

Correcting breaks in time series in Eurostat's main indicator series: evaluation for the Swiss labour Force Survey

The new recommendations from the Integrated European Social Statistics lead to the revision of the Swiss Labour Force Survey (SLFS) questionnaire from the first quarter of 2021. Therefore, risk of break in series for some of the main SFLS estimates (number of economically active and unemployed persons at the national level, by sex and by age groups) has been assessed¹.

Methodology

This evaluation is based on modeling the probability of being economically active, respectively unemployed, before the revision of the questionnaire, as a function of socio-demographic variables contained in the sampling frame, as well as auxiliary information from Old Age and Survivors's Insurance (OASI) data, or from labor market placement agencies and unemployment funds.

Model parameters "before change" are estimated based on the pooling of the 2015-2020 quarterly data. They are then applied to obtain quarterly predictions, including those of 2021. The predictions obtained for the years 2015-2020 allow for an assessment of the quality of the model by comparing them to the survey estimates, while the predictions obtained for 2021 allow for a decision regarding the appropriateness or not of applying series break factors.

As for the model's performance, the C-statistic (based on the rate of concordance/discordance) is of 82.4% for unemployment, highlighting thus a good explanatory power. However, comparisons between estimates and predictions for the quarters 2015-2020 have shown that a bias risk cannot be excluded for some of the predictions. This element must be taken into account when assessing the relevance of applying adjustment factors.

Results

The series of predicted values (PRED) for the first quarters of the years 2015 to 2021 for the number of unemployed shows more regularity than the situation described by the SLFS estimates. In 2021, the coefficient of variation of the prediction is 0.9% at national level, which corresponds to a confidence interval (95%) of +/- 4800 persons. Broadly speaking, the confidence intervals of the estimates and those of the predictions are overlapping, as shown by figure 1 for the first trimesters at national level, so that no break in series could be detected based on the applied method. The results are similar for the number of active persons.

Concluding remarks

Comparative analyses of the results before (until the 4th quarter of 2020) and after questionnaire's revision (from the 1st quarter of 2021) have detected no break in series. Evolutions are consistent with the economic and social situation at the time, i.e. a very high level of unemployment at the time of the introduction of the new questionnaire, largely due to the health situation (covid-19). The major change of context due to the health crisis should be taken into account when interpreting the results.

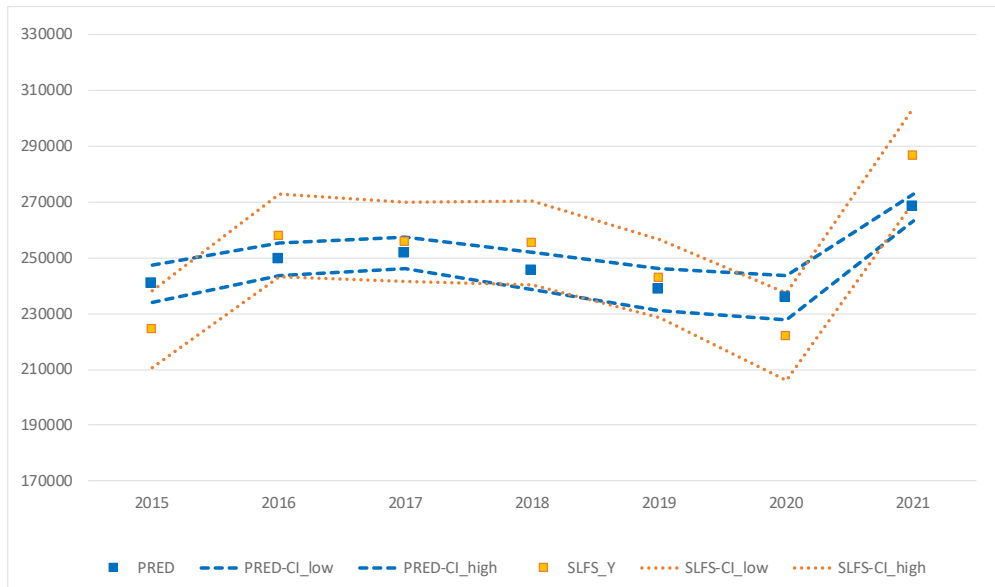
To conclude, taking into account the error margins of the estimates and predictions, in terms of variance but also sometimes of bias for the predictions, the developed method only allows the detection of breaks of great magnitude, which was not the case.

We therefore propose not to apply retrospective corrective factors.

If this method were to be used again in the future, a re-evaluation of the auxiliary information (introduction of new variables, new sources) should be carried out; this could enhance the model performance resulting in break detection of lower intensity.

¹ « Facteurs de ruptures de série dans l'ESPA: méthode et résultats », working paper, SFO, 2022

Figure 1: Number of unemployed persons and confidence interval, model-based predicted values and SLFS-values, 2015-2021



The depicted confidence intervals take only into account the variance, but not a possible bias.