

---

## **Application of the MissForest algorithm for imputing income variables in the Survey on Income and Living Conditions**

Blandine Bianchi & Daniel Kilchmann (Swiss Federal Statistical Office, Switzerland)

[blandise.bianchi@bfs.admin.ch](mailto:blandise.bianchi@bfs.admin.ch)

### *Abstract*

The Survey on Income and Living Conditions (SILC) is a yearly household survey. The household and its members, based on a random sample, are yearly interviewed and are followed for four years for longitudinal analysis. As income variables belong to the main survey objective variables, an appropriate processing is crucial to improve the quality and to increase the reliability of published results derived from the distribution of incomes. The MissForest algorithm, a non-parametric method based on random forests for imputing missing values that allows the use of mixed data type (categorical and quantitative), has been tested on the SILC 2020 income variables. We first imputed individual income variables, followed by household income variables. Socio-demographic variables and household characteristics were used as auxiliary variables for imputing individual income variables. The imputed individual income variables were then added to auxiliary variables on the household level for the imputation of household variables. We ran simulations to evaluate the imputations obtained with MissForest. Studying the non-response model, we simulated partial non-response on survey respondents. Therefore, we were able to assess the quality of the imputation in terms of the estimation of accuracy and imputation error based on the comparison of the imputed values with the answered values. Beside these results, we will also show the imputation impact on the results and discuss the next steps of this highly promising imputation approach.