studies of residuals and outliers. The exact cause for the decrease in the value of the FTE coefficient could not be identified. However, for the vast majority of enterprises in sector 2, the number of men in the first salary class is very low (even equal to 0 in 95.36% of cases). This may explain a certain instability in the estimated coefficient for m/s2, but also the fact that its impact on the total number of FTEs at the different levels considered, as well as on the quality measures, remains small.

3.4.3 Model for NOGA 78

For STATENT 2019 and the revised STATENTs 2015-2018, we used the nested model for NOGA 78 with adjustment by GREG described in Section 2.6. The results obtained show that the evolution of the coefficients remains relatively stable, with the exception of the coefficient estimated for men in STATENT 2016, where the coefficient for the first salary class is 0.27, while for 2015 and 2017 it is equal to 0.17 for all enterprises. A summary of the median of the estimated parameters by salary class, for men and women, STATENTs 2015-2019 is given in Table 10.

3.4.4 Quality measures

In order to assess the quality of the model we use the quality measures described in Nedyalkova, D., and Assoulin, D. (2017). The boxplots of the error distribution of the model fit are a visual tool that allows us to quickly identify if there is a specific problem. The boxplots of the *reldif* quality measure (see Figure 6, A.4) showed extreme values for the model m/s3, for years 2018 and 2019. A more detailed analysis showed that the extreme values can, to a wide extend, be explained by two (in 2018), respectively one (in 2019) enterprise(s) with extreme weighted prediction errors in NOGA 53 (Postal and courrier services). It should be noted that while the quality measures are based on the weight $w_i^{(2)}$, model estimation is based on a robust version of the weight $w_i^{(1)}$. This may explain the results of the subsequent analyses, which showed that

Table 10 Median of the estimated model parameters for NOGA 78, data as of 31.05.2021

Variable	Year	men	women
V1	2015	0.17	0.18
V1	2016	0.27	0.20
V1	2017	0.17	0.21
V1	2018	0.16	0.20
V1	2019	0.16	0.19
V2	2015	0.39	0.33
V2	2016	0.38	0.33
V2	2017	0.39	0.32
V2	2018	0.43	0.34
V2	2019	0.37	0.37
V3	2015	0.76	0.59
V3	2016	0.75	0.57
V3	2017	0.73	0.57
V3	2018	0.76	0.59
V3	2019	0.74	0.57
V4	2015	0.97	0.87
V4	2016	0.97	0.87
V4	2017	0.97	0.87
V4	2018	0.97	0.89
V4	2019	0.97	0.88

the impact of the three extreme units on the estimated model parameters remained rather moderate. For this reason, we decided to simply take note of these phenomena without modifying the model.

Overall, the analyses of the quality measure (see Figures 5, 6, 7, 8) showed that there are indeed differences between years, but given the small scale in the distribution of these measures and the small number of observations for some of the considered domains, we have concluded that the quality of the model remains good and stable.

3.5 Monitoring changes in total employment

An important factor in examining the results of STATENT is to compare the evolution of the number of employees, EMPTOT, with the evolution of the total FTE. Table 11 shows the totals, absolute and relative differences (in percentages) for STATENT 2015-2019 on the overall level of total employment and FTE as well as the values of the overall mean occupational level (MOL, where $\text{MOL} = \text{FTE}/\text{EMPTOT} \leq 1$) for each year. If the mean occupational level remained constant between two consecutive STATENTs, then the absolute difference in FTE should be smaller than that in EMPTOT, i.e.

$$|\text{FTE}_{18} - \text{FTE}_{17}| = \text{MOL} * |\text{EMPTOT}_{18} - \text{EMPTOT}_{17}| \leq |\text{EMPTOT}_{18} - \text{EMPTOT}_{17}|.$$

As can be seen from Table 11, although the value of MOL is relatively stable, this was not the case for the evolution between STATENT 17 and STATENT 18. This raises the question of whether the observed phenomenon would still be present if the FTE model parameters had remained unchanged.

In order to answer this question, a first idea was to represent the evolution (the difference) of FTE of two consecutive STATENTs as the sum of two terms. For the difference in the FTE

Table 11 Evolution of total employment

Year	EMPTOT	FTE	MOL	difference in EMPTOT	FTE	rel. difference in EMPTOT	FTE
2015	5 078 167	3 989 515	0.786				
2016	5 120 376	4 010 999	0.783	42 209	21 483	0.83%	0.54%
2017	5 180 644	4 039 658	0.780	60 268	28 659	1.18%	0.71%
2018	5 252 302	4 123 413	0.785	71 658	83 755	1.38%	2.07%
2019	5 310 225	4 156 341	0.783	57 923	32 928	1.10%	0.80%

of STATENT 2017 and 2018, denoted respectively by FTE_{17m17} and FTE_{18m18}, where the first suffix (17 or 18) denotes the year and the suffix m17 or m18 denotes the FTE model parameters used, either those of STATENT 2017 or those of STATENT 2018, we have:

$$\text{FTE}_{18m18} - \text{FTE}_{17m17} = (\text{FTE}_{18m18} - \text{FTE}_{17m18}) + (\text{FTE}_{17m18} - \text{FTE}_{17m17})\,.$$

The first term ($\text{FTE}_{18m18} - \text{FTE}_{17m18}$) represents the effect due to structural change (number of employees in the different salary classes, number of enterprises in the economic activity, etc.). The second term ($\text{FTE}_{17m18} - \text{FTE}_{17m17}$) represents the effect due to the change in parameters (same year, two different sets of parameters). We call this the model effect. The change (the difference) in FTE between two consecutive years is therefore given by the sum of these two terms, one coming from the structural effect and the other from the model effect. The aim of the following analyses was to compare the structural effect with the model effect. This has been done for the years 2015-2016, 2016-2017 and 2017-2018. The results were produced by the STATENT production team and the subsequent analyses were carried out jointly by METH and the STATENT team. Only the results for 2017-2018 are presented here.
Table 12 shows, for each combination of year and model, the following totals: number of enterprises, EMPTOT and FTE. It should be noted that the number of enterprises and EMPTOT does not depend on the model. The slightly different values between 17m17 and 17m18 are due to the fact that, in the process of generating a STATENT, data are regularly updated. This phenomenon also explains the fact that the values of the model effect for the EMPTOT evolutions (22, 32, -103) are not exactly zero, as might be expected.
It should be noted that Table 12 also presents results for the primary sector. Although the primary sector was not within the scope of this methodological report and the main developments made for the calculation of FTE for STATENT, it should be mentioned that for the primary sector most of the FTE values come from surveys, especially the Farm Structure Survey (also known as STRU), which is an exhaustive survey covering all farms. For those enterprises for which we had to estimate an FTE, a simple model was developed. The model is fitted to a very small number of enterprises in the primary sector, excluding NOGA '01' and the STRU survey. We used the PROG GLM procedure with two explanatory variables (V1 and V2 together, V3 and V4 together). No distinction by gender, no GREG or NOGA adjustments.
The results show that for all enterprises in STATENT 2017 and 2018, the effect of the model is generally smaller than the structural effect. A more detailed analysis by economic activity shows differences between sector 2 and sector 3. We can see that in sector 2, the difference in FTE due to the change in parameters (model effect) is equal to $1\ 432$ FTE and that the total difference in FTE between STATENT 2017 and 2018 is equal to $17\ 200$ FTE. This means that the model effect in sector 2 accounts for 8.3% of the total difference. In sector 3, we have a difference of $12\ 383$ FTE due to the change in parameters and a total difference of $63\ 883$ FTE between the two STATENTs. The model effect accounts for 19.4% of the differences and is therefore larger than in sector 2.

Table 12 Breakdown of change in employment and FTE by sectors for STATENT 2017 and 2018

		Totals			Evolution			
Year & model	Sec.	Number of enterprises	EMPTOT	FTE	effect	difference in EMPTOT	difference in FTE	ratio
17m17	1	53 285	160 006	104 221	model	22	38	-
17m18	1	53 287	160 028	104 259	structural	-1 568	-296	0.19
18	1	52 554	158 460	103 962	evolution	-1 546	-258	0.17
17m17	2	93 882	1 088 071	987 932	model	32	1 432	-
17m18	2	93 883	1 088 103	989 365	structural	13 210	15 768	1.19
18m18	2	93 341	1 101 313	1 005 133	evolution	13 242	17 200	1.30
17m17	3	479 270	3 934 480	2 921 013	model	-103	12 383	-
17m18	3	479 275	3 934 377	2 933 396	structural	58 194	51 500	0.88
18m18	3	481 338	3 992 571	2 984 896	evolution	58 091	63 883	1.10

Thus, if we examine the row with effect=structural in Table 12, we find that the answer to the question of whether the observed phenomenon between STATENT 2017 and 2018 would still be present if the FTE model remained unchanged is no. If we ignore the model effect, the ratio between the difference in FTE and the difference in total employment in sector 3 is 0.88. With the model effect, this ratio is 1.10. In sector 2, the ratios are 1.19 and 1.30 respectively. We can conclude that the effect of the FTE model is much larger in sector 3. In this way, we have illustrated the use of the decomposition in the analysis and interpretation of the results of STATENT, in particular of the evolution of total employment.

Another analysis, which aims at measuring the impact of the model on the evolution of total employment and total FTE, was carried out for the STATENTs 2011-2019. It was based on a panel of enterprises between two successive STATENTs that obtain their FTE only from the model and not from the surveys (we call them model units). The results are shown in Table 13.

Table 13 Evolution of total employment for model units

	Totals			Evolution		
Panel	Number of enterprises	EMPTOT	FTE	difference in EMPTOT	difference in FTE	ratio
11	341 721	1 763 762	1 316 807			
12	341 721	1 744 489	1 302 927	-19 273	-13 887	0.72
12	353 347	1 835 533	1 373 691			
13	353 347	1 827 805	1 372 109	-7 728	-1 583	0.20
13	373 654	1 957 898	1 473 720			
14	373 654	1 938 923	1 458 109	-18 975	-15 611	0.82
14	378 397	2 005 264	1 508 684			
15	378 397	1 983 048	1 490 816	-22 216	-17 868	0.80
15	380 813	2 185 416	1 662 701			
16	380 813	2 165 045	1 644 056	-20 371	-18 645	0.92
16	397 517	2 344 405	1 789 769			
17	397 517	2 327 192	1 768 004	-17 213	-21 765	1.26
17	423 743	2 527 928	1 920 276			
18	423 743	2 527 171	1 935 055	-757	14 778	-19.52
18	450 670	2 714 927	2 080 468			
19	450 670	2 705 130	2 069 831	-9 797	-10 637	1.09

We can see that, taking into account only model units, there was an increase in FTE for the 2017-2018 evolution, while total employment showed a decrease. It should be noted that between 2012 and 2013, even though the model coefficients are the same, the difference in FTE is much smaller than the difference in total employment (a ratio of 0.20).
We also proposed to examine panels of enterprises, categorised according to whether total employment remained stable, increased or decreased between years. We found that in cases where employment increased or decreased, there was a consistency between the increase or decrease in total employment and FTE, i.e. when EMPTOT increased, the increase in FTE was less and vice versa.

3.6 Conclusions following the revision

These various analyses have shown that the FTE model, while providing generally consistent results, is sensitive to the data on which the estimates are based. The developments made for STATENT 2016 and the new modifications to the model and calibration applied for STATENT 2015-2019 ensure relative stability in the construction of FTE. The analyses carried out - statistics on FTE coefficients, quality measures, graphs of the evolution of FTE and total employment, and the breakdown of changes into structural and model effects - are valuable tools that allow us to monitor these changes, to assess the quality of the model and also to put into perspective its impact on the whole STATENT.

4 Variance estimator of the FTE estimate

4.1 FTE in STATENT

In the previous sections, we have described the methodological developments in the FTE model first due to changes in the JobStat statistics and then following the revision. However, a missing part of our analyses was a variance estimator of the FTE estimate for STATENT. In principle, in STATENT, we use survey FTE when available. For enterprises not covered by a complementary survey, FTE are constructed using the FTE prediction model. There is also the case of public administration enterprises for which FTE are imputed. In order to ensure consistency between sources and with the variable number of employees (EMPTOT), survey and imputed FTE can be adjusted subsequently. Thus, the estimator of total FTE for STATENT is a complex mixture of survey FTE, imputed FTE and predicted FTE based on variables coming from the OASI register which is the reference source for STATENT.
The integration of administrative and survey data has revealed the existence of inconsistencies in the variable EMPTOT. In such cases, the survey FTE, FTE_S, had to be adjusted to reflect the total number of employees according to the register, EMPTOT_R. The aim was to harmonise the survey FTE and total employment variable (EMPTOT_S) with the variable EMPTOT_R for modelling purposes and for the calculation of STATENT FTE totals, see Nedyalkova, D., and Assoulin, D. (2017), Sections 4.2 and 4.3 for a detailed description of the harmonisation procedure and its impact on the dependent and independent variables of the model and Section 4.4 for the adjustment of survey FTE for the calculation of STATENT totals. We will only note here, that for the calculation of totals, survey FTE are adjusted according to the following two cases:

- case 1: EMPTOT_R $>$ EMPTOT_S $=>$ FTE_S is adjusted,
- case 2: EMPTOT_R $<$ EMPTOT_S $=>$ FTE_S is adjusted,

When EMPTOT_R is equal to EMPTOT_S then FTE_R is equal to FTE_S. The adjustment of FTE_S is different for case 1 and case 2 (see Nedyalkova, D., and Assoulin, D., 2017). This illustrates the complexity of the estimator which has influenced the choice of the variance estimation method.

4.2 Bootstrap variance estimation method

Linearisation is often the method of choice for estimating the variance of complex estimators. However, in some cases, and particularly in the case of estimating total FTE for STATENT, due to the complexity of the estimator, an analytical form of the estimator, which is a necessary condition for linearisation, is not easily obtainable. In such cases, the use of a bootstrap variance estimation technique may be an appropriate choice. However, bootstrap estimation can sometimes be very time-consuming, depending on the complexity of the estimators, the population and sample sizes, and the number of replications used.
The bootstrap technique has been developed in the context of an infinite population, Efron (1979). Since the bootstrap aims to approximate the distribution of an estimator, it can be used to obtain a consistent estimate of the variance and confidence intervals.
The adaptation of the bootstrap technique to the sampling theory based on a finite population has been the subject of a considerable amount of research. For an equal probability sampling design, Gross (1980) suggests the construction of a pseudo-population U^* by means of the sample, in which we can resample according to a simple random sampling without replacement. In the case of unequal probability designs, and in particular a Poisson sampling design, a generalisation of this method based on the idea of the Horvitz-Thompson estimator has been proposed by Chauvet (2007). An important result, shown in Chauvet (2007) and Mashreghi et al. (2016), is that the bootstrap variance estimator of the Horvitz-Thompson estimator of the total is asymptotically unbiased under Poisson sampling.

4.3 Variance estimation of the total FTE for STATENT

To account for the sampling design and the complexity of the FTE estimator, the bootstrap technique was used to develop a procedure for estimating the variance (under the design) of the estimates of the total FTE for STATENT. Here, we adapted the algorithm proposed in Chauvet (2007, p. 174).

Algorithm for bootstrap variance estimation

The FTE model is estimated on a sample composed of the single-establishment enterprises from JobStat and ERST surveys. This will be the sample from which the pseudo-population will be constructed. Note that the public administration and the primary sector are excluded from the pseudo-population. Let p_k be the probability of selection for each unit of the initial sample s. For the ERST survey, considered as a census, we set $p_k = 1$, and for the Jobstat survey we used the sampling probability corrected for non-response.

Step 1 : Construction of the fixed part U_f^* of the pseudo-population.

(a) Let $r_k = 1/p_k$ and let $\lfloor r_k \rfloor$ be the integer part of r_k.

(b) U_f^* is obtained by replicating each unit k in the initial sample $\lfloor r_k \rfloor$ number of times.

Step 2 : Construction of the random part U_r^* of the pseudo-population.

(a) We select from s a pseudo-population U_r^* according to a Poisson sampling design with selection probabilities: $r_k - \lfloor r_k \rfloor$.

(b) The pseudo-population U^* is obtained by combining U_r^* and U_f^*.

(c) The selection of the random part is repeated C times in order to obtain C different pseudo-populations $U_c^*, c = 1, \ldots, C$.

Step 3 : Selection of bootstrap samples and estimation.

Let y_k, $k \in U_c^*$, denote the observed values (survey FTE). For each of the C pseudo-populations $U_c^*, c = 1, \ldots, C$, we repeat B times the following steps:

(a) We select a sample $s_b^*, b = 1, \ldots, B$, from the pseudo-population U_c^* with selection probabilities p_k.

(b) We perform data harmonisation as described in Nedyalkova, D., and Assoulin, D. (2017), Section 4.3 and fit the model described in Section 2.3 on the selected bootstrap sample.

(c) We adjust inconsistent survey FTE on the sample s_b^*, which gives us a new variable denoted by $\tilde{y}_k$.

(d) We calculate an estimator of the total FTE, $\hat{Y}_{c,b}^*, b = 1, \ldots, B$, based on the bootstrap sample s_b^* and the part of the pseudo-population U_c^* not included in the sample, denoted by $\bar{s}_b^*$, according to the equation:

$$\hat{Y}_{c,b}^* = \sum_{k \in s_b^*} \tilde{y}_k + \sum_{k \in \bar{s}_b^*} \hat{y}_k, \quad (1)$$

where $\hat{y}_k$ denotes the model-estimated FTE for unit k.

The variance of the estimates obtained on the basis of all the samples drawn from the pseudo-population U_c^* is given by:

$$\hat{V}_c^* = \frac{1}{B-1} \sum_{b=1}^{B} \left(\hat{Y}_{c,b}^* - \hat{Y}_c^* \right)^2, \quad (2)$$

where $\hat{Y}_c^* = B^{-1} \sum_{b=1}^{B} \hat{Y}_{c,b}^*$.

Finally, the variance estimate (based on all pseudo-populations) is obtained by taking the average of the variance estimates corresponding to each of the C pseudo-populations:

$$\hat{V}^* = \frac{1}{C} \sum_{c=1}^{C} \hat{V}_c^*. \quad (3)$$

4.4 Results

In this section, we present the results obtained for the bootstrap variance estimation of the total FTE for the STATENT 2020 and 2021. For STATENT 2020, the size of the initial sample is $48\ 022$ and for STATENT 2021 - $37\ 361$. For illustration, for STATENT 2020, the average bootstrap sample size across all pseudo-populations is $48\ 031$ and the average pseudo-population

size is $497\,352$. For both STATENT 2020 and 2021, we created 10 pseudo-populations and selected 100 bootstrap samples from each pseudo-population following the procedure described in Section 4.3. The total number of bootstrap samples was set to 1 000. This guarantees a sufficient number of replications to obtain a good balance between precision and computational time, see e.g. Chauvet (2007). This choice was also confirmed by our experience and by the complexity of the computations we had to perform for each bootstrap sample, i.e. data harmonisation, model fit, treatment of inversions, final adjustment of the total FTE. It should be noted that even with this number of replications, the computational time was in the order of 10 to 12 hours and the calculations were performed by groups of samples, often having to restart the process due to server interruptions.
We then calculated coefficients of variation (CVs) by sector of economic activity (industry and services), by GREG and by different NOGA levels. The calculation of a variance estimator for the primary sector did not make sense, as all the enterprises in the sample had sampling weights of 1. These calculations were made for the estimation of total FTE by gender (men and women) as well as for the total. The results for STATENT 2020 and 2021 are presented in Tables 14, 15, 16, 17 and 18. Note that for the results by NOGA OFS50, we present only the minimum, maximum and median values of the distribution of CVs obtained by sector.

Table 14 CVs (in %) by economic sector of activity

Year	2020	2021	2020	2021	2020	2021
Sector	Men	Men	Women	Women	Total	Total
2	0.09	0.09	0.21	0.20	0.09	0.08
3	0.13	0.15	0.18	0.20	0.12	0.13

Table 15 CVs (in %) by GREG in sector 2

Year	2020	2021	2020	2021	2020	2021
GREG	Men	Men	Women	Women	Total	Total
1	0.19	0.17	0.38	0.40	0.17	0.16
2	0.18	0.19	0.27	0.31	0.15	0.17
3	0.21	0.24	0.60	0.36	0.21	0.21
4	0.16	0.13	0.77	0.52	0.20	0.15
5	0.17	0.17	0.42	0.48	0.16	0.17
6	0.19	0.20	0.64	0.64	0.20	0.21
7	0.31	0.31	1.13	0.96	0.35	0.34

Table 16 CVs (in %) by GREG in sector 3

Year	2020	2021	2020	2021	2020	2021
GREG	Men	Men	Women	Women	Total	Total
1	0.20	0.21	0.30	0.29	0.19	0.20
2	0.36	0.41	0.44	0.47	0.32	0.35
3	0.32	0.29	0.55	0.50	0.34	0.31
4	0.20	0.22	0.26	0.26	0.17	0.18
5	0.28	0.33	0.48	0.46	0.28	0.28
6	0.32	0.35	0.49	0.56	0.31	0.33
7	0.57	0.73	0.88	0.86	0.53	0.58

Table 17 Distribution of CVs (in %) by NOGA OFS50 in sector 2

Year	2020	2021	2020	2021	2020	2021
Value	Men	Men	Women	Women	Total	Total
min	0.020	0.03	0.12	0.13	0.06	0.06
median	0.11	0.12	0.30	0.30	0.11	0.11
max	0.54	0.31	1.24	1.19	0.50	0.38

Table 18 Distribution of CVs (in %) by NOGA OFS50 in sector 3

Year	2020	2021	2020	2021	2020	2021
Value	Men	Men	Women	Women	Total	Total
min	0.034	0.037	0.005	0.003	0.021	0.039
median	0.324	0.377	0.492	0.490	0.294	0.293
max	1.806	1.828	1.538	1.712	1.396	1.449

4.5 Discussion

This bootstrap technique has already been used successfully for several FSO projects. The bootstrap estimation of the variance allowed us to efficiently obtain an estimator of the variance of the total of FTE for STATENT which is a complex estimator combining survey and administrative data. The implemented method allows to take into account the sampling design with unequal selection probabilities and the complexity of the estimator, including the treatment of inconsistencies due to the use of different data sources and the treatment of parameter inversions. The CVs obtained are generally small, indicating a good precision of the estimator of total FTE in terms of variance. Note that, in some cases, the variance estimator for the total, which combines the results for men and women, may be larger than the one for men. This could be explained by the larger variance of the estimator of FTE for women. Intuitively, the smaller size of the initial sample in 2021 should generally lead to larger CVs. However, the evolutions of the sample sizes within different domains are not homogeneous which could explain for example, that the expected CV increase can be observed in sector 3, but not in sector 2.

5 Final remarks and conclusions

The first methodological report Nedyalkova, D., and Assoulin, D. (2017) has introduced the basis of the methodology for estimating FTE based on matched data coming from the JobStat and ERST surveys and the OASI register. The treatment of data inconsistencies, the development of the FTE prediction model and its application to FTE for STATENT were described in detail. However, in 2017, following the integration of STATENT into the sampling framework of the JobStat survey and its new sampling design for the reference year 2015, the methodology for estimating FTEs for STATENT 2015 had to be reviewed.
This report documents the changes that have been made to the FTE prediction model and the initial data used for modelling, first due to the changes in JobStat for STATENT 2015 and 2016 and, next, following the subsequent revision of the whole series of STATENT 2011-2018 which intervened in 2020. The results have been backed up by various monitoring analyses, e.g.:

- model quality,
- model stability,
- the number of units with robustified weights

over the period 2015-2019. The new FTE model and calibration approach have been applied to the whole series of STATENT and also to STATENT 2020 and 2021. Our analyses led to the conclusion that the model quality could be improved while the estimated model parameters show a satisfying stability over the years.

In Section 3.5 we have proposed a method for decomposing the differences between total FTE for two consecutive years into a model effect and a structural effect. The first one allowing to measure the impact of the model, i.e. the changes in parameters, and the second one allowing to observe an evolution that is not impacted by the re-estimation of the model.

The last part of the report proposes a bootstrap method for estimating the precision, in terms of variance, of the estimator of total FTE based on a pseudo-population and a simulation on $1'000$ bootstrap samples, for STATENT 2020 and 2021. The obtained results show the good precision of the estimator as well on the overall level (by economic sector of activity) as for smaller breakdowns. Variance estimation has proved to be a useful tool for evaluating and interpreting STATENT results on FTE. It would be interesting to continue the study for the subsequent STATENTs in order to evaluate the stability of the bootstrap variance estimation.

This report provides a complete overview of the developments in the estimation and its precision of FTE for STATENT over the last few years.

A Appendix

A.1 List of abbreviations

Table 19 Table of abbreviations

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

A.2 List of the different variants of the FTE model

- "16fch": original FTE model as described in Nedyalkova, D., and Assoulin, D. (2017). A model allowing separate slopes for some specific combinations of GREG and NOGA. Adjustment on a dataset containing $BESTA\_EUNT$ and $ERST\_MINI$ enterprises.
- "16a": A model allowing separate slopes for some specific combinations of GREG and NOGA (same as for the 16fch model). Adjustment on a new dataset comprising $BESTA\_EUNT$ and $ERST\_EUNT$ enterprises. Small changes in the weighting scheme.
- "16an": A "nested" model. Adjustment on a new dataset containing $BESTA\_EUNT$ and $ERST\_EUNT$ enterprises. Small changes to the weighting scheme.
- "16angr" : A "nested" model with grouping of NOGA sections B with C. Adjustment on a new dataset containing $BESTA\_EUNT$ and $ERST\_EUNT$ enterprises. Small changes to the weighting scheme.
- "16angra" : A "nested" model with grouping of NOGA sections B with C, D with E, K with L and R with S. Adjustment on a new dataset containing $BESTA\_EUNT$ and $ERST\_EUNT$ enterprises. Small changes to the weighting scheme.
- "16angrc": A "nested" model with grouping of NOGA sections B with C. Treatement of inversed model parameters using the "maximum value replacement" method. Adjustment

on a new dataset containing $BESTA\_EUNT$ and $ERST\_EUNT$ enterprises. Small changes to the weighting scheme.

- "16angrcm": A "nested" model with grouping of NOGA sections B with C. Treatement of inversed model parameters using the "mean-value replacement" method. Adjustment on a new dataset containing $BESTA\_EUNT$ and $ERST\_EUNT$ enterprises. Small changes to the weighting scheme.

A.3 Description of the classifications of economic activities

Table 20 NOGA Sections

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 21 NOGA Divisions 01 to 43

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

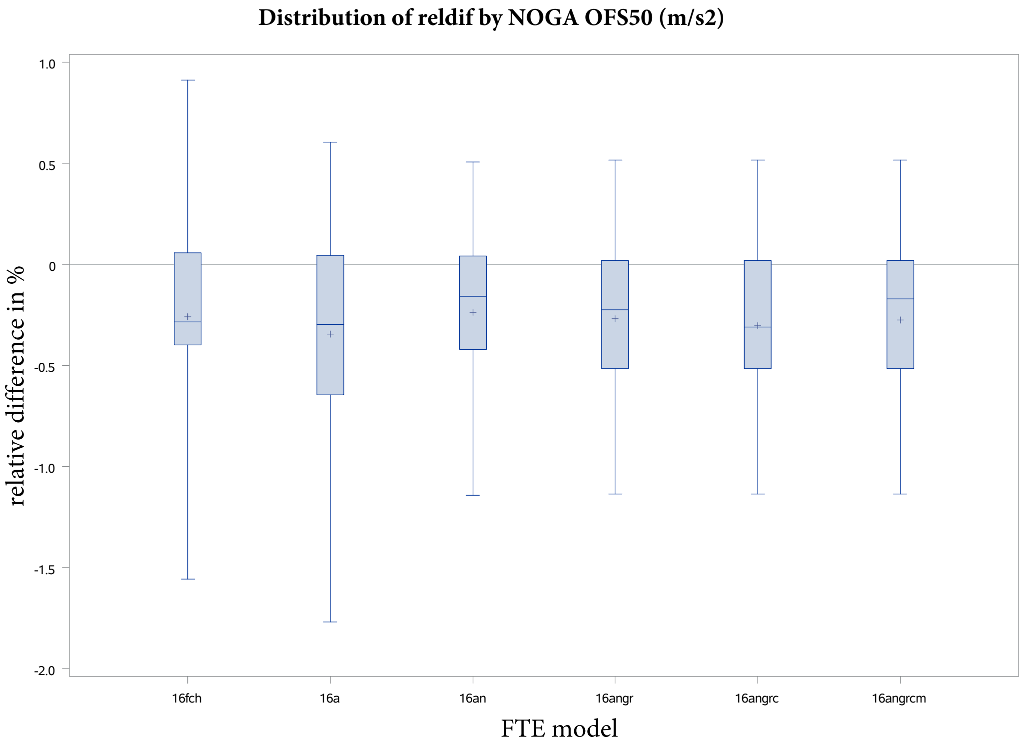
Table 22 NOGA Divisions 45 to 99

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 23 NOGA OFS50

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

A.4 Boxplots of the quality measure



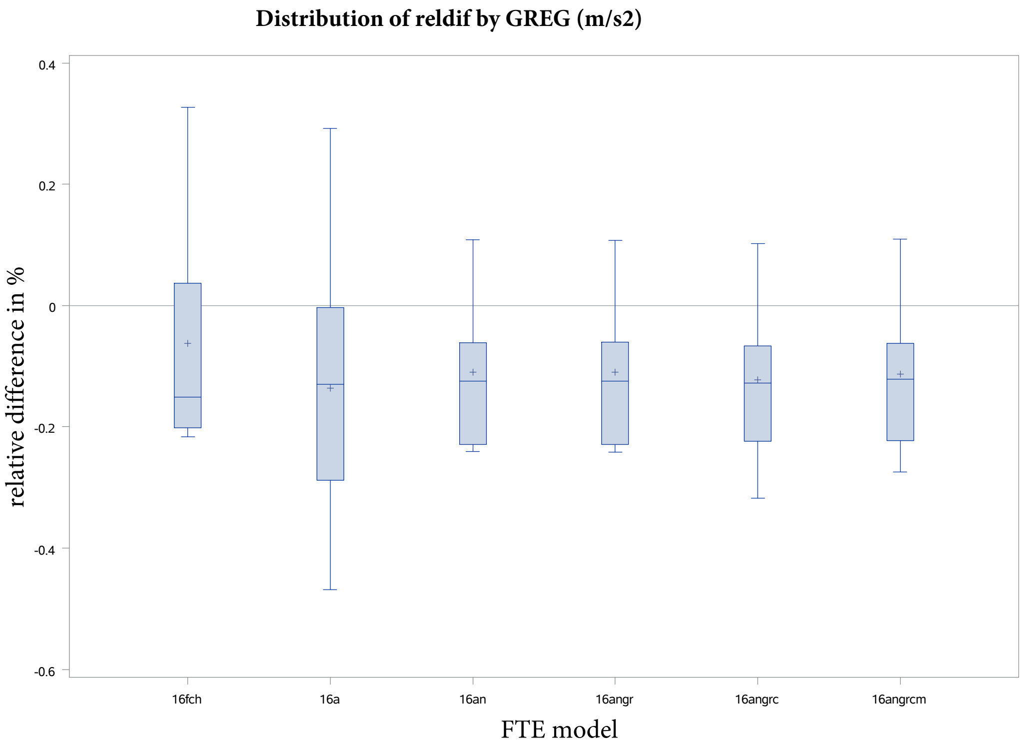


Figure 1 Relative differences for men/sector2

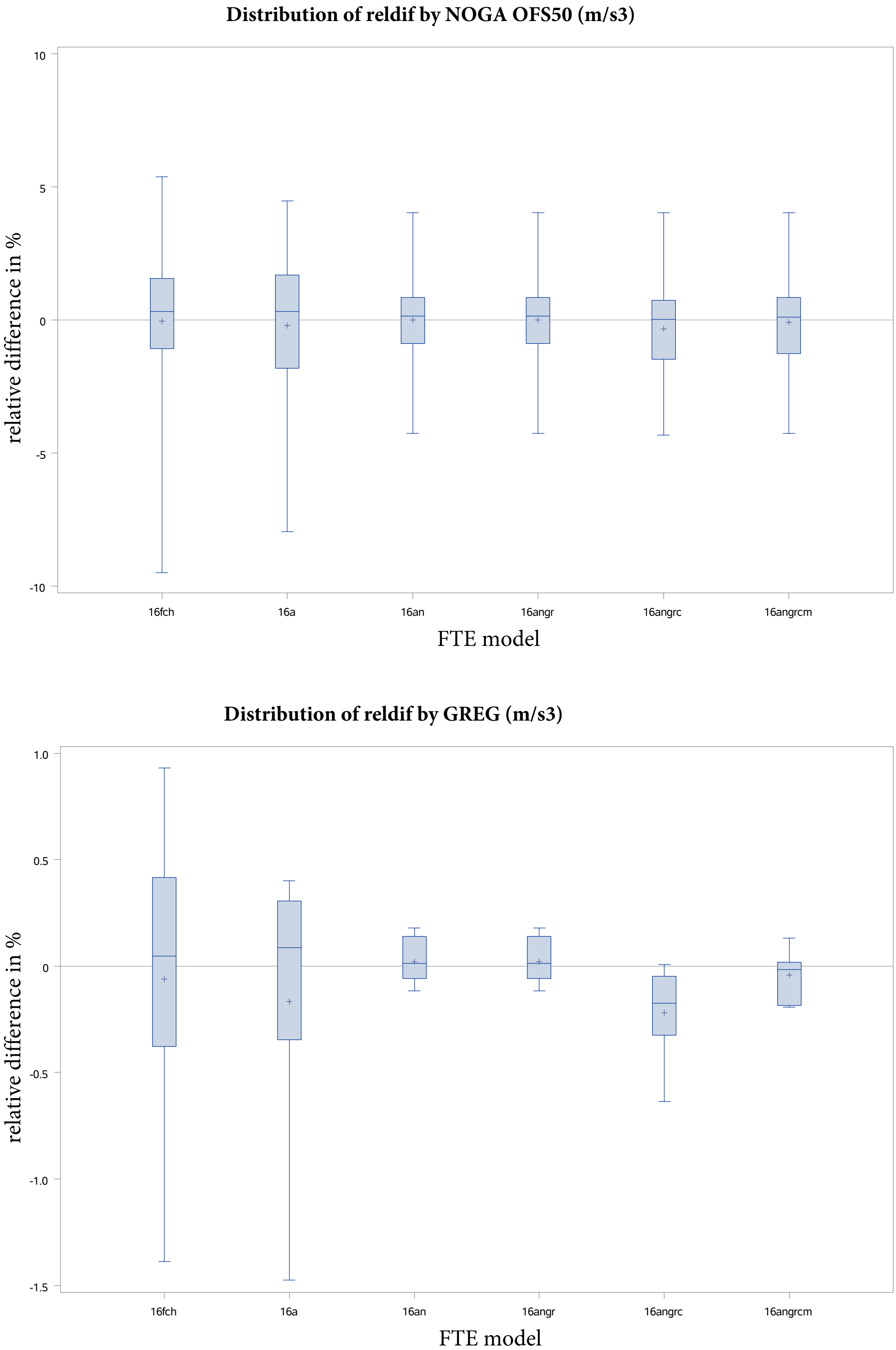


Figure 2 Relative differences for men/sector 3

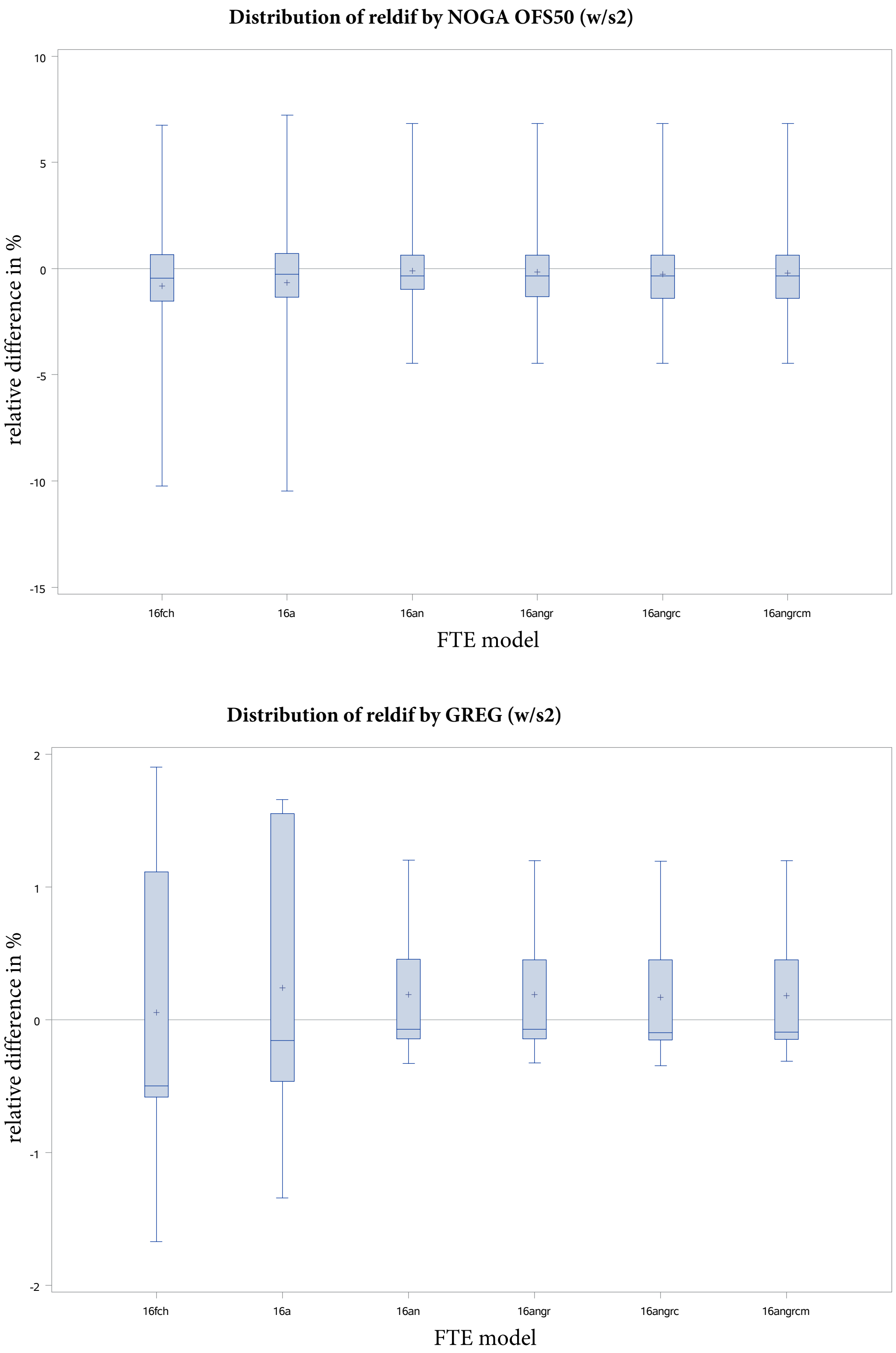


Figure 3 Relative differences for women/sector 2

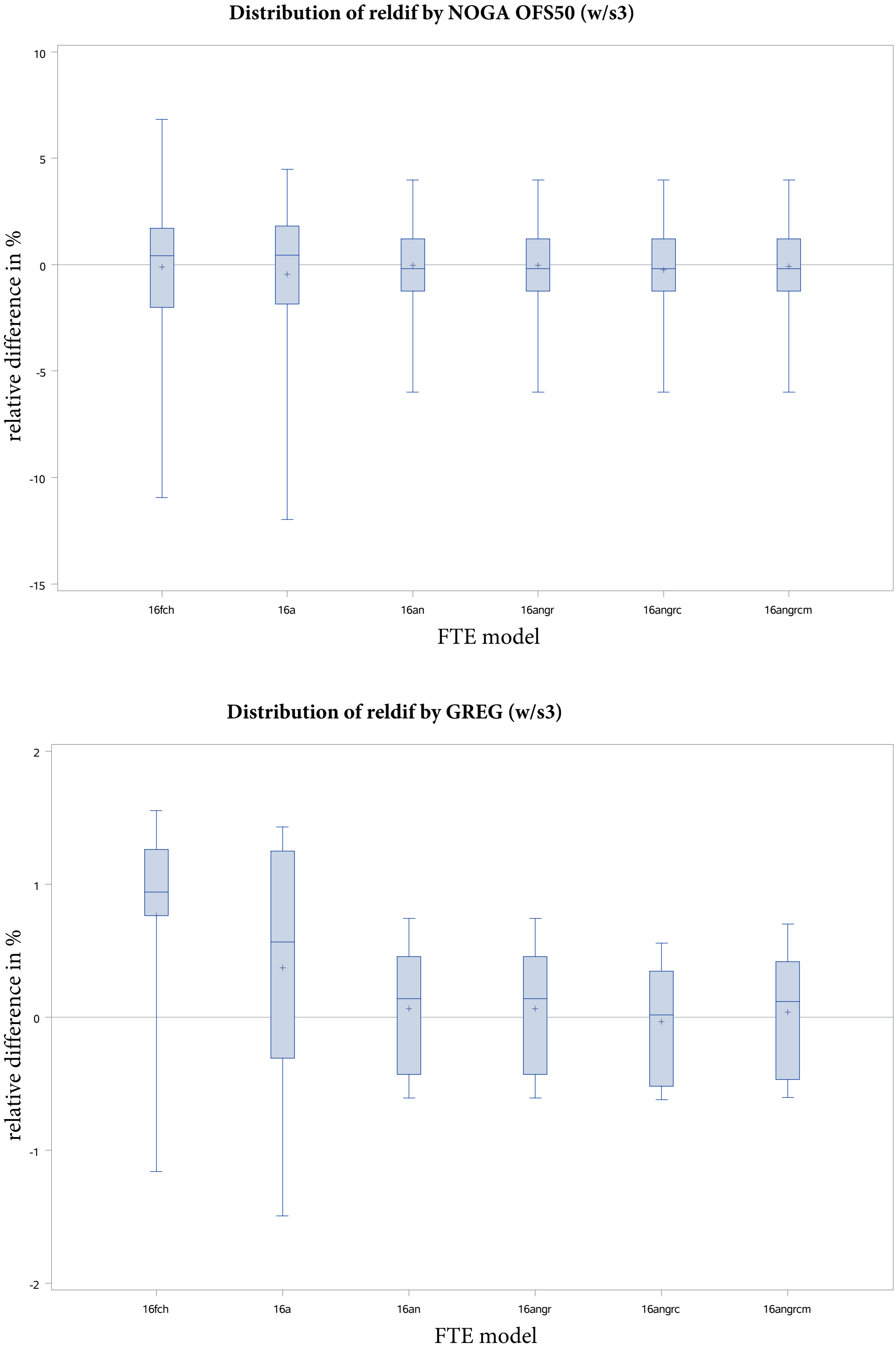
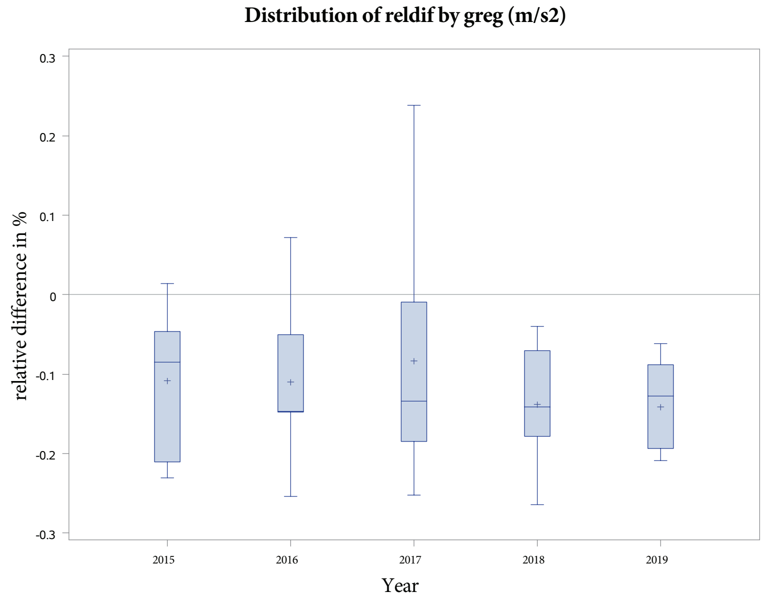


Figure 4 Relative differences for women/sector 3



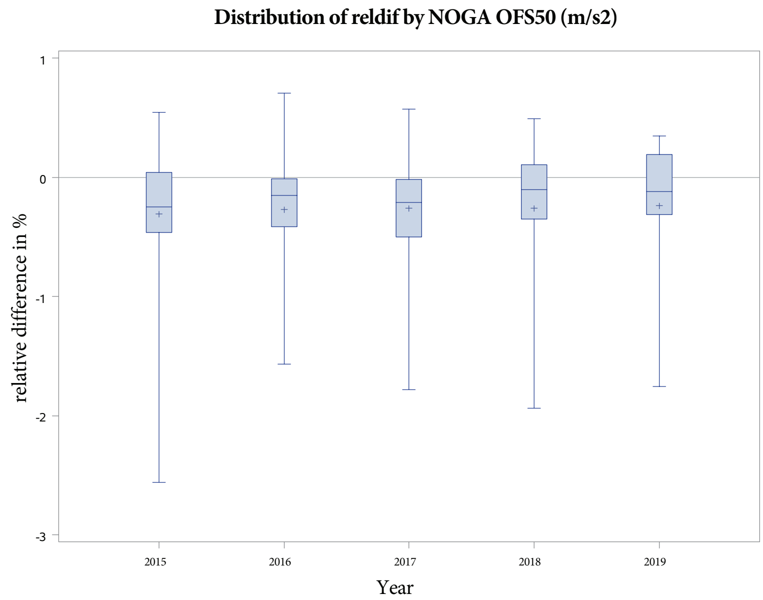
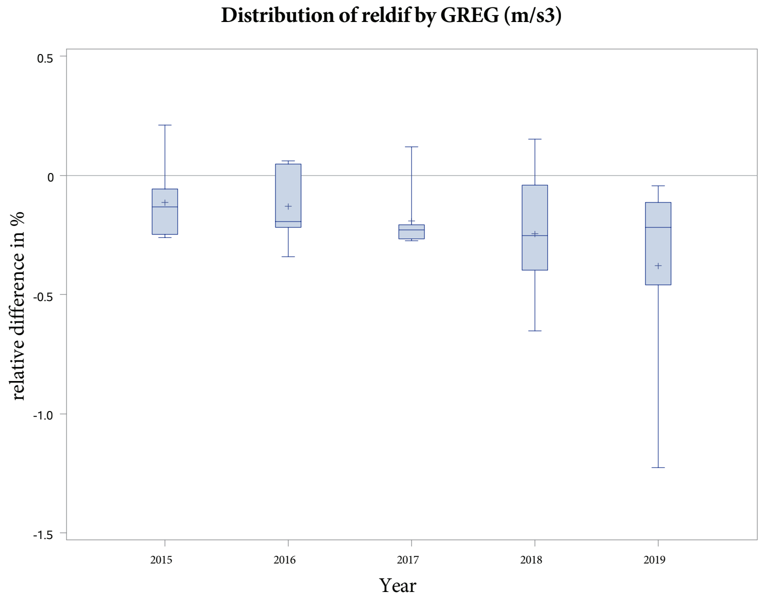


Figure 5 Quality measure *reldif* STATENT 2015-2019, m/s2



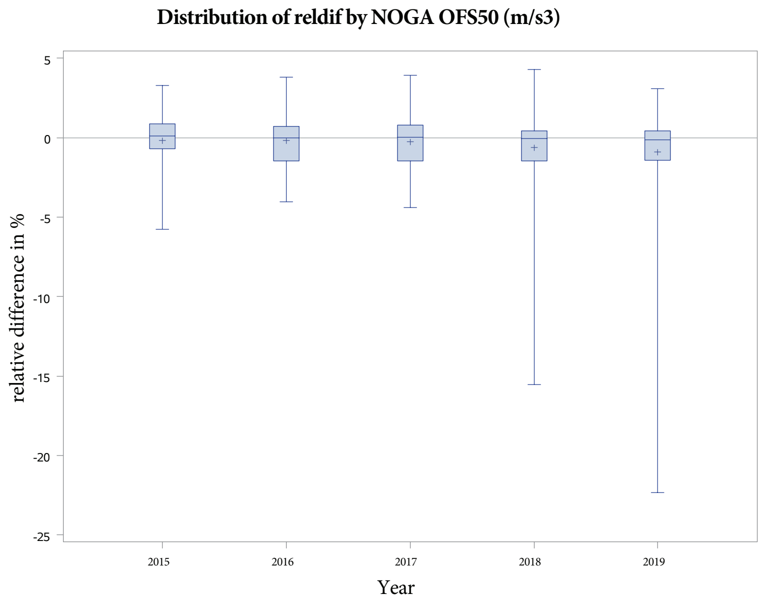
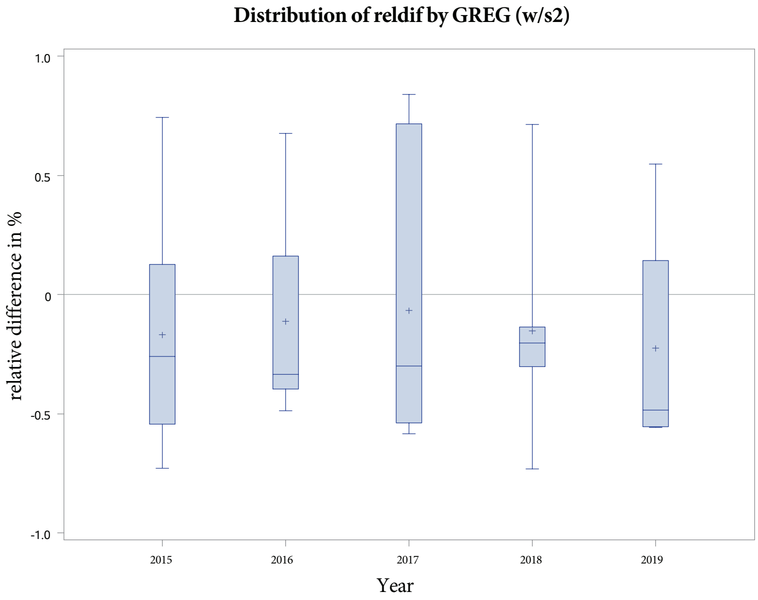


Figure 6 Quality measure *reldif* STATENT 2015-2019, m/s3



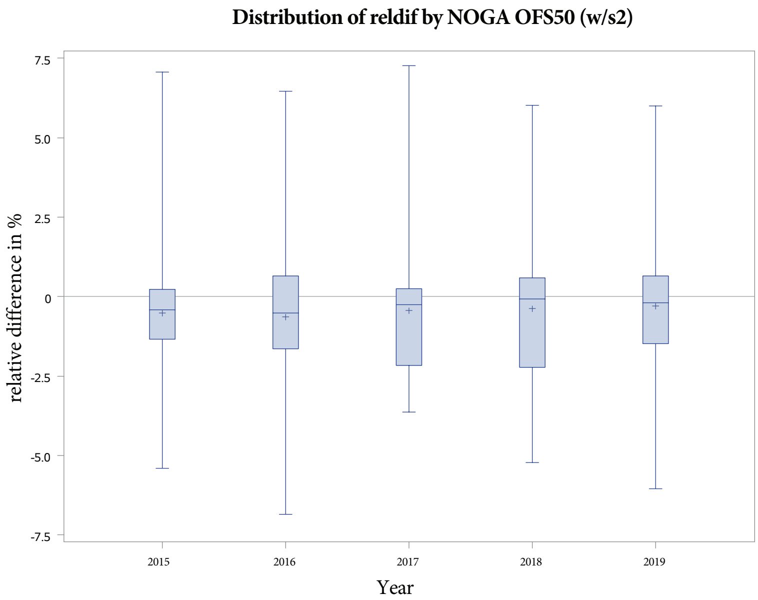
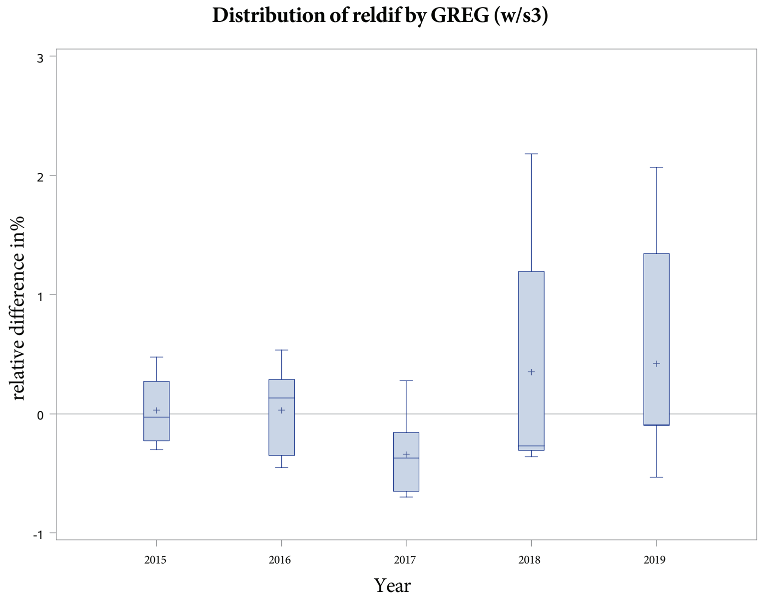


Figure 7 Quality measure *reldif* STATENT 2015-2019, w/s2



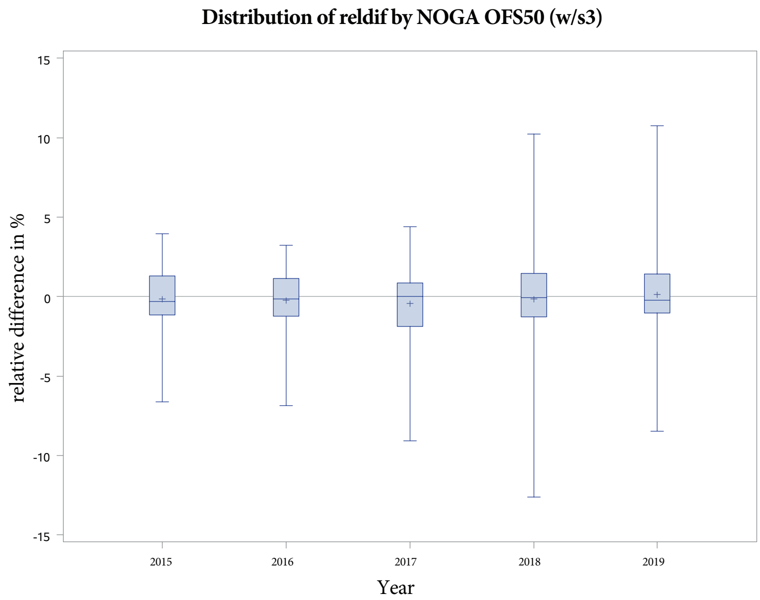


Figure 8 Quality measure *reldif* STATENT 2015-2019, w/s3

References

Chauvet, G. (2007). *Méthodes de Bootstrap en population finie*. PhD thesis, ENSAI.

Deville, J.-C. and Särndal, C.-E. (1992). Calibration Estimators in Survey Sampling. *Journal of The American Statistical Association*, 87:376–382.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1):1 – 26.

Gross, S. (1980). Median Estimation in Sample Surveys. In *Proceedings of the Section on Survey Research Methods*, pages 181–184. American Statistical Association, Alexandria, VA.

Mashreghi, Z., Haziza, D., and Léger, C. (2016). A survey of bootstrap methods in finite population sampling. *Statistics Surveys*, 10(none):1 – 52.

Nedyalkova, D., and Assoulin, D. (2017). Construction of full-time equivalents for the Swiss structural business statistics 2011. Methodological report, Swiss Federal Statistical Office.

Potterat, J., Assoulin, D. and J.-M. Nicoletti (2019). Beschäftigungsstatistik BESTA. Revision 2015: Stichprobenrahmen, Stichprobenplan und Hochrechnung. Methodological report, Swiss Federal Statistical Office.

SAS Institute, I. (2018). *SAS/STAT® 15.1 User's guide*. Cary, NC: SAS Institute Inc.

Sautory, O. and Le Guennec, J. (2003). La macro CALMAR2 : redressement d'un échantillon par calage sur marges – documentation de l'utilisateur. Technical report, Insee, Paris.

Methodology reports published by the FSO's Statistical Methods Section
Methodenberichte der Sektion Statistische Methoden des BFS

Nedyalkova, D., Assoulin, D. (2024). Swiss structural business statistics (STATENT) 2015-2021. Estimation and precision of the full-time equivalents. FSO number: 338-0088

Qualité, L. (2023). Relevé structurel de la population suisse 2018. Révision de la pondération et sélections répétées. Numéro OFS: 338-0087

Potterat, J., Assoulin, D. (2023). Forschung und Entwicklung 2021. Stichprobenplan, Gewichtung und Schätzverfahren. BFS-Nummer: 338-0086

Chevalier, C. (2023). Méthodes de correction des effets calendrier en usage à l'OFS en 2023. Construction, choix et application des régresseurs mis à disposition. Numéro OFS: 338-0085

Mariétan, R. (2022). Comportement en matière de voyages 2020. Révision de la pondération. Numéro OFS: 338-0084

Massiani, A. (2021). Enquête suisse sur la population active 2015-2020 : Révision de la pondération et de l'estimation de la variance. Numéro OFS: 338-0083

Chevalier, C., Qualité, L. (2021). Enquête suisse sur la structure des salaires 2018. Révision de la pondération. Numéro OFS: 338-0082

Qualité, L., Potterat, J. (2020). Enquête suisse sur la structure des salaires 2018. Révision du plan de sondage. Numéro OFS: 338-0081

Potterat, J., Qualité, L., Assoulin, D. (2019). Strukturerhebung der eidgenössischen Volkszählung. Stichprobenplan und -ziehung, Gewichtung, Schätzverfahren und Pooling 2010-2018. BFS-Nummer: 338-0080

Potterat, J., Assoulin, D., Nicoletti, J.-M. (2019). Beschäftigungsstatistik BESTA. Revision 2015: Stichprobenrahmen, Stichprobenplan und Hochrechnung. BFS-Nummer: 338-0079

Massiani, A. (2017). Estimation de la couverture du nouveau recensement en Suisse en 2012. Numéro OFS: 338-0078

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

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

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

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

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

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

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

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

Potterat, J. (2012). Use of conversion keys for NOGA2002-2008. FSO 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 OFS: 338-0066

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

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

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

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

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

Eichenberger P., Hulliger B., Potterat J. (2010). Describing the Anticipated Accuracy of the Swiss Population Survey. FSO 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 OFS: 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. FSO 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 OFS: 338-0057-05

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

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

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

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

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

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

Kilchmann, D. (2008). Statistik der sozialmedizinischen Institutionen 1999-2004 und Krankenhausstatistik 1999-2002. Einsetzungen für fehlende Daten. BFS-Nummer: 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 OFS: 338-0049

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

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

Gabler, S., Häder, S. (2007). Haushalts- und Personenerhebungen. Machbarkeit von Random Digit Dialing in der Schweiz. BFS-Nummer: 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 OFS: 338-0045

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

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

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

Renfer, J.-P. (2006). Enquête sur les chiffres d'affaires du commerce de détail. Plan d'échantillonnage et méthodes d'estimation. Numéro OFS: 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 OFS: 338-0040

Renaud, A. (2006). Statistique suisse des bénéficiaires de l'aide sociale. Pondération des communes 2004. Numéro OFS: 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. FSO number: 338-0038

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

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

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

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

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

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

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

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

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

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

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

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

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

Renaud, A. (2004). Analyse de données d'enquêtes. Quelques méthodes et illustration avec des données de l'OFS. Numéro OFS: 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 OFS: 338-0023

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

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

Potterat, J. (2003). Landwirtschaftliche Betriebszählung 2003. Stichprobenplan der Zusatzerhebung. BFS-Nummer: 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 OFS: 338-0019

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

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

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

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

Potterat, J., Salamin, P.A. (2002). Betriebszählung 2001. Methoden für die Datenbereinigung. BFS-Nummer: 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 OFS: 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 OFS: 338-0012

Potterat, J., Salamin, P.A. (2002). Betriebszählung 2001. Stichprobenplan und Schätzverfahren für die provisorischen Ergebnisse. BFS-Nummer: 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 OFS: 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 OFS: 338-0009

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

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

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

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

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

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

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

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

Hulliger, B., Eichenberger, P. (2000). Stichprobenregister für Haushalterhebungen: Umstellung auf Telefonnummern ohne Namen und Adressen, Abläufe für Erstellung und Stichprobenziehung. BFS-Nummer: 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 +41 58 463 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 4500 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 inquiries

Statistical information centre
+41 58 463 60 11, info@bfs.admin.ch

The Swiss structural business statistics (STATENT) is published yearly starting with the reference year 2011. 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. The FTE model was developed on matched data coming from the register and two surveys – the quarterly survey of employment (JobStat) and the quarterly survey of new enterprises (ERST). In this report, we first describe the modifications in the methodology used for the construction of FTE for STATENT following the revision of the JobStat survey in 2015. Then, we describe subsequent changes in the model and in the data following the revision of STATENT in 2021. Finally, we describe a bootstrap method for estimating the variance of total FTE and show its results for STATENT 2020 and 2021.

Online
www.statistics.admin.ch

Print
www.statistics.admin.ch
Federal Statistical Office
CH-2010 Neuchâtel
order@bfs.admin.ch
tel. +41 58 463 60 60

FSO number
338-0088

ISBN
978-3-303-00737-2

The information in this publication contributes to the measurement of the Sustainable Development Goals (SDGs).

The MONET 2030 indicator system
www.statistics.admin.ch → Look for statistics → Sustainable development → The MONET 2030 indicator system