

ABSTRACT

Participant:	Eliette Castelain
Contact e-mail address:	eliette.castelain@insee.fr
Organisation:	INSEE
Subject:	INCGROSS
Title:	INCGROSS methodology in French LFS

Abstract:

INCGROSS is a variable requested by Eurostat for the Labour Force Survey (LFS) data submission. The aim is to provide information on the gross monthly pay including other amounts of payments of a higher periodicity (for example, exceptional payments). Imputation for this variable is compulsory when item non-response before imputation is over 5%, which is the case in the French LFS.

Since 2021, in France, we collect the net amount because respondents are more used to this concept. Before 2021, they could answer either by the net or the gross amount and they mainly answered by the net amount. Respondents can give their monthly net income either by clear amount or by range. They can also refuse to answer or not know their income. In 2021, 75% of the average monthly net incomes collected correspond to clear amounts, 17% to ranges and 8% to missing. In order to deliver INCGROSS variable to Eurostat, the variable collected by the survey needs imputation and conversion into gross amount. The imputation and conversion methods require a quarterly processing and to be reproducible.

Non-response and response by income range frequencies vary with the employment status. In particular, salaried business leaders indicate their income much less often, in the same way as temporary workers and fixed-term contracts employees to a lesser extent.

In line with the 2021 redesign, we took the opportunity to review the imputation method, in order to simplify it and to reduce the processing time. In 2021, we implemented two types of imputation for the INCGROSS variable: a multiple linear regression model for non-response and a hot-deck imputation for response by income range.

To impute non-response, we estimate two multiple linear regressions: one for the executives and another for the rest of the employees. The choice of explanatory variables relies on the imputation model used in the French SILC by adapting it and enriching it. The explanatory variables retained are sex, age, socio-professional category, level of education, status, two indicators of residency (agglomeration of Paris or not and Metropolitan France or not), seniority, number of hours usually worked, sector or economic activity, number of employees in the company, management tasks indicator. In addition, we also impute with this method negatives incomes and outliers: it represents 0.08% of income values in 2021.

The distribution of net monthly income imputed by regression extends towards the bottom of the distribution. According to the model, non-respondents have a slightly lower income than those who indicate a clear amount.

To impute a clear amount using income range, we apply a random Hot-Deck method by groups with replacement and unweighted. Each group is composed of donors whose income range and socio-professional category is the same as the recipients. It allows to obtain a distribution of imputed amounts within each range similar to the one observed among respondents by clear amount. In particular, there is no smoothing of the distribution. As accumulation points can mark the distribution, the Hot-Deck method leads to reproduce them for the missing values. If there is no donor in the income range and the socio-professional category of the recipient, the median value of the range is imputed. In 2021, it concerns 0.05% of income values.

Because respondents who give a range amount are more often at the top of the income distribution, the distribution imputed by Hot-Deck extends more upwards than the final distribution of the income variable. In the end, on average, observed and imputed distributions of net incomes are close.

In France, social contributions depends on the status of the employee (civil servants, executives, apprentices, etc.). Maintaining a contribution rate for each type of social contributions is difficult over time. In order to convert net income into gross income, we apply ratios ($\text{Ratio} = \text{Net income} / \text{Gross income}$) estimated by a micro-simulation model (INES) carried out within the institute. This model's first aim is to simulate the impact of fiscal and social policies. Every year, the INES' team estimates five different ratios for five groups of employees with a delay of one year (it appears that the ratios are stable over time). The distribution of ratios by quintiles of income is stable within the groups justifying the use of a single ratio per group.

To conclude, the chosen method of imputation and conversion from net to gross amounts meets production constraints with no long calculation times and submits a complete INCGROSS variable closest to the collected data in the French LFS. However, French LFS is not the reference source on income for data exploitation. Exhaustive administrative data on income exist, in particular from the nominative social declarations. In the long term, we plan to initiate some works on the pairing of French LFS with administrative sources in order to recover more reliable information on income data.