Annex 6 - Estimation and Imputation

2024-07-15

Imputation procedure used

With effect from 2008, households are no longer asked about their income in SILC survey. These administrative data are obtained on the basis of the tax files held by the Directorate General for Public Finance (DGFiP) or management bodies for social income benefits (CNAF, CNAV, CSMSA). The matching of the survey file and the tax file takes place using the address of the household and information on the declarant (surname, first name, date of birth, department of birth, country of birth). If identifying data do not match, the income will not be found. The family allowances and benefits and minimum social security benefits are obtained from the organisations holding them. It is always the address of the household that is used for the search, together with the demographic characteristics of the recipient (gender, date of birth). For various reasons, some addresses do not match. The associated incomes are therefore imputed.

PY010 - Wage or similar income

A wage was imputed to 448 unmatched survey respondents who declared that they are in receipt thereof (2.7%). The fact of being in receipt and the amount are imputed simultaneously using a Tobit2 econometric model. The amount of the wage is explained in particular by seniority, gender, socio-professional category and activity status. The following components (tax-free wages that are not declared and that cannot be retrieved from administrative data) are added to PY010:

- Tax-free remuneration declared in SILC survey (AESNDIR variable)
- The imputed amounts of the prime de partage de la valeur (exceptional purchasing power bonus). The bonus is imputed to those eligible based on the average, by strata defined by the type of activity undertaken by the employer, the income and the size of the company, based on the share of persons in receipt of a wage in each stratum.
- Exempt overtime hours (same method, with strata defined by profession and socio-economic category, activity and wage bands)

PY050 – Self-employed income

Self-employed income was imputed to 85 unmatched survey respondents who declared that they were in receipt thereof (5.5%). The amounts are imputed using a hot deck procedure with the detailed socio-professional category for self-employed categories.

PY090 – Unemployment benefits

Unemployment benefits were imputed to 104 unmatched survey respondents who declared that they were in receipt thereof (4.2% of respondents in receipt of unemployment benefits). The fact of being in receipt and the amount are imputed simultaneously using a Tobit2 econometric model. The amount of unemployment benefits is explained in particular by the amount of time spent in employment, gender, the highest level of education attained and the duration of receipt.

PY100 et PY110 – Pensions and surviving dependants' pensions

Pensions were imputed to 128 unmatched survey respondents who declared that they are in receipt thereof (1.2% of respondents in receipt of a pension). The fact of being in receipt and the amount are imputed

simultaneously using a Tobit2 econometric model. The amount of the pension is explained in particular by the amount of time spent in employment, gender, the highest level of education attained and the types of pensions received. No distinction is made between pensions and surviving dependants' pensions in the tax return. The distribution between the aggregates PY100 and PY110 takes place based on the declaration of receipt of these two types of income by the respondents in the questionnaire.

The minimum old-age pension is also paid into aggregate PY100. This benefit may be paid to either a single person or a couple. The minimum old-age pension is also paid into aggregate PY100. This benefit may be paid to either a single person or a couple. 59 (19.4 %) households or individuals not matched to social sources but having declared themselves to be in receipt thereof or not in receipt thereof, but presenting similar characteristics to households in receipt of old age pensions. Without forcing such additional receipt, the coverage rate of the minimum old-age pension (the ratio of the weighted survey data to the framing data) was too low. Imputation is performed by means of a hot deck procedure, distinguishing between cases in which the recipient lives alone or with a partner. The stratification variable is the ratio of income to the eligibility threshold, in bands.

PY120 – Health insurance-related benefits

The amount of sickness and parental benefits is not included in the questionnaire, as respondents do not know how to distinguish these from their wage; it is also not possible to isolate these benefits based on tax returns. These benefits are therefore not recorded in PY120, but in PY010. PY120 is comprised of benefits and annuities paid following an occupational accident. A value within the band is derived for individuals who have not indicated a clear amount. For those who have not indicated an amount within a band, the monthly amount of benefits or annuities is imputed using a hot deck procedure stratified by age group. This was the case for 145 individuals (27.4%).

PY130 – Disability benefits

The benefits included in the PY130 aggregate are: - Disabled Adult Allowance (Allocation Adulte Handicapé – AAH) (Amount retrieved from administrative data) - Military disability or war victim's pension (Amount declared in the survey) - Disability pension (Amount retrieved from administrative data and declared in the survey) - Personal autonomy benefits (Aide personnalisée à l'autonomie – APA) (Amount declared in the survey) - Specific dependency benefit (Prestation spécifique dépendance – PSD) (Amount declared in the survey) - Compensatory disability benefit (Prestation de compensation du handicap – PCH) or third party assistance compensation (Allocation compensatrice pour tierce personne – ACTP) (Amount declared in the survey)

Income was imputed to 176 unmatched survey respondents who declared that they are in receipt thereof(33.3%). Imputation took place using a hot deck strategy stratified by age group.

HY040 – Property income

Property income was imputed to 14 households, i.e. to 0.5% of households that received property income. - We start by imputing to those that entered minimum and maximum thresholds for the value of real estate assets (VIMMOMIN and VIMMOMAX variables) a value within that interval, using an econometric model that makes use of indicators for the type of real estate property owned in the form of explanatory variables. - Once the value of the property has been imputed, the property income is imputed by means of a hot deck procedure. For recipients with imputed property value a hot deck is created and the stratum is based on the property value deciles. For recipients who do not have an imputed property value a non-stratified hot deck procedure is used, limiting the donors to matched households with property income of between 1,000 and 10,000 euros that have not declared the value of their real estate assets.

HY090 – Financial income

- We start by imputing the holding of different financial products when the household has not responded to the questions concerning holdings. Using a hot deck strategy stratified by income decile and age

group, we simultaneously impute the receipt of the 7 financial products. Where necessary, we correct the holding where it has been left blank, but financial income is found in the tax return.

- We impute the amounts of financial products held, by product type, by taking a value from within the amount band indicated.
- The amounts are then calibrated on the basis of the amounts derived from Banque de France data and the national accounts for 4 of the 7 financial products (tax-free savings accounts, taxable bank accounts, homebuyer savings plans, life insurance income).
- Finally, based on the imputed property value, we estimate the financial income generated by that property by applying the average rates of return for the year under review, by product type.

HY080 et HY130 – Transfers paid and received between households

Transfers received: 41 amounts to be imputed (4.4%)

Transfers paid: 105 amounts to be imputed (4.7%)

In both cases, the amounts are imputed using a hot deck strategy, stratified by benefit type (maintenance payments, rental assistance, other) and by 8 socio-professional categories for the household concerned.

HY070 – Housing benefits

Housing benefits are imputed to households not matched to social sources and declaring that they are in receipt thereof. Imputation is also performed for households not declaring that they are in receipt of housing benefits but that have similar characteristics to households in receipt thereof. The amount of the benefit is imputed using a hot deck strategy, stratified by the number of children and the primary income by consumption unit.

HY050 – Family allowances and benefits

Family allowances and benefits are imputed one by one using a hot deck procedure. The stratification variables differ from one allowance or benefit to the other, since they depend on the conditions under which each allowance or benefit is awarded. We imputed 336 households (8.5 %).

Allowances and benefits	Stratification variable(s) for imputation by hot deck
Family allowances and benefits	number of children (0-1, 2, 3, 4+) x household income > EUR 75,000 indicator
Family supplement	Additional family household income > EUR 30,000 indicator
Back-to-school allowance	number of children (0, 1, 2, 3, 4+)
Education allowance for disabled children (Allocation d'éducation de l'enfant handicapé – AEEH)	No stratification variable
Daily Parental Attendance Allowance	No imputation
Child benefit (Prestation d'accueil du jeune enfant – PAJE)	indicator of the presence of an adult in the household x (household income > EUR 50,000 indicator)
Shared Child Education Benefit (Prestation partagée d'éducation de l'enfant – PreParE)	No imputation

HY063 – Minimum social security benefits

An active solidarity income (Revenu de solidarité active – RSA) and/or in-work benefit amount was imputed to 601 households (8.5%). The imputation was performed using a hot deck stratified by household type. The HY063 aggregate also includes social welfare not included in the administrative data retrieved (MTAIDAUT variable), the young person's guarantee (MGARJEUNE variable), inflation allowance and a specific allowance given to some households in 2022 (exceptional solidarity support - aide exceptionnelle de solidarité).

HY120 – Taxes on wealth

The aggregate is comprised of property tax and wealth tax. The tax on real estate assets (Impôt sur la Fortune Immobilière – IFI) is calculated on the basis of a scale based on the amounts of real estate assets declared in the tax return. A question concerning the amount of property tax is included in the questionnaire (TAXFON variable). The amount is imputed for households that have not been able to answer this question using a hot deck stratified by the income band declared.

HY140 – Income tax and social security contributions

The aggregate is comprised of income tax, residence tax, the Generalised Social Contribution (Contribution sociale généralisée – CSG) and the Social Debt Repayment Contribution (Contribution pour le remboursement de la dette sociale – CRDS). - When the income tax amount is missing. The rate of taxation is imputed using a hot deck procedure stratified by taxable income deciles. The imputed tax amount is then calculated as the product of the rate of taxation and the taxable income declared. - For residence tax: This tax is being gradually phased out in France; in 2023, no households will pay residence tax on their primary residence. Residence tax has been permanently removed for almost 80% of tax households in 2022. In order to perform the imputation, it must first be determined whether or not the household is eligible for the suppression of the residence tax; the amounts are then drawn using a hot deck procedure, stratified by French department. - The CSG and CRDS are calculated on an aggregate-by-aggregate basis (wages, self-employed income, pensions, unemployment, property and financial income, benefits). For employment income, average contribution rates are used for each income decile and each type of income, calculated using the INES microsimulation model.

PY021 – Company car

Respondents are asked what benefits in kind they received (AVANAT variable) and then for an estimate of the amount represented by the company car benefit (MS1AVNAT). However, a large proportion of respondents do not know how to quantify this benefit. We therefore imputed this income for 265 respondents using the declared median amount.

HY030 – Imputed rent

This variable is calculated every 3 years. It is transmited in 2023. A fictitious rent is imputed to owners, homebuyers, beneficial owners, those provided with housing free of charge and social housing tenants paying an amount of rent that is lower than the market price (14 519 households in 2023). A "kit", developed within INSEE by the team responsible for the housing survey, is used to estimate this variable. This kit: - Performs matching of statistics on income and living conditions with administrative auxiliary information about the dwelling and its area. - Imports equations estimated in the housing survey in order to impute a fictitious rent (8 equations: apartment or house x occupancy status) – 20 explanatory variables are used that relate to housing (surface area, date of construction), its environment (IRIS, size of the urban unit, climatic zone, etc.) and its occupants (age, education, employment, household type, etc.). - Supplements the adjusted value with a residual drawn using a stratified hot deck procedure.

When auxiliary information about the dwelling is not available there is no imputation (811 households in 2023).

HY100 – Interest repayments on mortgage

For the three main loans, the questionnaire collects the start and end dates of the loan, the amount initially borrowed and the interest rate. This information is used to calculate, for each loan, firstly the amount of the monthly repayment, then the monthly interest charge and finally the annual interest repayments on mortgage.

HH070 – Total housing cost

The total cost of the home is equal to the sum: - For owners specifically, co-ownership charges and monthly interest repayments on mortgage - For tenants specifically, rent and rental charges - For everyone, water,

electricity, gas, heating, home insurance and household waste collection charges.

Most of these amounts are collected in the questionnaire. However, some amounts are not well known by respondents: this is the case for water, home insurance and household waste collection charges. If respondents do not answer these questions, these amounts are imputed, according to: - for water costs, the number of people living in the home, - for home insurance, the decile of the household's standard of living, - for household waste collection tax, the département to which the surveyed dwelling belongs.

Type of income	number of perceptive observations	number of imputed observations	percent
PY010 Net salary received	16643	448	2.7
PY020 Benefits in kind (incl. company car)	4277	1061	24.8
PY050 Net profit or loss from self-employed activity	85	1536	5.5
PY090 Net unemployment benefits	2505	104	4.2
PY100 Pensions (including old-age pension, excluding reversion)	10634	128	1.2
PY130 Disability pensions	529	176	33.3
PY140 Scholarships	435	0	0.0
HY040 Net income from rental of property or land	2852	14	0.5
HY050 Adjusted net family/children benefits	3931	336	8.5
HY060 Net social exclusion benefits	11349	601	5.3
HY070 Net housing benefit	2842	264	9.3
HY080 Net transfers received from other households (rent, alimony)	940	41	4.4
HY090 Financial income	15389	13197	85.8
HY110 Net income from persons under 16 years of age	293	0	0.0
HY120 Net wealth tax (IFI and property tax)	11220	11220	100.0
HY130 Net transfers to other households (rent, alimony)	2223	105	4.7
HY140 Income tax and social security contributions	17010	7264	42.7