

STATEC



Conference Paper

INCGROSS variable in Luxembourgish LFS

Hans Neumayr

17th Workshop on Labour Force Survey Methodology

Neuchâtel, 25–26 April 2024

Abstract

The revised methodology of the Labour Force Survey (LFS), implemented in 2021, mandates the annual inclusion of the INCGROSS variable. This paper will explore the revised production process of the income variable within the Luxembourgish LFS.

It will specifically focus on the new imputation method introduced in 2023 to address non-responses and compare its impact with that of the outlier treatment.

Introduction

Until 2021, only net income was collected through the LFS questionnaire but reported in deciles only. Starting from 2021, the practice of collecting net income data was discontinued and gross income was asked directly in the questionnaire. In order to comply with the regulation, in particular to the fact that imputation for this variable is compulsory when item non-response is over 5%, a standardized imputation procedure was introduced in 2023, with a general review of the production process including a more detailed analysis of the outliers.

Sources and chosen methods to estimate INCGROSS:

As outlined in the Eurostat-table below, Luxembourg, along with Denmark, Norway, and the Netherlands, was one of the few countries that collected gross income without performing any imputation in 2021. However, with a non-response rate approaching 10%, Luxembourg was no longer in compliance with regulations. Consequently, we had to adopt an additional method to estimate INCGROSS and ensure compliance with the regulation.

	11	12	13	21	22	25
Grand Total	41.7%	2.6%	10.7%	11.4%	12.2%	21.4%
AT	92.1%		7.9%			
BE	96.0%		4.0%			
BG	29.6%	31.1%	3.2%	5.9%	30.2%	
CH	62.7%	1.0%		36.0%	0.3%	
CY	66.6%			33.4%		
CZ		100.0%				
DE						100.0%
DK	100.0%					
EE	88.8%		9.8%	1.4%		
EL			0.0%	54.2%	45.8%	
ES			100.0%			
FI	81.7%		18.3%			
FR				75.4%	24.6%	
IE	70.8%		29.2%			
LU	100.0%					
LV	17.9%	10.3%	71.8%			
MT		100.0%				
NL	100.0%					
NO	100.0%					
PL				100.0%		
PT				71.7%	28.3%	
RO					100.0%	
SE	99.3%		0.7%			
SP			100.0%			
SK						100.0%

Source: EUROSTAT

Why did we opt solely for imputation? In Luxembourg, the Labor Force Survey (LFS) is not the primary reference source for income, resulting in limited resources to estimate INCGROSS. There is comprehensive administrative income data available from social security, which includes cross-border workers who constitute half of the workforce (but are not part of the LFS sample). However, at present, data linking to these administrative sources is too labor-intensive for the intended purpose. Additionally, it was crucial to develop a functional solution in time for the 2023 data collection. Consequently, we opted for a “minimalistic approach” to implement imputation for item non-response within the Labour Force Survey.

Item non response behavior for INCGROSS

Overall total non-response is about 9%. However the rate differs for certain subgroups such as age and citizenship as shown in the table.

age	NR in %
[15,28]	12.71
(28,42]	5.97
(42,55]	9.25
(55,69]	12.18
(69,82]	17.65

CITIZENSHIP	NR in %
BE	6.44
DE	7.64
EU	5.83
FR	5.89
LU	11.08
non-EU	5.71
PT	6.07

Modelling of the Imputation Process

In a first step all variables were imputed using single imputation and the default methodological settings of the mice package in R.

Within this subset, the Random Forest method was utilized as an automatic variable selection technique to identify the 10 most important predictor variables of missing values in INCGROSS for the imputation model:

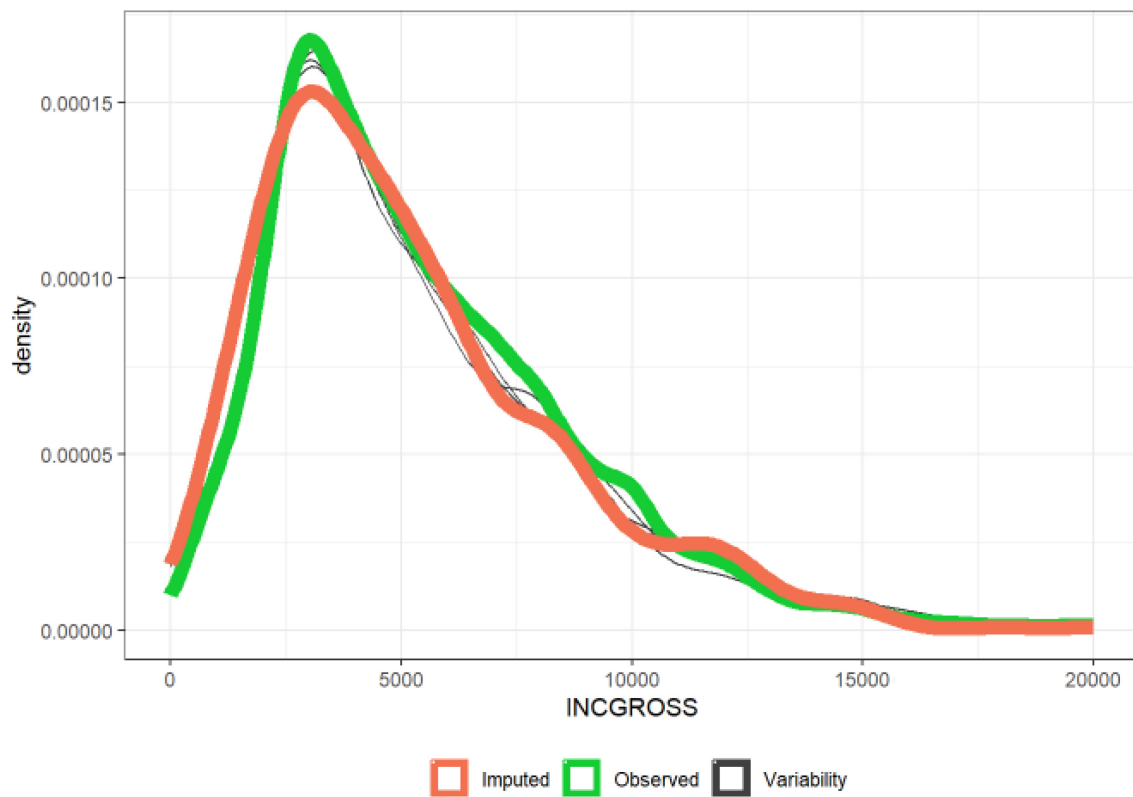
FTPT, ISCO4D, YEARBIR, NACE3D, MIGREAS, HATLEVEL, HATYEAR, TEMP, CITIZENSHIP, SEX

Methodology:

This involves determining the appropriate techniques for handling missing data in each variable type: Default imputation methods of the mice package in R are being used. This includes Predictive Mean Matching (pmm) for numeric variables, Polytomous Regression Imputation (polyreg) for variables with more than two categories, and Logistic Regression Imputation (logreg) for binary variables.

Variable	Method	Specification
FTPT	polyreg	Default
ISCO4D	polyreg	Default
YEARBIR	polyreg	Default
NACE3D	polyreg	Default
MIGREAS	polyreg	Default
HATLEVEL	polyreg	Default
HATYEAR	polyreg	Default
TEMP	polyreg	Default
CITIZENSHIP	polyreg	Default
SEX	logreg	Default
INCGROSS	pmm	Default

Result and Analysis of the Imputation Process:



As expected with “only” about 10% missing values, the impact of the imputation is minor on the distribution of INCGROSS.

before imputation:				after imputation:			
income				INCGROSS			
Percentiles	Smallest			Percentiles	Smallest		
1%	600	250		1%	580	250	
5%	1490	250		5%	1421	250	
10%	2100	256	Obs 7,378	10%	2100	256	Obs 7,378
25%	3000	260	Sum of wgt. 7,378	25%	3000	256	Sum of wgt. 7,378
50%	4795	Mean	5559.916	50%	4784.5	Mean	5536.871
	Largest	Std. dev.	3579.755		Largest	Std. dev.	3568.298
75%	7300	35000		75%	7200	35000	
90%	10000	35000	Variance 1.28e+07	90%	10000	35000	Variance 1.27e+07
95%	12000	35000	Skewness 1.973445	95%	12000	35000	Skewness 1.985599
99%	17266	37000	Kurtosis 11.29387	99%	17068	37000	Kurtosis 11.44545

Impact of revised treatment of outliers:

before imputation:				after imputation:				before outlier "treatment":			
income				INCGROSS				INCGROSS			
Percentiles	Smallest			Percentiles	Smallest			Percentiles	Smallest		
1%	600	250		1%	580	250		1%	341	1	
5%	1490	250		5%	1421	250		5%	1,248	1	
10%	2100	256	Obs 7,378	10%	2100	256	Obs 7,378	10%	2,000	1	Obs 7,378
25%	3000	260	Sum of wgt. 7,378	25%	3000	256	Sum of wgt. 7,378	25%	3,000	1	Sum of wgt. 7,378
50%	4795	Mean 5559.916		50%	4784.5	Mean 5536.871		50%	4,700	Mean 6,235.47	
		Largest	Std. dev. 3579.755			Largest	Std. dev. 3568.298			Largest	Std. dev. 20,883.5
75%	7300	35000		75%	7200	35000		75%	7,254	410,364	
90%	10000	35000	Variance 1.28e+07	90%	10000	35000	Variance 1.27e+07	90%	10,000	434,235	Variance 4.36e+08
95%	12000	35000	Skewness 1.973445	95%	12000	35000	Skewness 1.985599	95%	12,000	1000000	Skewness 43.11367
99%	17266	37000	Kurtosis 11.29387	99%	17068	37000	Kurtosis 11.44545	99%	20,000	1200000	Kurtosis 2,197.37