CENTRAL STATISTICAL BUREAU OF LATVIA
EUROPEAN COMMISSION
EUROSTAT GRANT

Project No — 101195467 — 2024-LV-BOP

Project name: ESA2010 and BPM6 implementation, quality improvements, and execution of the GNI control cycle

WP1, deliverable D1.2 (D2)
Final report

Submitted via CIRCABC on 30 June 2026

June 2026

Table of Contents

ABBREVIATIONS

COFOG	Classification of the Functions of Government
COICOP	Classification of Individual Consumption According to Purpose
CSB	Central Statistical Bureau of Latvia
EKK	Economic Classification Code
EU-SILC	European Union Statistics on Income and Living Conditions
ESA 2010	European System of Accounts 2010
FISIM	Financial Intermediation Services Indirectly Measured
GMI	Guaranteed Minimum Income
HBS	Household Budget Survey
HDA	Household Distributional Accounts
HFCS	Household Finance and Consumption Survey
ISCED	International Standard Classification of Education
ISCO	International Standard Classification of Occupations
NACE Rev.2	Statistical Classification of Economic Activities (Revision 2)
NA	National Accounts
OECD	Organisation for Economic Co-operation and Development
SPKC	Centre for Disease Prevention and Control
SRS	State Revenue Service
STiK	Social Transfers in Kind
UAA	Utilised Agricultural Area

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The methodology for the compilation of Household Distributional Accounts (HDA) was developed in accordance with the recommendations set out in the OECD Handbook on Distributional National Accounts and the related guidance on the distribution of household income, consumption and saving.

The objective was to produce distributional information fully consistent with National Accounts aggregates while reflecting the characteristics of individual households observed in microdata sources.

The development of the methodology involved several key steps:

- First, the latest HDA reporting templates and methodological guidance were reviewed to ensure compliance with international recommendations and reporting requirements.
- A detailed assessment of National Accounts income and consumption variables was then carried out to identify components that include institutional households, which are not covered by household surveys. Where relevant, additional data sources were identified to enable the separate treatment and allocation of institutional households.
- For household consumption expenditure, adjustments were made to ensure consistency with the household sector concept used in HDA. This included separating the consumption expenditure of non-resident households on the domestic territory and incorporating the expenditure of resident households abroad, resulting in adjusted National Accounts aggregates for detailed consumption components.
- For each National Accounts income and consumption item, corresponding microdata sources were identified. Where no direct microdata source was available, additional administrative information, auxiliary datasets, or estimation methods were used to support the allocation process.
- To combine information from different sources, statistical matching techniques were applied, linking household-level information from several datasets into a unified analytical framework.
- Appropriate imputation methods were then selected for items not directly observed in the available microdata.
- Following the imputation stage, microdata was aligned with National Accounts totals. Any differences between macroeconomic aggregates and microdata estimates were allocated to households or household groups using the allocation methods recommended in the HDA framework. The choice of method depended on the characteristics of the specific income or consumption component and the availability of supporting information.
- Households were subsequently grouped according to equivalised disposable income, allowing the production of distributional results for different income groups.
- Finally, data reliability checks and sensitivity analyses were carried out to assess the robustness of the estimates and the impact of methodological assumptions on the final results.

The methodological report is supplemented by the following documents, which are provided separately:

• D1.3 – Excel template with quantitative HDA results for the 2019 reference year, based on the OECD HDA template.

• D1.4 – Microsoft PowerPoint presentation delivered at the Eurostat Expert Group on Sector Accounts meeting held on 2–3 June 2026.

METHODOLOGICAL DESCRIPTION OF INSTITUTIONAL HOUSEHOLD DATA FOR INCOME

The methodology is based on **2019 data** and are applied to all subsequent years. Due to the limited availability of data sources for institutional households, an item-specific approach was chosen to assess their impact on various income components. The largest impact on both gross disposable income and adjusted gross disposable income comes from wages and salaries, social benefits, and social transfers in kind.

Calculations are not performed for the following indicators for institutional households: **B2g, B3g, D4, D71, and D72**. It is highly unlikely that data for residents in institutions are available for these specific indicators.

Calculations are performed for the following indicators: D11, D121 (D611), D122 (D612), D613, D51, D62, D63, and D75.

WAGES AND SALARIES (D11)

Information on the number of employed individuals residing in institutional households is obtained from census data, which in the case of Latvia is based on administrative data sources. The data is categorized by the type of institutional household and the NACE Rev. 2 classification group they worked in as of January 1, 2020. A condition for inclusion is that they have resided in the specific institution for at least 12 months. The CSB census data is available as of January 1st of each year.

The chart below shows the number of employed people living in institutional households in 2019 (Figure 1).

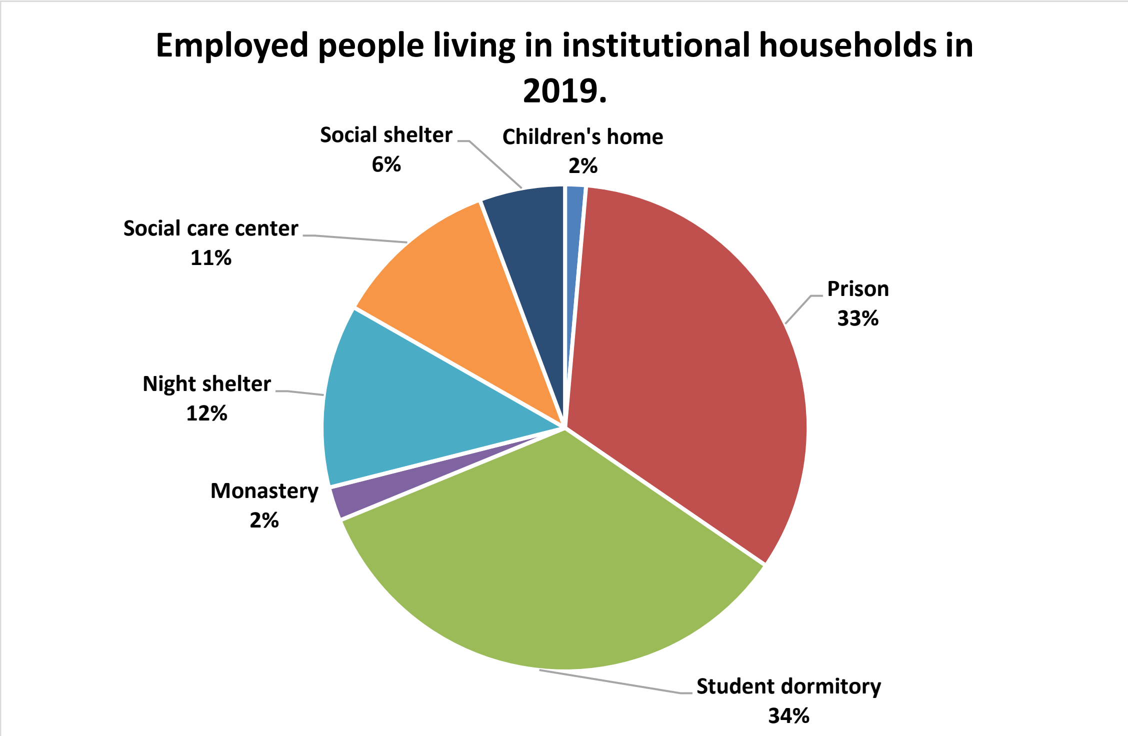

Figure 1 Employed people living in institutional households in 2019.

Source: CSB census data *2019*

The chart shows that the largest share of employed residents living in institutions are in prisons and student dormitories (67%), followed by those in night shelters and social care centers.

Data from the Prison Administration's annual public report is used to determine the number of inmates employed for payment in prison maintenance and commercial workplaces in 2019. The report also provides the average monthly salary for a single convicted person (after taxes and other deductions). The gross wage for prisoners is calculated using an in-house "net-to-gross" wage calculator. This public report is available online at:

https://www.ievp.gov.lv/lv/gada-publiskie-parskati?utm_source=https%3A%2F%2Fwww.google.com%2F.

Due to the lack of precise income data, we are forced to use assumptions. These assumptions are based on the socio-economic situation of these population groups. It's assumed that residents of night shelters, who often have unstable employment, receive the minimum wage. Conversely, residents of children's homes, social care centres, and social shelters, whose income is very limited and often depends on state support, are assumed to receive half of the minimum wage.

To calculate the wages and salaries (D11) for residents of monasteries, we use **census data** containing information on the number of employed residents and their sector of employment. The wages are calculated based on the average gross wage in the relevant sector, which provides a more accurate income estimate than general assumptions. This data is available on the CSB website at: https://data.stat.gov.lv/pxweb/lv/OSP_PUB/START__EMP__DS__DSV/DSV030/.

Most residents in student dormitories are students. To determine their average weekly working hours, we used data from the "Social and Economic Conditions of Students in Latvia 2022" study, available at:

https://www.izm.gov.lv/lv/jaunums/eurostudent-aptauja-sniedz-ieskatu-par-studentu-dzives-apstakliem-latvija.

This study was part of the eighth **Eurostudent** survey, which involved nearly 30 European countries. The analysis of the situation in Latvia is based on the online survey responses of 2,646 students from 52 state and private higher education institutions.

EU-SILC survey experts assume that 80-90% of all students living with their parents have a common budget, so their data are attributed to the parental household.

Students' wages are calculated based on the average gross wage in the sector where the student works, with the following adjustments:

- Based on information from the "Social and Economic Conditions of Students in Latvia 2022" study, students work an average of 28.9 hours per week (2017: 31.3 hours). An average coefficient of 0.75 ((28.9+31.3)/2) is applied.
- Considering that 80-90% of student wages are attributed to the parental household, it is assumed that only 15% of working student wages are relevant to dormitories, and a coefficient of 0.15 is applied.

PERSONAL INCOME TAX (D51)

Personal Income Tax (PIT) is calculated in parallel with the wage calculations for the institutions. The calculation formula is:

$\text{PIT} = (\text{Gross Wage} - \text{Employee SSC} - \text{Non} - \text{Taxable Minimum}) * \text{PIT Rate} * \text{Number of Employees in the Institution}$

EMPLOYERS' ACTUAL SOCIAL CONTRIBUTIONS (D121, D611)

Employers' actual social contributions are calculated in parallel with the wage calculations for the institutions. The calculation formula is:

$\text{D121} = \text{Gross Wage} * \text{Employer SSC Rate}$

EMPLOYERS' IMPUTED SOCIAL CONTRIBUTIONS (D122, D612)

To calculate employers' imputed social contributions for employed residents in institutions, a coefficient is first determined by dividing the total employers' imputed social contributions in Latvia by the total gross wages in Latvia. Data from the Central Statistical Bureau (CSB) are used for this calculation. This coefficient shows what percentage of the average national wage is made up of employers' imputed social contributions. This coefficient is then multiplied by the institutional gross wages to obtain the approximate employers' imputed social contributions for the specific institution.

HOUSEHOLDS' ACTUAL SOCIAL CONTRIBUTIONS (D613)

Households' actual social contributions are calculated in parallel with the wage calculations for the institutions. The calculation formula is:

$$D613 = \text{Gross Wage} * \text{Employee SSC Rate}$$

SOCIAL BENEFITS (D62)

To calculate social benefits for people living in institutions, it's essential to first identify which residents are most likely to receive them. Figure 2 shows Institutional residents by type of instituions, 2019.

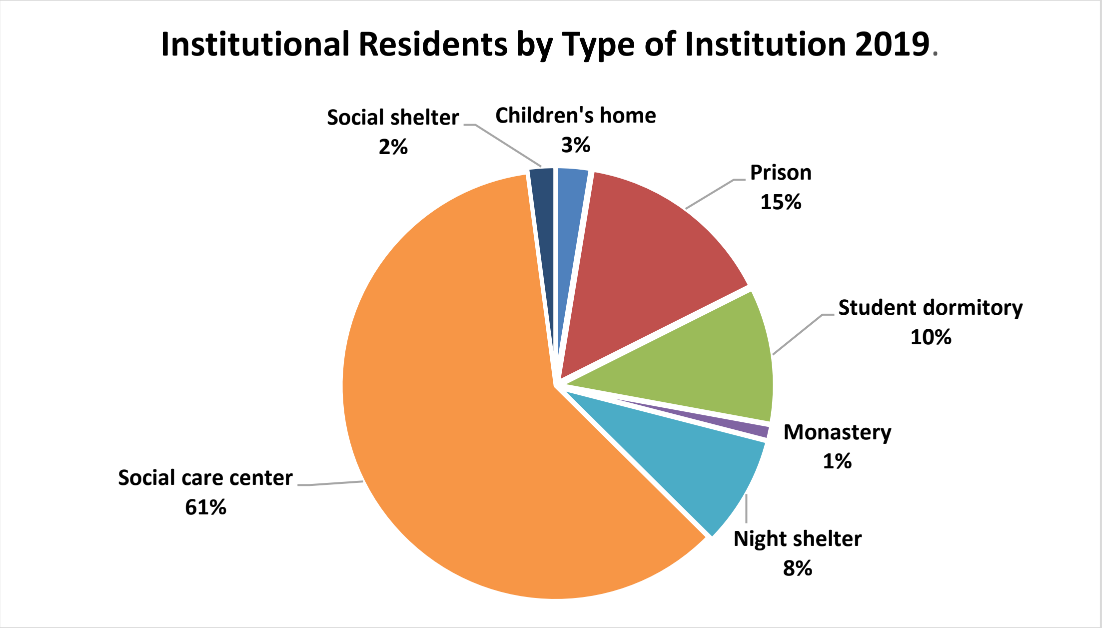

Figure 2 Institutional Residents by Type of Institution 2019

Source: CSB census data 2019

Social benefits are most likely to be received by residents in the following institutions:

- Social care centres
- Children's homes
- Social shelters
- Night shelters

Prisoners typically do not receive social benefits, as their upkeep is provided by the state. Students in dormitories generally don't require social benefits. Residents of monasteries usually rely on community support and rarely receive state social benefits.

According to census data, the majority of social benefit recipients (61%) live in social care centres.

To calculate the benefits received by residents of long-term social care institutions and children's homes, we used information from the Ministry of Welfare (MoW) report titled "Report on the Provision of Long-Term Social Care and Social Rehabilitation Services" (available at: https://www.lm.gov.lv/lv/gada-dati). This report provides data on the number of people in long-term care, categorized by age group and type of benefit received.

Additional information on the average amount of state social benefits and pensions is obtained from the Central Statistical Bureau (CSB) database. The MoW website also provides data on the amount of social security benefits, including those for people with disabilities.

In accordance with **Section 29 of Chapter IV of the Social Services and Social Assistance Law**, individuals living in long-term care institutions are entitled to receive at least 15% of their due pension or benefit. Similarly, children aged seven and older are entitled to receive a cash amount equivalent to 15% of the state social security benefit. Therefore, to calculate the amount of social benefits or pensions for residents in these institutions, a **coefficient of 0.15** is applied.

In 2019, the total amount received by residents in these institutions was **3,215 thousand EUR**.

Individuals in night shelters (8%) and social shelters (2%) may receive social assistance, such as the Guaranteed Minimum Income (GMI) benefit. According to data from the Ministry of Welfare, the average GMI paid per person in Latvia in 2019 was **215.12 EUR per year** (available at: https://www.lm.gov.lv/lv/gada-dati). Since 81% of night shelter residents are unemployed, they are likely to receive GMI from the state. To calculate the total GMI benefit paid to shelter residents per year, we multiply the average GMI benefit in Latvia by the number of unemployed individuals in shelters:

$$215.12 * 1,853 = 399 \text{ thousand EUR}$$

In total, residents of institutions received **3,614 thousand EUR** in social benefits in cash:

$$3,215 + 399 = 3,614 \text{ thousand EUR}$$

SOCIAL TRANSFERS IN KIND (D63)

Social transfers in kind are in-kind expenditures by the government (S13) and non-profit institutions serving households (S15) received by households (S14). The calculation is based on available expenditure data.

The main breakdown of expenditures by **T24 Questionnaire on GDP Expenditure Weights** in 2019:

- Education (40%)
- Health (33%)
- Recreation and culture (10%)
- Housing (4%)
- Social protection (7%)
- Other (7%)

HEALTH AND EDUCATION

Data on the number of people in long-term care by age group, covering both social care institutions and children's homes, is obtained from the Ministry of Welfare's report, "Report on the Provision of Long-Term Social Care and Social Rehabilitation Services" (available at: https://www.lm.gov.lv/lv/gada-dati). The total expenditures on health by **S13** and **S15** were **1,007,087 thousand EUR**, obtained from the **Eurostat T24 "Questionnaire on GDP Expenditure Weights"**.

Average health expenditures per capita by 5-year age intervals are obtained from the National Health Service (NHS) data. Based on the NHS age-group expenditure structure, the total health expenditures (1,007,087 thousand EUR) are distributed, and the average per capita expenditure for each age group is calculated. To calculate health expenditures for residents in social care centers and children's homes, the average per capita expenditure is calculated for the age group breakdown provided in the Ministry of Welfare report. The health expenditure for residents in institutions for each age group is calculated using the following formula:

$$D63 = \text{Expenditure per capita} * \text{Number of residents}$$

Data on the number of pre-school and school-age children and their corresponding education levels are obtained from the previously mentioned data source. State expenditures on education-by-education level (**ISCED**) are obtained from the CSB statistical portal. This data is used to structure the education level breakdown. To determine the portion of education expenditures related to unclassified education, data on government expenditures by functional categories (**COFOG**) from the CSB statistical portal are used. Unclassified education (adult education programs) accounts for 11% of total education expenditures according to COFOG. State social transfers in kind for education in 2019 amounted to **1,182,408 thousand EUR**.

Expenditures for both unclassified and classified education are calculated as follows:

$$1{,}182{,}408 * 11\% = 130{,}451 \text{ thousand EUR}$$

$$1{,}182{,}408 - 130{,}451 = 1{,}051{,}976 \text{ thousand EUR}$$

The number of students by education level is obtained from the CSB statistical portal. The average expenditure per student is calculated for each education level.

Expenditures on education for children in institutions by education level are calculated as:

$$D63 = \text{Expenditure per student} * \text{Number of children}$$

Information on inmate participation in education programs by level and the number of students is available from the **Prison Administration's annual public report** (publicly available online at: https://www.ievp.gov.lv/lv/gada-publiskie-parskati?utm_source=https%3A%2F%2Fwww.google.com%2F).

To calculate education expenditures for inmates, the previously calculated average expenditure per student for each education level is multiplied by the number of inmates participating in each level. For those involved in non-formal and hobby education, the average expenditure for vocational education per capita is applied, as these expenditures most accurately reflect the costs associated with education provision outside the formal education system.

To calculate **D63 for health** for inmates, the number of inmates by age group is obtained from the Prison Administration's annual public report. The average national health expenditures per capita by age group are calculated as described above. Health expenditures for inmates in each age group are calculated using the following formula:

$$D63 = \text{Expenditure per capita} * \text{Number of residents}$$

SOCIAL PROTECTION

Data on social services provided by social and night shelters are obtained from the Ministry of Welfare's "Report on Social Services and Social Assistance in the Municipality" (available at: https://www.lm.gov.lv/lv/gada-dati).

Social and night shelters provide various services, including short-term night or 24-hour accommodation, personal hygiene facilities, meals, social worker consultations, and assistance. These expenditures relate to **D63 for social protection**. A total of **3,111 thousand EUR** was spent on these services for shelters in 2019. These services were received by a total of **6,864 clients**, of whom 38 were children (0.6%). The average cost per person is calculated as:

$$3{,}111/6{,}864 = 0.45 \text{ thousand EUR}$$

According to census data, 2,384 people were long-term residents in social and night shelters in 2019. Expenditures for social protection for institutional residents are calculated using the following formula:

$$0.45 * 2{,}384 = 1{,}081 \text{ thousand EUR}$$

The report does not provide information on service recipients by age group, only the number of children (0.6%) and adults (99.4%) is known. These percentages are used to divide the long-term residents of shelters into children and adults:

$$0.6\% * 2,384 = 13 \quad 99.4\% * 2,384 = 2,371$$

To calculate the **D.63 for health** received by residents of shelters, the average national health expenditures per capita for children (0-17 years old) and adults (18+) are used. The average health expenditure per capita for children and adults is presented in Table2.

Table 1 Average Health Expenditure per Capita by Age Group

Variable	Children 0 – 17 years	Adults
Population	358 813	1 555 020
Total Health Expenditures, EUR	116 342 329	890 744 668
Expenditure per capita, EUR	324	573

The calculation for health services received by shelter residents is:

$$(13*324+2,371*573)/1000=1,362 \text{ thousand EUR}$$

Monasteries and Student Dormitories

Monasteries house 255 residents, which is only about 1% of all institutional residents. The closed and independent life of monasteries makes it difficult to obtain accurate statistical data. Data by age group and other information on life in monasteries are not available. For residents of monasteries, the average health expenditure per adult in the country is used:

$$255*573/1000=146 \text{ thousand EUR}$$

Student dormitories primarily house students aged 18 to 24, who comprise the majority of bachelor's and master's students. According to census data, 2,332 students live in dormitories. However, considering that 80-90% of students' income is attributed to their parental household, it is assumed that only 15% of students are relevant to the dormitory.

Of the total expenditures on education, 97% is covered by the government and 3% by MABO. In 2019, state expenditures on higher education were **118,136 thousand EUR**. On average, 61,156 students were enrolled in state universities in 2019.

To calculate the social transfers in kind for education that can be attributed to student dormitories, the following calculation is performed:

$$(118,136/61,156) * (2,332*0.15) = 676 \text{ thousand EUR}$$

To calculate the social transfers in kind for health received by students living in dormitories, data on the average annual per capita health expenditures for age groups 15-19 (342 EUR) and 20-24 (191 EUR) from the National Health Service are used. The average health expenditure for one resident in the 18-24 age group is calculated as:

$$((342*2) + (191*5))/7=234 \text{ EUR}$$

The calculation for social transfers in kind for health received by students in dormitories per year is:

$$234 * 2{,}332 * 0.15/1000 = 82 \text{ thousand EUR}$$

In total, the amount attributed to student dormitories is:

$$676 + 82 = 757 \text{ thousand EUR}$$

From the municipal budget reports collected by function (**COFOG 10.000**), information on services provided for social protection is obtained. **EKK code 6412**, "Payment for services provided by long-term social care and social rehabilitation institutions," includes the provision of housing, social care, and social rehabilitation for individuals unable to care for themselves due to age or health conditions, as well as for orphaned and parentless children. The total expenditure for these services in 2019 was **8,098 thousand EUR**. This entire amount is attributed to institutional residents.

According to the Eurostat format, the calculated social transfers in kind for institutions are broken down into the main groups (Table 3).

Table 2 Breakdown of Social Transfers in Kind (D63R) by Main Category

D63R	STiK	27 662
D63R1	Education	4 973
D63R2	Health	13 511
D63R3	Other	9 178

OTHER CURRENT TRANSFERS (D75)

Other current transfers received by **households (S14)** come from the **government (S13)** and the **rest of the world (S2)**. Transfers from the rest of the world can be in cash or in kind, received from emigrants or those working abroad. It's assumed that transfers from the government are not directed to residents in institutions.

We assume that other current transfers could be received by residents in prisons, social care centres, and students living in dormitories. Since specific data on the cash amounts received by residents of these institutions are not available, we calculate a proportional coefficient based on the share of the total population residing in these institutions as an estimation method. This coefficient is then used to multiply the total S14 D75 receipts.

A total of 17 487 individuals is attributed to these institutions. The total population of Latvia in 2019 was 1 913 188. The coefficient is calculated as:

$$\text{Coefficient} = 17\ 487 / 1\ 913\ 188 = 0{,}009$$

The D75 transfers received by institutions are calculated as:

$$\text{D75 transfers rec. by institutions} = \text{Coefficient} * \text{Total S14 D75 rec. from S2}$$

$$0.009 * 436{,}613 = 3{,}991 \text{ thousand EUR}$$

It may be difficult for children in children's homes to receive various regular transfers from emigrants or workers abroad, as money transfers are typically managed by the institution to ensure funds are used for the children's needs. Residents of monasteries are often part of religious communities where personal finances are collectively managed and used for community needs. Residents of night shelters and social shelters are often homeless or in a crisis situation and may not have access to a permanent bank account.

All various current transfers paid by households are attributed to private households. Residents in institutions often cannot make regular transfers to other households, as their finances are frequently managed by the institutions or communities, which limits their access to personal funds. Moreover, residents in institutions may not have access to bank accounts or other financial instruments required to make transfers. Students typically have insufficient funds for money transfers, as they often have limited financial resources.

STATISTICAL MATCHING AND CONSTRUCTION OF THE FUSED EU-SILC, HBS AND HFCS DATASET

To combine the EU-SILC, HBS and HFCS microdata and produce a synthetic fused dataset, a statistical matching approach was adopted. Statistical matching is one of the recommended methods for integrating microdata from independent surveys when no common identifiers are available. The method links records using common socio-economic characteristics observed in both datasets.

EU-SILC was selected as the recipient dataset because it contains the richest set of socio-economic variables and has the largest sample size among the three surveys. HBS and HFCS served as donor datasets providing additional information on household consumption and household wealth that is not available in EU-SILC.

Statistical Matching between EU-SILC and HBS

The matching was performed using the 2019 EU-SILC data and the 2019 HBS data. A nearest neighbour statistical matching approach based on the Gower distance was applied. Matching was carried out at the household level.

The statistical matching process consisted of the following stages:

- data preparation and initial quality assessment;
- selection and harmonisation of common variables;
- aggregation of person-level variables to the household level where required;
- selection of matching variables;
- execution of the statistical matching model;
- validation of the matching results.

The matching was implemented in R using the NND.hotdeck function from the **StatMatch** package.

The NND.hotdeck algorithm performs nearest neighbour hot-deck imputation. For each household in the recipient dataset (EU-SILC), the algorithm identifies the most similar household in the donor dataset (HBS) based on the selected matching variables and transfers the donor variables to the recipient. Gower distance was selected because it can simultaneously accommodate categorical, ordinal and continuous variables. The algorithm allows donor households to be matched multiple times when appropriate.

Harmonised Matching Variables

The statistical matching relied on a set of harmonised household and person-level variables available in both EU-SILC and HBS. Where necessary, variables were recoded into common categories to maximise comparability across surveys.

Table 4 summarises the harmonised household and person-level variables used in the statistical matching between EU-SILC and HBS, including the applied recoding and aggregation procedures.

Table 3 Harmonised matching variables used for EU-SILC and HBS statistical matching

Level	EU-SILC variable	Description	HBS variable	Description	Harmonisation
Household	Laupil	Urban/rural area	laupil	Urban/rural area	Direct correspondence.
Household	per_sk / per_sk_gr	Household size (grouped into 1–2, 3–4, 5+)	per_sk / per_sk_gr	Household size	Direct correspondence after grouping.
Household	HY020 / income_decile	Household disposable income (weighted deciles)	income / income_decile	Household income (weighted deciles)	Harmonised using weighted income deciles because raw income distributions differ between surveys.
Household	HH021	Housing tenure status	xC1	Housing tenure status	Recoded into common categories.
Household	HH030	Number of rooms	xC6	Number of rooms	Direct numerical comparison.
Person → Household	PL051	Occupation (1-digit ISCO)	xB5	Occupation (1-digit ISCO)	Aggregated to household level using indicator variables.
Person → Household	Age group	Age group	xA5	Age group	Harmonised age groups.
Person → Household	PE040	Highest educational attainment	xA10	Highest educational attainment	Recoded into ISCED groups 0–2, 3–4 and 5–8.
Person → Household	RB090	Sex	xA1	Sex	Direct correspondence.
Person → Household	PL031	Economic activity status	xB6_B9	Economic activity status	Harmonised categories.
Person → Household	PL111	Industry (broad NACE groups)	xB3	Industry	Harmonised into common NACE sections.
Person → Household	PB190	Marital status	xA6	Marital status	Recoded into comparable categories.

Validation of the Statistical Matching

The quality of the statistical matching was evaluated using the Hellinger distance (HD). Smaller Hellinger distance values indicate greater similarity between the distributions observed in the donor survey and those reproduced in the fused dataset.

The validation also included:

- comparison of weighted frequency distributions;
- graphical comparison of variable distributions;
- assessment of household-level matching quality;
- review of matching distances generated by the algorithm.

Together, these diagnostics confirmed that the statistical matching preserved the main distributional characteristics of the donor variables.

Deliverables

The statistical matching outputs are organised in a structured directory containing the following files:

Data

- fused_data.csv – final fused dataset containing EU-SILC households with transferred HBS variables.
- matching_ids.csv – donor–recipient linkage file.

Model

- model_vars.csv – variables used in the statistical matching model.
- Model_result_data.xlsx – summary of matching outputs.

Validation

- HD_model.xlsx
- HD_model_weighted.xlsx

These files contain the statistical matching validation results based on the Hellinger distance.

Metadata

- variable_metadata.xlsx – metadata for all variables.
- variable_mapping.csv – correspondence between EU-SILC and HBS variables.

Documentation

- README
- Process description and methodological documentation.

Description of the Main Files

matching_ids.csv

Contains the linkage between EU-SILC recipient households and HBS donor households.

Variables include:

- rec_id – technical record identifier;
- id_ms – EU-SILC household identifier;

- ms_id_n – matched HBS donor household identifier;
- distance – Gower distance between the matched households.

fused_data.csv

Contains the final fused microdata set, including all original EU-SILC variables together with the HBS variables transferred through statistical matching.

model_vars.csv

Lists all variables included in the statistical matching model and documents the matching specification.

Model_result_data.xlsx

Contains several worksheets summarising the matching process:

- matching identifiers;
- fused dataset;
- matching variables.

variable_mapping.csv

Provides the correspondence between EU-SILC and HBS variables used during harmonisation and matching.

variable_metadata.xlsx

Contains metadata for each variable, including:

- dataset;
- variable name;
- description;
- variable type;
- observation level;
- harmonisation status;
- use in statistical matching;
- applied transformations.

HD_model.xlsx

Contains the statistical matching quality assessment based on the Hellinger distance.

The **Summary** worksheet reports:

- analysed variable;
- variable type;
- Hellinger distance.

The **Distributions** worksheet reports:

- category or value;
- weighted frequency or proportion;
- distribution before and after matching;
- corresponding variable.

HD_model_weighted.xlsx

Provides the weighted version of the Hellinger distance validation.

The file contains both summary indicators and detailed weighted distributions for each matching variable.

Statistical Matching between EU-SILC and HFCS

The statistical matching between the European Union Statistics on Income and Living Conditions (EU-SILC) and the Household Finance and Consumption Survey (HFCS) was performed using the 2019 EU-SILC data and the 2020 Latvian HFCS data collected by Latvijas Banka.

A statistical matching approach was selected, as it is one of the recommended methods for integrating independent microdata sources that do not share common identifiers. The methodology is based on linking records using a set of harmonised socio-economic characteristics observed in both surveys.

EU-SILC was selected as the recipient dataset because it contains a comprehensive set of socio-economic variables and has a substantially larger sample size than HFCS. The HFCS served as the donor dataset, providing detailed information on household financial assets, liabilities and net wealth that is not available in EU-SILC.

The matching was performed at the household level.

The statistical matching procedure included the following stages:

- data preparation and initial quality assessment;
- selection and harmonisation of common variables;
- aggregation of person-level variables to the household level where necessary;
- selection of matching variables;
- estimation of statistical matching models;
- validation of the matching results.

The matching was implemented in R using the NND.hotdeck function from the **StatMatch** package. This function performs nearest neighbour hot-deck imputation by identifying, for each recipient household, the most similar donor household according to the selected matching variables and transferring the donor information to the recipient. Household similarity was measured using the Gower distance, which is suitable for datasets containing both categorical and continuous variables. Donor households were allowed to be matched to multiple recipient households.

Harmonised Matching Variables

The statistical matching relied on a set of harmonised household- and person-level variables available in both EU-SILC and HFCS. Where necessary, variables were recoded into common classifications to maximise comparability across the two surveys. Table 5 summarises the harmonised household and person-level variables used in the statistical matching between EU-SILC and HFCS, including the applied recoding and aggregation procedures.

Table 4 Harmonised matching variables used for EU-SILC and HFCS statistical matching

Level	**EU-SILC variable**	**Description**	**HFCS variable**	**Description**	**Harmonisation**
Household	HH021_j	Housing tenure status	DHHST	Main residence tenure status	Direct harmonisation; included in the final matching model.
Household	hh_j	Household type	DHHTYPE_j	Household composition	Harmonised household categories.
Household	ienakumu_kvint	Household income quintile	ienakumu_kvint	Household income quintile	Retained in the initial matching evaluation.
Person → Household	vec_gr_1–vec_gr_4	Number of household members by age group	vec_gr_1–vec_gr_4	Number of household members by age group	Constructed as count indicators and aggregated to household level.
Person → Household	RB090_1, RB090_2	Sex indicators	RA0200_1, RA0200_2	Sex indicators	Direct correspondence; aggregated to household level.
Person → Household	PL040_2, PL040_3	Employment status indicators	PE0200_j_2, PE0200_j_3	Employment status indicators	Harmonised and aggregated.
Person → Household	PL031_j_1–PL031_j_6	Economic activity indicators	PE0100a_j_1–PE0100a_j_6	Economic activity indicators	Harmonised into common categories and included as structural matching variables.
Person → Household	PB190_j_*	Marital status indicators	PA0100_*	Marital status indicators	Harmonised into comparable categories.
Person → Household	PE040_j_*	Educational attainment indicators	PA0200_*	Educational attainment indicators	Harmonised and aggregated to household level.
Person → Household	PL051_1digit_*	Occupation (ISCO major groups)	PE0300_1digit_*	Occupation (ISCO major groups)	Dummy indicators aggregated to household level.
Person → Household	PL111_j_*	Industry (broad NACE groups)	PE0400_j_*	Industry (broad NACE groups)	Harmonised into common NACE sections.

Level	EU-SILC variable	Description	HFCS variable	Description	Harmonisation
Person → Household	PL140_1	Full-time/part-time employment indicator	PE0500_1	Full-time/part-time employment indicator	Direct correspondence and aggregation to household level.

Evaluation of Candidate Matching Variables

Several income variables were initially considered for inclusion in the matching model. These included household income quintiles and deciles as well as gross income variables derived from both EU-SILC and HFCS.

Following an assessment based on the Hellinger distance, these variables were excluded from the final matching model because their distributions differed substantially between the two surveys. In particular, variables describing self-employment income exhibited a high proportion of zero values, indicating unstable distributions and reducing their suitability for statistical matching.

The variables excluded from the final model included:

- bruto_i_kvint
- PY050G_kvint
- bruto_i_decil
- PY050G_decil
- ienakumu_decil

The evaluation also considered the proportion of missing and zero values together with Hellinger distance statistics at both quintile and decile levels. Variables with large distributional discrepancies (HD > 0.10) were excluded from the final specification.

Validation of the Statistical Matching

The quality of the statistical matching was assessed using the Hellinger distance (HD), complemented by comparisons of weighted distributions before and after matching.

As a general guideline, Hellinger distance values below 0.10 were considered indicative of satisfactory agreement between the original and matched distributions.

The validation process also included:

- comparison of weighted frequency distributions;
- graphical assessment of variable distributions;
- evaluation of household-level matching quality;
- review of donor-recipient matching distances.

Deliverables

The statistical matching outputs are organised in a structured directory containing the following files.

Data

- fused_data.csv – final fused EU-SILC/HFCS dataset;
- matching_ids.csv – donor-recipient linkage file.

Model

- model_vars.csv – variables included in the matching model;
- Model_result_data.xlsx – summary of the matching outputs.

Validation

- HD_model.xlsx;
- HD_model_weighted.xlsx.

Metadata

- silc_hfcs_metadata.xlsx;
- variable_mapping.csv.

Documentation

- Process description and methodological documentation.

Description of the Main Files

matching_ids.csv

Contains the donor-recipient linkage between EU-SILC and HFCS households.

model_vars.csv

Lists all variables included in the statistical matching model and documents the final model specification.

Model_result_data.xlsx

Provides a summary of the matching process, including:

- matching identifiers;
- fused dataset;
- matching variables.

variable_mapping.csv

Documents the correspondence between EU-SILC and HFCS variables used during harmonisation and statistical matching.

silc_hfcs_metadata.xlsx

Contains metadata for each variable, including:

- dataset;
- variable name;

- description;
- variable type;
- observation level;
- harmonisation status;
- use in statistical matching;
- applied transformations.

HD_model.xlsx

Contains the quality assessment of the statistical matching based on the Hellinger distance.

The **Summary** worksheet reports:

- analysed variable;
- variable type;
- Hellinger distance.

The **Distributions** worksheet reports:

- category or value;
- weighted frequency or proportion;
- distribution before and after matching;
- corresponding variable.

HD_model_weighted.xlsx

Provides the weighted version of the Hellinger distance validation.

The file contains summary indicators together with weighted distributions for all variables included in the validation process.

DISTRIBUTION OF MICRO–MACRO INCOME GAPS ACROSS HOUSEHOLDS

GENERAL METHODOLOGICAL APPROACH

The objective of the income allocation is to reconcile household microdata with National Accounts aggregates while preserving the observed distribution of income across households. Depending on the availability and quality of microdata, four methodological approaches recommended in the OECD *Handbook on Distributional Results on Household Income, Consumption and Saving* were applied.

Method A – Proportional allocation (Scaling).

Applied when a survey variable represents the same economic concept as the National Accounts transaction. Household values are proportionally scaled to match the National Accounts aggregate while preserving the original distribution.

Method B – Proxy allocation.

Applied when no direct microdata are available. A conceptually related survey variable is used as a proxy to distribute the National Accounts total across households.

Method C – Allocation using external information.

Applied when neither direct microdata nor an appropriate proxy is available. Household allocation is based on external data sources or additional statistical information.

Method D – Residual allocation.

Applied to balancing items that cannot be allocated directly. Household values are derived as residuals from other National Accounts transactions to preserve accounting consistency.

All allocated variables were calibrated to exactly reproduce the corresponding National Accounts aggregates.

The methodology was developed and implemented in Python using 2019 reference data.

B2G – GROSS OPERATING SURPLUS

This item in the national accounts represents the surplus generated by households from owner-occupied housing services and the leasing of dwellings.

Imputed rent for owner-occupied dwellings (B2R1g)

Methodological Approach

In national accounts, imputed rent is estimated using the User cost method reflecting the full housing stock. Microdata in EU-SILC is estimated using a rental equivalence method. The primary microdata source is the **EU-SILC** indicator **HY030G** (Imputed rent). Due to conceptual differences between the NA methodology and EU-SILC estimates, microdata values exceed the NA benchmark.

Gap Allocation

To ensure consistency with national accounts aggregates, **Method A** is applied. Microdata values are proportionally scaled down across all households using the ratio of NA imputed rent to SILC imputed rent. This adjustment reconciles the level while preserving the distributional pattern observed in the microdata. The coverage gap reflects **conceptual differences**, not data quality issues.

Technical implementation

Microdata values (HY030G) were aggregated by household and aligned to the NA benchmark using a proportional scaling factor. The scaling factor was calculated as the ratio of adjusted NA imputed rent (B2R1g) to the corresponding SILC total for each year. Adjusted micro values were obtained by multiplying all household-level SILC imputed rent amounts by this factor, preserving the distribution while matching the NA aggregate.

Verification and Robustness Checks

A set of standard plausibility and internal-consistency checks was carried out to validate the distribution of imputed rents after proportional scaling:

- **Tenure Consistency:** Imputed rent was examined against tenure status (HH021). As expected, the majority of households with non-zero imputed rent were owner-occupiers. A limited number of tenant households displayed positive values, which is consistent with EU-SILC reporting practices (e.g. co-residency, secondary-property ownership). No systematic inconsistencies were identified.
- **Structural Consistency with Dwelling Value:** Given that imputed rent is derived directly from HY030G via proportional scaling, a perfect linear relationship was confirmed (correlation = 1.00), demonstrating that the scaling procedure preserves the underlying microdata structure.

- **Outlier Assessment:** No extreme imputed rent values were observed after scaling. In particular, no households recorded imputed rents exceeding EUR 30,000 annually, indicating that the application of the scaling factor did not distort the upper tail of the distribution.
- **Dwelling-Type Comparison:** Average imputed rents remained higher for detached houses relative to apartments, consistent with the higher average value of owner-occupied single-family dwellings. This confirms that the distribution maintains expected differences across dwelling categories.

Leasing of dwellings (B2R2g) and Rent Received (D45)

Methodological Approach

To ensure consistent data handling, the **Leasing of dwellings** (B2R2g) is processed together with **Rent received** (D45rec), as the corresponding microdata source does not allow for a granular distinction between these two components.

The primary microdata source is the **EU-SILC variable HY040G** (Income from rental of a property or land). Since EU-SILC does not allow a clear separation between these components, **National Accounts (NA) proportions** are used to disaggregate HY040G into:

- D45rec (Rent received): **81%**
- B2R2g (Leasing of dwellings): **19%**

This approach ensures consistency with the NA classification while preserving the SILC distributional structure.

Gap Allocation

The combined micro-macro coverage rate is **135%**, resulting in a **negative gap** of EUR 33,7 mil. Given the high coverage and the small scale of these items (D45rec accounts for only 0.45% of gross disposable income), **Method A** is used.

The total gap is proportionally reduced across all households based on their HY040G income. Subsequently, the adjusted values are split according to the NA proportions (81% to D45rec and 19% to B2R2g). This approach ensures a methodologically sound and resource-efficient distribution.

Data Source Validation and Consistency

A comparative assessment using EU-SILC and HFCS (DI1300 - Income from real estate rental) was conducted for 2020. Although HFCS showed a closer alignment with NA (coverage 118% vs. 134% in SILC), EU-SILC was retained as the primary input to ensure cross-component consistency, avoid survey mixing, and prevent artificial breaks in the HDA time series. The allocation of rental income is summarised in Table 6.

Table 5 Summary of the Allocation for Rental Income (EUR thousand)

Component	Macro value	structure	Micro value	Gap	Method
D45R Rent received	79 484	81%	*106 930*	*-27 447*	**A**
B2R2g Leasing of dwellings	18 208	19%	*24 496*	*-6 288*	**A**
Total	97 692		*131 426*	*-33 734*	**A**

Technical Implementation

Data processing description

The alignment of the rental income was performed in two sequential steps:

- **Linear Gap Adjustment:** The original EU-SILC variable HY040G was reconciled with the National Accounts (NA) benchmark using a scaling factor k = 0.7433, calculated as the ratio between the combined NA target (97 692 EUR thousand) and the weighted SILC total (131 426 EUR thousand). This adjustment was applied to each household record to eliminate the 135% over-coverage gap.
- **Proportional Disaggregation:** Post-adjustment, the total value was decomposed into two distinct HDA components based on fixed NA proportions:
  D45rec = 0.81 × k × HY040G
  B2R2g= 0.19 × k × HY040G

This procedure ensures that while the absolute levels are benchmarked to NA totals, the underlying distributional characteristics (income concentration and inequality) remains consistent with the survey data.

Verification and Robustness Checks

A structured set of plausibility checks was performed to validate the distributional consistency of the scaled rental income components:

- **Tenure Consistency:** 93% of households receiving rental income were identified as homeowners (HH021 = 1–2). A small number of tenant households (23 cases) recorded rental income, which is consistent with sub-letting or secondary-property ownership. No systematic inconsistencies were detected.
- **Dwelling Structure vs. Land Rent (D.45):** Adjusted land rent (D45rec_adj) was primarily associated with detached houses (EUR 298 thousand), with a smaller share attributed to apartment dwellings (EUR 264 thousand), reflecting the expected link between land income and landed property types.
- **Outlier Assessment:** The scaling procedure did not generate implausible extremes. Only two households were identified with annual rental income above EUR 30,000. No intervention was deemed necessary.

B3G - MIXED INCOME

Mixed income is a balancing item in the national accounts that combines both wages and business profits for unincorporated enterprises owned by households. Due to its complex nature, it is divided into three main components, each requiring a specific distributional approach based on the availability and quality of microdata. Table 7 shows the decomposition of household gross mixed income (B3g) by components in the HDA framework.

Table 6 B3g income composition

Code	Description	Macro value	Proportion
B3R1g	Own account production	207 261	15%
B3R2g	Underground production	499 813	36%
B3R3g	Mixed income excluding underground and own account production	693 211	49%
Total B3g		**1 400 295**	100%

B3R1G OWN ACCOUNT PRODUCTION

Methodological Approach

The own account production component in the national accounts includes three main elements:

- Value of goods produced for own consumption.
- Major housing improvements (capital repairs) carried out by households themselves.
- Self-built single dwelling houses.

Table 8 compares ESA 2010 own-account production and housing-related national accounts items with their corresponding microdata proxies, highlighting coverage rates and methodological gaps.

Table 7 Coverage of own-account production components

National account item	Macro value	Micro item	Micro value	Coverage rate	Method
Own account agricultural products produced by households	146 394	HY170G: Value of goods produced for own consumption	182 341	125%	A
Own account major improvements of dwellings by households	11 424	Cadastral proxy (property value)	N/A	NA-only	B
Own account single dwellings houses produced by households	49 444	Cadastral proxy (property value, new-construction)	N/A	NA-only	B

Own-account agricultural products (HY170G)

The microdata source for the **own acount agricultural products** produced by households is the EU-SILC variable **HY170G** (Value of goods produced for own consumption). The micro–macro coverage rate is **125%**, indicating that SILC overstates the national accounts benchmark. Therefore, **Method A** is applied. HY170G values are proportionally scaled down using the ratio Macro / Micro to ensure the weighted micro-total aligns with the NA aggregate while preserving the original household distribution.

Validation – Own-production from Agriculture (B3R1g)

Basic plausibility checks confirm the consistency of the results. Around 41% of households report own-production (HY170G > 0), which is in line with the broad EU-SILC definition. The distribution is concentrated in detached houses and rural areas, with a smaller but plausible presence in urban settings. No extreme values were observed, and zero-value records remained unchanged after scaling. The adjusted weighted total fully aligns with National Accounts.

Own-account major improvements of dwellings by households

Own-account capital repairs were distributed using administrative cadastral information linked to EU-SILC households through pseudonymised person identifiers.

Since no direct survey variable captures major dwelling improvements or self-built dwellings, a **Method B** approach was applied, using dwelling characteristics as structural proxies for the distribution of investment activity.

Data Sources

The allocation combines three sources:

Cadastral database;

EU-SILC household and person files.

The cadastral records contain dwelling characteristics including total residential floor area (TG_FLAT_TOTAL_AREA), building type (PREGUT_ID), and year of commissioning (EXPLOIT_YEAR).

Cadastral records were first linked to persons using the pseudonymised identifier PK. Household identifiers (id_ms) were subsequently derived through the EU-SILC person file, allowing aggregation to household level while preserving confidentiality. Only successfully matched records were used in the allocation.

B3g Major Improvements of Dwellings

Methodological Approach

National Accounts expenditure on major improvements was allocated using total residential floor area as the distribution key. The underlying assumption is that larger dwellings require larger renovation expenditures and are therefore more likely to account for a greater share of capital improvements.

Distributional Proxy

For each household, the allocation key was calculated as the sum of **TG_FLAT_TOTAL_AREA** across all matched cadastral properties. The resulting household-level area measure was calibrated to the National Accounts total.

National Accounts Target

B3g Major Improvements of Dwellings: 11.424 million EUR

Validation Results

The allocation covered 3,931 households. Distributional concentration remained moderate:

Top 1% of recipient households received 23.14% of the allocated amount;

Top 10% received 46.78%.

The largest household allocation amounted to approximately 1.36 thousand EUR. Additional validation based on household-level income quintiles confirmed that the allocation produced zero household movements of two or more quintiles, ensuring high stability of the baseline distribution.

B3g Own-account single dwelling houses produced by households

Methodological Approach

National Accounts expenditure related to self-built dwellings was allocated using information on recently commissioned private houses recorded in the cadastral database. Because the exact year of construction is not available in EU-SILC, the cadastral commissioning year (EXPLOIT_YEAR) was used as a proxy for recent construction activity.

Selection Criteria

Only cadastral records satisfying both conditions were included:

PREGUT_ID = 1110 (single-dwelling private houses);

EXPLOIT_YEAR between 2008 and 2019.

This produced 97 eligible cadastral records corresponding to 93 households.

Distributional Proxy

The allocation key was based on residential floor area adjusted by a commissioning-year score:

Key = TG_FLAT_TOTAL_AREA × YearScore

where newer dwellings receive higher weights. The scoring scheme was applied as follow in Table 9.

Table 8 Commissioning year weighting scheme for own-account dwelling allocation

Year	Score
2008	0.6
2009	0.65
2010	0.7
2011	0.75
2012	0.8
2013	0.85
2014	0.9
2015	0.95
2016	0.98
2017–2019	1

This approach assumes that recently commissioned houses are more likely to reflect construction activity captured in National Accounts, while older dwellings receive progressively lower weights.

National Accounts Target

Own-account construction of owner-occupied dwellings: 49.444 million EUR

Validation Results

The allocation covered 93 households. Distributional concentration remained low despite the small target population:

Top 1% share: 1.25%

Top 10% share: 17.20%

The maximum household allocation was approximately 15.99 thousand EUR. A robustness check based on weighted household income quintiles showed exactly zero household movements of two or more quintiles after the inclusion of the imputed amounts, indicating that the allocation does not distort the macro-structural income distribution.

Overall Assessment

The methodology relies exclusively on observable cadastral characteristics and avoids model-based imputations. Residential floor area serves as the primary proxy for investment capacity, while commissioning year provides additional information for identifying recently constructed dwellings.

The resulting distributions are fully consistent with the National Accounts aggregates, exhibit structurally sound concentration profiles, and pass rigorous household-level plausibility checks based on income-quintile stability.

B3G SELF-EMPLOYMENT

The self-employment component in national accounts is divided into three subcategories:

- Registered self-employed (based on administrative data and annual reports)
- Self-employed farmers (NACE 01)
- Adjustment for non-registered employment (NA estimates based on the labour input method)

Coverage and Scaling

A comparative analysis reveals a micro–macro coverage rate of **79%** when comparing EU-SILC data (PY050G) exclusively against the official registered self-employment component in National Accounts.

To align with National Accounts, **Method A** is applied.

Technical Implementation

The estimation of **PY050G_HDA** was based on the EU-SILC person-level microdata. The source variables used were **PY050G** (observed self-employment income), **RB050** (person weight), **id_pers** (person identifier), and **id_ms** (household identifier).

The adjustment was restricted to individuals with positive observed **PY050G**. Consequently, no new recipients of self-employment income were introduced; the allocation gap was distributed exclusively among individuals already reporting positive **PY050G** in the source data.

First, the weighted aggregate of observed **PY050G** was calculated using **RB050** and compared with the corresponding macro target. The difference between the macro target and the weighted EU-SILC aggregate was defined as the *allocation gap*.

The gap was allocated proportionally to observed **PY050G**, while applying a progressive brake to the upper tail of the distribution. Eligible individuals were ranked by **PY050G** and assigned to weighted percentile groups using **RB050**. A brake factor was then applied to each group to reduce the allocation weight of higher-income observations.

The brake schedule was defined as follows:

- P0–P80: 1.00
- P80–P90: 0.65
- P90–P95: 0.35
- P95–P99: 0.10
- P99–P100: 0.00

For each eligible individual, the allocation weight was defined as:

allocation weight = observed PY050G × brake factor

The total gap was then distributed proportionally to these adjusted allocation weights. The final adjusted value was calculated as:

PY050G_HDA = observed PY050G + allocated gap component

The resulting person-level variable was written to *HDA_Working_Person.csv*. For the household-level dataset, person-level **PY050G_HDA** values were aggregated by **id_ms** and written to *HDA_Master_HH.csv*. No other variables were modified.

Plausibility Checks

Several plausibility checks were conducted following the allocation.

First, it was verified that the weighted person-level total of **PY050G_HDA** exactly matched the macro target. This result was subsequently confirmed at the household level after aggregation.

Second, the pre-allocation coverage ratio was assessed by comparing the weighted observed **PY050G** total in EU-SILC with the macro target. This provided an indication of the magnitude of the allocation gap.

Third, the distributional impact of the adjustment was evaluated using the Top 1% share, Top 10% share, and the Gini coefficient, both before and after the allocation.

Finally, the results were reviewed across weighted percentile groups to confirm that the progressive brake operated as intended—namely, that the lower and middle segments of the distribution absorbed a larger share of the gap, while the highest-income groups received a substantially reduced share or no additional allocation.

B3R2G UNDERGROUND ECONOMY

In Latvian National Accounts, underground and illegal production is recorded through exhaustiveness adjustments that primarily affect the household sector's **Gross Mixed Income (B3g)**. The methodological focus is on three components:

- **N1**: Non-registered employment,
- **N6**: Misreporting of income by self-employed persons (separately available from 2022 onwards; previously included in total adjustments).
- **N2**: Illegal production.

The underground economy component accounts for **35%** of total mixed income.

N1 Non- registered self-employed + N6 Misreporting of income by self-employed persons

The non-registered self-employed are individuals operating informally.

The allocation was implemented as a hybrid observed-anchor and latent scoring C-method. The approach combines explicit self-employment signals already observed in the microdata with a residual latent allocation for households with indications of potential missing income.

Conceptual framework

The approach consists of three main steps:

Step 1: Observed self-employment anchor and sector alignment: The observed component is restricted to persons with explicit or implicit self-employment evidence in the microdata and to NACE sections with a positive benchmark value. NACE division 01 was explicitly excluded, while other eligible divisions within section A were retained where present in the benchmark. The observed component is aligned with the NACE structure of the external benchmark.

Step 2: Essential consumption–income discrepancy: Households are prioritised using an essential consumption–income gap derived from the fused SILC–HBS dataset. A positive and sufficiently large gap is interpreted as an indicator of potential missing income and increases the priority for latent allocation.

Step 3: Low formal income and socio-demographic scoring: Low or zero formal wage income is used as an additional proxy for potential informal activity.

The broad essential consumption aggregate is constructed using selected COICOP categories (food, housing, health, communication, education, clothing, transport, and basic services) to approximate a reasonable standard of living. The indicator is equivalised using the OECD equivalence scale to adjust for household size and ensure comparability across households.

This is combined with donor-based socio-demographic profile factors derived from observed self-employment cases.

The approach is conceptually consistent with National Accounts treatment but implemented as a household-level scoring model rather than a direct Labour Input Method estimation.

Macro allocation logic

The total macro benchmark is distributed across households so that the final weighted allocation matches the National Accounts control total.

For the observed component, the initial allocation is aligned with the NACE structure of the benchmark. After applying household-level caps, the realised NACE fit is evaluated using validation indicators rather than imposed as an exact constraint.

Technical implementation

The estimation of B3g_N1N6_HDA is based on the integrated use of:

- EU-SILC person file
- EU-SILC household file
- fused SILC–HBS household dataset
- NA N1+N6 benchmark by NACE

The allocation is implemented at the household level, using both person-level and household-level information.

Conceptually, the total target amount is decomposed into:

- an observed component
- a latent component

with an initial planning ratio of 50:50.

Operationally:

- the observed component is estimated first and capped
- the latent component is defined residually to reach the total target

Observed block

The observed block captures households with explicit or implicit self-employment evidence.

A person is considered to have an observed anchor if at least one of the following applies:

- self-employment status (PL031)
- current self-employment status (PL040)

- non-zero PY050G (positive or negative)

The observed allocation is restricted to NACE sections with a positive benchmark. NACE division 01 is excluded.

Person-level base

The observed allocation base combines:

- strength of the self-employment anchor
- sign of PY050G
- remaining distance to a section-specific cap

The magnitude of PY050G enters indirectly through the remaining capacity to the cap. Persons with high reported income therefore have less capacity for additional allocation.

The observed component is aligned with the benchmark by distributing the planned envelope across NACE sections proportionally, generating section-specific scaling factors.

Household-level caps

The allocation is aggregated to the household level and constrained using caps based on:

- essential consumption gap
- existing positive PY050G
- signals from negative PY050G
- allowances for households without a positive gap

This ensures that the observed component remains consistent with both the observed signal and the household's economic capacity.

Latent block

The latent block captures potential underreporting in households without observed self-employment signals.

Eligibility (person level)

A person is eligible if:

- adult
- no observed anchor
- not in excluded NACE division 01
- low or zero wage income
- low or zero self-employment income
- low or zero dividend income

Latent socio-demographic scores are derived using donor-based profiles from observed cases.

Household-level base

A household is eligible if:

- no observed anchor
- positive consumption gap above threshold
- low wage income
- low self-employment income
- low dividend income
- at least one positive latent score

The latent base combines:

- the sum of the two highest latent person scores in the household,
- gap-based scaling factor,
- additional factor for low wage households,
- downward adjustment for benefit-only households.

After finalising the observed component, the residual amount is distributed proportionally across latent households.

The latent component is not constrained by NACE and acts as a residual adjustment to reach the macro target.

Plausibility checks

A comprehensive set of plausibility and validation checks was carried out after the imputation.

First, it was verified that the weighted total allocation matched the target total and that the decomposition into observed and latent components was internally consistent.

Second, the realised split between the observed and latent components was compared with the initial 50:50 design. As the observed component was subject to household-level caps, the latent component absorbed the residual required to reach the overall target.

Third, the distribution of the observed component across NACE sections was evaluated against the external N1 benchmark using:

- section-level allocation shares,
- Spearman rank correlation between target and achieved shares,
- and mean absolute error (in percentage points).

Fourth, concentration measures were assessed for the total allocation and separately for the observed and latent components. The following indicators were used:

- Top 1% share,
- Top 10% share,

- weighted Gini coefficient (before and after imputation).

Finally, a set of household-level review flags was constructed to identify potentially implausible outcomes. These flags captured cases where the imputed amount was high relative to:

- household income,
- household consumption,
- or the number of adults in the household,

as well as households receiving substantial imputations despite having no observed self-employment anchor. Such cases were extracted for targeted review.

Key validation indicators

- Target vs allocated: 311,065.21 vs 311,065.21 thousand EUR
- Observed share: 29.6%
- Latent share: 70.4%
- Spearman (NACE): 0.94
- MAE: 2.72 p.p.
- Top 1%: 5.18%
- Top 10%: 30.6%
- Observed households: 333
- Latent households: 379
- Allocation to HH with positive gap: 81.79%
- Allocation without observed anchor: 70.45%

N2: Illegal Production

Illegal tobacco and alcohol trade

Methodological Approach

Illegal tobacco and alcohol were allocated using a **two-stage regional risk-based approach**.

First, national totals were distributed across regions using **externally defined regional structures**. This ensured that the final household-level allocations were fully consistent with the imposed regional distribution.

Two external data sources are used to derive **geographical weights**:

- **Illegal tobacco**: NielsenIQ Empty Pack Survey (regional share of detected illegal tobacco). Original city-level estimates were aggregated to regional level using population-weighted shares. The resulting regional structure was used as a geographical constraint in the household allocation.
- **Illegal alcohol**: Regional weights were derived from public health statistics published by the Centre for Disease Prevention and Control (CDPC), based on the prevalence of alcohol dependence and alcohol-induced psychosis diagnoses per 100,000 inhabitants. The combined indicator was normalised to obtain the regional distribution used as a geographical constraint in the allocation.These

indicators provide independent, non-survey-based measures of illicit market intensity and are used to constrain both income and consumption allocations geographically.

Second, within each region, potential household candidates were identified using the **equivalised essential gap (gap_essential_eq_broad)** derived from the SILC–HBS fused dataset. Only households with at least one person aged 16–65 and a **positive socio-demographic score** were considered eligible.

Thresholds were selected empirically following iterative plausibility testing in order to obtain a sufficiently narrow candidate base while maintaining adequate household coverage within each region.

This regional candidate filtering was introduced to avoid assigning illegal production or trade activity too broadly across the household population.

Within the selected candidate group, the final allocation was made **proportionally to a household risk score**. This score combined:

- household-level socio-demographic characteristics from EU-SILC household data
- person-level characteristics from EU-SILC personal data, aggregated as the **sum of the top two persons aged 16–65**
- only households passing the regional essential-gap selection remained eligible for allocation

The final regional allocation followed the formula:

HH allocation = Regional target × (HH risk score / Σ HH risk score within region)

This procedure was applied separately for tobacco and alcohol, using different regional candidate cut-offs while maintaining the same general scoring framework.

Plausibility Checks

The final weighted totals matched the National Accounts targets **exactly** for both products, and all regional targets were also fully met.

The allocation resulted in a relatively narrow candidate base, which was considered **economically more plausible** for illegal market activity than broader household-wide allocations:

- **169** households received positive illegal tobacco allocation
- **293** households potential involvements received positive illegal alcohol allocation

The resulting concentration remained moderate:

- **Illegal tobacco:** Top 1% share = 6.3%, Top 10% share = 22.5%
- **Illegal alcohol:** Top 1% share = 4.0%, Top 10% share = 27.0%

No regional fallback allocation was required, indicating that each region contained a sufficient number of positive-risk households to support the imposed regional targets.

Interpretation

The allocation does not attempt to identify actual participants in illegal activity. Instead, it distributes National Accounts totals to households displaying a combination of economic signals, socio-demographic characteristics and regional exposure indicators that are statistically more consistent with potential involvement in informal or illegal market activity. The allocation should therefore be interpreted as a statistical distribution mechanism rather than an identification of actual illegal behaviour.

Prostitution, Drug and Fuel Smuggling

The residual illegal-market component **B3gillegal_other** comprises income attributed to prostitution, smuggled fuel and narcotics. No direct microdata source or reliable external distributional information was available for these activities. Therefore, **Method D**, as described in the OECD Distributional National Accounts Guidelines, was applied.

A household-level **B5g proxy** was reconstructed from the already distributed HDA components representing the Balance of Primary Incomes. This proxy was selected as the most appropriate balancing item because the underlying activities are market-based income-generating activities and are therefore conceptually closer to primary income than to disposable income.

The allocation key was defined as the positive part of the reconstructed B5g proxy:

$$\boldsymbol{Key_i = max\,(B5g_i, 0)}$$

Only households with positive B5g values were eligible for allocation. The National Accounts total was distributed proportionally to the allocation key and subsequently calibrated to match the National Accounts aggregate exactly. Consistent with Method D, this approach preserves the distribution of the selected balancing item while providing a transparent allocation mechanism in the absence of direct distributional evidence.

The resulting household-level variable was recorded as: **B3gillegal_other**

Validation

Validation included calibration to the National Accounts total, recipient-household counts, concentration indicators and distributional consistency checks based on reconstructed B5g income quintiles.

The final allocation:

matched the National Accounts target exactly;

covered **6,034 households** with positive allocations;

produced a maximum household allocation of **EUR 3,702.77**;

resulted in a **Top 1% share of 12.40%** and a **Top 10% share of 42.19%**.

The allocation followed the existing primary-income structure of households and did not rely on unsupported behavioural, socio-demographic or regional assumptions. Given the absence of suitable microdata or external indicators, the resulting distribution was considered the most transparent and statistically consistent implementation of Method D for this component.

B3G OTHER

Methodological Approach

The residual B.3g Other component was allocated using a composite household-level proxy based on two mixed-income variables already available within the HDA framework:

- **PY050G_HDA**
- **B3g_N1N6_HDA**

As B3g Other represents a residual mixed-income component for which no direct microdata counterpart exists, households with positive self-employment or mixed-income-related amounts were considered the most plausible recipients. Following the OECD guidance for proxy-based imputations, the allocation was based on the distribution of the closest available household-level mixed-income variables.

Distributional Proxy

A composite allocation key was constructed as:

$$Key = \sqrt{\max\left(PY050G\_HDA, 0\right)} + \sqrt{\max\left(B3g\_N1N6\_HDA, 0\right)}$$

Only positive values contributed to the key.

The square-root transformation preserves the ranking of households according to mixed-income participation while reducing the influence of extreme observations in the upper tail of the distribution. This smoothing approach was adopted to avoid excessive concentration of the residual component among a small number of households with exceptionally large mixed-income amounts.

The use of a composite proxy also broadens the recipient base by capturing households represented in either of the two core mixed-income variables, thereby providing a more comprehensive approximation of the potential distribution of the residual mixed-income component.

Allocation Procedure

The National Accounts total for B.3g Other was distributed proportionally to the household allocation key and subsequently calibrated to match the National Accounts aggregate exactly.

The resulting household-level allocation was recorded as: **B3gother_HDA**

Validation

Several plausibility checks were performed, including:

- calibration to the National Accounts total;
- number of recipient households;
- maximum household allocation;
- concentration indicators (Top 1% and Top 10% shares).

As a sensitivity test, an alternative allocation based solely on **PY050G_HDA** was evaluated. This specification resulted in a substantially smaller recipient population and a considerably more concentrated distribution. The composite proxy therefore was retained as the preferred specification, as it produced a broader and more stable allocation while remaining closely linked to the underlying mixed-income structure.

Alternative specifications based on the full sum of previously distributed B.3g components were evaluated. These produced substantially higher concentration levels and introduced economically heterogeneous allocation drivers (e.g. owner-occupied housing services). Consequently, these specifications were not retained.

The final allocation achieved full consistency with National Accounts totals and generated a distribution considered plausible for a residual mixed-income component lacking a direct microdata counterpart.

D11 – WAGES AND SALARIES

The microdata source for wages and salaries in cash is EU-SILC indicator PY010G (Employee cash or near-cash income). The micro/macro coverage rate for cash wages is 95%.

The microdata sources for wages in kind were EU-SILC indicators PY020G (Non-cash employee income) and PY021G (Company car), with a coverage rate of 95%.

Given the strong link between micro and macro data, the gap in wages is distributed proportionally across households using **Method A**.

Technical Implementation

Data Processing

Wages and salaries were aligned with adjusted National Accounts (NA) totals using **Method A (proportional scaling)**, given the high coverage rate (95%). The alignment is performed separately for cash and in-kind components:

- **Cash:** Based on EU-SILC variable **PY010G**.
- **In-kind:** Based on the sum of **PY020G** and **PY021G**.

Scaling factors (k) are derived by comparing the adjusted NA targets against the weighted microdata totals ($\sum = \mathrm{Variable} \times \mathrm{DB090}$)

Algorithm:

$$\mathrm{Adjusted\_Value}_h = (\mathrm{PY010G}_h \times \mathrm{k}_{\mathrm{cash}}) + ((\mathrm{PY020G}_h + \mathrm{PY021G}_h) \times k_{\mathrm{kind}}$$

(where h denotes the individual household).

Imputation and Gap Distribution:

- **No Synthetic Imputation:** Due to high coverage, no new recipients are imputed. The 5% gap is distributed proportionally among existing recipients to maintain the survey's internal structure.
- **Zero-values:** Households with no reported wage income remain at zero, as the discrepancy is attributed to underreporting or valuation differences rather than missing records.
- **Weighting:** All aggregations utilize cross-sectional weights (**DB090**) to ensure the scaled microdata reflects the total resident private household population.

Plausibility checks

The plausibility assessment for D11 covered both the cash and in-kind components of compensation of employees. The underlying microdata were derived from the EU-SILC variables **PY010G, PY020G and PY021G**, while the aligned HDA aggregates were taken from the household-level variables **D11_HDA** and **D11inkind_HDA**.

The reconciliation results indicate close alignment between the raw microdata and the national-accounts-consistent HDA aggregates. The weighted raw cash component (PY010G) amounted to **EUR 11,968.8 million**, and the raw in-kind component (PY020G + PY021G) to **EUR 81.3 million**, giving a combined raw D11 total of **EUR 12,050.1 million**. The aligned HDA totals were **EUR 12,636.8 million** for the cash component and **EUR 90.2 million** for the in-kind component, resulting in a total D11_HDA of **EUR 12,726.9 million**. The corresponding alignment factors were **1.0558** for cash, **1.1092** for in-kind, and **1.0562** for total D11 — indicating that only a moderate upward adjustment was required.

Technical validation checks did not reveal any anomalies. No negative values were present in the raw or aligned wage components, and no zero-base inconsistencies were observed (i.e. no positive HDA values were assigned where the underlying PY010G/PY020G/PY021G observations were zero).

A consistency check against the self-defined activity-status variable **RB210** identified some cases where individuals reported a status other than “employed” while still receiving positive D11 income. These cases were retained without adjustment, as **RB210 is a subjective self-classification** and does not necessarily correspond to annual employee income as recorded in PY010G.

The recipient structure appeared plausible. **51.50% of persons** had a positive raw D11 amount, while **70.24% of households** recorded a positive D11_HDA value, consistent with the broad coverage expected for compensation of employees.

Overall, the plausibility checks confirm that the D11 allocation — including both cash and in-kind components — is technically robust, internally consistent, and substantively credible for use in the HDA compilation.

D121 – EMPLOYERS’ ACTUAL SOCIAL CONTRIBUTIONS

The microdata source for employers’ actual social contributions is EU-SILC indicator PY030 (Employers’ social insurance contributions). The micro/macro coverage rate is 99%.

Due to the high coverage and close alignment between micro and macro data, the gap was distributed proportionally using **Method A**.

D122 – EMPLOYERS’ IMPUTED SOCIAL CONTRIBUTIONS

Methodological Approach

Since no direct microdata source for employers' imputed social contributions exists in EU-SILC, a **top-down imputation** (Method B) is applied. This component represents benefits provided directly by employers to their employees (the counterpart of which is recorded in **D.612**). In line with the economic mechanism where such contributions typically scale with remuneration, the synchronized wage variable **PY010G_adj** (already reconciled with National Accounts) is used as the distributional proxy.

Gap allocation

As this is an NA-only item without a direct SILC equivalent, **Method B** is applied:

- No scaling of existing micro-values is performed.
- The National Accounts total is distributed across households proportionally to their share of the adjusted wage fund (PY010G_adj).
- This approach ensures that the imputed contributions follow the macro-consistent distribution of wages, while the aggregate sum matches the NA benchmark by construction.

Technical implementation

The National Accounts total for **D122** is allocated using a weight-based approach. For each individual record, the value is calculated as:

$$D122i = Target_{D122} \times (\frac{Py010G\_adj_i}{\sum PY010G\_adj})$$

Households with no wage income receive a zero allocation.

To ensure full accounting symmetry within the HDA framework, the identical distribution is applied to **D612** (Social benefits from unfunded schemes), as it represents the counterpart of the imputed contribution. The final weighted total (using **DB090**) aligns with the National Accounts figure with a 0.0% discrepancy.

Quality Assurance and Validation

Pre-scaling validation includes:

- non-negativity checks for PY030G and wage variables;
- verification that D121 applies only to wage earners;
- internal consistency between D122 and the D11 structure;
- coverage checks comparing weighted SILC totals to NA benchmarks.

Post-scaling validation confirms that:

- weighted totals for D121 (and D611) match the NA target;
- weighted totals for D122 (and D612) match the NA target;
- the resulting D61 corresponds exactly to the combined NA benchmark.

No adjustments or exclusions beyond proportional scaling were required.

D41 - INTEREST

Interest paid

Initially, interest paid (excluding FISIM) is allocated into three main types based on national accounts data:

- Mortgage interest - **75%**
- Consumer loan interest - **16%**
- Other interest - **9%.**

For **mortgage interest**, the initial microdata source is the **EU-SILC** survey variable **HY100G**, but its coverage rate is only 19% of total mortgage interest. To improve coverage and consistency with National Accounts, HFCS mortgage-related information was integrated into the allocation framework.

Technical implementation

D41use was estimated using **fused_data_HFCS_SILC.csv** as the micro-level source and written back to **HDA_Master_HH.csv** at household level using **id_ms**.

Three components were compiled separately:

- **D41use_maja** — housing loan interest
- **D41use_pater** — consumer credit interest
- **D41use_other** — other loan interest

Housing loan interest

The allocation key combines information from both EU-SILC and HFCS:

- HY100G
- HB1701 × HB1901 / 100
- HB1702 × HB1902 / 100

This combined approach substantially improves coverage while preserving the underlying distribution.

Consumer credit interest

The allocation key is based on HFCS consumer-credit variables:

- HC0802 × HC0902 / 100
- HC0801 × HC0901 / 100

Other loan interest

Direct micro-level information was not available for this component. Therefore, **Method B** was applied, using HFCS loan-related expenditure variables as proxy allocation indicators:

- HB37011 × HB39011 / 100
- HB37012 × HB39012 / 100
- HB37021 × HB39021 / 100
- HB37022 × HB39022 / 100

Each component was calibrated separately to its corresponding National Accounts target using official survey weights (**DB090**).

The final aggregate was defined as:

$$D41use\_HDA = D41use\_maja + D41use\_pater + D41use\_other$$

The resulting variables were written to **HDA_Master_HH.csv**:

- D41use_maja
- D41use_pater
- D41use_other
- D41use_HDA

Validation

Initial micro–macro coverage reached **61.4% overall**, with component coverage ranging from:

- Housing interest: 48.7%
- Consumer credit interest: 108.0%
- Other interest: 84.8%

Separate calibration factors ensured exact reconciliation with National Accounts totals (zero residual difference).

The final distribution remained economically plausible without excessive upper-tail concentration:

- Top 1% share: 13.2%
- Top 10% share: 53.2%

for total **D41use**.

The allocation was constrained by an income/resource-based household-level cap. The cap was based on available income-related HDA components, such as wages, self-employment income, property income received and other relevant income proxies. This was introduced to avoid implausibly high interest payments for households with very low observed income capacity. Any amount restricted by the cap was redistributed to other eligible households, preserving the national accounts totals for housing loans, consumer credit and other loans.

Interest received (D41rec)

The observed microdata source for interest received is the administrative **Code 19** amount, available in the household file as **HY090G_kods19**.

This source provides approximately **50% micro–macro coverage**, implying that a substantial residual component remains to be allocated.

The observed administrative component was preserved unchanged, while only the uncovered residual was modelled.

To estimate the residual component, auxiliary household-level financial asset information from the fused survey dataset was used.

Two variables were selected as indicators of potential interest receipt:

- **HD1210** — balance of deposits and savings accounts
- **DA2103** — investments in debt securities

These variables were not treated as direct measures of interest income but as auxiliary indicators identifying households more likely to receive interest income outside the observed administrative source.

Methodological approach

D41rec was estimated using a two-step allocation framework.

First, the observed administrative component (**HY090G_kods19**) was retained unchanged and treated as a fixed observed block of the distribution. This component covered approximately 50% of the National Accounts benchmark.

Second, only the remaining gap to the National Accounts target was imputed.

The uncovered component was allocated exclusively to households without observed administrative interest income.

An auxiliary household-level allocation signal was constructed using deposits, savings accounts, and debt-security investments. The signal was designed to identify households more likely to receive interest income while reducing excessive influence from very large asset holdings.

Method B was applied, using household financial assets as auxiliary proxy indicators.

The allocation was performed only outside the observed administrative block, ensuring that the original administrative distribution remained fully preserved.

The final estimate was defined as:

$$D41rec\_HDA = \textbf{observed administrative block} + \textbf{imputed component}$$

The final weighted aggregate was calibrated to match the National Accounts target exactly.

Plausibility checks

The final distribution remains highly concentrated, reflecting primarily the concentration structure observed in the administrative data rather than the imputation procedure.

The observed administrative component shows:

- Top 1% share: 51%
- Top 10% share: 88%

confirming a strong concentration of received interest income.

In contrast, the imputed component is distributed substantially more broadly:

- Top 1% share: 3.35%
- Top 10% share: 19.19%

These results indicate that the imputation procedure did not artificially amplify upper-tail concentration.

Instead, it preserved the concentration structure observed in administrative data while allocating the uncovered component across households without observed administrative interest income.

D41 FISIM

Methodological approach

FISIM was allocated proportionally to the already compiled household interest flows at household level.

For **interest use**, total FISIM on loans was allocated according to the structure of **D41use_HDA**. For **interest received**, total FISIM was allocated according to the structure of **D41rec_HDA**.

In both cases, the same general principle was applied:

$$\boldsymbol{FISIM_{HDA} = D41_{HDA} \times k}$$

where **k** is the calibration coefficient obtained as the ratio of the relevant aggregate FISIM target to the weighted total of the corresponding D41 component.

This proportional approach was chosen to ensure full consistency between interest flows and FISIM adjustments at household level, while avoiding implausible results from more detailed subcomponent splits.

As a result, the household distribution of FISIM follows the same structure as the already allocated D41 flows, and the combined totals after FISIM adjustment remain fully consistent with the corresponding National Accounts aggregates.

Plausibility

Because FISIM was allocated strictly proportionally to the corresponding D41 distributions, the concentration profile remained unchanged. The FISIM adjustment therefore did not introduce additional distortions into the household distribution but only scaled the already established D41 structure to the required macro totals.

D421 - DIVIDENDS

Data Source Assessment and Justification

A comparative assessment of dividend income revealed substantial discrepancies between survey-based estimates (EU-SILC HY090G) and administrative tax records. The weighted SILC total amounted to **117 972 thousand EUR**, corresponding to **20%** of the National Accounts (NA) benchmark. This confirms the well-known limitation that capital income—particularly dividends—is substantially under-reported in household surveys.

In contrast, administrative tax records captured **1 882 075 thousand EUR** in dividends (codes **1011** and **3011**) prior to outlier treatment. Relative to the NA benchmark, this corresponded to **317% over-coverage**, primarily driven by a small number of exceptionally high dividend declarations. When combined with survey weights (DB090), these extreme values disproportionately inflated the aggregate totals and masked the underlying distributional structure.

Methodological Approach

Although dividend income is inherently skewed, the principal methodological issue arises from the interaction between extreme observations and survey weights, which magnifies the impact of rare high-value records

Allocation Logic

- **Administrative baseline**: Dividend recipients and amounts are identified directly from the State Revenue Service (SRS) records using income codes **1011** (taxable dividends) and **3011** (tax-exempt dividends). Combined, these administrative sources provide complete micro-level coverage of the National Accounts **D421R (dividends)** aggregate.
- **Outlier mitigation:** A capping threshold of **EUR 333 000 (≈p99.5)** was applied to reduce the influence of extreme values. Two records exceeded the threshold and were capped.
- **Gap reconciliation:** After capping, the remaining gap relative to the NA benchmark was closed via a uniform scaling factor (**k ≈ 0.73**), ensuring exact macro–micro alignment while preserving the overall distributional pattern.

Technical Implementation

- **Data Processing and Linking:** Administrative records were linked to the SILC sample via anonymised personal identifiers, enabling the identification of actual dividend recipients in SILC.
- **Gap allocation and scaling**: Outlier-adjusted dividend amounts were multiplied by *k*, ensuring exact reconciliation with the NA aggregate.
- **Survey weights**: Official SILC survey weights (**DB090**) were retained without recalibration, preserving the demographic representativeness of the SILC sample.

Macro alignment was achieved entirely through income value adjustments (capping + scaling), without altering the weighting structure.

Plausibility checks

Administrative dividends were introduced because survey-based coverage was insufficient for HDA purposes. Although “paid dividends” differ from the accrual concept in ESA 2010, they were considered a substantially better proxy under **Method B** for the recipient distribution.

After capping, weighted totals fell from **EUR 1 882.1 million** → **EUR 808.1 million**, and subsequent scaling aligned the value with the NA target (**EUR 593.7 million**).

Recipient Profile

The validation results indicate that the final distribution is economically plausible. Dividend recipients remain a relatively small share of the population, with **3.38%** of households receiving a positive D421 amount. At the same time, the allocation preserves the expected concentration of dividend income among the highest-income recipients.

Concentration among positive recipients (individual level):

- **Top 1%** of positive individuals: **40.58%** of total
- **Top 5%:** 59.11%
- **Top 10%:** 77.68%

Concentration at the household level:

- **Top 1%:** 40.58%
- **Top 5%:** 61.79%
- **Top 10%:** 79.75%

This concentration is economically plausible and aligns with international evidence on dividend distribution.

No zero-consistency problems were observed, meaning that no positive D421 amount was assigned where the underlying proxy base was zero. No negative values were generated at either person or household level.

Sensitivity analysis was carried out using alternative capping thresholds (no cap, p99.0, p99.5 and p99.9). The p99.5 specification was retained as the preferred option because it reduced the influence of extreme observations while preserving the strongly top-concentrated profile expected for dividend income. The selected specification therefore represents a balanced compromise between limiting undue outlier influence and maintaining a realistic distributional structure.

Overall Assessment

The final D421 allocation:

- is fully consistent with national accounts aggregates;
- improves the coverage of dividend income through the use of administrative microdata;
- limits the influence of extreme administrative observations through the application of a household-level cap;
- preserves the highly concentrated distribution that is characteristic of dividend income; and;
- and passes all substantive and technical plausibility checks.

The resulting distribution is **appropriate for use within the HDA framework** and represents a significant improvement over survey-only information.

D44 - OTHER INVESTMENT INCOME

Methodological approach

Property income attributable to insurance policy holders, pension entitlements and collective investment fund shareholders represents a relatively small but conceptually important component of household property income and gross disposable income (B6g). Direct information on these income flows is not available in EU-SILC microdata. Therefore, Method B is applied, using information on underlying financial assets and entitlements from the HFCS survey as distributional proxies.

The allocation is performed at household level and calibrated to National Accounts (NA) totals. HFCS variables are used exclusively to determine the relative distribution across households, while aggregate amounts are constrained to the corresponding National Accounts benchmarks.

D441 Investment Income Attributable to Insurance Policy Holders

Methodological approach

Investment income attributable to insurance policy holders mainly relates to life insurance products with a savings component. As these income flows are not directly observed in EU-SILC, Method B is applied.

Distributional proxy

The allocation is based on household holdings of life insurance products recorded in the HFCS survey.

Insurance products are identified using:

PFA090x = LV_INSUR

The corresponding asset values are obtained from:

PFA0801–PFA0807

Household-level proxy values were constructed by summing the relevant insurance asset values across all household members.

Allocation procedure

The allocation follows four steps:

- Household-level insurance assets are aggregated.
- Survey weights were incorporated when calculating household shares.
- Relative household shares are calculated.
- The National Accounts total for D.441 were distributed proportionally to these shares.

The resulting household-level variable is:

D441_HDA

Validation

The final allocation reproduces the National Accounts aggregate exactly. No households without life insurance assets participated in the allocation.

Concentration indicators show:

- Top 1% share: 11.7%
- Top 10% share: 41.3%

These results are consistent with the expected concentration of life-insurance-related financial assets.

D442 Investment Income Attributable to Pension Entitlements

Methodological approach

National Accounts distinguish three separate pension-related investment income components:

- State-funded pension schemes (Level 1 and Level 2)
- Voluntary private pension funds (Level 3)
- Life pension products managed by insurers

As these components represent different institutional arrangements and ownership structures, each component is allocated separately before constructing the final D.442 aggregate.

Distributional proxy

HFCS pension asset information is used as the allocation key.

For each household member:

- Pension scheme type is identified using **PFA090x**.
- Asset values are taken from the corresponding **PFA080x** variables.

Component 1: State-funded pension schemes (Level 1 and Level 2)

Eligible observations:

- **PFA090x = LV_LEVEL_1**
- **PFA090x = LV_LEVEL_2**

Proxy:

- Corresponding **PFA080**x asset values

National Accounts target:

54,904 thousand EUR

Output variable:

D442_level1_2

Component 2: Voluntary private pension funds (Level 3)

Eligible observations:

- **PFA090x = LV_LEVEL_3**

Proxy:

- Corresponding **PFA080x** asset values

National Accounts target:

1,901 thousand EUR

Output variable:

D442_level3

Component 3: Life pension products

Direct information on life pension entitlements is not available in HFCS.

In the absence of direct information on life pension entitlements in the HFCS, participation in funded pension schemes (Level 2) was used as the closest available proxy, reflecting the institutional design of the Latvian pension system.

Therefore, the allocation uses the same proxy as the funded pension component:

- **PFA090x = LV_LEVEL_2**

Proxy:

- Corresponding **PFA080x** asset values

National Accounts target:

804 thousand EUR

Output variable:

D442_lifepens

Final D.442 aggregate

The three components are combined:

D442_HDA = D442_level1_2 + D442_level3 + D442_lifepens

National Accounts target:

57,609 thousand EUR

The final aggregate reproduces the National Accounts total exactly.

Validation

Coverage checks indicate strong consistency between pension participation and labour market activity:

- Households with Level 1/2 pension assets: 4,406
- Of these, households with positive employment income (D11_HDA): 3,713

This confirms that pension-related assets are predominantly concentrated among economically active households. This supports the economic plausibility of the selected proxy.

Concentration indicators for the final D.442 allocation:

- Top 1% share: 6.5%
- Top 10% share: 33.3%

The separate treatment of Level 3 pension funds preserves the higher concentration typically associated with voluntary pension savings while maintaining consistency with National Accounts aggregates.

D443 Investment Income Attributable to Collective Investment Fund Shareholders

Methodological approach

Investment income attributable to collective investment fund shareholders is not directly observed in EU-SILC. Therefore, **Method B** is applied using household holdings of collective investment funds recorded in the HFCS survey.

Distributional proxy

The allocation is based on:

DA2102 – Investments in collective investment funds

This indicator reflects both participation in investment funds and the scale of household holdings and therefore provides the closest available proxy for the associated investment income.

Allocation procedure

A direct proportional allocation based on DA2102 generated an excessively concentrated distribution due to a small number of households holding very large investment fund positions.

To reduce the influence of a small number of extremely large HFCS investment holdings while preserving the ranking of households, a logarithmic transformation was applied:

Proxy = log(1 + DA2102)

Only households with positive DA2102 values participate in the allocation.

The transformed proxy is then used to calculate household shares and distribute the National Accounts total.

National Accounts target:

1,150.9 thousand EUR

Output variable:

D443_HDA

Validation

The logarithmic specification substantially reduces excessive concentration while preserving the underlying ownership structure.

Final concentration indicators:

- Positive households: 109
- Top 1% share: 3.2%
- Top 10% share: 13.7%

Compared with the untransformed allocation, the resulting distribution is considerably more stable and less sensitive to extreme HFCS asset values while remaining fully consistent with National Accounts totals.

General consistency checks

For all D.44 components:

- Household-level distributions are derived from observed HFCS asset holdings.
- National Accounts totals are matched exactly.

- Survey weights are incorporated during calibration.
- All allocations are performed at household level.

Separate component-level allocations preserve the institutional structure of the National Accounts while maintaining coherence with observed financial asset ownership patterns in the HFCS survey.

Overall, the D.44 allocation framework combines observed household financial asset information from the HFCS with National Accounts control totals, producing household-level distributions that are both economically plausible and fully consistent with macroeconomic aggregates.

D45 - RENT

Received rent

The allocation is described is section **Leasing of dwellings (B2R2g) and Rent Received (D45)** (see above)

Rent paid

The National Accounts aggregate D45 Use of Land and Other Natural Resources is not directly observed in EU-SILC. Therefore, a **Method B** approach was applied using observable indicators of agricultural activity as distributional proxies.

Additional data sources from the **CSB database** are available:

- The price of agricultural land rental per region (euros per ha, without VAT).
- The area of agricultural land used per region, municipality, and parish (ha).

Average regional land rental prices were multiplied by the utilised agricultural area (UAA) to derive a hypothetical regional rental value assuming full rental of all agricultural land. This benchmark was subsequently used to estimate the proportion of land that is effectively rented at the national level.

To obtain the approximate cost of land rental for one farm in each region, we must first calculate what percentage of the total rental fee, if all UAA were rented, is actually paid for rent, since a portion of the land may be owned by the farm.

Real total land rental fee (D.45use) / Total amount if all UAA were rented

20,758 / 124,988 = 17%

The National Accounts D45 aggregate (EUR 20.8 million) represents only a fraction of the hypothetical rental value that would arise if all utilised agricultural land were rented. The ratio between these two amounts equals approximately **17%**, which was interpreted as the effective share of utilised agricultural land that is rented rather than owner-occupied.

Calculation for the paid rental fee for one farm in each region:

Average UAA per farm (ha) * UAA price (EUR per ha) * 17%

Table 9 Distribution of D45 Use of Land by Region

Estimated average annual land rental payment per farm, thsd. EUR	D45 use (paid rental fee), thsd. EUR	Average rental fee per farm, thsd. EUR
Pierīga	2 652	0.28
Vidzeme	3 604	0.25
Kurzeme	4 244	0.41
Zemgale	6 814	0.61
Latgale	3 445	0.15

Table 9 presents the regional distribution of land use (D45 – paid land rents), showing total expenditure on land rental, the estimated average annual rental payment per farm, and the average rental fee per farm across Latvian regions.

Methodological Approach

The allocation was carried out in two stages:

- Allocation of the National Accounts total across regions;
- Distribution of regional totals to eligible agricultural households.

This approach ensures consistency with the regional structure of the National Accounts aggregate while concentrating the allocation among households most likely to use agricultural land.

Regional Allocation

The National Accounts total for 2019 amounted to EUR 20.8 million.

As a first step, the aggregate was distributed across regions using externally derived regional control totals based on agricultural land-use statistics. The resulting regional targets were treated as fixed calibration constraints throughout the allocation process.

The household allocation was subsequently performed independently within each region. This guarantees that both regional and national totals match the National Accounts benchmark exactly.

Identification of Eligible Households

Eligible households were identified using the EU-SILC person-level variable **PL111.**

A household was considered eligible if at least one household member reported economic activity in **NACE Rev.2 01** (Crop and Animal Production, Hunting and Related Service Activities). This restriction ensures that D45 is allocated only to households with observable agricultural activity.

The identification was performed at person level and subsequently aggregated to the household level using the household identifier.

This procedure identified:

302 eligible individuals;

244 eligible households.

Distributional Proxy

Because no direct information on paid land rent exists in EU-SILC, agricultural activity was used as a proxy for the use of agricultural land.

The household allocation key combines two complementary indicators:

PY050G – gross self-employment income;

HY170G – value of goods produced for own consumption.

PY050G captures market-oriented agricultural activity, while HY170G captures subsistence and semi-subsistence agricultural production. The inclusion of HY170G allows households actively using agricultural land, but reporting little or no self-employment income, to participate in the allocation.

To reduce the influence of extreme observations, square-root transformations were applied to both variables.

The household allocation key was defined as:

$Key_h = \sqrt{\max(PY050G_h, 0)} + 0.5 \times \sqrt{\max(HY170G_h, 0)}$

where:

h denotes the household;

only positive values of PY050G and HY170G are considered.

The weighting applied to HY170G reflects its supporting role in the allocation, while maintaining self-employment income as the primary indicator of agricultural activity.

Calibration

Within each region, the household allocation keys were calibrated to the corresponding National Accounts regional target using household survey weights (DB090).

For region r:

$k_r = NA_r / \Sigma(Key_h \times DB090_h)$

The final household allocation was calculated as:

$D45use_h = Key_h \times k_r$

This calibration procedure ensures exact consistency with both regional and national National Accounts totals.

Validation and Plausibility Checks

Several validation checks were performed.

Of the 244 eligible agricultural households, 207 received a positive allocation.

The remaining households had neither positive self-employment income nor positive own-consumption agricultural production and therefore received no allocation.

All regional control totals were reproduced exactly.

The sum of regional allocations matched the National Accounts aggregate without residual differences.

The resulting distribution remained moderately concentrated:

- Top 1% share: 4.7%;

- Top 10% share: 35.2%.

These results indicate that the allocation is not dominated by a small number of households while still reflecting differences in agricultural activity.

The maximum household allocation remained moderate and no implausible outliers were generated.

Summary

D45 use of Land and Other Natural Resources was allocated using a regionalised Method B approach based on observed agricultural activity. The distribution combines information on self-employment income (PY050G) and own-consumption agricultural production (HY170G), thereby capturing both commercial and subsistence farming households. Regional calibration ensures full consistency with National Accounts aggregates, while the use of square-root transformations reduces sensitivity to extreme observations and produces an economically plausible household-level distribution that is fully consistent with the National Accounts aggregates.

D422 – HIDDEN-PROFIT DIVIDENDS

Methodological approach

No direct microdata source was available to identify recipients of hidden-profit dividends. Therefore, D422_HDA was allocated proportionally to households with positive observed D42_HDA values, assuming that existing dividend recipients are the most plausible recipients of hidden-profit income.

Allocation logic

As this is an NA-only item without a direct SILC equivalent, **Method B** is applied:

- D42_HDA used as a structural donor variable
- Allocation restricted to households with positive observed dividend income
- Separate calibration to the D422 macro target
- No new dividend recipients introduced

Plausibility and validation

The weighted aggregate of D422_HDA matched the macro target exactly (EUR 825.2 million). The allocation covered 206 households with positive observed dividend income.

The final distribution remained strongly top-concentrated:

- Top 1%: 40.58%
- Top 10%: 79.75%

This concentration pattern is economically plausible and consistent with the known distributional structure of dividend income.

D51 TAXES ON INCOME, D613 HOUSEHOLDS' ACTUAL SOCIAL CONTRIBUTIONS & D614 HOUSEHOLDS' SUPPLEMENTARY SOCIAL CONTRIBUTIONS

The EU-SILC variable HY140G (Taxes on income and social contributions) serves as the primary microdata source for these components. As HY140G aggregates several types of taxes and social contributions, it is not possible to separately identify direct income taxes (D51), actual household social contributions (D613), and supplementary contributions (D614) at the microdata level.

To ensure consistency with National Accounts (NA), a combined alignment and decomposition procedure is applied.

Methodological Approach: Linear Scaling and Component Decomposition

Combined National Accounts Target

To reflect the structure of HY140G, the relevant NA aggregates are combined into a single net target defined as:

D51 + D613 + D614 − D61SC,

representing the total net tax and social contribution burden borne by households.

Scaling and Gap Allocation

The initial micro-to-macro coverage rate is 91%, indicating a high degree of alignment between EU-SILC and NA aggregates. The remaining gap of approximately 9% is addressed through proportional allocation across households by applying a uniform linear scaling factor (**k = 1.1027**) to HY140G values.

This approach preserves the observed distributional structure in the microdata while ensuring full consistency with macro-level totals.

Component Decomposition

Following scaling, the adjusted household-level totals are decomposed into the corresponding NA components (D51, D613, D614, D61SC) based on their proportional shares in the NA aggregates.

This ensures that each component is consistently represented at the micro level while maintaining alignment with the overall NA structure.

Validation

A weighted aggregation of the resulting microdata confirms full consistency with National Accounts totals for each component. EU-SILC sampling weights (DB090) are applied without recalibration, preserving demographic representativeness while achieving macro-level alignment through value adjustments.

D59 OTHER CURRENT TAXES

This indicator in national accounts includes several methodologically distinct components:

- Vehicle excise duty;
- Fees for keeping animals;
- Income from the rental of bodies of water and fishing rights (fishing licenses);
- Taxes paid abroad.

Its total share in gross disposable income (B6g) is very small — only 0.35%, which makes the methodological choice less critical.

To ensure accuracy, the distribution is carried out by individual components.

Vehicle operating tax for vehicles which have been registered on natural personsaccounts for approximately 98% of the total D59 use aggregate. The remaining components (e.g., fishing licences, animal ownership fees, and other minor administrative taxes) are considered marginal in magnitude and are not separately observable at the micro level.

Given the absence of direct microdata on tax payments and vehicle-level taxation, a constrained allocation framework is applied.

Methodological Approach

D59 use is allocated using a constrained proxy-based allocation framework (Method B). The approach assumes that vehicle-related tax exposure is primarily driven by vehicle ownership status and ownership intensity.

The allocation is performed at household level using EU-SILC (DB090 weighted) microdata, complemented by harmonised vehicle asset information from HFCS.

The model does not attempt to identify exact vehicle tax payments at micro level but constructs a latent exposure index consistent with observed ownership patterns and the National Accounts aggregate.

The allocation is based on the integrated HFCS_SILC fused dataset.

Only households reporting vehicle ownership are considered eligible for receiving D59 use

HS110 = 1

(i.e., household reports ownership of at least one vehicle).

The allocation key combines a basic vehicle ownership component with a limited intensity adjustment derived from vehicle-related assets:

Key= 1 + 0.05 × √(HB4400) + 0.5 × HB4500_flag

where:

HB4400 = total reported market value of household vehicles

HB4500_flag = indicator for additional transport assets beyond standard passenger cars

The square-root transformation of HB4400 reduces sensitivity to extreme values and ensures diminishing marginal impact of high-value vehicles. HB4400 is used solely as a heterogeneity proxy and is not interpreted as a determinant of tax liability or vehicle count.

Allocation Procedure

The allocation is performed as follows:

- Define National Accounts benchmark for D59 use.
- Identify eligible households (HS110 = 1).
- Compute allocation key for each household.
- Apply EU-SILC household weights (DB090).
- Allocate D59 proportionally to weighted keys.
- Apply a calibration factor to ensure exact consistency with the National Accounts total:

$$\sum_i D\,59_i = D59^{NA}$$

Calibration is performed through proportional scaling of the allocation key without modifying survey weights.

Validation and Quality Checks

The resulting allocation is evaluated using the following checks:

- Top 1% share: **3.89%**
- Top 10% share: **22.36%**

No implausible redistribution across income quintiles is observed, and no extreme allocation concentration arises from vehicle value heterogeneity.

Rationale

Given the dominance of vehicle-related taxation within D59, a vehicle ownership-based allocation is considered the most appropriate proxy in the absence of direct tax payment microdata.

The inclusion of vehicle asset information provides additional heterogeneity within the vehicle-owning population while preserving alignment with macroeconomic constraints.

Overall Assessment

The final D59 use allocation:

- is fully consistent with National Accounts aggregates;
- reflects the dominant role of vehicle ownership in D59;
- applies a constrained proxy-based methodology under data limitations;
- ensures exact macro–micro consistency through calibration;
- preserves stable and plausible distributional patterns;
- and passes all internal validation checks.

The resulting estimates are suitable for use in Household Distributional Accounts and provide a coherent approximation of vehicle-related tax incidence under conditions of limited microdata availability.

D611 USE EMPLOYERS' ACTUAL SOCIAL CONTRIBUTIONS

This component represents the social contributions paid by employers on behalf of their employees. It's a non-cash form of compensation.

The microdata source for this component is the same as for **D121**, since both indicators represent the same economic transaction. Therefore, the distribution of **D611** is followed by the identical methodology used for **D121**.

The total national accounts value for **D611** is distributed proportionally to households using the microdata from the EU-SILC survey, specifically the indicator **PY030** (Employers' social insurance contributions).

Given the high micro/macro coverage rate of **99%** for this indicator, the gap between the microdata and national accounts data is minimal. This justifies the use of **Method A**, which allocates the gap proportionally to the microdata values. This approach ensures that the distribution of these contributions is directly aligned with the distribution of wages and salaries.

D612 USE EMPLOYERS' IMPUTED SOCIAL CONTRIBUTIONS

This indicator reflects social benefits that employers pay directly to their employees, bypassing social security funds. Its distribution is identical to that of employers' imputed social contributions (D122). See D122 for methodological details.

D612 RES EMPLOYERS' IMPUTED SOCIAL CONTRIBUTIONS

The D612res variable represents the total resource flow of household unincorporated enterprise income, including imputed contributions and income flows attributable to self-employment and mixed income activities. It serves as the macro-consistent control total within the HDA framework.

D62 SOCIAL BENEFITS RECEIVED

The micro-to-macro reconciliation for social benefits is performed at the most detailed level possible, in line with best statistical practices. This approach more accurately separates pension and non-pension benefits, as a single EU-SILC variable may sometimes cover both types.

The core microdata source for social benefits received is the EU-SILC survey. National accounts data for benefit types are obtained from state and local government budget execution reports, where information is available according to **EKK codes** (Budget expenditure classification according to the function categories). The macro data are grouped to align with the available microdata distribution.

The overall micro-to-macro coverage for social benefits from the government sector (S.13) is **99.7%**. This high coverage rate indicates that the EU-SILC data provide a strong basis for distributional analysis for this item, as the primary source for both micro and macro data is the State Social Insurance Agency. However, granular analysis of specific components, particularly Private Pension Funds (PY080G), showed significant discrepancies.

The EU-SILC variable **PY080G** captures income from private pension funds based on SRS income codes 1024 and 3024, which include a mixture of pension-related withdrawals and investment income. As such, the variable does not fully correspond to the ESA 2010 definition of private pension benefits (due to differences between ESA 2010 concepts and tax-based administrative definitions), leading to a micro–macro coverage ratio of approximately 363%. Given this conceptual mismatch, Method A is not considered appropriate.

Instead, **Method B** is used, with PY080G serving exclusively as a distributional proxy. To improve conceptual alignment, the allocation is restricted to households where at least one individual is **aged 55 or above**, which is the standard pension eligibility age for private pension funds in Latvia. This approach aims to ensure that NA private pension benefits are assigned only to those households that are realistically eligible to receive such payments.

Allocation Logic

- **Eligibility Filtering (Age 55+):** To ensure conceptual alignment for **PY080G**, a legal eligibility filter was applied using the SILC variable **RB081** (Age). All reported private pension values for individuals under the age of 55 were set to zero for analytical consistency, as 3rd-pillar pensions are legally accessible only from this threshold.

- **Method A (Linear Scaling):** For standard social benefits (e.g., PY100G, PY110G, PY130G, HY050G), where survey definitions closely match NA concepts, the remaining gap is closed using uniform scaling factors ($k \approx 1.02$ for old-age pensions).

- **Method B (Proxy-based Scaling):**
    - **Private Pensions (PY080G):** After the age filter was applied, the remaining micro-total was scaled down using a $k \approx 0.28$ to match the administrative (SRS/VID) benchmark.
    - **Employer Sick Pay (A_lapa):** Since EU-SILC does not explicitly capture employer-paid short-term sick leave, these values were imputed using adjusted employment income (D.11_adj) as a distribution proxy. This proxy is chosen because these benefits are intrinsically linked to employment status and earnings level.
    - **Non-pension Other:** Distributed based on the combined structure of all non-pension benefits (HY050G, PY090G, PY120G, PY140G, HY060G, HY070G).
    - **Other (Residual):** Remaining small gaps, including flows with the rest of the world (S.2), were distributed proportionally to households' total benefit profiles.

Technical Implementation

- **Data Processing and Aggregation:** Calculations were performed at the individual level and subsequently aggregated to the household level.

- **Gap Allocation and Scaling:** Proportional scaling factors k were computed using automated Python procedures, ensuring exact reconciliation of weighted micro totals with NA aggregates.

- **Survey Weights:** Official SILC weights (DB090) were retained without recalibration. All alignment was done by adjusting values, not weights.

Validation: The final results show full alignment between the adjusted microdata aggregates and national accounts totals, with no residual discrepancy at the aggregate level.

Table 11 presents the calibration of social benefits received between microdata and national accounts.

Table 10 Micro-Macro Reconciliation and Calibration Coefficients for Social Benefits

Micro source	Component	Micro value	NA adjusted	koeficient (k)	Method
PY100G	Old_age	1 995 865	2 036 137	1.02	A
PY080G	Private pension	24 621	6 788	0.28	B
PY110G	Survivor	38 450	61 888	1.61	A
PY130G	Disability	226 946	264 723	1.17	A
HY050G	Family	473 525	437 553	0.92	A
PY090G	Unemployment	174 749	138 081	0.79	A
PY120G	Sickness	209 225	199 639	0.95	A
PY140G	Education	15 588	13 561	0.87	A
HY060G	Social excl.	14 774	15 844	1.07	A
HY070G	Housing allow.	13 563	12 433	0.92	A
D11_adj	Employer Sick Pay	12 637 243	134 031	0.01	B
HY050G+PY090G+PY120G+PY140G+HY060G+HY070G	Other non-pension	901 418	19 702	0.02	B
Residual allocation	Other pension and non-pension	3 340 398	143 180	0.04	B
Total			3 483 560		

Plausibility Checks and validation results

The plausibility assessment of the D62 allocation was carried out after completing the distribution procedure. The validation framework consisted of two core elements:

- reconciliation of all D62 components with national accounts control totals;
- plausibility checks focusing on target-group coverage, leakage, income-decile profiles, and internal technical consistency.

Reconciliation with national accounts totals

The reconciliation exercise confirmed full consistency between the allocated microdata and the national accounts targets. For all D62 sub-components included in the allocation process, the final weighted HDA totals matched the corresponding national accounts control totals with effectively zero deviation. This result was obtained for all directly allocated components (HY050G, HY060G, HY070G, PY090G, PY100G, PY110G, PY120G, PY130G, PY140G), as well as for the nationally derived components A_lapa, Other, and PY080G.

The reconciliation results therefore indicate that the D62 module is fully aligned with the macroeconomic totals and that no part of the D62 gap remained unallocated after the implementation of the module.

Residual component

The residual component (D62other_HDA) represented **4.68%** of total D62. This is a relatively limited share and suggests that the bulk of the D62 aggregate was allocated using identifiable micro-level information rather than through residual balancing. Consequently, the final result is not driven by an excessive reliance on mechanically distributed residual amounts.

Coverage and leakage analysis

Coverage and leakage indicators were used to evaluate whether the allocated amounts were concentrated in economically plausible recipient groups

- **The pension-related component** (D62pens_HDA) showed very high coverage, with **99.52%** of households containing at least one person aged 65 or over receiving a positive pension-related amount. Leakage outside this simplified reference group amounted to 19.27%. This result should be interpreted with caution, as the pension block in the present implementation includes not only old-age pensions, but also survivor-related, disability-related and proxy-allocated private pension flows, which may legitimately accrue to households without a 65+ member.
- **Child-related benefits** (HY050G_HDA) showed a high degree of plausibility, with coverage of 91.18% among households with children and leakage of only 3.64% outside this group. This indicates that the allocation is strongly concentrated among the expected recipients.
- **Educational benefits** (PY140G_HDA) displayed weaker alignment with the simplified student indicator used for validation. Coverage amounted to 6.19%, while leakage reached 77.20%. This suggests that the available proxy for student status does not fully identify the relevant beneficiary population and that this component remains one of the weaker parts of the D62 allocation from the plausibility perspective.
- **Sickness-related benefits** (PY120G_HDA) showed coverage of 30.32% among households with at least one person of working age and leakage of 6.70% outside that group. This result is not unexpected, since sickness benefits are episodic in nature and are not received by all households containing working-age individuals. The relatively low leakage nevertheless supports the substantive plausibility of the allocation.
- **The earnings-related sickness benefit** component (A_lapa) showed 100.00% coverage and 0.00% leakage, reflecting the fact that it was allocated directly in proportion to the labour income base (D11_HDA).
- **The private pension component** (PY080G) was allocated using a proxy-based approach and showed 100.00% coverage and 0.00% leakage relative to the defined proxy population. In the present implementation, no separate administrative primary proxy was supplied; therefore, the allocation relied on the observed PY080G structure restricted to persons aged 55 and over.

Distribution across income deciles

As an additional plausibility check, the allocated D62 components were analysed across household income deciles derived from an equivalised income proxy. At the current stage of the overall HDA compilation, these deciles should be interpreted as an interim validation tool rather than as final disposable-income deciles, since the full B6g distribution was not yet completed.

- The decile profile of **social assistance** (HY060G_HDA) is economically plausible. The first decile received **22.31%** of the total amount, the second decile **14.57%,** and the third decile **8.04%.** This indicates a clear concentration in the lower part of the distribution, as would be expected for means-tested or strongly targeted support. At the same time, a non-negligible presence in the middle deciles can be considered plausible given that some underlying elements may not be strictly income-tested.
- **Housing support** (HY070G_HDA) was even more strongly concentrated in the lower-income part of the distribution. A total of **84.79%** of the allocated amount was recorded in deciles 1 to 3, which provides strong support for the plausibility of the allocation of this highly targeted component.
- **The residual component** (D62other_HDA) was not expected to be evenly distributed, as it follows the structure of its allocation base rather than being imposed uniformly across households. Its

distribution was relatively spread across the income range, with the highest share observed in the top decile (13.70%). Given the limited size of the residual component in total D62, its influence on the overall distribution remains moderate.

Technical consistency checks

- No negative values occurred for any D62 components.
- Only **1.26%** of households received zero D62.
- Total D62_HDA matched exactly the sum of its components (machine-precision zero deviation).

These results confirm full internal consistency of the D62 module.

Overall assessment

Taken together, the validation results indicate that the D62 allocation is both technically robust and substantively plausible. Aggregate consistency with national accounts totals was achieved exactly, the residual share remained limited, and the main benefit components were concentrated to a large extent in the expected target groups. The distribution across interim equivalised income deciles also follows an economically credible pattern, particularly for the most strongly targeted components of D62.

The main caveat concerns the educational benefit component **(PY140G**), where the available student proxy appears insufficient to capture the full beneficiary population. In addition, the leakage of the broader pension block outside the 65+ group should be interpreted in light of the broader composition of that block, which includes more than old-age pensions alone.

Subject to these limitations, the plausibility checks support the conclusion that the D62 allocation is suitable for use within the HDA compilation and sufficiently robust for inclusion in the project results.

D62 SOCIAL BENEFITS PAID

The D62use_HDA variable represents the distribution of D612res across households. It is constructed using a micro-level allocation key based on SILC income components, capturing both self-employment income and non-observed entrepreneurial activity.The allocation key is defined at household level as the sum of square-root transformed income components:

$$\boldsymbol{Key_i} = \sqrt{\mathbf{max}\,(\boldsymbol{PY050G_i}, \mathbf{0})} + \sqrt{\mathbf{max}\,(\boldsymbol{B3g\_N1N6_i}, \mathbf{0})}$$

The square-root transformation reduces the influence of extreme values while preserving the relative distribution of entrepreneurial activity across households.

The resulting index is weighted using EU-SILC household survey weights (DB090) and proportionally scaled to match the National Accounts aggregate total (**EUR 451.6 thousand)**. The scaling factor ensures full consistency between micro-level distribution and macroeconomic totals.

The final output is a distributional micro-level estimate consistent with National Accounts concepts of mixed income and household entrepreneurial activity.

D71 USE NON-LIFE INSURANCE PREMIUMS

Methodological Approach

The household distribution of non-life insurance premiums (D71P) was derived using expenditure information from the fused EU-SILC–HBS dataset. As direct information on insurance premiums paid is available in the Household Budget Survey, a direct expenditure-based allocation approach (Method B23) was applied.

National Accounts distinguish four categories of non-life insurance premiums:

- Dwellings insurance
- Health insurance
- Transport insurance
- Other non-life insurance

Household allocation proxies were available for the first three categories:

National Accounts category	**HBS proxy**
Dwellings insurance	COICOP 12.5.2.0
Health insurance	COICOP 12.5.3.2
Transport insurance	COICOP 12.5.4.1 + COICOP 12.5.4.2

Travel insurance expenditures (COICOP 12.5.4.2) were combined with transport insurance expenditures (COICOP 12.5.4.1), as National Accounts provide a single aggregate for transport-related non-life insurance premiums.

The residual "Other non-life insurance" category represented approximately 2% of total D71 premiums and could not be matched to a specific expenditure category in the microdata. Therefore, this residual amount was redistributed proportionally across the dwellings, health and transport insurance components prior to household allocation.

Adjusted National Accounts totals for each insurance category were then allocated proportionally to households reporting positive expenditure in the corresponding HBS category.

For each component, household allocation followed:

$$\boldsymbol{Allocation_i = NA\,Target \times \frac{Proxy_i}{\sum Proxy_i}}$$

where the proxy corresponds to the relevant HBS insurance expenditure.

The resulting household-level variables were:

- D71_dwellings
- D71_health
- D71_transport

The total non-life insurance premiums variable was calculated as:

$$\boldsymbol{D71use\_HDA = D71\_dwellings + D71\_health + D71\_transport}$$

Validation

All three insurance components were calibrated exactly to their adjusted National Accounts targets, and the aggregated household total matched the National Accounts D71 total without discrepancy.

Coverage of the allocation reflected the observed incidence of insurance expenditures in the HBS data:

Component	Positive households
Dwellings insurance	1,451
Health insurance	1,567
Transport insurance	2,837
Total D71	3,602

Distributional concentration remained moderate across all insurance types:

Component	Top 1% share	Top 10% share
Dwellings insurance	6.1%	31.0%
Health insurance	8.0%	27.3%
Transport insurance	6.9%	36.7%
Total D71	7.6%	35.5%

The resulting distributions closely follow observed household insurance expenditure patterns while remaining fully consistent with National Accounts totals. The use of direct expenditure proxies provides a transparent and economically interpretable allocation of non-life insurance premiums across households.

D72 REC NON-LIFE INSURANCE CLAIMS

Methodological approach

The household distribution of non-life insurance claims (D72R) was allocated using **Method D** with the previously distributed household-level **D71use_HDA** as the proxy distribution. No direct microdata source for insurance claims was available at the household level, and the National Accounts aggregate for claims was not split by insurance type in a way that could be matched to the microdata.

Given the close conceptual link between non-life insurance premiums and claims, households with higher non-life insurance premium expenditure were treated as the most plausible recipients of insurance claims. The household distribution of **D71use_HDA** therefore served as the balancing-item proxy for the allocation of **D72rec_HDA**.

Allocation key

The allocation key was defined as the positive part of the household **D71use_HDA** distribution:

$$\boldsymbol{Key_i} = \mathbf{max}\,(\boldsymbol{D71use\_HDA_i}, \mathbf{0})$$

Only households with positive **D71use_HDA** values were eligible for allocation.

The National Accounts total for D72 claims was then distributed proportionally across households and calibrated to match the macro total exactly:

$$D72rec\_HDA_i = NA\,Target \times \frac{Key_i}{\sum Key_i}$$

This approach ensures consistency with the distribution of non-life insurance premiums while avoiding unsupported assumptions about claim incidence at the household level.

Resulting variable

The resulting household-level variable was recorded as:

D72rec_HDA

Validation

Validation included calibration to the National Accounts target, recipient-household count, maximum household allocation, median and mean positive allocations, and concentration indicators.

The final allocation:

- matched the National Accounts target exactly;
- covered **3,602 households** with positive allocations;
- produced a maximum household allocation of **EUR 3,173.24**;
- resulted in a **Top 1% share of 7.55%** and a **Top 10% share of 35.50%**.

The resulting distribution mirrors the household pattern of non-life insurance premium expenditure and provides a transparent and reproducible allocation of insurance claims in the absence of direct microdata.

D75 MISCELLANEOUS CURRENT TRANSFERS

D75rec: Transfers received

The National Accounts aggregate for **D75rec** consists primarily of two components:

- transfers received from abroad and from other households;
- government payments related to project participation and similar activities.

The primary microdata source for transfers received from other households is the EU-SILC variable **HY080G (regular inter-household cash transfers received)**. This variable provides direct micro-level information for part of D75rec but covers only approximately **42.7%** of the corresponding National Accounts benchmark, indicating the need for additional allocation methods.

Methodological approach

D75rec was estimated using a split allocation approach with two components.

The **D752rec** component was allocated using the EU-SILC variable **HY080G** as the primary micro-level proxy and calibrated to the corresponding National Accounts target.

The residual component (D75recOther_HDA) was allocated separately using auxiliary information from the fused **EU-SILC and HFCS** survey dataset, in particular the household-level variables **DI1700 and HG0260**, which capture regular private transfers and irregular financial support from outside the household. **Method B** was applied for this residual part. The residual component was calibrated to the remaining National Accounts target:

$$D75recOther = D75rec - D752rec$$

The final weighted aggregate matched the National Accounts target exactly.

Validation and plausibility checks

The D752rec component showed partial micro support, with HY080G covering approximately 42.7% of the National Accounts benchmark before calibration.

The residual component relied on auxiliary proxy information rather than directly observed recipient data and therefore remains subject to greater uncertainty. Despite this limitation, the final distribution remained economically plausible. The combined D75rec allocation showed moderate concentration (**Top 1%: 7.7%; Top 10%: 31.0%),** avoiding the overly flat distribution observed under alternative specifications.

D75use: Transfers Paid

The national accounts value for **D75use** consists of transfers to other households abroad (D752use) and membership fees or donations to various organizations.

Methodological approach

D75use was estimated using a split allocation approach with two components.

The **D752use** component was allocated using the EU-SILC variable **HY130G** (Regular inter-household cash transfers paid) as the primary micro-level source and calibrated to the corresponding National Accounts target. The coverage rate is **81%**, which justifies using **Method A** to distribute the gap proportionally to households that reported a payment.

The residual component, **D75useOther_HDA**, was allocated separately using the HFCS-based variable **HI0310 (**measures the approximate average monthly amount of material support provided by the household in cash to persons outside the household or to private organisations) from the fused survey dataset. Since the variable is recorded on a monthly basis, values were annualised before allocation and calibration.

To avoid identical allocations for households with identical HI0310 values, a small deterministic tie-break factor was applied. In addition, a B6g-formula-based household-level cap was introduced. The cap was calculated using the available B6g components excluding D75use itself. Amounts exceeding the cap were redistributed to other households with remaining capacity, while preserving the national accounts D75useOther target.

For the residual component, **Method B** was applied, using HI0310 as an auxiliary proxy variable to distribute the remaining National Accounts amount:

$$D75useOther = D75use - D752use$$

The final estimate was defined as:

$$D75use\_HDA = D752use\_HDA + D75useOther\_HDA$$

Both components were calibrated separately using official survey weights to ensure exact consistency with National Accounts totals.

The final weighted aggregate matched the National Accounts target exactly.

Validation and plausibility checks

The **D752use** component showed strong micro support, with **HY130G covering approximately 81.1%** of the corresponding National Accounts benchmark prior to calibration.

The residual component also demonstrated relatively strong auxiliary support, with annualised **HI0310 covering approximately 67.6%** of the residual National Accounts target before calibration.

Calibration coefficients remained moderate for both components (**k = 1.23** for D752use and **k = 1.48** for D75useOther), indicating that no excessive scaling adjustments were required.

The final distribution remained economically plausible. The combined **D75use_HDA** allocation showed moderate concentration (**Top 1%: 5.9%; Top 10%: 30.7%**), while both subcomponents also retained balanced concentration profiles.

These results suggest that the allocation preserves realistic distributional patterns while maintaining exact National Accounts consistency.

D63 SOCIAL BENEFITS IN KIND

Social benefits in kind (**D63**) represent expenditures by the government (S13) and non-profit institutions serving households (S15) on services received by households. The distribution is conducted by type of service using **Method C** (imputation based on socio-demographic profiles).

Health Services (D63R2)

Description

The distribution of health care services (D63R1) follows the "Insurance Value" approach, which is the international standard recommended by Eurostat and OECD. Instead of allocating expenditures based on actual individual usage during the survey year, this method treats the public health system as an insurance scheme where the benefit is the value of the "premium" provided by the state to individuals within specific risk groups (age and gender). This ensures a stable and fair distribution that reflects the social protection provided to all residents.

Data Sources

1. **Macro Data:** National Accounts total for Government Social Transfers in Kind for Health. **Total Target Value:** € 993.576 million.

2. **Profile Data:** Detailed administrative data from the **National Health Service (NVD)** providing average per capita health expenditures.

3. **Micro Data: EU-SILC 2020** (reference year 2019) database, specifically the Person file for demographic variables (Age - RB081, Weights - RB050) and Household file for aggregation (id_ms).

Methodology and Python Implementation

The distribution was executed using a custom Python script to ensure precision and reproducibility. The technical process consisted of the following steps:

1. **Data Harmonization and Cleaning:** The administrative NVD profile (provided in 5-year intervals) was processed using regular expressions (regex) to clean and standardize age labels (e.g., removing whitespace, handling "85+" categories). This ensured a perfect join between the administrative categories and the numerical age values in the SILC database.

2. **Imputation Profile Mapping:** Each individual in the HDA_Base_Person file was assigned a base health expenditure value based on their age (RB081). The script mapped every respondent to their corresponding 5-year age group expenditure level defined by the NVD data.

3. **Weighted Scaling (Calibration to Macro Totals):** To ensure 100% alignment with National Accounts, a scaling factor was calculated.

- The script first calculated the total weighted expenditure using EU-SILC person weights (RB050).
- The initial weighted total was compared to the NA target of € 993.576 million.
- A **scaling factor of 0.9973** was derived, indicating an exceptionally high correlation (99.7%) between the SILC population structure and the administrative health data.
- All individual values were then multiplied by this factor to reach the exact macro target.

4. **Aggregation from Micro to Macro Level:** The final individual health benefits (D63_health_individual) were aggregated at the household level (id_ms) to be integrated into the HDA_Master_HH file. This allows for the calculation of **Adjusted Disposable Income (B.7g)** at the household level.

Plausibility Checks and Robustness Assessment

To ensure the credibility of the imputed distribution of Government Social Transfers in Kind for Health (D63R1), a series of plausibility checks were carried out in accordance with Eurostat/OECD HDA quality guidelines.

Consistency with National Accounts (macro validation)

- The final weighted micro-total matches the National Accounts benchmark of **€ 993.576 million** with a residual difference of **0.00 EUR**.
- The calibration factor (**k = 0.9973**) indicates a near-perfect alignment between the EU-SILC demographic structure and the administrative health-expenditure profile.

Internal consistency (micro validation)

- No negative or missing values were generated at either the individual or household level.
- Age-group consistency was verified: average D63 allocations increase monotonically with age (0–18 < 41–60 < 61–80 < 80+), consistent with the expected pattern of healthcare utilization across the life cycle.
- All age groups contain an adequate number of observations, ensuring stable estimates.

Distributional plausibility

- Households were ranked by total income (deciles), and mean D63 values were examined across groups. The distribution shows a smooth upward gradient, with higher-income households receiving somewhat higher imputed health benefits due to larger household sizes and different age compositions — a pattern consistent with demographic structure rather than income itself.
- Top- and bottom-1% quantile checks revealed no outliers or unrealistic spikes.

Robustness checks (sensitivity analysis)

- Minor adjustments to age-band definitions (±5 years) and alternative cut-offs (e.g., merging 75–80 and 80+) resulted in negligible differences in the final distribution (<1%).
- Recalculation using narrower/wider age bands did not materially change the distributional pattern or macro-alignment.
- The scaling factor remained stable under alternative specifications, demonstrating methodological robustness.

Conclusion

The plausibility tests confirm that the D63 distribution is **statistically coherent**, **economically meaningful**, and **fully aligned with National Accounts**, allowing for reliable integration into the household accounts framework (B6g → B7g).

Education Services (D63R1)

This category is divided into two sub-categories based on the type of education.

- **Macro Total:** €1,220.073 million.
- **Formal Education (86%):** The amount is €1,047.169 million. The distribution uses previously calculated average per-student expenditure for each level of formal education. Appropriate sums is allocated to students living in households. The imputed sums is proportionally adjusted to the national accounts total for formal education.
- **Non-formal Education (14%):** The amount is €172.904 million. This sum is distributed among adults and children who are most likely to be involved in non-formal education (e.g., children's music, art, or sports schools). As there is no direct microdata source, a **Method C** profile is used, based on household composition and likely participation in these activities.

Methodological approach for formal education

D63 education was compiled at the individual level using information from the EU-SILC personal dataset and subsequently aggregated to the household level.

The allocation was based on an external table of annual public expenditure per student by ISCED level and age group. The assignment followed a hierarchical approach:

- for ISCED levels 0–2, expenditure was assigned primarily on the basis of age groups;
- from age 16 onwards, the allocation relied primarily on the EU-SILC variable PE020 to identify the relevant education level;
- where no direct PE020 information was available, the allocation reverted to the corresponding age-based profile.

This produced an initial estimate of education expenditure for each individual. The resulting individual-level amounts were then proportionally calibrated to match the National Accounts total for D63 education.

Finally, the calibrated individual values were aggregated to the household level using the common household identifier *id_ms*, producing:

- **D63_educ_individual** (individual level);
- **D63_educ_HH** (household level).

Calibration

The initial micro-based estimate covered **96.15%** of the National Accounts target.

To ensure full consistency with National Accounts, a proportional calibration factor was applied:

k = 1.0401

This represents only a minor adjustment, indicating that the underlying micro-allocation already provided a very close approximation to the macro benchmark.

Plausibility checks

The final weighted aggregate matched the National Accounts target exactly, and no negative values were generated at either the individual or household level.

The aggregation process was internally consistent: the unweighted sum of individual allocations matched the unweighted sum of household allocations exactly after aggregation by *id_ms*.

The resulting age profile was economically plausible:

- higher average values were observed among school-age children and adolescents;
- the highest average allocations occurred in the 16–18 age group, consistent with participation in upper-secondary education;
- lower or zero values at older ages reflect the fact that the allocation captures primarily formal education participation identified through age and PE020.

The final distribution was dominated by ISCED levels 0–3, while substantially smaller amounts were allocated to ISCED levels 4–8. This is consistent with the underlying structure of educational participation observed in the microdata.

Methodological approach for informal education

D63 Informal Education was allocated at individual level using EU-SILC personal data. The allocation is based on a simple score model combining age-group weights and the PE040 indicator as a proxy for participation or intensity of informal education activities. Children aged 1–15 are included with positive age-based weights to reflect publicly financed extracurricular and informal learning activities, while persons aged 16–64 are allocated according to age weights and PE040 intensity. Persons aged 65+ receive zero allocation.

The resulting individual scores were weighted and proportionally calibrated to match the National Accounts target exactly. Individual values were then aggregated to household level using id_ms. The checks confirm that there are no negative values, the NA target is matched exactly, and individual-to-household aggregation is internally consistent.

Other Social Benefits in Kind (D63R3)

This diverse category accounts for 27% of the total D63. Table 12 presents other social benefits in kind by COFOG.

Table 11 Other Social Benefits in Kind by components

Other social benefits in kind	Private HH
Housing	132 823
Recreation and culture	296 144
Social protection	205 361
OTHER SERVICES	200 863
Total	835 191

Due to time constraints, the residual component of social transfers in kind, D63other_HDA, was allocated using B6g_HDA as the distributional proxy (**D method).** The national accounts target for D63 other was 993,576.138 thousand EUR. Household-level values were derived by applying a proportional calibration factor to B6g_HDA, using EU-SILC household weights (DB090) so that the weighted total matched the national accounts control total.

This approach assumes that the distribution of the residual D63 other component broadly follows the distribution of household disposable income after HDA adjustments. It should therefore be interpreted as a pragmatic proxy allocation rather than a component-specific microdata-based distribution. The main advantage of this method is that it preserves consistency with the national accounts aggregate while providing a complete household-level distribution for template purposes. The limitation is that the true distribution of D63 other may differ from the income-based proxy.

D8 ADJUSTMENT FOR THE CHANGE IN PENSION ENTITLEMENTS

The adjustment for the change in pension entitlements (D8) was estimated at the household level using a distributional proxy based on the underlying social contribution and pension benefit flows available in the HDA master file.

Since a fully separated pension and non-pension split of D61 was not available at the micro-data level, the following household-level proxy was constructed:

D8_proxy_HDA = D611_HDA + D612_HDA + D613_HDA + D614_HDA − D61sc_HDA − D62pens_HDA

This proxy reflects the underlying economic logic of D8. Social contributions and contribution-related components increase households' pension entitlements, while pension benefits reduce previously accumulated pension entitlements. Consequently, positive D8 values are expected primarily among economically active households, whereas negative D8 values are expected among pension-receiving households.

For institutional households, D8 was estimated using the available institutional flows. Since a detailed pension/non-pension split of D61 was not available, social contribution flows were used as a proxy for the active population component, while the pension part of D62 was used for the pension-benefit component. The institutional D8 proxy was therefore calculated as pension-related contribution proxy minus pension benefits. D8 for institutional households **– 420 thousand EUR.** This amount was deducted from the total S14 D8 control total to derive the private-household D8 target used for micro-level calibration.

The resulting proxy distribution was subsequently calibrated to the National Accounts D8 control total. A calibration factor of **0.4392** was therefore applied to the proxy distribution, resulting in a final weighted **D8_HDA** total of **621 550 thousand EUR**, fully matching the National Accounts target.

As part of the validation process, the calibrated D8_HDA distribution was examined. It contained **3,167 households with positive values, 2,832 households with negative values, and 96 households with zero values**. All households with negative D8 values were confirmed to be pension recipients, validating that the negative part of the distribution is correctly linked to pension benefit flows and is fully consistent with the economic interpretation of D8. Furthermore, the maximum difference between the stored D8 proxy and the manually reconstructed proxy was negligible, confirming the internal consistency of the calculations.

This approach preserves the structural distinction between economically active households, whose social contributions increase pension entitlements, and pension-receiving households, whose pension benefits reduce accumulated pension entitlements, while ensuring full consistency with the National Accounts aggregates.

CONSUMPTION GAP DISTRIBUTION METHODOLOGY

METHODOLOGICAL APPROACH

General Approach (Applicable to all COICOP Categories)

Before proceeding with the micro-macro gap distribution, it is crucial to evaluate and, if necessary, perform imputations for household groups or individuals who may be missing or insufficiently represented in the Household Budget Survey (HBS) data, but who are logically expected to have expenditure in the relevant category. This primarily addresses cases where:

- Specific household groups (e.g., the very wealthy) may not be adequately covered.
- Households that logically should have expenditure in a particular COICOP category have no recorded expenditure in the HBS data (e.g., households with children showing no expenditure on children's clothing).

While international guidelines do not set specific coverage thresholds, the following **indicative criteria** will be used to select the most appropriate distribution method, depending on microdata quality. Following micro/macro coverage assessments have been assigned:

- **High Coverage (>70%)**
- **Medium Coverage (30%–70%)**
- **Low Coverage (<30%)**

Low coverage indicates serious discrepancies where the HBS data is likely not representative for the specific category. The gap may be formed by significant definitional differences, major under-coverage issues, or very low reporting rates (e.g., by high-income households).

Methodological Steps

Step 1: Imputation for Missing Population Segments

This step is carried out only for **3-digit COICOP categories** with low coverage or known structural issues suggesting missing household groups.

In such cases, the following approach is applied:

Identification of **"Potential consumers"** and expenditure imputation

Identify households that, based on logical criteria, should have expenditure in the relevant COICOP category, even if none are recorded in the HBS data. Criteria may include: presence of children, demographics, income level, car ownership, type of residence, etc.

- **Define the "Potential consumer" group:** Establish clear criteria for identifying these households within the HBS data.
- **Estimate the imputable amount:**
  - External survey data: use other available surveys (e.g., tourism surveys for travel expenditures, which may cover wealthy households better).
  - Data from Similar HBS Households: If there are HBS households with a similar profile that do report the relevant expenditure, their average expenditure can be used.
  - Expert knowledge: utilize expert knowledge and experience regarding the frequency and amount of consumption for specific goods/services.
- **Impute expenditures to this Group:** The missing expenditure amount is distributed among these identified "potential households." Imputation can be done:
  - Evenly among all these "potential" households.

    - Proportionally based on simplified criteria (e.g., the number of households in the relevant income or demographic groups), if there is a reason to believe that consumption differs across these groups (e.g., allocate higher amounts to higher-income households).

After performing this imputation, the HBS total for the respective COICOP category will be increased. Consequently, the micro-macro gap (between the NA total and the adjusted HBS total) will decrease, and the remaining portion of the gap will be distributed according to the methods described in **Step 2** of this Annex.

Step 2: Gap Distribution

The international guidelines include four methods (**A, B, C, D**) to reconcile microdata with the respective National Accounts (NA) totals. The priority and suitability of their application depend on the HBS data coverage and the availability of supplementary information.

The core principle of gap distribution is to proportionally **add or subtract** the discrepancy from HBS household expenditures to ensure data consistency with National Accounts values. Distribution is performed based on weighting coefficients that reflect a household's expenditure share in the relevant COICOP category. The choice and priority of the gap distribution methodology depend on the HBS data coverage and representativeness within each COICOP category. The distribution will primarily use the 3-digit COICOP categories, and if a deeper analysis of the gap's "nature" is necessary, the COICOP 4- or 5-digit level will be utilized.

If microdata coverage is insufficient, alternative distribution methods (**B, C, or D**) will be used, such as:

- Utilizing similar COICOP categories.
- Distribution based on income groups or household types.
- Uniform distribution.

Method A: Proportional Distribution for High Coverage (>70%)

- Principle: Distribute the gap proportionally to households' actual expenditures in that category.
- Assumption: Microdata adequately reflects the consumption of the entire population in this category.
- Condition: If a household initially has zero expenditure in the specific COICOP category, it will not be allocated any part of the gap.

Application for Medium Coverage (30%–70%): Although Methods B or C are generally preferred for medium coverage, **Method A may be used** if representativeness checks (e.g., verifying that expenditures are not concentrated in only a few households) provide confidence that the HBS distribution still accurately reflects the consumption pattern in the population (e.g., for universal consumption categories where the main issue is under-reporting, such as food). If bias is detected, alternative methods (B or C) will be applied.

Application for Negative Gap (HBS > NA): If the gap is negative (HBS > NA), Method A (proportional reduction) will always be applied, assuming that over-reporting is proportional to consumption.

Method B: Substitution using Distribution of other Income or Consumption Components

This method is prioritized if direct microdata is not available for the specific item, or if coverage remains low or data is not representative after the imputations in Step 1. It replaces the missing information by using the distribution of another income or consumption component in the HBS data as a proxy.

Application: The remaining gap is distributed using weighting coefficients derived from:

- **Best Proxy**: Another very similar COICOP category (e.g., using the distribution of "gas" consumption to distribute the gap for "electricity"). This relies on the assumption that the consumption structure is similar across these related categories.

- **Next Level:** A broader, aggregating category (e.g., using the total distribution of "Clothing and footwear" (COICOP 03) for the gap of "Clothing" (COICOP 03.1).

Method C: Imputation using Exogenous Socio-Demographic Data

This method is used if Method B is not applicable or does not yield satisfactory results. It imputes the missing distribution of information based on **exogenous data** (e.g., socio-demographic information) available at the individual or household level.

- **Grouping:** HBS households are grouped into **income quintiles** or defined **household types**.

- **Determining Distribution Proportions between Groups:** Based on general consumption patterns or expert judgment, the share of the remaining COICOP gap to be allocated to each income quintile or household type will be determined. The rationale for choosing these proportions (e.g., why the gap is distributed precisely this way across income quintiles) will be documented.

- **Gap Distribution within the Group:** Once each group has received its share of the gap, this amount is distributed among the individual households within that group:

  - **Evenly:** Each household in that group is allocated an equal share of the gap.

  - **Proportionally to existing HBS expenditures/income within the group:** The gap can be distributed proportionally to households' original expenditures or income within the specific group, if this yields a credible distribution (e.g., allocating larger amounts to higher-income households).

Method D: General Imputation (last resort)

This method is used only when no information is available (neither from other components nor socio-demographic data), and the previous methods are not applicable. Imputations are performed such that the inclusion or exclusion of the component does not affect the distribution results of the main indicators. Most often, this involves distributing the gap **proportionally to total household consumption expenditure**, assuming that households with higher total consumption are also likely to consume more of this hard-to-distribute component.

This method is used only as a last resort.

Quality Control

Several checks will be performed to ensure the quality and plausibility of the final data set:

- Does the expenditure of any household in a specific category exceed its total income after the gap distribution?

- Have the distribution indicators (e.g., Gini coefficient, quintile ratios) changed drastically after the correction? The goal is to reconcile the level with the NA total, not to completely distort the original HBS distribution.

- Are the results economically logical (e.g., is the consumption of luxury goods genuinely concentrated among the wealthiest households)?

Documentation: The reasoning for every decision made (why one method was chosen over another), the proportions used, and all assumptions will be clearly documented.

COICOP 01: Food and Non-Alcoholic Beverages

Overall Micro-Macro Coverage Rate: 65%:

- Purchased Food and Non-Alcoholic Beverages: 63% (Medium Coverage)
- Self-Produced Food and Non-Alcoholic Beverages for Own Consumption: 106% (High Coverage)

Distribution Strategy

This category is specific because the NA total includes not only goods purchased on the market but also household **self-produced products for own consumption**. Since the NA data allows these two components to be separated, the distribution approach is split into two parts, applying a method appropriate to each component's coverage level.

Self-Produced Food and Non-Alcoholic Beverages (own Consumption)

The value of self-produced food and beverages is separated from the total NA sum.

Methodology: Method A (proportional Distribution) will be used.

Application: The gap is distributed proportionally among those HBS households that report own consumption ("potential producers/consumers"). If HBS lacks precise data on the specific consumption item but holds information on the existence of self-production (e.g., farm, garden, mixed income (B3g)), this will serve as the basis for identification. The gap is allocated proportionally to the volume of own consumption already reported in the HBS by these households.

Purchased Food and Non-Alcoholic Beverages

- **Coverage:** The coverage for purchased food and beverages is **medium (63%)**, suggesting potential under-reporting, a common issue in universal consumption categories.
- **Procedure:** A detailed representativeness check on HBS data for purchased food will first be performed. This check ensures that expenditures are not overly concentrated in only a few households and that the distribution reflects the entire population's structure.

If confidence is gained after the representativeness check that the HBS distribution accurately reflects the consumption pattern in the population (e.g., if the main issue is assumed to be universal under-reporting), **Method A** can be utilized.

Otherwise (if bias or major structural issues are detected), **Method C** is applied to allocate the gap based on socio-demographic factors (e.g., income quintiles or household type).

COICOP 02: Alcoholic Beverages, Tobacco, and Narcotics

Overall Micro-Macro Coverage Rate: Low (21%)

- Self-Produced Alcohol: 100% (high coverage)
- Purchased Alcohol and Tobacco: 25% (low coverage)

COICOP 02.1: Alcoholic Beverages

Self-Produced Alcohol (Own consumption)

Methodology: Method A (proportional distribution) will be used.

The gap for self-produced alcohol is distributed proportionally among those HBS households that report self-production or are identified as potential producers/consumers (e.g., rural households, households with mixed agricultural income (B3g)). The gap is allocated proportionally to the volume of own consumption already reported in the HBS by these households.

Purchased Alcohol (legally)

The low coverage (23%) for purchased alcohol indicates significant under-reporting in the HBS. The gap is distributed using **Method C**, primarily based on supplementary data from the Centre for Disease Prevention and Control (SPKC).

Data Sources for Distribution Profiles:

- **SPKC Study ("Latvian Population Health-Influencing Habits Study, 2022"):** Data on the frequency of alcohol use by gender and age groups (%) will be utilized.

Respondents with an alcohol consumption frequency of at least once per month (including "2–3 times a month," "once a week," and "2–3 times a week") is counted. This helps establish a more precise "conditional consumer count" reflecting the demographic profile of households that contribute a significant share to total alcohol consumption and, potentially, the micro-macro gap. Frequencies such as "a few times a year" is not considered, as they do not provide a representative insight into regular and significant consumption.

- **CSB Official Data:** Official data on the Latvian population size and demographic distribution (by gender and age) for 2022 will serve as the basis for extrapolating the SPKC study percentages to the total population.

Distribution of Purchased Alcohol Gap: The remaining gap for purchased alcohol is distributed proportionally to the **estimated number of regular consumers** within the HBS households, grouped by their respective gender and age profiles (as derived from the SPKC and CSB data).

Illegal Alcohol

The total value of illegal alcohol is treated as part of the total purchased alcohol gap and generally distributed using the **same methodology** (Method C, based on regular consumers by gender and age).

Types of alcohol analyzed:

• Spirits

• Wine

• Beer

Table 12 Beer gap distribution by gender and age group

Age group	Men (%)	Women (%)	Number of men, 2022	Number of women, 2022	Estimated Number of Consumers (men)	Estimated Number of Consumers (women)	*Structure men*	*Structure women*
15–24	53,5	11,3	90 403	85 327	48 366	9 642	*0.10*	*0.02*
25–34	65,6	14,3	119 369	110 747	78 306	15 837	*0.16*	*0.03*
35–44	67,7	21	130 780	126 586	88 538	26 583	*0.18*	*0.05*
45–54	62,1	18,8	124 169	132 681	77 109	24 944	*0.16*	*0.05*
55–64	53	12,4	118 774	145 306	62 950	18 018	*0.13*	*0.04*
65–74	42,3	6,6	80 198	128 068	33 924	8 452	*0.07*	*0.02*
Total					389 193	103 476		

Table 13 Wine gap distribution by gender and age group

Age group	Men (%)	Women (%)	Number of men, 2022	Number of women, 2022	Estimated Number of Consumers (men)	Estimated Number of Consumers (women)	*Structure men*	*Structure women*
15–24	21.6	20.9	90 403	85 327	19 527	17 833	*0.06*	*0.05*
25–34	19.3	35.3	119 369	110 747	23 038	39 094	*0.07*	*0.12*
35–44	21.5	41.1	130 780	126 586	28 118	52 027	*0.09*	*0.16*
45–54	14.7	36.8	124 169	132 681	18 253	48 827	*0.06*	*0.15*
55–64	11.1	30.1	118 774	145 306	13 184	43 737	*0.04*	*0.13*
65–74	8.4	14.8	80 198	128 068	6 737	18 954	*0.02*	*0.06*
Total					108 856	220 472		

Table 14 Spirits gap distribution by gender and age group

Age group	Men (%)	Women (%)	Number of men, 2022	Number of women, 2022	Estimated Number of Consumers (men)	Estimated Number of Consumers (women)	*Structure men*	*Structure women*
15–24	36.3	11.8	90 403	85 327	32 816	10 069	*0.08*	*0.02*
25–34	40.6	15.4	119 369	110 747	48 464	17 055	*0.11*	*0.04*
35–44	55	20.3	130 780	126 586	71 929	25 697	*0.17*	*0.06*
45–54	50.5	19	124 169	132 681	62 705	25 209	*0.15*	*0.06*
55–64	52.2	12.6	118 774	145 306	62 000	18 309	*0.14*	*0.04*
65–74	46.9	13	80 198	128 068	37 613	16 649	*0.09*	*0.04*
Total					315 527	112 987		

Methodological calculation steps:

For each sex and age group, a "**conditional number of consumers**" was calculated by multiplying the total population in that demographic group by the alcohol consumption prevalence rate (converted to a decimal). This was used to construct a distribution structure by age and sex.

The same methodology was applied to allocate the gap for wine and spirits. Since publicly available data on alcohol consumption prevalence by region are not available, the distribution was based on the number of newly registered patients diagnosed with "Alcohol dependence" and "Alcohol-induced psychosis" per 100,000 inhabitants (SPKC, 2022: https://www.spkc.gov.lv/lv/datu-vizualizacijas-par-alkohola-lietosanu.)

These two diagnoses reflect different but related manifestations of alcohol use intensity at the regional level. Their combined impact was summed for each region. The resulting values were then normalized to obtain a unified percentage-based regional distribution. This combined indicator serves as an indirect but justified proxy for alcohol consumption intensity across Latvian regions by sex, enabling a more targeted gap allocation in the absence of direct regional and gender-specific consumption data (see table below).

Table 15 Regional Distribution of Alcohol-Related Diagnoses and Consumption Intensity Proxies by Sex

Region	Women Diagnoses	Men Diagnoses	Women Structure	Men Structure
Rīga	113	405.4	0.18	0.14
Pierīga	69.4	311.7	0.11	0.11

Vidzeme	83.9	507.5	0.13	0.17
Kurzeme	121	597.4	0.19	0.20
Zemgale	112.7	377.2	0.18	0.13
Latgale	130.9	747.9	0.21	0.25

For the distribution of the gap for each type of alcohol, using the micro HBS data on each household's person's age, gender and region, combined weights are created for each household.

COICOP 02.3 Tobacco (Cigarettes, Cigars, etc.)

Overall micro-macro coverage rate: Low (25%).

The National Accounts (NA) data allows for the separation of illegally and legally purchased tobacco. HBS coverage for legally purchased tobacco is 32% (medium-low). The significant remaining gap is distributed using **Method C**, based on supplementary data from the Centre for Disease Prevention and Control (SPKC).

Tobacco Illegal is distributed using the same method as legally purchased tobacco.

To accurately estimate the distribution of tobacco consumption, two main data sources are used:

- **SPKC study "Latvian Population Health-Influencing Habits Study, 2022":** This provides data on smoking prevalence (percentage of smokers) by demographic group (sex and age, 15–74 years). *Available at:* https://www.spkc.gov.lv/lv/datu-vizualizacijas-par-smekesanu
- **Official CSB data** on the population size and demographic breakdown (by sex and age) in Latvia in 2022: These data are used to extrapolate the SPKC prevalence rates to the total population.

Methodological calculation steps:

First, a "**conditional number of smokers**" was calculated for each sex and age group by multiplying the total population in that group by the smoking prevalence rate from the SPKC study. This figure is theoretical and reflects the distribution of smoking across demographic groups, serving as a weighted value rather than an exact count of smokers.

Example:
Men aged 15–24 = Number of men aged 15–24 × Smoking prevalence among men aged 15–24

Next, the proportion of each sex and age group's conditional number of smokers was calculated relative to the total conditional number of smokers (men + women).

Unlike alcohol, there are no publicly available or sufficiently detailed regional data on tobacco consumption. Therefore, educational attainment is used as the third demographic dimension (in addition to sex and age).

Smoking habits are often more closely linked to education level than to geographic region. Thus, based on available data, education level was selected as the third imputation dimension.

Table 16 Proportion of daily smokers 15-74 years old, %

Age group	Men (%)	Women (%)	Number of men, 2022	Number of women, 2022	Estimated Number of Consumers (men)	Estimated Number of Consumers (women)	*Structure men*	*Structure women*
15-24	38.7	12.7	90 403	85 327	34 986	10 837	*0.09*	*0.03*
25-34	44.2	15.7	119 369	110 747	52 761	17 387	*0.13*	*0.04*
35-44	46	22.1	130 780	126 586	60 159	27 976	*0.15*	*0.07*
45-54	49.7	22.3	124 169	132 681	61 712	29 588	*0.15*	*0.07*
55-64	43.1	15.1	118 774	145 306	51 192	21 941	*0.13*	*0.05*
65-74	30.5	8.2	80 198	128 068	24 460	10 502	*0.06*	*0.03*
Total					285 270	118 230		

Table 17 Daily smoking rate 15-74 years by education level, %

Education level	Men (%)	Women (%)	Number of men, 2022	Number of women, 2022	Estimated Number of Consumers (men)	Estimated Number of Consumers (women)	*Structure men*	*Structure women*
Primary education	49,1	28,9	136 308	124 108	66 927	35 867	0.15	0.08
General secondary	46,9	21,3	318 818	308 251	149 526	65 657	0.34	0.15
Special secondary	47,7	16,7	85 762	119 263	40 908	19 917	0.09	0.05
Higher	22,2	7,5	170 768	307 203	37 910	23 040	0.09	0.05
Total					295 272	144 482		

To identify HBS households, their sex, age group, and education level are used. For gap allocation, microdata from the HBS survey on each household's age, sex, and education level are used to create combined weights for each household. These weights reflect the household's relative likelihood of tobacco consumption based on demographic characteristics and serve as the basis for imputing both legal and illegal tobacco consumption.

COICOP 02.4 Narcotics

Narcotics consumption is not available in HBS microdata. Therefore, data is imputed using **Method C**, based on the most recent publicly available study conducted by the Centre for Disease Prevention and Control (SPKC) in 2020.

To accurately estimate the distribution of narcotics consumption, two main data sources are used:

- **SPKC data**: In 2020, SPKC commissioned a study on the prevalence of addictive substance use among the population **"Prevalence of substance use among the population in 2020**" (*available at:* https://www.spkc.gov.lv/lv/media/15536/download?attachment). The aim was to obtain nationally and internationally comparable data on alcohol and drug use patterns among residents aged 15–64 in Latvia. The final sample size (valid responses) was 4,616 respondents. The sample was designed to be representative of the entire Latvian population, covering different regions, sexes, age groups, and education levels. Data were collected using stratified random sampling, ensuring proportional representation of each region and social group. Weighting was applied to reflect the actual population structure.

- **CSB official data**: Population size and demographic breakdown (by sex and age) for Latvia in 2020. These data will be used to extrapolate the SPKC prevalence rates to the total population.

Table 18 Percentage of respondents who used illegal narcotic substances in the past 12 months, by sex and age group (%).

Age group	Men (%)	Women (%)	Number of men, 2022	Number of women, 2022	Estimated Number of Consumers (men)	Estimated Number of Consumers (women)	Structure men	Structure women
15-24	13.8	5	89 671	83 847	12 375	4 192	0.31	0.29
25-34	12.4	4.3	130 909	122 431	16 233	5 265	0.41	0.36
35-44	6.6	2.1	127 972	126 494	8 446	2 656	0.21	0.18
45-54	1.5	1	124 813	134 808	1 872	1 348	0.05	0.09
55-64	0.7	0.7	120 626	149 503	844	1 047	0.02	0.07
			593 991	617 083	39 770	14 508		

Table 19 Proportion of recent and current users of any drug by gender and region, %

Regon	Men	Women	Men 15-64g	Women 15-64g	Estimated Number of Consumers (men)	Estimated Number of Consumers (women)	Structure men	Structure women
Rīga	11.6	4.6	184 846	209 161	21 442	9 621	0.69	0.31
Kurzeme	4.3	0.4	74 558	75 119	3 206	300	0.91	0.09
Pierīga	5.3	2.2	118 371	119 739	6 274	2 634	0.70	0.30
Vidzeme	8.1	1.4	59 224	58 330	4 797	817	0.85	0.15
Zemgale	4.3	0.7	74 321	72 605	3 196	508	0.86	0.14
Latgale	1.0	1.0	82 671	82 129	827	821	0.50	0.50
			593 991	617 083	39 741	14 702		

COICOP 03 Clothing and Footwear

Overall micro-macro coverage: 44% (clothing: 41%; footwear: 55%).

In both HBS data and NA, it is not possible to separately identify purchases of new and used items, nor household sales of clothing and footwear. In both sources, consumption in this category is recorded as total expenditure on clothing and footwear, regardless of the item's status (new or used) or its source of acquisition. Therefore, the gap between NA and HBS data reflects general underreporting or insufficient coverage in HBS data.

Since the coverage level differs between footwear and clothing, and consumption patterns may be specific, each subcategory is analyzed and the gap allocated separately. **Method C** is used, based on the socio-demographic and income indicators of HBS households. In this case, external studies on clothing/footwear consumption habits by demographic groups are limited, so weights were constructed based on consumption trends observed in HBS data. That is, a similar logic to the "conditional consumer" calculation is applied, using the average expenditure shares or expenditure distribution on clothing and footwear reported in HBS data within each demographic group (income, gender, age, household type/size) to create allocation weights.

COICOP 04 Housing, Water, Electricity, Gas and Other Fuels

COICOP 041 Actual Rent

Coverage for actual rent is 86% (high). Both HBS and National Accounts include actual rental payments in this subcategory (both those paid by tenants themselves and rent covered by social assistance). HBS also includes payments for secondary residences and garages. The remaining gap is allocated using **Method A.**

COICOP 042 Imputed Rent

Coverage for imputed rent is 80% (high). The gap should be allocated only among owners, as only they can generate imputed rent. Allocation can be done proportionally to imputed rent in HBS or alternatively based on:

- Property value
- Living area
- Region or urban/rural status

COICOP 043 Maintenance and Repair of the Dwelling

Coverage is 363%, meaning microdata are three times higher than macrodata. National Accounts include only minor repairs (e.g., wallpapering, painting, small electrical work, replacement of plumbing parts, etc.). HBS also includes only minor repair costs. Conceptually, the data sources are aligned. The significantly higher coverage in HBS compared to NA suggests a large discrepancy in expenditure amounts, even though both sources conceptually include “minor repairs.” This difference is likely due to adjustments made in NA to exclude intermediate consumption from final consumption (e.g., expenditures by owner-occupiers related to the production of housing services rather than final consumption), as outlined in methodological guidelines.

In HBS data, it is necessary to exclude owner expenditures related to dwelling maintenance (both in the primary residence and secondary homes), as these expenses represent intermediate consumption in the production of housing services rather than final consumption. Only tenant expenditures on minor repairs should be retained.

Since HBS-reported expenditures significantly exceed the NA total, **Method A** is used, and HBS data is proportionally reduced to match the NA total.

COICOP 044 and 045 Water, Electricity, Gas

COICOP 044 has a micro-macro coverage of 193% (high). Since HBS data cover these expenditures and the coverage is high (excessive), **Method A** is used. The amounts recorded in HBS for CP044 was proportionally reduced so that the total matches the National Accounts figure. This method preserves the household expenditure distribution structure observed in HBS while ensuring consistency with NA values.

COICOP 045 has a micro-macro coverage of 87% (high). Since HBS data cover these expenditures and the coverage is high, **Method A** is used. The gap between NA and HBS data is distributed among households proportionally to their expenditure structure on electricity, gas, and other fuels as observed in HBS. This approach is based on the assumption that HBS data accurately reflect household consumption patterns in these categories.

COICOP 05 Furniture, Household Equipment and Routine Household Maintenance

Coverage: 67% (medium)

As in other categories (e.g., clothing and footwear), neither HBS data nor National Accounts (NA) for the CP050 category allow for separate identification of new vs. used purchases, nor adjustments for the provision of durable goods as “services” or household sales. Therefore, both HBS and NA data in this category are treated as total expenditures on these goods and services, regardless of the item's status (new or used) or its acquisition/use type. The gap between NA and HBS data most likely reflects general underreporting or insufficient coverage in HBS data.

HBS data typically contain information on household expenditures in this broad category. Although the coverage is 67% (medium and below 70%), it is assumed that HBS data are representative in terms of consumption distribution across households. **Method A** is used. The gap between NA and HBS data is distributed among households proportionally to the expenditure structure already observed in HBS for furniture, household equipment, and routine maintenance.

This is a comprehensive category in which households regularly incur expenses for various types of goods and services. These monetary expenditures are generally well captured in HBS data. Although the total HBS volume is lower than NA, this likely reflects general underreporting rather than systematic differences in consumption patterns across household groups. Method C would require external and reliable data on consumption patterns for this broad category, which are difficult to obtain and may not be representative. Therefore, given the representativeness of HBS data in terms of expenditure structure, the internal structure of HBS data is the most appropriate solution for allocating the gap.

COICOP 06 Health

COICOP 061 Medicines and health products

The micro-macro coverage rate is 97% (high). Since the coverage is high, the gap is distributed proportionally to HBS expenditure in this category – **Method A**.

COICOP 062 Outpatient care services

The micro-macro gap is 51% (medium-low). A more detailed analysis found that such low coverage for this group is primarily caused by Outpatient therapeutic and rehabilitation services, whose coverage is only 20%. This indicates significant underreporting or a lack of data in this subcategory. The consumption of healthcare services is closely linked to demographic and socio-economic factors, therefore **Method C** ensures a more accurate and reliable distribution of the gap. The gap between the NA and HBS data is distributed among households using socio-demographic weights. These weights were calculated based on the information available in the HBS data and additional assumptions about the consumption patterns of these services, taking into account the existing underreporting.

Key criteria for gap distribution and weight creation:

- **Age:** Older people use therapeutic and rehabilitation services more often due to an increased frequency of health problems.
- **Gender:** Women use preventive and rehabilitation services more often, as well as services related to reproductive health.
- **Income level:** Higher income generally provides greater access to paid and private medical services, thus influencing the volume of consumption.
- **Health status (if available):** Households with people with chronic diseases, disabilities, or other specific health impairments will incur significantly higher expenditures.

COICOP 063 Hospital care services

The micro-macro coverage is 117% (high – over-coverage). Since HBS data cover these expenditures and the coverage is high, **Method A** is used. The data recorded in the HBS was **reduced proportionally** so that the total corresponds to the NA indicator. This method preserves the structure of the distribution of household expenditures observed in the HBS data while ensuring consistency with the NA value.

COICOP 064 Other health care services

Coverage is only 42% (medium-low).

This subcategory includes the services diagnostic imaging services and medical laboratory services. The medium-low coverage rate (only 42%) indicates underreporting or data collection shortcomings in this subcategory in the HBS data. Similar to COICOP 062 (Outpatient care services), the consumption of healthcare services is closely linked to demographic and socio-economic factors. Therefore, to ensure a more accurate and reliable distribution of the gap among households, **Method C** is used. The gap between the NA and HBS data is distributed among households using **socio-demographic weights**. These weights were calculated based on the information available in the HBS data and additional assumptions about the consumption patterns of these services, taking into account the existing underreporting. **Key criteria for gap distribution and weight creation:**

- **Age:** Older people undergo diagnostic procedures and analyses more often due to the increased risk of health problems and the need for regular health status monitoring.
- **Gender:** Women use preventive and reproductive health-related diagnostic services more often.
- **Income level:** Higher income generally provides greater access to paid and private diagnostic and laboratory services, thus influencing the volume of consumption.
- **Health status (if available in HBS data or can be linked):** Households with persons with chronic diseases, disabilities, or other specific health impairments will incur significantly higher expenses for diagnostics and analyses.

COICOP 07 Transport

COICOP 071 Purchase of transport equipment

Overall coverage rate: 92% (high). Separate information on the purchase of new and used transport vehicles is available in HBS data. The total NA final consumption expenditure in the CP071 category includes purchases of both new and used vehicles. HBS data on the purchase of used vehicles are already reported on a **net basis**.

- Coverage for new vehicles is 128% (high – over-coverage).
- Coverage for used vehicles is 76% (high).

Since both the NA and HBS include the purchase of both new and used vehicles (with HBS data for used vehicles already at the net level), reconciliation are carried out separately for each subcategory, taking into account their differing coverage levels and consumption patterns.

For the purchase of new vehicles, **Method A** is used. The amounts recorded in the HBS data for the purchase of new vehicles is **proportionally reduced** so that the total corresponds to the National Accounts indicator for new vehicles. The over-coverage suggests that the HBS may report slightly more or classify some costs differently, but the HBS structure in this category is reliable for distribution, as new vehicle purchases are usually well-documented and associated with higher incomes.

For the purchase of used vehicles, HBS reports data already on a net basis (i.e., trade between households is excluded). Given that car purchases are large, irregular purchases where underreporting may occur, **Method A is still justified** if the HBS structure is considered representative of those who buy used cars.

Coverage for **motorcycles** is 338% (over-coverage) and for **bicycles** is 89% (high). **Method A** (proportional approach) is used for both means of transport.

COICOP 072 Operation of personal transport equipment

The micro-macro coverage is 58% (medium). All subcategories included in the National Accounts data are covered in the micro data. Coverage is between 38%–64% in practically all subcategories. However, the volume of operating expenses for personal transport is closely linked to household vehicle ownership and the intensity of their use, which, in turn, is most accurately reflected in household expenditures on vehicle purchases (CP071). Taking into account the close link between the purchase of vehicles (CP071) and their operation (CP072), as well as the relatively high coverage of the household balance sheet (MBS) in the CP071 category (after relevant adjustments), **Method B** is used to split the difference between the NA and HBS data. The gap is distributed using the expenditure structure observed in the HBS data for the purchase of vehicles (CP071), as well as household spending on technical inspections, repairs, maintenance of personal vehicles, car tires, spare parts, and all types of fuel.

COICOP 073 Passenger transport services

The micro-macro coverage is 41% (medium). Although all subcategories corresponding to the NA concept are available in HBS data, there is a significant gap between micro and macro data. Coverage is between 39%–56% in practically all subcategories. The gap is distributed among households using **Method C** (socio-demographic weights). These weights were calculated based on the information available in the HBS data and additional assumptions about the consumption patterns of these services, taking into account the identified underreporting.

Key criteria for gap distribution and weight creation:

- **Household income level (or income quintile/decile):** Income affects both the frequency and the type (e.g., public transport vs. taxis, airplanes) of transport service usage.
- **Place of residence (urban/rural):** Public transport availability is higher in cities, while rural areas may require other types of transport services or a greater dependence on personal transport.
- **Household type/size:** Larger households or those with children may use transport services more frequently, for example, for travel or daily errands.
- **Age:** Older people may use public transport more often, while younger people may use travel services more frequently.
- **Employment and education level:** Can affect the frequency and type of daily commutes (e.g., commuting to work/school).

COICOP 074 Freight transport services

The coverage rate is 34%, indicating substantial underreporting or data collection shortcomings in the HBS data for this expenditure category. These services are often irregular and linked to specific household events (e.g., moving house or delivery of large purchases), which may not be fully captured in household surveys. As there is no direct and evident relationship between the use of these services and the consumption patterns observed in other transport categories (CP071, CP072 and CP073), Method B is not considered appropriate. Therefore, Method C is identified as the most suitable approach for allocating the gap between National Accounts and HBS data. Under this approach, the gap is allocated across households using socio-demographic weighting factors derived from information available in the HBS and supported by assumptions regarding the consumption patterns of these services, while accounting for the identified underreporting.

Key criteria for gap distribution and weight creation:

- **Household income level (or income quintile/decile):** Higher-income households may more often choose paid delivery services, while lower-income households may use alternatives, such as their own transport or assistance from friends.
- **Household type/size:** Larger households or those actively renovating their homes are more likely to make large purchases that require delivery.

- **Place of residence (urban/rural):** Households living in rural areas may have a greater need for freight transport services if they are far from stores.

- **Housing ownership status / Change of residence:** If data on a recent household move or furnishing a new home are available, this is a strong indicator of the need for freight transport services. (This could be a potentially very strong factor if data is available.)

- **Purchase of durable goods (CP050 category, if reported simultaneously):** If a household reports large purchases of furniture or household appliances, there is a high probability that they also used delivery services.

COICOP 08 Information and Communication

The micro-macro coverage rate is 65%. The HBS microdata source does not include information on software expenditures (COICOP 08.2). However, calculations show that excluding software expenditures results in a coverage rate of 66%, indicating that software has only a marginal impact on the overall volume of COICOP 08 in the National Accounts. Consequently, a separate treatment of software expenditures is not considered necessary, and the gap between National Accounts and HBS data is allocated across the entire COICOP 08 category.

Given the close relationship between communication services expenditure and household demographic and socio-economic characteristics, Method C is considered the most suitable approach for allocating the gap. Under this approach, the allocation is based on socio-demographic weighting factors derived from available HBS information and supported by assumptions regarding consumption patterns, while accounting for the identified underreporting.Key criteria for gap allocation and weight construction:

- **Household income level (e.g., income quintile/decile):** Higher income levels often correlate with more expensive internet plans, greater mobile data usage, and smartphone purchases.

- **Household type/size:** Larger households typically incur higher communication expenses (e.g., multiple mobile phones, larger internet packages).

- **Age:** Younger households tend to use mobile internet and smartphones more frequently, while older households may rely more on landlines or use internet services less intensively.

- **Place of residence (urban/rural):** Differences in service availability and costs may exist between urban and rural areas.

COICOP 09 Recreation, Sport and Culture

The overall coverage rate for this group is 44%. However, given the broad and heterogeneous nature of the expenditure categories included in COICOP 09, a single allocation approach is not considered appropriate. Therefore, each COICOP 09 subcategory is assessed separately, and the most suitable allocation method is identified based on its coverage rate, expenditure characteristics, and observed consumption patterns.

COICOP	Title	Coverage	Method
09.1.	Durable recreational goods	8%	
	– Photographic, cinematographic, and optical equipment	3%	**C**
	– Major durable recreational goods	373%	**A**
09.2.	Other recreational goods	88%	**A**
09.3.	Garden products and pets	67%	**C**
09.4.	Recreational services	28%	**C**
09.5.	Cultural goods	30%	**C**
09.6.	Cultural services	73%	**A**
09.7.	Newspapers, books, and stationery	32%	**C**
09.8.	Package holidays	48%	**C**

COICOP 09.1 Durable Recreational Goods

This category includes audio-visual, photographic, and information processing equipment, as well as other large durable goods for recreation and culture. The overall coverage rate is very low (8%), but detailed analysis reveals significant differences between subcategories. Therefore, allocation is done at two levels using different methodologies.

CP09.1.1 Photographic, Cinematographic, and Optical Equipment

Includes cameras, accessories, and optical instruments. Coverage is very low (3%), and HBS data are not representative for gap allocation.
Method C is used, applying socio-demographic weights based on HBS data and additional assumptions about consumption patterns, considering the identified underreporting.

Key criteria for gap allocation and weight construction:

- **Household income level**: Higher income often correlates with the purchase of more expensive equipment.
- **Age and household type**: Younger households and those with children may be more likely to purchase photo/video equipment.
- **Education level and interests**: May be linked to hobbies and leisure activities, especially if relevant professional education is present (if available in HBS).

CP09.1.2 Major Durable Recreational Goods

Includes camper vans, trailers, aircraft, ultra-light planes, boats, yachts, outboard motors, horses, ponies, camels, and other large recreational items.
Coverage is extremely high (373%), with specific items showing wide variation (e.g., 0% for campers, 247% for aircraft, 498% for horses).
Method A is used: HBS-reported expenditures is proportionally reduced to match the National Accounts total, preserving the internal distribution structure observed in HBS.

COICOP 09.2 Other Recreational Goods

This category includes games, toys, sports equipment, and camping gear. The coverage rate is 88%, indicating that these expenditures are well captured in the HBS data and that the expenditure structure observed in the survey is broadly consistent with that recorded in the National Accounts.

Given the high level of coverage and the absence of significant indications of underreporting, **Method A** is considered the most appropriate approach for allocating the gap, relying on the expenditure distribution observed in the HBS.

COICOP 09.3 Garden Products and Pets

This category includes garden tools, fertilizers, seeds, flowers, as well as pets and pet-related products such as food, accessories, and veterinary services. The overall coverage rate is 67% (65% for garden-related expenditures and 68% for pet-related expenditures), indicating a moderate level of underreporting in the HBS despite the relatively regular nature of these purchases.

Because the expenditure patterns and household characteristics associated with garden-related and pet-related consumption differ, a common allocation approach is not considered appropriate. Therefore, each subcategory is assessed separately. **Method C** is identified as the most suitable approach for allocating the gap, using socio-demographic weighting factors derived from HBS information and supported by assumptions regarding the consumption patterns of the respective household groups, while accounting for the identified underreporting.

Key criteria for gap allocation and weight construction:Garden Products, Plants, and Flowers (coverage: 65%)

- **Place of residence**: Rural households or those in private homes with gardens likely spend more.
- **Housing type/ownership**: Households owning or renting homes with garden areas.
- **Income level**: Higher income may correlate with more expensive equipment or exotic plants.
- **Age group**: Older households often engage more in gardening.

Pets and Pet Products (coverage: 68%)

- **Pet ownership indicator**
- **Household type/size**: Both large families (with children) and single-person households (especially elderly) may own pets.
- **Income level**: Affects ability to purchase premium food, accessories, or veterinary services.
- **Age group**: May correlate with pet type and related expenses.

COICOP 09.4 Recreational Services

This category includes a wide range of services such as rental of photographic and cinematographic equipment, rental of recreational goods, sports, camping, games and toys, veterinary services, recreational and sports services, and gambling services.

The overall coverage rate for this category is 28%, indicating substantial underreporting in the HBS data. A more detailed assessment shows that gambling services have an exceptionally low coverage rate of only 2%. As gambling services account for 58% of total COICOP 09.4 expenditure, the expenditure structure observed in the HBS is not considered sufficiently representative for allocating the gap between National Accounts and HBS data. Consequently, the allocation of the gambling-related gap is based primarily on external data sources combined with supporting assumptions regarding participation patterns across population groups.

Method C is identified as the most suitable approach for allocating gambling expenditures. The allocation is based on socio-demographic weighting factors derived from external evidence on gambling participation and complemented by official demographic statistics.

Data sources for gambling expenditure allocation (Method C):

- **Ministry of Health's 2019 study**: *"Gambling habits and behaviours in Latvian society"* – the latest available study, based on a survey of 4,912 respondents aged 15–64.
  - 66% of men and 50% of women have ever played gambling or lotteries.
  - Among those who have ever played: 45% have primary or lower education, 52% secondary, and 49% higher education.
  - Regional breakdown of those who have ever played: Vidzeme (73%), Zemgale (64%), Kurzeme (56%), Riga (50%), Latgale (58%).
- **CSB official demographic data (2019)**: Used to extrapolate the study's percentages to the total population.

The Table 21 presents the weighting coefficients derived from this socio-demographic profile and used for allocating the gap across age and gender groups.

Table 20 Estimation of Participant Population and Structural Allocation Weights for Gambling by Age Group and Sex

Age Group	Men Count (2019)	% Playing (Men)	Estimated Male Players (66%)	Women Count (2019)	% Playing (Women)	Estimated Female Players (50%)	*Male Structure*	*Female Structure*
15-24	90 767	0.45	26 958	85 237	0.45	19 178	0.12	0.11
25-34	135 275	0.65	58 033	127 534	0.65	41 449	0.26	0.23
35-44	127 310	0.61	51 255	126 969	0.61	38 726	0.23	0.22
45-54	125 348	0.6	49 638	136 323	0.6	40 897	0.22	0.23
55-64	120 568	0.51	40 583	150 578	0.51	38 397	0.18	0.21
Total	599 268		226 467	626 641		178 647		

Table 21 Regional Breakdown and Proxy Structure for Gambling Engagement

Region	% Ever Played	Population (2019)	Estimated Players	*Structure*
Rīga	0.50	399 126	199 563	*0.28*
Pierīga	0.58	236 747	137 313	*0.20*
Vidzeme	0.73	119 759	87 424	*0.12*
Kurzeme	0.56	152 431	85 361	*0.12*
Zemgale	0.64	149 013	95 368	*0.14*
Latgale	0.58	168 833	97 923	*0.14*
Latvija		1 225 909	702 953	

Combined weights are constructed for each household in HBS based on **region, gender, and age group**, to allocate the gap proportionally.

Table 22 Other subcategories in COICOP 09.4 and gap allocation methods used

Subcategory	Coverage	Method	Notes

CP09.4.2 Hire, maintenance and repair of major durables for recreation	33%	B	Allocate proportionally to CP09.1.2 expenditures
CP09.4.3 Hire and repair of games, toys and hobbies	1%	B	Allocate proportionally to CP09.2.1 expenditures
CP09.4.4 Hire and repair of equipment for sport, camping and open-air recreation	703%	A	Proportional reduction of HBS values
CP09.4.5 Veterinary and other services for pets	55%	C	Allocate among households reporting pet food purchases in HBS
CP09.4.6 Recreational and sporting services	81%	B	Allocate proportionally to CP09.2.2 expenditures

COICOP 09.5 Cultural Goods

Coverage: 30% (low).

This category includes relatively specific and often high-value purchases, such as musical instruments and audiovisual media products. These expenditures are typically concentrated among particular household groups and are not evenly distributed across the population.

The low coverage rate indicates that the HBS does not fully capture expenditure patterns for these items and that the observed expenditure structure is not sufficiently representative for direct gap allocation. Given the concentration of consumption within specific socio-demographic groups and the limitations of the HBS data, **Method C** is identified as the most suitable approach. Under this approach, the gap is allocated using socio-demographic weighting factors derived from available HBS information and supported by assumptions regarding the consumption patterns of the household groups most likely to purchase these goods.

Key criteria:

- Income level (very high income)
- Education level (e.g., music school background)
- Relevant profession
- Age

COICOP 09.6 Cultural Services

Includes theatre, concerts, opera visits, museum services, libraries, photography services, etc.

Coverage: **73% (high)**.

The high coverage indicates that HBS data are sufficiently representative. **Method A** is used, preserving the expenditure structure observed in HBS.

COICOP 09.7 Newspapers, Books and Stationery

Includes newspapers, books, printed materials, and stationery. Coverage: **32% (medium)**.

Method C is more appropriate due to low representativeness in HBS.

Key criteria:

- **Age**: Younger individuals may purchase textbooks; older individuals may prefer newspapers.
- **Household type**: Households with school-age children or students may spend more on stationery and textbooks.
- **Income level**: Higher income correlates with greater spending.

If data show that many households are represented, proportional allocation may be considered. However, if expenditures are concentrated in a few households, **Method C** should be applied.

COICOP 09.8 Package Holidays

This category includes holiday trips or tours that provide travel, food, accommodation, guides, and other services.

Coverage: **48% (moderately low)**. Since these are large and irregular expenditures, HBS data alone are not sufficiently representative to accurately allocate the observed gap. Therefore, **Method C**, based on socio-demographic data, is recommended for gap allocation.

To build a precise profile of the "potential consumer," two main data sources is used:

- **CSB data** on multi-day trips abroad by Latvian residents, broken down by gender and age group.
- **CSB data** on multi-day trips within Latvia by Latvian residents, also broken down by gender and age group.

Based on CSB's published data on total expenditures made by Latvian residents on multi-day trips abroad and within Latvia, an expenditure structure is calculated by gender and age group (see Table 24 and Table 25).

Table 23 Allocation Structure for Multi-Day Travel Expenditures by Gender (2019–2023)

Gender	2019	2020	2021	2022	2023
Men	0.56	0.57	0.63	0.50	0.54
Women	0.44	0.43	0.37	0.50	0.46

Table 24 Allocation structure for Multi-Day Travel Expenditures by age group (2019–2023):

Age group	2019	2020	2021	2022	2023
15–24	0.15	0.03	0.14	0.11	0.11
25–34	0.24	0.26	0.21	0.20	0.23
35–44	0.24	0.25	0.27	0.26	0.26
45–54	0.13	0.26	0.19	0.18	0.18
55–64	0.16	0.16	0.15	0.14	0.13
65+	0.07	0.04	0.03	0.11	0.09

Additional criteria for profiling potential consumers:

- **Household income level**: Since package holidays are large and irregular expenses, they are closely tied to the household's financial situation and disposable income. Households with higher income are more likely to use such services.

- **Household composition and age**: Different household types and age groups have varying travel habits. Families with children often choose "all-inclusive" or organized holidays. Older individuals may prefer guided tours.
- **Education level and employment status**: Higher education may correlate with higher income, and those with stable employment and vacation rights are more likely to afford travel.
- The combination of these criteria with relevant data from the Central Statistical Bureau enables the development of a representative socio-demographic profile of households most likely to incur these expenditures. This profile forms the basis for the allocation of the gap between National Accounts and HBS data, ensuring that unreported expenditures are distributed across households in a manner that reflects the most plausible consumption patterns.

Combining these criteria with relevant CSB data makes it possible to create a representative household profile and allocate the gap to the households most likely to incur these expenditures.

COICOP 10 Education

The overall coverage for this COICOP category is **37%**. HBS data primarily include paid education, meaning that conceptually, HBS and National Accounts (NA) data are aligned. However, the low overall coverage indicates the need to allocate the gap using external data sources.

Primary and Secondary Education

These subcategories have the **lowest coverage** (8% and 11%, respectively). Since paid education at these levels is most likely used by households with higher incomes, **Method C** is applied. The gap is distributed proportionally among households meeting the following combined criteria:

- Households with the highest income (**4th and 5th income quintiles)**
- Households with children aged **7–18 years** who indicated in question A12 of the 1-MBA survey that they are currently enrolled in an educational program
- **Geographic location** (urban/rural): Paid education is more commonly available in urban areas

Tertiary Education

Coverage: **32% (medium)**. Although paid tertiary education may be accessible to households across income levels (e.g., via student loans), HBS data may underreport these expenditures.

To address this, it is important to identify **"potential consumers"** using **Method C**, based on demographic characteristics.

The gap is distributed evenly among households identified as having students, with additional criteria:

- Households with members aged **18–25 years** who indicated in question A12 of the 1-MBA survey that they are currently enrolled in an educational program
- Households matching the **student household type**
- Households with **higher education levels** (more likely to pursue or encourage further education)
- **Income level** (especially 4th and 5th quintiles, as higher education often requires significant financial resources, even with loans)

Pre-primary Education

Coverage: **44% (medium)**.

A **combined approach (Method A + C)** is used. Part of the gap can be allocated proportionally using HBS expenditure data (**Method A**), given the relatively good coverage. The remaining unreported portion is allocated using **Method C**, based on the following criteria:

- Households with children aged **1.5–6 years** who indicated in question A12 of the 1-MBA survey that they currently attend a pre-primary institution (kindergarten)
- **Parental employment status** (working parents are more likely to use paid services)
- **Income level** (4th and 5th quintiles)
- **Regional factors** (priority given to urban households, where paid pre-primary services are more likely and potentially underreported)

Education Not Defined by Level

Coverage: **69% (relatively high)**.

Method A is recommended, with additional validation. If expenditures are not highly concentrated in a few households, it is reasonable to allocate the gap proportionally based on existing HBS expenditures.

If expenditure concentration is observed or a more precise consumer profile is needed, **Method C** may be applied using the following socio-demographic criteria:

- **Active adults** (especially aged 25–55) investing in career or personal development
- **Income level**: Although affordable courses exist, many continuing education programs, language schools, and hobby courses are paid services. Households in the **3rd–5th income quintiles** are more likely to afford such expenses
- **Education level**: Individuals with higher completed education are more likely to pursue lifelong learning and seek courses to enhance their skills or knowledge (e.g., professional development)
- **Employment status**: Employed individuals are more likely to use continuing education and qualification courses related to their job or career change aspirations

COICOP 11 Catering Services and Accommodation

This category covers catering services provided by restaurants, cafés, bars, canteens, etc., as well as accommodation services offered by hotels, guesthouses, motels, holiday villages, and similar establishments.

The overall coverage for this category is **47%**, indicating a significant volume of unreported expenditures. It is important to note that both National Accounts (NA) and HBS data for this COICOP category **do not include rent for secondary residences** (reported under COICOP 04) or expenses related to **package holidays** (reported under COICOP 09). Therefore, NA and HBS data are conceptually aligned in terms of included items.

Given the low coverage, a **combined approach** is used for gap allocation, analyzing catering services and accommodation services separately.

COICOP 11.1 Catering Services

Coverage: **48% (moderately low)**, indicating substantial underreporting.

Part of the observed gap may be allocated **proportionally to the actual expenditures reported in HBS** (**Method A**), assuming that HBS data are representative of the relevant consumer segments.

The remaining unreported portion is allocated using **Method C**, based on a “potential consumer” profile defined by the following socio-demographic criteria:

- **Household income**: Households in the 4th and 5th income quintiles are more likely to dine out.
- **Household type and composition**: Single-person households, couples without children, and smaller households tend to use catering services more frequently.
- **Age group**: Young adults and working-age individuals (e.g., 18–44 years) are more likely to visit restaurants and cafés.
- **Employment status**: Employed individuals may eat out more often due to work or social activities.
- **Geographic location**: Urban residents have greater access to and more frequent use of catering services.

COICOP 11.2 Accommodation Services

Coverage: **46% (medium)**, similar to the overall category.

Given the link between accommodation services and travel, the same approach used for **COICOP 09.8 Package Holidays** is applied (**Method C**). To define the “potential consumer” profile, CSB data on multi-day trips abroad and within Latvia by Latvian residents is used (see COICOP 09.8).

From CSB’s published data on total expenditures made during multi-day trips, the expenditure structure by gender and age group is used to identify which demographic groups are most likely to incur accommodation-related expenses.

Additional criteria to refine the profile:

- **Household income level**: Higher-income households are more likely to afford travel involving accommodation.
- **Household composition**: Couples without children and individuals living alone may travel more frequently, while families with children may choose specific types of accommodation.
- **Employment status**: Employed individuals may travel for work or leisure more often.
- **Geographic location**: Residents of urban areas may travel more frequently or use accommodation services more often.

Using these criteria and data sources will enable accurate allocation of unreported expenditures to the appropriate HBS households, forming a robust imputation model.

COICOP 12 Insurance and Financial Services

COICOP 12.1 Insurance

This category includes service fees for life, housing, health, and transport insurance. The overall coverage rate is **70%**, but allocation is complicated by the fact that HBS microdata do not contain separate service fee data, only premium payments. Additionally, there are differences in data availability and alignment with National Accounts (NA) between **life and non-life insurance**.

Life Insurance Services

To allocate the total NA value for life insurance service fees, data from the **Latvian Bank’s HFCS survey** on household contributions to life insurance is used. These contributions (e.g., from HFCS variables such as DA2109 or PFA080x, filtered by PFA090x = LV_WH_LIFE) is aggregated across all household members. The total NA value is distributed proportionally to the contributions reported in HFCS data for households making such payments. This approach aligns with handbook recommendations to use premium payments as a proxy for service fee allocation.

Non-Life Insurance Services

HBS data on non-life insurance premiums **exceed macro values**, mainly due methodological differences between both sources. Since HBS covers all types of non-life insurance and the coverage exceeds NA values, **Method A** is used. This means that HBS-reported expenditure on non-life insurance are proportionally reduced to match the total NA value for non-life insurance service fees.

COICOP 12.2 Financial Services

This includes various financial services and FISIM. Overall coverage: 3%.

FISIM (Financial Intermediation Services Indirectly Measured)

Method B is used for the distribution of FISIM. FISIM related to final consumption (CP12.2.1) is distributed among households **proportionally** to the **actual interest payments (D.41pay) on consumer loans** and **interest receipts (D.41rec) on deposits** reported by households in the **EU-SILC data**. The share of each household out of the total absolute value of interest payments and receipts in the EU-SILC data is calculated. The total NA FISIM final consumption amount is distributed proportionally to these weights.

Verification: CP12.2.1 (at the household level) ≤ |D41rec_FISIM| (for recipient households) + |D41pay_FISIM| (for paying households), taking into account that a part of FISIM is accounted for as intermediate consumption.

Other Financial Services

Other financial services cover direct household payments for various financial services provided by deposit-taking corporations and other financial institutions. Unlike FISIM, these are directly stated charges, such as for account maintenance, transfers, card servicing, and other banking services. The distribution is carried out based on the **"potential consumer" profile (Method C)**. Given the widespread use of banking services (bank accounts, payment cards) in modern households, the expenditure for these services is distributed **among all households**. An exception or reduced distribution value is applied to households consisting solely of **elderly people** (e.g., all adults' ages exceed a certain threshold, indicating limited or minimal use of banking services). This approach will reflect the reality of the near-universal use of banking services, except for specific demographic groups.

COICOP 13 Personal Care, Social Protection, and Miscellaneous Goods and Services

This is a broad and diverse category that requires an individual approach for each sub-item. The method is chosen based on the HBS data coverage level compared to the NA data.

13.1. Personal Care

Coverage is 78% (**high**). Since the coverage for this subcategory is relatively high, **Method A** is used.

- For **Hairdressing services (13.1.3.)**, HBS data have high over-coverage (148%), so the HBS data will be **proportionally reduced** to match the NA macro value.
- For the remaining personal care goods subcategories (**13.1.1. and 13.1.2.**) with medium-high coverage (70% and 64%), HBS data is used to **proportionally distribute the gap** between HBS and NA, based on the expenditures reported by households in the HBS.

13.2. Other Personal Effects

Overall coverage is medium (**68%**), where **Method A** could generally be used.

- **13.2.1. Jewellery and watches** (coverage: medium – 38%): **Method C** also is applied here, and the distribution concentrated more towards the higher income quintiles, as these items are more of a luxury good.

- **13.2.2. Articles for religious and spiritual practices** (coverage: low – 30%): For articles for religious and spiritual practices with low coverage, where HBS data may not be sufficiently representative, **Method C** is used. Distribution is based on demographic indicators, such as age (older persons) or regional affiliation, if other additional information is available. If no specific data is available, a broader COICOP aggregate is used as a proxy, e.g., total final consumption expenditure (**Method D**).

- **13.2.9. Other personal effects not elsewhere specified** (coverage: high – 82%): **Method A** is applied.

13.3. Social Protection

Coverage is **low (21%)**. **Method C** is applied for the distribution. Given the very low coverage in HBS data, the distribution is based on socio-demographic criteria that characterize the recipients of social protection services. This would include households with elderly people, persons with disabilities, low-income households, or those receiving specific benefits or identified in other supplementary data sources.

13.9. Miscellaneous Services

Coverage is **very low (19%)**. This category also includes subcategories: 13.9.0.1. Prostitution, 13.9.0.2. Religious services, and 13.9.0.9. Other services n.e.c. Given the very low overall coverage, HBS data are most likely not sufficiently representative.

- **Prostitution services (13.9.0.1.)**: Given that this is part of the **non-observed economy** and there is no direct information in HBS data, the distribution will be carried out based on a broader indicator, such as proportional to households' total final consumption expenditure (**Method D**).

- **Religious services (13.9.0.2.)**: A profile could be created based on age structure, nationality or regional distribution (**Method C**).

- **Other services n.e.c. (13.9.0.9.)**: If it is not possible to identify a specific profile, a broader distribution base may be used, such as proportional to total household consumption expenditure (**Method D**).

TECHNICAL IMPLEMENTATION

General Approach

This methodological framework describes the reconciliation of Household Budget Survey (HBS) microdata with National Accounts (Household Distributional Accounts, HDA) control totals. The primary objective is to ensure full consistency with macroeconomic aggregates while preserving the underlying distributional properties observed in microdata. The methodology was implemented in **R using 2019 reference data** as the benchmark year.

Depending on data quality, coverage, and item-specific characteristics, the following alignment methods are applied:

- **Method A – Proportional Scaling**

- **Method B – Residual Allocation (Delta Approach)**

- **Method C – Within-Group Proportional Allocation**

- **Propensity Method – Calibrated Consumption Probability Imputation**

Method A: Proportional Scaling

Description

Method A is applied when the microdata and macro-level National Accounts targets are conceptually well aligned, and only a level adjustment (scale correction) is required.

Mathematical Formulation

The adjustment coefficient $(\alpha)$is calculated as:

$$\alpha = \frac{\text{Target}_{\text{NA}}}{\sum_i (\text{HBS}_i \times w_i)}$$

The adjusted value for each household $(HDA_i)$is then defined as:

$$HDA_i = \text{HBS}_i \times \alpha$$

Where:

- HBS_i— initial microdata expenditure value for household i
- w_i— survey expansion weight for household i
- $\text{Target}_{\text{NA}}$— National Accounts macro control total

Extension: Winsorization

To prevent extreme values from distorting the distribution, an upper-bound threshold $(p_{99})$is applied in specific cases:

$$HDA_i = \min(\text{HBS}_i \times \alpha, p_{99})$$

Interpretation and Validation

- **Preservation:** Perfectly preserves the relative micro-distributional structure.
- **Macro Consistency:** Ensures an exact match with macro control totals.
- **Extensive Margin:** Does not generate new consumers; zero values remain unchanged.
- **Validation Metrics:** High correlation with the initial data, minimal decile rank mobility, and a stable Gini coefficient.

HDA Variables Aligned via Method A

Direct Baseline Variables

HDA Category	Multiplier (α)
HDA041	0.895985
HDA042	1.160928
HDA044	0.553838

HDA Category Multiplier (α)

HDA045	1.065659
HDA051	1.537753
HDA052	1.038951
HDA053	1.276076
HDA054	1.741969
HDA055	0.499052
HDA056	1.613973
HDA061	0.963974
HDA063	0.537072
HDA071	0.821405
HDA083	1.142524
HDA092	0.849458
HDA093	1.420338
HDA096	1.454293
HDA101	1.487634
HDA105	0.483136

Aggregated Components (Sum of Method A Sub-items)

- HDA011 (derived from HDA0111–HDA0119 Method A components)
- HDA012 (derived from HDA0121–HDA0129 Method A components)
- HDA043 (derived from HDA0431–HDA0432 Method A components)
- HDA062 (sum of HDA0621 Method A + HDA0622/HDA0623 Method C components)
- HDA091 (partially Method A + Method C)
- HDA111 (partially Method A)

HDA Category Multiplier (α)

HDA0111	1.849551
HDA0112	1.218975

HDA Category	Multiplier (α)
HDA0113	1.166086
HDA0114	1.321842
HDA0115	1.691718
HDA0116	1.229763
HDA0117	1.276806
HDA0118	1.662154
HDA0119	1.250307
HDA0121	2.130586
HDA0122	2.324877
HDA0123	2.006635
HDA0124	1.287294
HDA0125	1.278467
HDA0126	1.357546
HDA0129	1.125800
HDA0431	0.481810
HDA0432	0.910994
HDA0621	0.593429
HDA0911	1.081290
HDA1112	1.192855

Method B: Residual Allocation (Delta Approach)

Description

Method B is used when a specific macro-level gap must be allocated across the microdata while explicitly anchoring the distribution to a predefined external auxiliary variable profile.

Mathematical Formulation

The total macro-micro residual (Δ)is computed as:

$$\Delta = \text{Target}_{\text{NA}} - \sum_{i} (\text{HBS}_i \times w_i)$$

The allocation proportion (p_i)based on an external conditioning variable $(X_i)$is defined as:

$$p_i = \frac{X_i \times w_i}{\sum_j (X_j \times w_j)}$$

The final baseline micro-adjustment is implemented as:

$$HDA_i = \text{HBS}_i + \frac{\Delta \times p_i}{w_i}$$

Interpretation and Validation

- **Targeted Control:** Allows precise, targeted control over the conditional allocation structure.
- **Targeting:** Adjustments can be restricted to specific socio-demographic sub-populations.
- **Non-linear:** Avoids flat, uniform scaling effects across the entire distribution.
- **Validation Metrics:** Macro-level totals match exactly.

Note: This method requires close monitoring to prevent potential localized structural distortions within sub-groups.

Implementation Case Study

- **HDA0943:** Distributed using the underlying socio-demographic structure derived from HBS0921.

Method C: Within-Group Proportional Allocation

Description

Method C decomposes the macro-micro discrepancy and distributes it conditionally across granular socioeconomic groupings (e.g. income quintiles, age bands, household types).

Calculation Steps

1. Group Aggregate Totals

Compute the weighted sum for group g:

$$G_g = \sum_{i \in g} (\text{HBS}_i \times w_i)$$

2. Total Discrepancy Allocation

Determine the group-specific macro gap:

$$\Delta_g = \text{Target}_g - G_g$$

3. Group Allocation Proportion

Define proportional shares based on group profiles.

4. Within-Group Proportional Adjustment

Re-scale variables within each partition:

$$HDA_{i \in g} = \text{HBS}_i \times \frac{\text{Target}_g}{G_g}$$

Interpretation and Validation

- **Structure:** Strongly preserves between-group macro structures.
- **Smoothing:** Moderates extreme variability within specific socio-demographic clusters.
- **Micro Reallocation:** Frequently triggers structural shifts and rank mobility at the individual micro level.
- **Validation Dynamics:** Higher decile rank mobility, potential introduction of extensive-margin shifts, and a marginal reduction in overall inequality metrics.

Table 25 Conditional Factors for Method C Clustering

HDA Category	Method Type	Conditioning / Grouping Variables	Methodological Justification
HDA031C	(Personal)	kvpc.i, Fam_v1, vec_gr	Income quintile + household structure + age profile
HDA032C	(Personal)	kvpc.i, Fam_v1, vec_gr	Aligned with the structure of HDA031
HDA0622C	(Personal)	kvpc.i, xA1, vec_gr	Income quintile + gender + age profile
HDA0623C	(Personal)	kvpc.i, xA1, vec_gr	Identical approach to HDA0622
HDA064C	(Personal)	kvpc.i, xA1, vec_gr	Socio-demographic conditional allocation
HDA073C	(Personal)	kvpc.i, Fam_v1, vec_gr, laupil, strad_mac	Extended socio-economic profile mapping
HDA081C	(Personal)	kvpc.i, Fam_v1, vec_gr, laupil	Household structure + regional urbanization factor
HDA0911C	(Personal)	kvpc.i, Fam_v1, vec_gr, laupil, ed_lev	Inclusion of educational attainment control
HDA095C	(Personal)	kvpc.i, vec_gr, ed_lev	Streamlined socio-demographic structure
HDA097C	(Personal)	kvpc.i, vec_gr, ed_lev	Analogous approach to HDA095

HDA Category	Method Type	Conditioning / Grouping Variables	Methodological Justification
HDA098C	(Personal)	kvpc.i, vec_gr, ed_lev, Fam_v1, strad	Expanded, multi-dimensional socio-economic model
HDA102C	(Personal, Selective)	Children present (7–18), 4th or 5th quintile, laupil	Highly targeted allocation (dependent children + high-income bands)
HDA104C	(Personal, Selective)	Higher education students, 4th or 5th quintile, laupil, reg_st	Selective allocation based on higher education status
HDA1111C	(Personal)	vec_gr, kvpc.i, strad, Fam_v1, laupil	Broad socio-economic composite structure
HDA112C	(Personal)	kvpc.i, vec_gr, ed_lev, Fam_v1, strad	Combined educational attainment and employment matrix
HDA072C	(Household)	Vehicle usage (HBS071 > 0 or HBS072 > 0), kvpc_i, Fam_v1	Selective household-level asset usage filter
HDA074C	(Household)	Length of residency < 1 year, significant capital outlays (HBS0511 or HBS0531 > 500), kvpc_i, Fam_v1, laupil	Expanded household moving/capital acquisition model
HDA0945C	(Household)	Existing positive outlays in HBS0932 or HBS0945, kvpc.i	Streamlined past-expenditure baseline model
HDA133C	(Household, Selective)	Disability status within household OR 1st income quintile placement, Fam_v1, reg_st	Target-specific social investment support group

Propensity Score Method (Calibrated Consumption Probability)

Description

The Propensity Score Method is implemented to address extensive-margin missingness, specifically in situations where microdata reporting exhibits severe undercoverage or zero inflation despite known macro-level outlays.

Econometric Specification

A logistic regression model is fitted to predict the consumption probability:

$$\ln\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}$$

The calibrated response probability $(P_i)$is derived as:

$$P_i = \frac{1}{1 + e^{-\mathbf{X}_i \boldsymbol{\beta}}}$$

Adjustment Execution

Expenditures are imputed or re-weighted conditionally based on the estimated propensity scores, shifting zero observations into positive expenditure statuses until macro-level target compliance is reached.

Interpretation and Validation

- **Data Quality:** Significantly improves microdata population coverage rates.
- **Behavioral Adjustment:** Corrects for behavioral consumption non-response.
- **Rank Swapping:** Modifies household consumption rankings.
- **Validation Metrics:** Broadens extensive-margin coverage, triggers structural wealth reallocation, and modifies relative distribution ranks.

Core Empirical Examples

- HDA0211, HDA0212, HDA0213 (Alcoholic beverages)
- HDA023 (Tobacco)
- HDA0947 (Specific recreational/cultural outlays)

Special Case Study: HDA024 Consumption Allocation (Illicit Transactions Only)

This item captures unobserved or illegal transactions through exogenous adjustment mechanisms, adding external components onto the baseline microdata structure. This approach functions structurally similarly to Method C.

$$HDA024 = \text{Illicit Outlays} \times \text{Proportional Share of Target Sub-group}$$

Interpretation: Successfully incorporates unobserved exogenous macro-components into microdata. **Structural Property:** Modeled to reflect the structural mechanics of Method C.

Combined and Hybrid Frameworks

For complex expenditure categories, multi-method hybrid approaches are implemented sequentially to ensure consistency between microdata and macro aggregates.

- **HDA021 (Alcoholic Beverages):** Propensity Method for 0211–0213; Method C for 0219
- **HDA062 (Outpatient Care):** Method A for 0621; Method C for 0622–0623
- **HDA091 (Information Processing Equipment):** Method C for 0911; Method A for 0912
- **HDA094 (Recreational and Cultural Services):** Methods A, B, C, and Propensity Method combined across sub-components
- **HDA111 (Catering Services):** Method C for 1111; Method A for 1112

Consumption information from the Household Budget Survey (HBS) was integrated into the HDA household income master file using a pre-defined bridge table linking EU-SILC/HDA households to their corresponding HBS donor households. The income-side HDA master file remained the core analytical dataset, while the bridge table served exclusively as a linking mechanism between matched household identifiers.

The bridge linked each EU_SILC/HDA household identifier (id_ms) to a single HBS donor household. The final dataset includes 6,095 EU-SILC/HDA households. HBS donor households could be used multiple times. Following the linkage, HBS consumption variables were transferred to the HDA master file and labelled as CPxxx_HBS, ensuring clear separation from calibrated consumption variables.

The transferred HBS variables provide the household-level distributional structure for each COICOP component. Final consumption aggregates were constructed as CPxxx_HDA. For each COICOP item, SILC-weighted totals were calculated using DB090, and a single calibration factor was applied uniformly across households:

CPxxx_HDA = CPxxx_HBS × calibration factor

The calibration factor is defined as the ratio between the National Accounts COICOP control total and the corresponding EU-SILC-weighted microdata total. This ensures exact alignment with macroeconomic targets while preserving the underlying household-level distribution derived from the HBS-based structure.

HDA082 was allocated according to the combined household distribution of **HDA081** and **HDA083**, as these components belong to the same broader communication-related consumption group and were considered the closest available HBS-based distributional proxies.

HDA103 was allocated according to the combined distribution of **HDA102**, **HDA104**, and **HDA105**. These components belong to the same education-related COICOP domain and therefore provide the most relevant available household-level structure for imputing CP10.3.

HDA139 was allocated using total household adjusted disposable income, represented by **B6g_HDA**. This proxy was selected because not sufficiently close HBS-based COICOP component was available, and income was considered a reasonable general indicator of households’ capacity to consume miscellaneous goods and services.

HDA1221 was allocated using the absolute values of FISIM-related interest flows, specifically **ABS(D41useFisim_HDA)** and **ABS(D41recFisim_HDA)**. This proxy was selected because Coicop12.2 is conceptually linked to financial-service-related consumption, while the use of absolute values ensures that both paid and received FISIM-related flows contribute positively to the allocation base. HDA122 other was allocated with B6g_HDA.

Total household consumption was defined as:

CP_TOTAL_HDA = Σ CPxxx_HDA

After applying EU-SILC household weights, the COICOP component totals were already broadly aligned with the National Accounts benchmarks, although some items still required calibration. The median calibration factor was 1.055590, indicating a modest overall upward adjustment across components.

The post-calibration check confirms exact reconciliation with the National Accounts targets within numerical tolerance for all COICOP items, with effectively zero remaining differences. Most calibration factors were close to unity, which suggests that the HBS-derived distribution was generally consistent with the macro benchmarks after EU-SILC weighting.

The largest upward adjustments were observed for CP091 (1.753), CP023 (1.377), and CP096 (1.304), followed by CP052, CP053, CP081, and CP093. These items were below their National Accounts targets in the EU-SILC-weighted HBS totals and therefore required stronger scaling. By contrast, CP043 (0.564) and

CP101 (0.716) required downward adjustment, indicating that their EU-SILC-weighted HBS totals exceeded the corresponding National Accounts benchmarks.

Overall, the results show that the fused file provides the distributional structure, while the calibration step ensures consistency with National Accounts control totals without altering the within-component household distribution, since only a component-specific multiplicative factor is applied.

FINAL VALIDATION OF THE HDA MASTER DATASET

Following the incorporation of the D8 adjustment, the construction of **CP_TOTAL_HDA** and the calculation of **B8g_HDA**, a final validation of the HDA master dataset was carried out prior to completing the OECD HDA template. The objective of the validation was to verify the technical integrity of the dataset, ensure consistency with the National Accounts benchmarks, and assess the plausibility of the distribution of income, consumption, D8 and saving across the newly constructed equivalised disposable income (B6g_eq) quintiles.

Technical integrity checks confirmed that the dataset contains **6,095 unique households**, with no duplicated household identifiers (*id_ms*), no missing or negative **DB090** household weights, and a total sum of weights identical to that used in the original EU-SILC dataset. The **B6g_eq** variable was independently recalculated as **B6g_HDA / cons_units_oecd**, with only negligible numerical differences attributable to floating-point precision. Likewise, **CP_TOTAL_HDA** was verified against the sum of all fifty **CPxxx_HDA** consumption components for each household, while **B8g_HDA** was independently recalculated as:

$$B8g\_HDA = B6g\_HDA + D8\_HDA - CP\_TOTAL\_HDA$$

In all cases, the recalculated values matched the stored variables.

Validation of the consumption block confirmed that all **50 CPxxx_HDA** variables were present in the dataset. No missing or negative household consumption values were identified, and the weighted aggregate of each COICOP consumption category exactly matched its corresponding National Accounts control total. Consequently, the household consumption block is fully consistent with the National Accounts at the aggregate level.

The newly constructed **B6g_eq** quintiles were also verified. Each quintile represents approximately 20% of the weighted household population, and the equivalised disposable income boundaries do not overlap. A transition analysis comparing the original SILC income quintiles (**kvpc.i**) with the new **B6g_eq_quintile** classification showed that only **0.858%** of the weighted household population moved by more than two quintiles. This indicates that the revised HDA income distribution preserves the overall household ranking and does not introduce excessive redistribution across income groups.

Following the inclusion of the D8 adjustment, household saving (**B8g_HDA**) was calculated as the difference between disposable income including D8 and final household consumption expenditure. The corresponding household saving rate was calculated as:

$$Saving\ Rate = \frac{B8g\_HDA}{B6g\_HDA + D8\_HDA}$$

The resulting distribution by income quintile is economically plausible. The first three quintiles exhibit negative aggregate household saving, the fourth quintile records a small positive saving balance, and the highest quintile shows substantially positive saving. The overall household saving rate is positive, amounting to approximately **8.6%.** This outcome is considered methodologically more appropriate than calculations excluding D8, as the adjustment captures changes in pension entitlements, which form part of household saving under the National Accounts framework.

Negative household saving remains common at the individual household level, particularly within the lower income quintiles. However, negative saving for individual households or lower-income groups should not automatically be interpreted as an error. Such observations may reflect the use of accumulated assets, credit-financed consumption, major one-off expenditures, life-cycle effects, temporarily low income, or the effects of matching between EU_SILC and HBS information. The OECD HDA methodological guidance recognises negative household saving as a common validation issue and recommends assessing not only average saving by income group but also the prevalence of negative saving and the distribution of household saving rates.

Additional conceptual consistency checks further confirmed the internal coherence of the dataset. Variables **D121** and **D611**, as well as **D122** and **D612**, matched exactly, reflecting the conceptual relationship between employers' social contributions and the corresponding household accounts transactions. The **D8 proxy** was independently reconstructed and matched the stored **D8_proxy_HDA** values. Pension benefits (**D62pens**) displayed the expected inverse relationship with **D8**, as pension benefit payments reduce accumulated pension entitlements. The distribution of **CP122** retained an almost perfect correlation with the FISIM allocation proxy, while differences in levels reflect the subsequent calibration to the National Accounts COICOP benchmark.

Overall, the validation results do not indicate any material technical or methodological inconsistencies. The HDA master dataset is considered suitable for completing the OECD HDA template. Although a small number of households exhibit exceptionally high household consumption or strongly negative household saving, these observations are consistent with the underlying microdata and the applied allocation methodology. They merit monitoring but do not constitute grounds for re-estimation of the household accounts. Given that the National Accounts control totals, accounting identities, household weights, quintile distributions and D8 allocation have all been successfully validated, the final HDA dataset is considered internally consistent and fit for distributional analysis.

INEQUALITY ASSESSMENT OF DISPOSABLE INCOME (B6G)

Income inequality was assessed before and after the HDA adjustment using the household disposable income indicator (B6g). The pre-adjustment results were based on the original EU-SILC disposable income variable, HY020, while the post-adjustment results were based on the final HDA-adjusted disposable income variable, B6g_HDA.

To ensure comparability, both income measures were equivalised using the modified OECD equivalence scale:

HY020_eq = HY020 / cons_units_oecd

B6g_eq = B6g_HDA / cons_units_oecd

The Gini coefficient and the quintile share ratio (Q5/Q1) were calculated using EU-SILC household weights (DB090). For the baseline estimates, income quintiles were derived from the equivalised EU-SILC income distribution (HY020_eq), whereas the post-adjustment estimates were based on the final HDA equivalised income distribution (B6g_eq).

A small number of households reported negative equivalised disposable income values. For the purpose of inequality measurement only, these observations were bottom-coded to zero prior to calculating the Gini coefficient and Q5/Q1 ratio. This treatment affected only the diagnostic indicators and did not modify the underlying income variables used in the HDA dataset.

The results indicate a moderate reduction in measured disposable income inequality following the HDA adjustment. The Gini coefficient decreased from 36.76% in the original EU-SILC distribution to 35.92% in the HDA-adjusted distribution, corresponding to a reduction of 0.84 percentage points. Similarly, the Q5/Q1 ratio declined from 6.969 to 5.966, indicating a lower disparity between households in the highest and lowest income quintiles.

These changes suggest that the HDA calibration process resulted in a somewhat more balanced distribution of disposable income across households while remaining consistent with the macroeconomic totals incorporated through the HDA framework.

The aggregate disposable income total increased from EUR 13,183.3 million in the original EU-SILC data to EUR 17,526.5 million after the HDA adjustment. This increase reflects the alignment of the household income distribution with the HDA and national accounts benchmarks.

In the original income distribution, 25 households had negative equivalised disposable income values, representing 0.436% of the weighted household population. As noted above, these observations were recoded to zero exclusively for the computation of the inequality indicators and were retained unchanged in the underlying microdata.

WAYS FORWARD

In the next iteration of the exercise, the allocation of household consumption will be refined by fully leveraging the results of the statistical matching process. Instead of relying on separate survey weights, the matched EU-SILC–HBS microdata file will be used directly, and EU-SILC household weights will be applied consistently across both income and consumption domains. This approach will improve internal consistency and ensure a unified weighting structure.

For the component D.63 (Other social benefits in kind), the current simplified proportional allocation will be replaced with a more granular, framework-driven approach. The component will be disaggregated by COFOG functions and allocated using targeted socio-demographic profiles:

Housing: to be distributed using a combination of microdata and administrative information on housing-related benefits to better target low-income and vulnerable households.

Social protection: to be further disaggregated according to COFOG functions (sickness, old age, family, unemployment, social exclusion) and allocated based on explicit eligibility criteria derived from social indicators.

Recreation, culture, and other services: alternative proxy indicators and distributional models will be explored to better identify likely beneficiary groups.

Overall, the next phase will further strengthen the use of microdata-based allocation methods, moving from a framework that combines proportional allocation techniques with socio-demographic modelling towards a more fully integrated microdata-based approach. This development is expected to improve both the distributive accuracy and methodological consistency of the estimates.

Work on the HDA is planned to continue from October 2027, following the finalisation of the National Accounts Non- financial Sector Accounts for the 2025 reference year. This timeline is aligned with the currently planned ESA 2030 requirements for HDA data transmission, which foresee mandatory reporting starting with the 2025 reference year onwards. In Latvia, the latest available Household Budget Survey (HBS) data relate to 2025, while EU-SILC data are available annually. This provides a strong basis for producing HDA results from the 2025 reference year onwards and for further strengthening the quality and robustness of the estimates.