

Statistical matching of EU-SILC and the Household Budget Survey to compare poverty estimates using income, expenditures and material deprivation

2013 edition

**Statistical matching of EU-SILC and the Household
Budget Survey to compare poverty estimates using
income, expenditures and material deprivation**

2013 edition

***Europe Direct is a service to help you find answers
to your questions about the European Union.***

**Freephone number (*):
00 800 6 7 8 9 10 11**

(*) Certain mobile telephone operators do not allow access to 00 800 numbers or these calls may be billed.

More information on the European Union is available on the Internet (<http://europa.eu>).

Cataloguing data can be found at the end of this publication.

Luxembourg: Publications Office of the European Union, 2013

ISBN 978-92-79-22844-5


ISSN 1977-0375

doi:10.2785/4151

Cat. No KS-RA-13-007-EN-N

Theme: Populations and social conditions
Collection: Methodologies & Working papers

© European Union, 2013
Reproduction is authorised provided the source is acknowledged.



Eurostat is the Statistical Office of the European Union (EU). Its mission is to be the leading provider of high quality statistics on Europe. To that end, it gathers and analyses data from the National Statistical Institutes (NSIs) across Europe and provides comparable and harmonised data for the EU to use in the definition, implementation and analysis of EU policies. Its statistical products and services are also of great value to Europe's business community, professional organisations, academics, librarians, NGOs, the media and citizens.

In the field of income, poverty, social exclusion and living conditions, the EU Statistics on Income and Living Conditions (EU-SILC) is the main source for statistical data at European level.

Over the last years, important progress has been achieved in EU-SILC as a result of the coordinated work of Eurostat and NSIs.

In June 2010, the European Council adopted a social inclusion target as part of the Europe 2020 Strategy: to lift at least 20 million people in the EU from the risk of poverty and exclusion by 2020. To monitor progress towards this target, the 'Employment, Social Policy, Health and Consumer Affairs' (EPSCO) EU Council of Ministers agreed on an 'at risk of poverty or social exclusion' indicator. To reflect the multidimensional nature of poverty and social exclusion, this indicator consists of three sub-indicators: i) at-risk-of-poverty (i.e. low income); ii) severe material deprivation; and iii) living in very low work intensity households.

In this context, the Second Network for the Analysis of EU-SILC (Net-SILC2) is bringing together National Statistical Institutes (NSIs) and academic expertise at international level in order to carry out in-depth methodological work and socio-economic analysis, to develop common production tools for the whole European Statistical System (ESS) as well as to ensure the overall scientific organisation of the third and fourth EU-SILC conferences. The current working paper is one of the outputs of the work of Net-SILC2. It was presented at the third EU-SILC conference (Vienna, December 2012), which was jointly organised by Eurostat and Net-SILC2 and hosted by Statistics Austria.

It should be stressed that this methodological paper does not in any way represent the views of Eurostat, the European Commission or the European Union. This is independent research which the authors have contributed in a strictly personal capacity and not as representatives of any Government or official body. Thus they have been free to express their own views and to take full responsibility both for the judgments made about past and current policy and for the recommendations for future policy.

This document is part of Eurostat's Methodologies and working papers collection, which are technical publications for statistical experts working in a particular field. These publications are downloadable free of charge in PDF format from the Eurostat website:

http://epp.eurostat.ec.europa.eu/portal/page/portal/income_social_inclusion_living_conditions/publications/methodologies_and_working_papers.

Eurostat databases are also available at this address, as are tables with the most frequently used and requested short- and long-term indicators.

TABLE OF CONTENT

1. Introduction	9
2. Statistical matching	11
2.1 Overview of statistical matching	11
2.2 Reconciliation of the data sources	11
2.2.1 Household	12
2.2.2 Household reference person	12
2.2.3 Population	12
2.2.4 Reference Period	12
2.3 Harmonization of variables	12
2.4 Choosing the matching variables	13
2.4.1 Coherence of distributions	13
2.4.2 Explanatory power of the variables	19
2.5 Matching methods	19
3. Results of statistical matching	21
3.1 Comparison of mean expenditure by expenditure deciles – EU-SILC imputed versus HBS observed	21
3.2 Comparison of joint distributions of matching variables with expenditure deciles	22
3.3 Comparison of expenditure by matching variables - EU-SILC imputed versus HBS observed	23
3.3.1 Comparison of expenditure by matching variables – observed versus imputed HBS	25
3.4 Comparison of expenditure by variables not used in statistical matching	26
4. Conditional independence assumption	27

5. Income and expenditure poverty and material deprivation.....	29
5.1 Headline poverty indicators	29
5.2 Overlap of income and expenditure poverty and material deprivation.....	30
6. Conclusions and next steps	35
6.1 Recommendations for future data collection.....	35
7. References	37
8. Annexes.....	39
Annex 1: Complete list of common variables EU-SILC and HBS, 2005.....	39
Annex 2: Complete list of derived variables.....	48
Annex 3: Output for material deprivation logistic regression model, EU-SILC 2005	51
Annex 4: Output for expenditure regression model, HBS 2005	52

Statistical matching of EU-SILC and the Household Budget Survey

Using synthetic matching techniques to facilitate the comparison of poverty estimates using income, expenditures and material deprivation

(Dominic WEBBER and Richard TONKIN⁽¹⁾)

Abstract: The Europe 2020 social inclusion target will be measured through work attachment, income and material deprivation indicators using the EU Statistics on Income and Living Conditions (EU-SILC). However, there has been increasing interest in recent years in whether expenditure and consumption provide more appropriate measures of standards of living than income. This Net-SILC2 work package therefore aims to compare people's exposure to poverty using three different measures: income, expenditure and material deprivation. However, no single data source provides joint information on all these variables. Therefore, this paper describes methodological work conducted to statistically match expenditure from the Household Budget Survey (HBS) with income and material deprivation contained within EU Statistics on Income and Living Conditions (SILC). The three matching approaches used are parametric, non-parametric and mixed. Overall, the mixed methods approach tends to perform slightly better at matching expenditure, based on a variety of measures. The implications of this work for the ongoing review of the EU-SILC legal basis are discussed.

⁽¹⁾ Richard Tonkin and Dominic Webber are from the UK Office for National Statistics (ONS). This work has been supported by the second Network for the analysis of EU-SILC (Net-SILC2), funded by Eurostat. The European Commission and ONS bear no responsibility for the analyses and conclusions, which are solely those of the authors. Email address for correspondence: richard.tonkin@ons.gov.uk. The authors would like to thank Anthony Atkinson, Eric Marlier, Marcello D'Orazio, Aura Leulescu and David Gordon along with Andrew Barnard and other ONS colleagues for their helpful comments and discussions.

1. Introduction

Most policy initiatives aimed at improving living standards tend to measure poverty relatively within the society, using income as a yardstick. However, there is an argument that income isn't sufficient as a sole measure of poverty, particularly if one considers poverty in terms of achieved standards of living⁽¹⁾.

It is the consumption of goods and services, along with other inputs such as time, which ultimately satisfies a household's wants. Because of this, consumption is arguably a more important determinant of economic well-being than income alone. Indeed, Brewer & O'Dea (2012) and others (see Noll, 2007, for a review) argue that it is preferable to consider the distribution of consumption rather than income on both theoretical and pragmatic grounds.

On a theoretical ground, income can be subject to fluctuations, due to such events as short-term unemployment. However, these fluctuations in income are not likely to be matched by corresponding downturns in living standards, as households are typically able to smooth consumption by drawing on savings or help from family members. This finding leads to Friedman's 'permanent income hypothesis', which suggests that decisions made by consumers are based on long-term income expectations rather than their current income. This view is supported in a number of studies (e.g. Cutler & Katz, 1991, and Jorgenson & Slesnick, 1987) which find stronger relationships between consumption and subjective well-being than between income and subjective well-being measures.

Beyond these conceptual arguments, there is also the practical consideration that evidence from a range of countries suggests a general tendency for income to be under-reported by households with low levels of resources, whilst reporting of expenditure by this group is relatively accurate (e.g. Meyer & Sullivan, 2011 and Brewer & O'Dea, 2012), though other evidence suggests that expenditure of higher income households may be under-reported (Sabelhaus, et al., 2011).

In economic and social research, data on household expenditure are typically used as a proxy for consumption. These data are often collected through the use of diary studies. However, it should be noted that expenditure is an imperfect measure of consumption as the amount spent by a household in a given month may differ from consumption, due to households making use of goods purchased previously or the purchase of consumer durables. In addition, consumption also includes inter-household in-kind transfers of gifts and services and social transfers in kind. However, these aspects of consumption are generally excluded from data due to the challenges of collecting this type of information.

Overall the evidence indicates that while income can be a good proxy for living standards, it is better when supplemented with a wider range of measures such as expenditure. This is consistent with the recommendations of the *Report by the Commission on the Measurement of Economic Performance and Social Progress* (Stiglitz, Sen, & Fitoussi, 2009).

This Net-SILC2 work package therefore aims to compare people's exposure to poverty using three different measures: income, expenditure and material deprivation, across countries of the EU. However, there is no data source which provides joint information on all of these variables for households or individuals. Therefore, a major component of this project is to statistically match expenditure from the Household Budget Survey (HBS) with income and material deprivation contained within EU Statistics on Income and Living Conditions (SILC). This paper reports results from a variety of different methods of statistically matching HBS and EU-SILC for the UK using 2005 data.

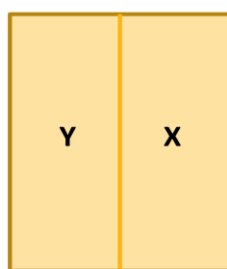
⁽¹⁾ As well as considering poverty in terms of an individual's standard of living, other approaches are possible, such as considering poverty in terms of a right to a minimum level of resources (see (Atkinson, Cantillon, Marlier, & Nolan, 2002) for a discussion).

2. Statistical matching

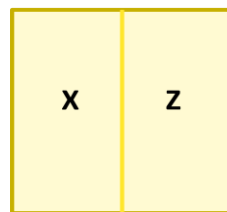
2.1 Overview of statistical matching

Statistical (or synthetic) matching is a broad term used to describe the fusing of two datasets. In this context, the datasets are of households sampled from the same population. The usual approach is to define one data set as the recipient, in this case EU-SILC, and one as the donor, HBS. The recipient data contains a variable Y, in this case material deprivation, which is not found in the donor, while variable Z, expenditure, is only contained within the donor. The aim is to use information contained within the set of variables common to both datasets, X, to link records from the donor to the recipient. Therefore, expenditure will be linked to EU-SILC, which contains information on income and material deprivation.

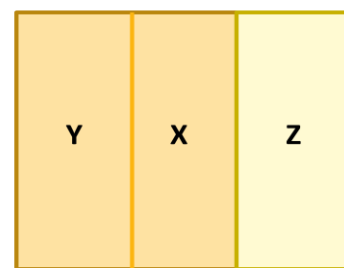
Recipient dataset (EU-SILC)



Donor dataset (HBS)



Matched dataset



Statistical matching approaches are typically classified into one of three categories:

- Non-parametric methods
- Parametric methods
- Mixed methods

These are described in more detail in section 2.5 below.

2.2 Reconciliation of the data sources

In order for statistical matching to be a success, it is vital that steps are taken to ensure the donor and recipient datasets, the variables and their distributions are comparable. D'Orazio, Di Zio, & Scanu (2006 pg 164) outline the following eight steps for achieving this:

1. Harmonization of the definition of units.
2. Harmonization of reference periods.
3. Completion of population.
4. Harmonization of variables.
5. Harmonization of classifications.
6. Adjustment for measurement errors (accuracy).
7. Adjustment for missing data.
8. Derivation of variables.

Both EU-SILC and HBS are produced from household surveys, and, in the UK are derived from the General Lifestyle Survey (GLF) and Living Costs and Food Survey (LCF) respectively. The LCF asks each individual aged 16 and over in the household visited to keep diary records of daily expenditure for two weeks. Information about regular expenditure, such as rent and mortgage payments, is obtained from a household interview along with retrospective information on certain large, infrequent expenditures such as those on vehicles. The fact that both the GLF and LCF are implemented by UK Office for National Statistics (ONS) and covered by the same standards and guidance is a significant advantage in this matching project. However, it is still a worthwhile and, indeed, necessary exercise to ensure that the key concepts of definition of household, household reference person and income reference period are comparable between the datasets.

2.2.1 Household

The concept of a household is similarly defined for both HBS and EU-SILC. This definition states that a household is constituted by a person or people living together in the same dwelling who share meals or joint provision of living conditions. This concept has been implemented fully on both the GLF and LCF.

2.2.2 Household reference person

Since 2001/2002 the concept of household reference person (HRP) has been adopted on all UK Government sponsored surveys. Therefore, the definition of HRP is the same on EU-SILC and HBS. The HRP is the householder who:

- owns the household accommodation, or
- is legally responsible for the rent of the accommodation, or
- has the household accommodation as an emolument or perquisite, or
- has the household accommodation by virtue of some relationship to the owner who is not a member of the household.

If there are joint householders the household reference person will be the one with the higher income. If the income is the same, then the eldest householder is selected.

2.2.3 Population

Both sources cover the same population (UK private households – excluding collective establishments) and use the same sampling frame (the Postcode Address File – a list of addresses provided by the UK Post Office).

2.2.4 Reference Period

EU-SILC in the UK measures current income, as opposed to income in the previous financial year, which tends to be the practice in other European States. Therefore, the 2005 UK EU-SILC dataset measures current annual income in 2005. The HBS also collects current annual income but its reference period covered the 2005/2006 financial year. However, HBS expenditure and income was deflated to 2005 prices prior to transmission to Eurostat.

The 2010 UK HBS, which will become available in the near future, will revert to calendar year for the income reference period.

2.3 Harmonization of variables

Annex 1 contains the full list of variables common to both EU-SILC and HBS. These variables needed to be harmonized across the two datasets in order to be used for the matching. This involved recoding of variables to the stage where they have the same degree of detail. The table in Annex 2 shows the codification of these derived variables.

The region variable, for instance, in EU-SILC is available at NUTS2 level, while in HBS it is only at NUTS1. In order to derive a variable that is harmonized between the two data sets, the higher level of detail in EU-SILC has to be reduced to resemble HBS. Similarly, the HBS variable which defines activity status is more detailed than the corresponding EU-SILC variable. The detail in HBS therefore has to be sacrificed to ensure that it is comparable with EU-SILC. This highlights a constraining factor in statistical matching – that detailed information on one survey is lost unless the corresponding variable on the other data set is available at the same level.

Once the variables have been harmonized a check for missing information was performed because some of the statistical matching methods used rely on regressions. If a variable has missing information in one case, that whole case is omitted from the regression, thereby losing potentially valuable information from the other variables. For this reason, a number of variables were excluded from further analysis - the most notable ones being occupation and employment status, due to missing information on the HBS.

2.4 Choosing the matching variables

The variables selected for matching must fulfil two criteria. First, there must be similarity in the distributions of the variables across the two surveys. Second, the variables must be significant in explaining variations in the target variables – in this case expenditure and material deprivation.

2.4.1 Coherence of distributions

The literature highlights two main methods for calculating the degree to which distributions of variables are similar across data sets. The first is a simple comparison of the weighted frequency distributions of the derived variables in the two datasets. The second is to use a measure such as the Hellinger Distance (HD). The HD is convenient because it provides a single number as a measure for the similarity in distribution of two variables. It is generally considered that an HD of over 5% should raise concerns about the similarities in distributions. The equation used to derive the HD is:

$$HD(V, V') = \sqrt{\frac{1}{2} \sum_{i=1}^K \left(\sqrt{\frac{n_{O_i}}{N_O}} - \sqrt{\frac{n_{P_i}}{N_P}} \right)^2}$$

Variable V is in the donor data set, V' in the recipient, K is the total number of cells in the contingency table, n_{O_i} is the frequency of cell i in the original data O , n_{P_i} is the frequency of cell i in the recipient and N is the total size of the specific sources.

Table 1 shows the weighted frequencies and HD of the common variables in the EU-SILC and HBS datasets. Tenure status, DV_TENURE, has an HD of 12.6% due to a large discrepancy in the proportion of people reporting paying rent at market price, and paying reduced or subsidized rent. This is likely to be due to differences in the way the question was surveyed. To resolve this issue, these two possible outcomes are recoded to one, simply representing respondents who pay rent. This reduces the HD for this variable to 1%, thereby ensuring that it is suitable as a potential matching variable. However, by limiting the possible outcome responses in the variable to three reduces its variation, thereby making it potentially less likely to be useful in explaining variations in material deprivation or expenditure.

Household type initially had an HD of 8.5%. Therefore, ‘All other households’ – the outcome with a high divergence – was recoded to be treated as a missing observation and therefore would be excluded from the analysis. Although this reduces the HD to 4.8%, it also removes over 20% of the observations from the HBS. For this reason, the household type variable was dropped from the analysis⁽²⁾.

⁽²⁾ Although this variable was not included in the final iteration of the statistical matching due to the dissimilarity of its distribution in the two data sources, work was carried out to determine whether it would have been useful as a matching variable or not. The original version of the variable was included in a run of the models, which are described in more detail in the next section. Although household type has some explanatory power in household expenditure, it is insignificant when considering material deprivation. This means that it wouldn't have been included as a matching variable, even if the distributions had been comparable between EU-SILC and HBS.

Other variables dropped due to a lack of similarity in distributions were marital status, DV_MARSTA and number of rooms, DV_ROOMS. Marital status had an HD of 9.0%, primarily driven by the difference between the two sources in terms of the proportion of people identified as being married. This was due to cohabitation being included in the ‘married’ category for HBS but not for EU-SILC. This highlights again the importance for effective statistical matching of ensuring that definitions for common variables being harmonised across data sources.

Table 1: Comparison of distributions of EU-SILC and HBS variables, 2005

Variable	EU-SILC		HBS		HD (%)
	Absolute frequency	Relative frequency (%)	Absolute frequency	Relative frequency (%)	
DV_SEX					0.15
Male	15,641	61.7	15,244	61.5	
Female	9,716	38.3	9,554	38.5	
DV_AGE					0.93
16-25	1,182	4.7	1,096	4.4	
26-35	3,892	15.4	3,857	15.6	
36-45	5,262	20.8	5,202	21.0	
46-55	4,659	18.4	4,459	18.0	
56-65	4,070	16.1	4,120	16.6	
66-75	3,271	12.9	3,233	13.0	
75 plus	3,021	11.9	2,832	11.4	
DV_MARSTA (Marital status)					9.01
Never Married	6,031	23.8	3,818	15.4	
Married	11,891	46.9	14,447	58.3	
Widowed	3,259	12.9	3,033	12.2	
Divorced or separated	4,175	16.5	3,500	14.1	
DV_CONUNI (Consensual union)					1.54
Yes, on a legal basis	11,891	46.9	12,073	48.7	
Yes, without a legal basis	2,336	9.2	2,375	9.6	
No	11,130	43.9	10,351	41.7	
DV_ACTSTA (Activity status)					2.45
At work	15,331	61.5	14,697	59.3	
Unemployed	455	1.8	459	1.9	
In retirement or early retirement or has given up business	6,379	25.6	6,335	25.5	
Other inactive person	2,777	11.1	3,308	13.3	
DV_LABOUR					0.51
Full time	12,010	85.5	12,650	86.0	
Part time	2,037	14.5	2,060	14.0	

Variable	EU-SILC		HBS		HD (%)
	Absolute frequency	Relative frequency (%)	Absolute frequency	Relative frequency (%)	
DV_REGION					1.24
North East	1,133	4.5	1,061	4.3	
North West	2,854	11.3	2,840	11.5	
Yorkshire and the Humber	2,270	9.0	2,194	8.9	
East Midlands	1,853	7.3	1,748	7.1	
West Midlands	2,156	8.5	2,168	8.7	
Eastern	2,219	8.8	2,213	8.9	
London	2,996	11.8	2,826	11.4	
South East	3,497	13.8	3,548	14.3	
South West	2,176	8.6	2,163	8.7	
Wales	1,242	4.9	1,278	5.2	
Scotland	2,254	8.9	2,142	8.6	
Northern Ireland	707	2.8	617	2.5	
DV_HHSIZE					2.12
1	7,864	31.0	7,090	28.6	
2	8,792	34.7	8,961	36.1	
3	3,804	15.0	3,964	16.0	
4	3,349	13.2	3,232	13.0	
5	1,099	4.3	1,048	4.2	
6	450	1.8	504	2.0	
DV_ROOMS					29.48
1 room	207	0.8	97	0.4	
2 rooms	2,343	9.2	323	1.3	
3 rooms	5,769	22.8	1,815	7.3	
4 rooms	7,595	30.0	4,672	18.8	
5 rooms	5,189	20.5	6,497	26.2	
6 rooms	2,444	9.6	5,238	21.1	
7 rooms	1,085	4.3	2,850	11.5	
8 rooms or more	727	2.9	3,305	13.3	

Variable	EU-SILC		HBS		HD (%)
	Absolute frequency	Relative frequency (%)	Absolute frequency	Relative frequency (%)	
DV_TENURE					12.62
Owner	17,406	68.7	17,371	70.1	
Tenant or subtenant paying rent at prevailing or market price	3,939	15.6	5,849	23.6	
Accommodation is rented at reduced rate (lower than market price)	3,626	14.3	1,259	5.1	
Accommodation is provided free	360	1.4	321	1.3	
DV_CAR					1.50
Yes	19,196	75.7	18,321	73.9	
No	6,157	24.3	6,478	26.1	
DV_TV					0.84
Yes	25,007	98.6	24,521	98.9	
No	350	1.4	278	1.1	
DV_PC					0.27
Yes	16,264	64.1	15,996	64.5	
No	9,093	35.9	8,803	35.5	
DV_DWELL					6.90
Detached House	5,725	22.6	5,688	22.9	
Semi detached or terraced house	14,799	58.4	14,376	58.0	
Apartment	4,778	18.8	4,240	17.1	
Other	54	0.2	495	2.0	
inc_band (Disposable income per year)					5.03
Less than £5,000 per year	981	3.9	978	4.0	
Between £4,999 and £10,000	4,330	17.1	3,831	15.5	
Between £9,999 and £15,000	4,690	18.5	3,904	15.7	
Between £14,999 and £20,000	3,580	14.1	3,105	12.5	
Between £19,999 and £30,000	5,210	20.5	5,127	20.7	
Between £29,999 and £40,000	3,010	11.9	3,464	14.0	
Between £39,999 and £50,000	1,670	6.6	1,953	7.9	
Greater than £49,999	1,890	7.5	2,437	9.8	

Variable	EU-SILC		HBS		HD (%)
	Absolute frequency	Relative frequency (%)	Absolute frequency	Relative frequency (%)	
DV_HHTYPE					8.52
ONE ADULT HOUSEHOLD					
One person, aged 65 years or more	3,334	13.2	3,183	13.0	
One person, aged 30 to 64 years	3,737	14.7	3,419	13.9	
One person, under 30 years	691	2.7	488	2.0	
One person with children up to 16 years old (exclusive)	2,028	8.0	1,260	5.1	
COUPLE WITHOUT CHILDREN					
Couple without children, older member aged 65 or more	2,646	10.4	2,001	8.1	
Couple without children, older member under 65	3,001	11.8	3,054	12.4	
COUPLE WITH CHILDREN AGED UP TO 16 YEARS OLD (EXCLUSIVE)					
One child	1,090	4.3	1,014	4.1	
Two children	1,380	5.4	1,176	4.8	
Three or more children	483	1.9	407	1.7	
OTHER					
Single parent or couple with at least one child of 16 years old or older	3,085	12.2	2,814	11.4	
All other households	3,882	15.3	5,771	23.5	

Source: UK Office for National Statistics/Eurostat, author's computation

2.4.2 Explanatory power of the variables

D’Orazio et al (2006) identifies the following method for choosing the matching variables from the set of common variables:

1. Let ψ_A consist of all the common variables such that ψ_A is independent of Y given the other common variables in the recipient data set.
2. Let ψ_B consist of all the common variables such that ψ_B is independent of Z given the other common variables in the donor data set.
3. Let $\psi = \psi_A \cap \psi_B$; then the other common variables define X, the matching variables.

Therefore, the common variables which should be used for matching are those that are statistically significant in explaining variations in both expenditure and material deprivation.

Predictors of material deprivation

Material deprivation was defined as a binomial variable, taking a value of 1 if the respondent was materially deprived, and 0 otherwise. The following logistic regression was fitted:

$$\text{Deprived} = \text{DV_REGION} + \text{DV_AGE} + \text{DV_ACTSTAT} + \text{DV_CONUNI} + \text{DV_TENURE} + \text{DV_SEX} + \text{DV_DWELL} + \text{DV_HHSIZE} + \text{DV_CAR} + \text{DV_TV} + \text{DV_PC} + \text{inc_band}$$

The full regression output is located in Annex 3, and reveals that the variables that significantly explain deprivation in EU-SILC were: DV_REGION, DV_AGE, DV_ACTSTAT, DV_CONUNI, DV_TENURE, DV_DWELL, DV_HHSIZE, DV_CAR, and inc_band.

Predictors of expenditure

Next, an expenditure model was estimated on HBS data. As expenditure is highly positively skewed, the stepwise regression model was estimated on the logarithm of expenditure.

$$\text{Logexp} = \text{DV_REGION} + \text{DV_AGE} + \text{DV_ACTSTAT} + \text{DV_CONUNI} + \text{DV_TENURE} + \text{DV_DWELL} + \text{DV_HHSIZE} + \text{DV_CAR} + \text{DV_TV} + \text{DV_PC} + \text{inc_band}$$

The full regression output is located in Annex 4. It reveals that the statistically significant predictors of expenditure were DV_REGION, DV_AGE, DV_TENURE, DV_DWELL, DV_SEX, DV_HHSIZE, DV_CAR, DV_TV, DV_PC, and inc_band. The adjusted R² for the final model was 0.56.

As stated above, the variables that should be selected for matching are those which are significant in explaining material deprivation and expenditure. Therefore, the final matching variables are:

- DV_REGION – The region in the household resides.
- DV_AGE – The age group of the HRP.
- DV_TENURE – The tenure status of the household.
- DV_DWELL – The dwelling type of the household.
- DV_HHSIZE – The household size.
- DV_CAR – Whether the household has a car.
- inc_band – The disposable income group of the household.

2.5 Matching methods

Three different matching methods were used in this analysis, covering the three broad categories of approaches mentioned in section 2.1.

The hotdeck method is a non-parametric approach. The procedure finds records in the donor file and matches them with records in the recipient file, based on a distance function. This results in actual observed values, for expenditure in this case, being imputed onto EU-SILC. A disadvantage of this

procedure, and especially relevant in this scenario, is that the multiple usage of donors is necessary as the donor dataset, HBS, is smaller than the recipient, EU-SILC. This can increase the risk that the distribution of the imputed variable does not reflect the original one.

The second (parametric) approach involves imputing predicted values obtained from a regression model. The reliability of this method is very much dependent on the accuracy of the model. In addition, regression towards the mean can be a potential problem with this approach.

Mixed methods, as the name implies, involves a combination of parametric and non-parametric techniques. A model is first fitted to the data to estimate an intermediate value of the variable to be matched. Then a distance function is used to locate a range of possible observations from the donor set which most closely resembles the intermediate value, with a value for imputation selected from that set. In the variant of the method used, this process is performed multiple times, producing multiple imputed datasets. This builds in some allowance for uncertainty in the model. Analysis was carried out on each imputed dataset, before the results were averaged across the imputed datasets produce one overall set of estimates.

3. Results of statistical matching

Testing the validity of matching procedures involves comparing the distributions of the matched variables against observed expenditure observed in the HBS. This is done in four ways:

- By comparing mean expenditure by expenditure decile to analyse the consistency of the overall expenditure distribution for each method.
- By using estimates based on the Hellinger Distance to provide a measure of the similarity of the joint distributions of the matching variables with expenditure (observed and imputed).
- By comparing the consistency of mean expenditure by variables used in the statistical matching for observed and imputed expenditure.
- By comparing the relationship between expenditure and variables in both datasets but not included in the model.

3.1 Comparison of mean expenditure by expenditure deciles – EU-SILC imputed versus HBS observed

Figure 1: Mean expenditure by expenditure decile for HBS and different matching methods, UK, 2005 (£ per week)

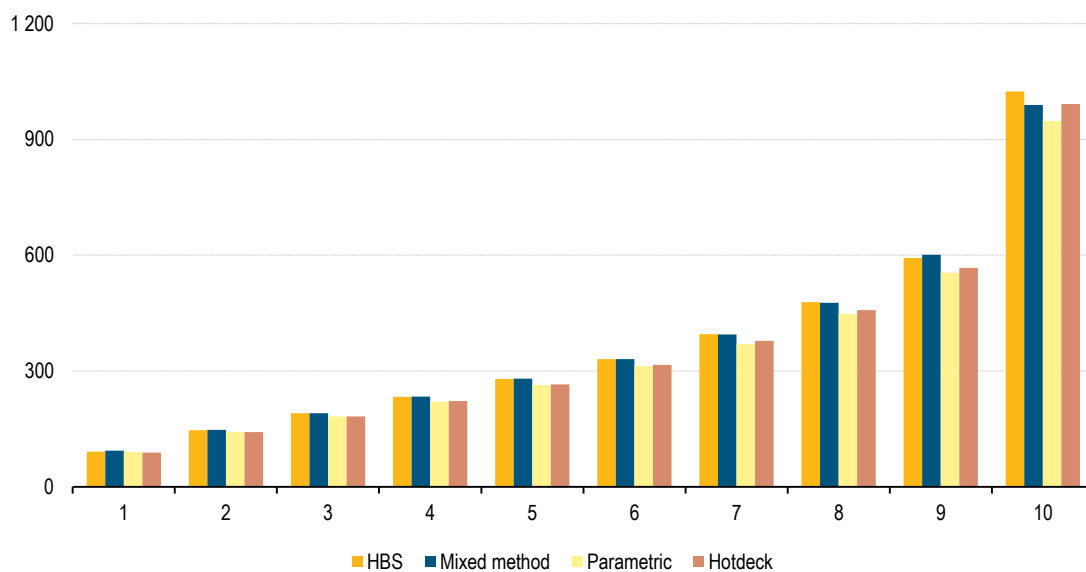


Figure 1 provides an indication of the performance of the different matching methods across the expenditure distribution. All three methods appear to be relatively effective in replicating mean expenditure by expenditure deciles. On average, there is a difference of £5 per week difference between the parametric approach and actual observed expenditure from the HBS. The average difference for the mixed methods and hotdeck approaches are £15 and £23 respectively, indicating that the values from the parametric method are closest to the observed, followed by mixed methods. In general, all three methods tend to underestimate mean expenditure by expenditure decile compared with observed expenditure from the HBS, particularly at the upper end of the distribution.

3.2 Comparison of joint distributions of matching variables with expenditure deciles

Figure 2: Comparison of joint distribution of imputed and actual expenditure deciles by matching variable, UK, 2005 (%)

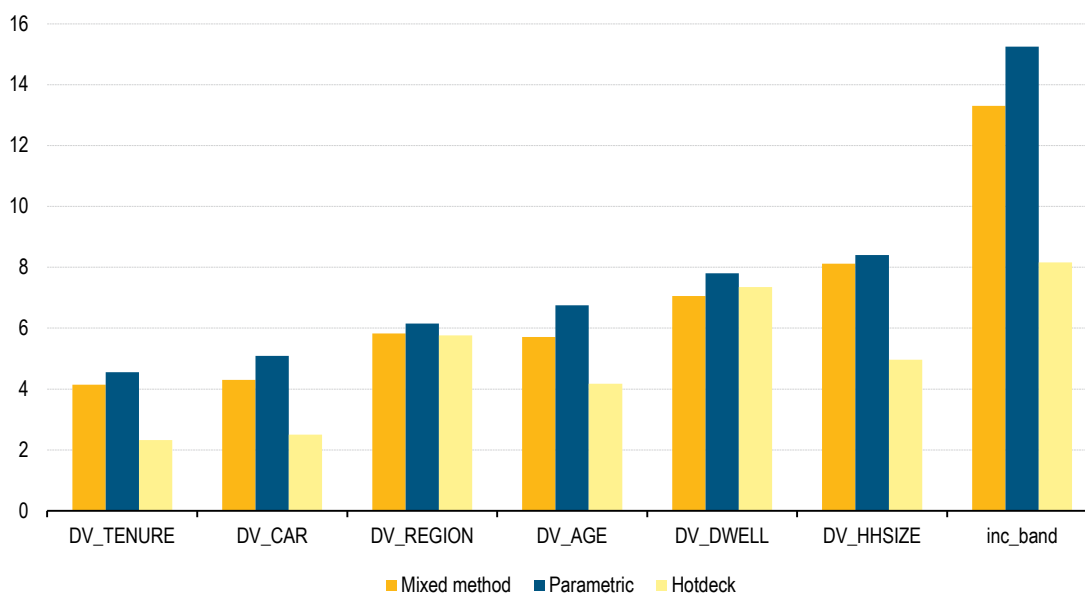


Figure 2, above, provides a measure of the similarity of the joint distributions of imputed and actual expenditure deciles, by each matching variable. It is based on the Hellinger Distance statistic presented earlier, where the lower the statistic, the greater the similarity of distributions.

Based on these measures, it would appear that both the mixed methods and hotdeck procedures are most effective in replicating the joint distribution of the HBS expenditure variable and the matching variable. The hotdeck approach performs the best across some variables, notably DV_TENURE2. However, the mixed method approach performs better across other variables, for instance DV_HHSIZE and inc_band. For all matching variables, the joint distributions are most dissimilar when the parametric approach is used.

3.3 Comparison of expenditure by matching variables - EU-SILC imputed versus HBS observed

Figure 3: Mean household expenditure by household disposable income band for HBS and matching methods, UK, 2005 (£ per week)



Figure 3 shows the distribution of actual expenditure, and expenditure derived from the matching methods across the income distribution. All three methods appear to perform well, in general. At the low end of the income distribution we see the expected expenditure ‘tick’ – higher average expenditure for the bottom income group than households in the second income group. However, the mixed method, and parametric approach seem to slightly underestimate the extent of the tick. In contrast, the hotdeck method appears to overestimate it.

Another way of looking at the extent to which the relationship between income and expenditure is preserved is to consider the correlation between positions in income and expenditure distributions using the Phi coefficient (Table 2). The Phi coefficient takes a value between 0 and 1, with 1 indicating perfect correlation and 0 representing no relationship. The correlation between income and expenditure quintile correlations is approximately 0.5. The mixed methods approach does the best at replicating the observed correlation between income and expenditure quintiles. The parametric approach is the least effective on this measure.

Table 2: Income and expenditure quintile correlations (Phi Coefficient), UK, 2005

Income/Expenditure quintile correlation	
HBS	0.49
Mixed method	0.46
Parametric	0.44
Hotdeck	0.45

Source: UK Office for National Statistics/Eurostat, author's computation

Figure 4: Mean household expenditure by age of HRP for HBS and matching methods, UK, 2005 (£ per week)

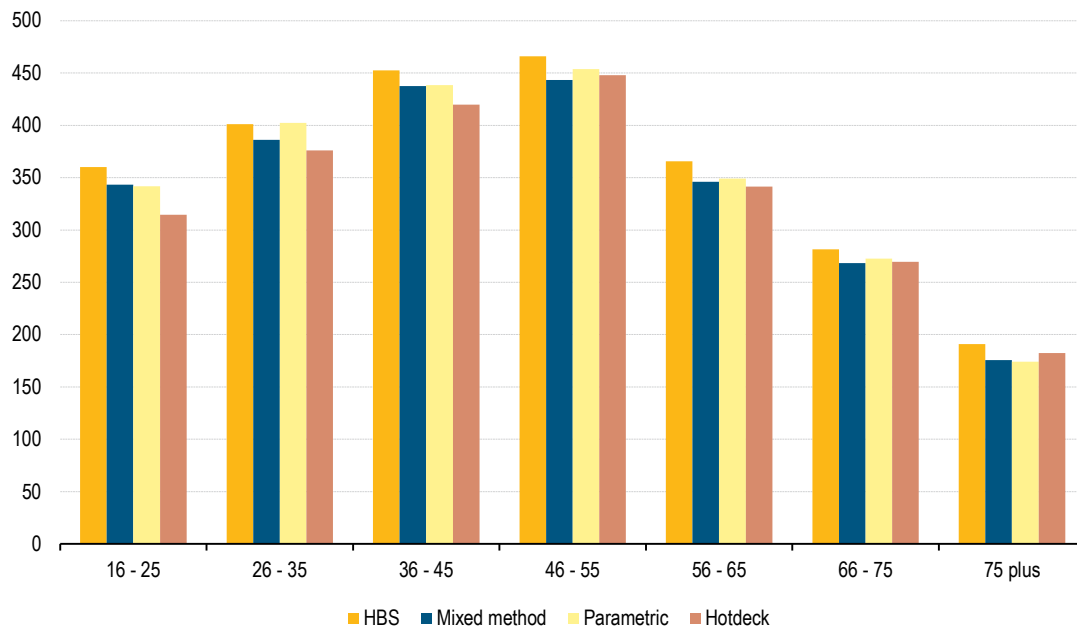
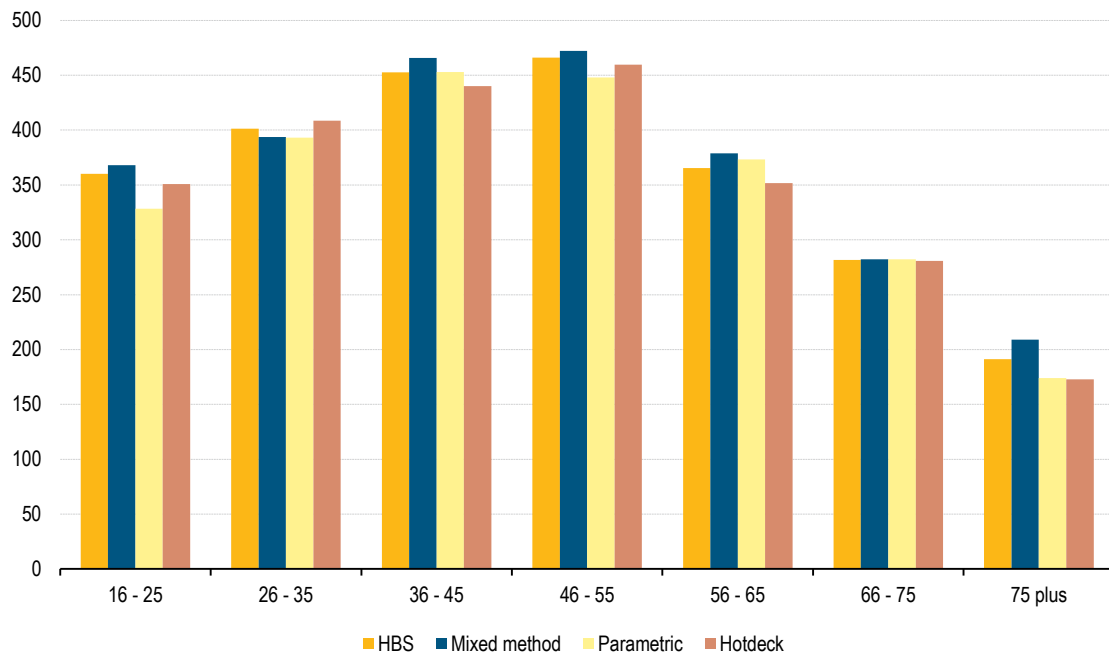


Figure 4 shows the distribution of actual expenditure and matched expenditure by age group. All three matching methods were broadly effective at replicating the pattern of expenditure across the different age groups, but the actual mean value of expenditure for each age group appears to be underestimated. A possible explanation for this is the difference in income between HBS and EU-SILC. Average weekly disposable income in the HBS was £500 per week, compared with £462 a week on EU-SILC. As income is being used as a matching variable, the lower values of disposable income in EU-SILC may help to explain this apparent systematic under-representation of expenditure.

3.3.1 Comparison of expenditure by matching variables – observed versus imputed HBS

One way of investigating this hypothesis is to artificially remove expenditure from a random selection of half the HBS sample and then impute expenditure back on using each of the three methods. Figure 5 shows the distribution of mean expenditure across the age distribution using this approach.

Figure 5: Mean household expenditure by age of HRP for HBS observed and HBS imputed, UK, 2005 (£ per week)



In this figure, the apparent underestimation of expenditure for each method is no longer present. This therefore supports the view that the observed under-estimation in the EU-SILC imputed data is due to the fact that the matching variables, most notably income, differ in their distributions between the sources, rather than any issues with the matching models themselves.

3.4 Comparison of expenditure by variables not used in statistical matching

Figure 6: Mean household expenditure by household type for HBS and matching methods, UK, 2005 (£ per week)

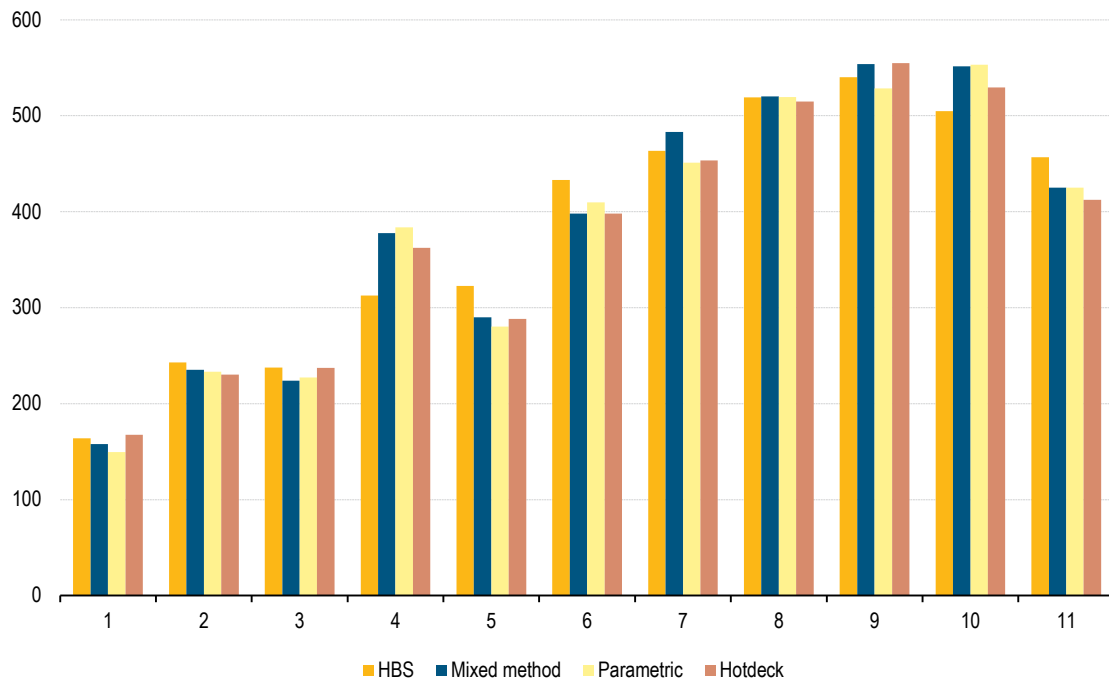


Figure 6 shows the relative performances of the matching methods at estimating expenditure across a variable not used in the matching process – household type. All three perform reasonably well, although there are some over/under-estimation of expenditure for certain types of household for all methods. In particular, expenditure appears to be overestimated for single adult/parent households (category 4), but underestimated for couples without children (categories 5 and 6).

4. Conditional independence assumption

All three statistical matching techniques described in this paper implicitly assume conditional independence, that is, given knowledge of X (matching variables), knowledge of Y (material deprivation) provides no information on the value of Z (expenditure) and vice versa. D’Orazio et al (2006) notes that, in statistical matching, this assumption is both particularly strong and, unfortunately, rarely holds in practice. The absence of conditional independence may result in incorrect inferences being made when analysing data produced through statistical matching.

Conditional independence cannot be tested from the matched datasets. It is possible to avoid making the conditional independence assumption by incorporating some auxiliary information (either at the micro or macro level). Therefore, for the purpose of studying the relationship between income and expenditure in the matched dataset, the CIA is avoided by the use of `inc_band` as a matching variable. However, such auxiliary information is not immediately available in the case of expenditure and material deprivation.

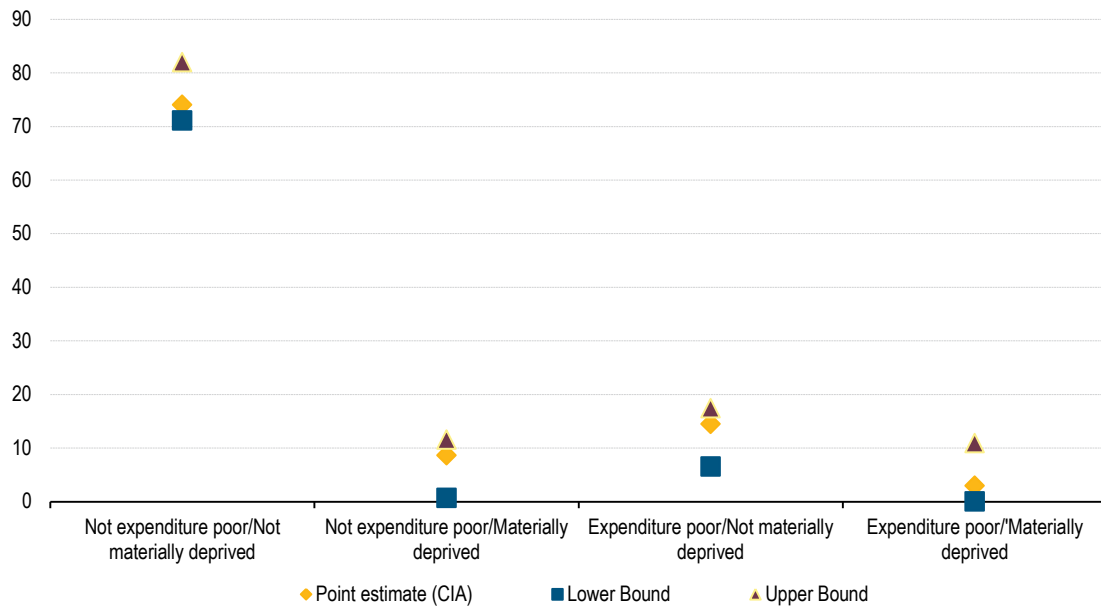
An alternative approach to statistical matching is to evaluate the uncertainty regarding an estimate of the parameter of interest. In particular the ESSNet on Statistical Integration highlighted the use of Fréchet bounds in order to estimate the range of plausible values that it can hold. The insight provided by this kind of uncertainty analysis can be useful to assess the plausibility of the conditional independence assumption.

Fréchet bounds have therefore been calculated for the contingency table between material deprivation and expenditure. As it is necessary for both the Y and Z variables to be categorical when calculating these bounds, expenditure has been recoded as a single binary variable relating to whether a household’s equivalised⁴⁽³⁾ expenditure is greater or less than 60% of the median.

In order to accurately calculate Fréchet bounds it is necessary to first to harmonise the joint distribution of the matching variables (Renssen, 1998). This harmonisation is extremely difficult to carry out successfully with a large number of matching variables. For this reason, just two matching variables were used in this process: `inc_band` and `DV_HHSIZE`.

⁽³⁾ Expenditure was equivalised using the modified-OECD equivalence scale. Although this and other scales were primarily developed for income analysis, they were designed to reflect the economies of scale achieved in consumption by households comprising more than one person. Therefore these scales are equally applicable for expenditure measures.

Figure 7: Fréchet bounds for severe material deprivation by expenditure poverty, UK, 2005 (%)



When using just these two variables, analysis of the intervals computed with Fréchet bounds shows that the uncertainty space is relatively large (Figure 7): the lower and upper bounds when conditioning on the common variables are approximately 10 percentage points apart. However, it is likely that the use of a greater number of matching variables would reduce this range of plausible values.

5. Income and expenditure poverty and material deprivation

The overall aim of this Net-SILC2 work-package is to conduct analysis of income, material and expenditure based poverty across various EU countries. Whilst this paper is primarily focused on the methodological work necessary to create the datasets for this analysis, it is interesting to show some initial analysis of the three different poverty measures and the degree of overlap between them.

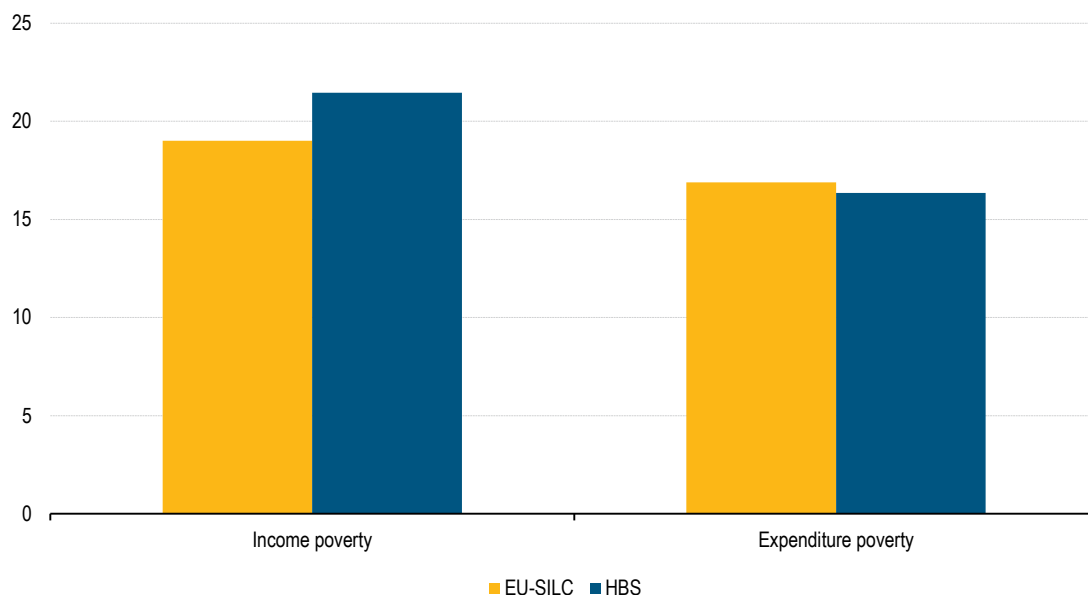
For the purpose of this analysis, income poverty is defined as having an equivalised household income below 60% of the national equivalised median income. This is in line with the definition used in the At Risk of Poverty or Social Exclusion (AROPE) indicator which is used to monitor progress towards the Europe 2020 headline target. Expenditure poverty is defined in comparable terms: equivalised household expenditure less than 60% of the equivalised median. Individuals are classed as being materially deprived if they have an enforced lack of at least three out of a list of nine material deprivation items⁽⁴⁾.

The results presented below used the mixed methods approach to statistical matching, as it was felt that this approach was the most effective method overall. However, the analysis has also been conducted using the data imputed from both the parametric and hotdeck methods and the findings are very similar for each.

5.1 Headline poverty indicators

Figure 8 shows the proportion of people at risk of income and expenditure poverty estimated using the matched data set and HBS.

Figure 8: Income and expenditure poverty, UK, 2005 (%)



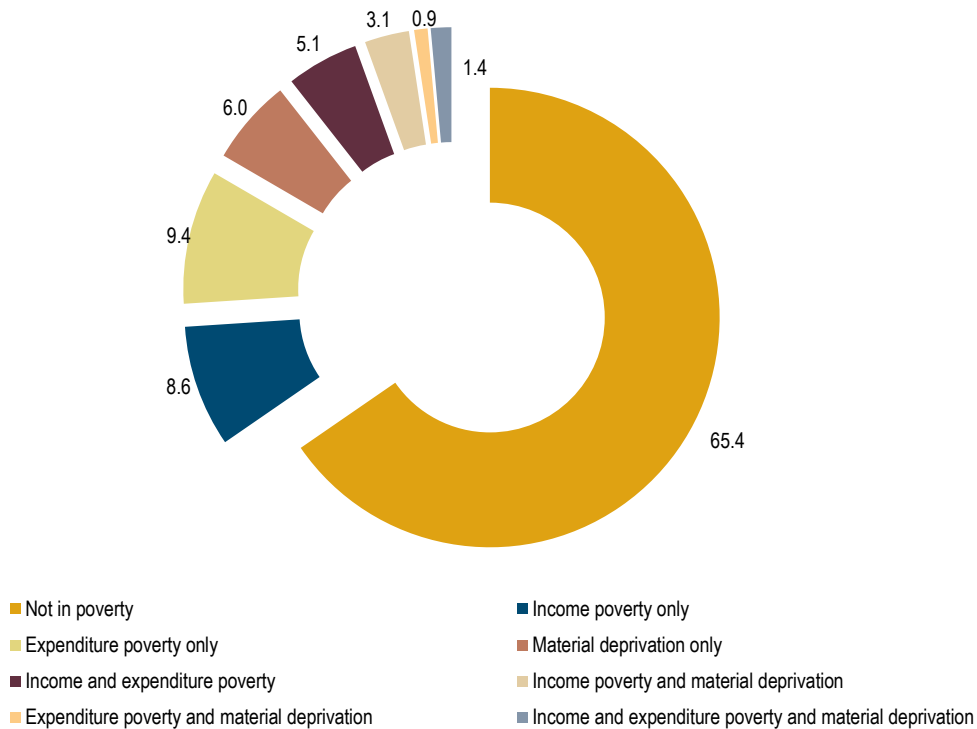
⁽⁴⁾ Currently these 9 items are: Arrears on mortgage or rent payments, utility bills, hire purchase installments or other loan payments; capacity to afford paying for one week's annual holiday away from home; capacity to afford a meal with meat, chicken, fish (or vegetarian equivalent) every second day; capacity to face unexpected financial expenses; household cannot afford a telephone (including mobile phone); household cannot afford a colour TV; household cannot afford a washing machine; household cannot afford a car; ability of household to pay for keeping its home adequately warm.

Figure 8 shows that according to EU-SILC, 19.0% of the population was at risk of income poverty while the HBS measures 21.5% at risk of income poverty. This highlights the difference between the two surveys in its measurement of income. However, the estimates of expenditure poverty between the two datasets are more closely aligned. EU-SILC estimates an at risk of expenditure poverty rate of 16.9%, while the HBS is 16.4%. This appears indicative of a good matching of expenditure from HBS to EU-SILC.

5.2 Overlap of income and expenditure poverty and material deprivation

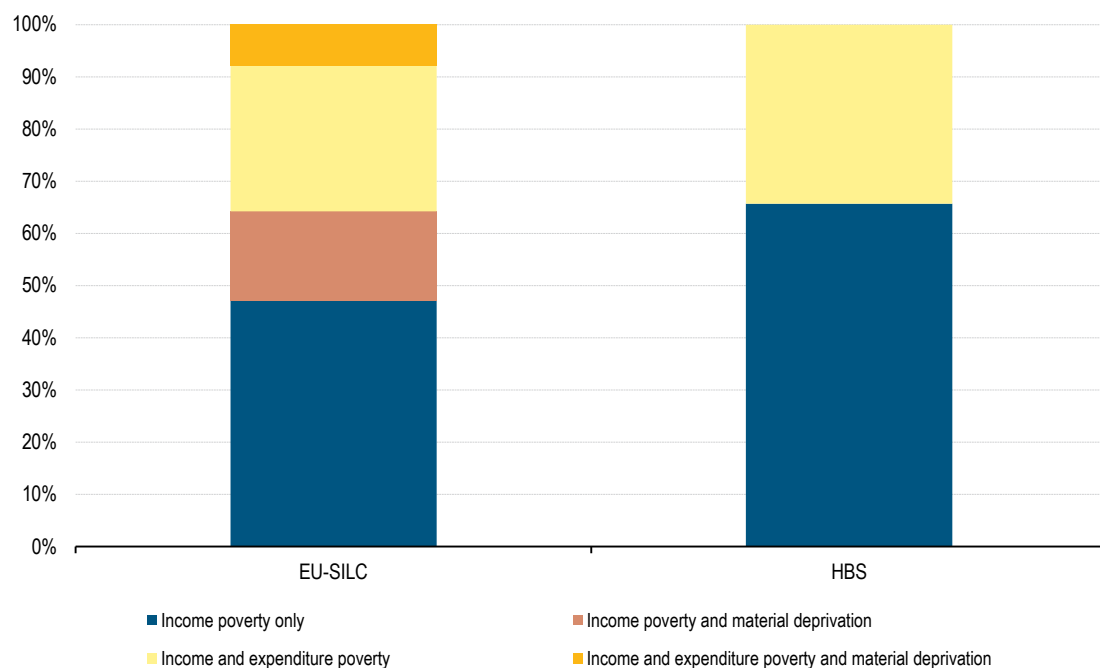
Figure 9 highlights the degree of overlap between different poverty statuses in 2005 in the UK. It shows that 35% of people were either in income poverty or expenditure poverty or were materially deprived, while 11% of people were in poverty on two or more of these measures and 1% were in poverty on all three.

Figure 9: Income poverty, expenditure poverty and material deprivation, UK, 2005 (%)



Looking more closely at the degree of overlap between the three measures, figure 10 shows that in the matched data, 36% of those who are income poor are also expenditure poor (shown by the top two bars in figure 10). As might be expected, the degree of overlap is similar to that observed in the HBS, where 34% of those who were income poor were also expenditure poor.

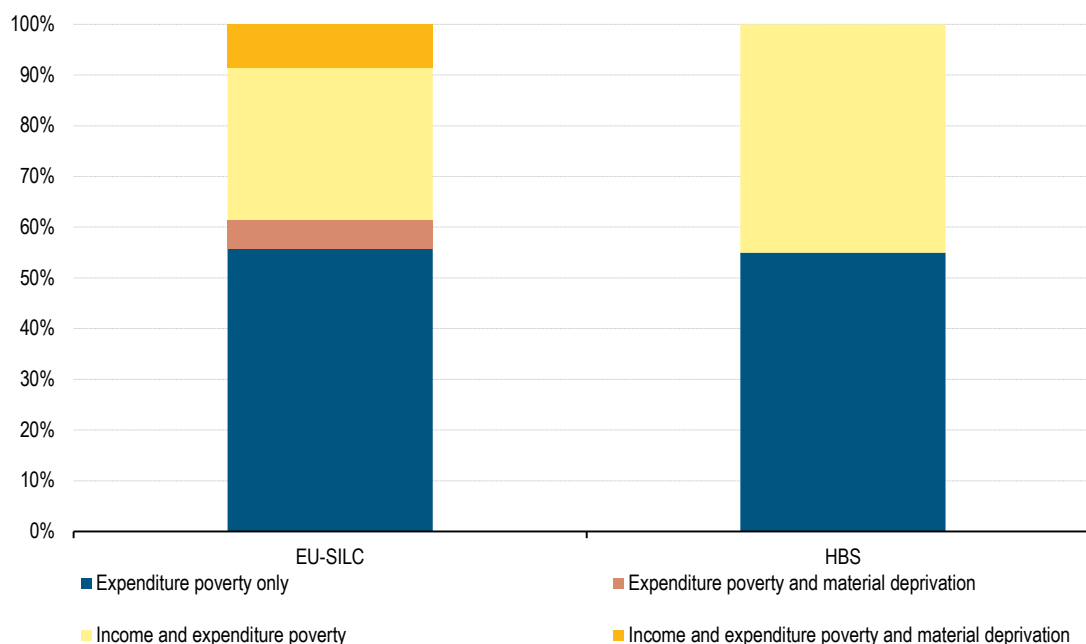
Figure 10: Percentage of income poor individuals experiencing other forms of poverty, UK, 2005 (%)



There was lower degree of complementarity in between income and material deprivation dimensions, with only 25% of those who were income poor also being materially deprived on this measure.

Figure 11 shows the proportion of expenditure poor individuals who are also in poverty on one or both of the other two measures. Overall, 17% of individuals were expenditure poor (cf 16% on the HBS). Of these individuals, 39% were also income poor. The level of overlap is slightly lower than for the comparable HBS figures (45%), but this is probably explained by overall estimated income poverty rates being higher on the HBS than EU-SILC.

Figure 11: Percentage of expenditure poor individuals experiencing other forms of poverty, UK, 2005 (%)

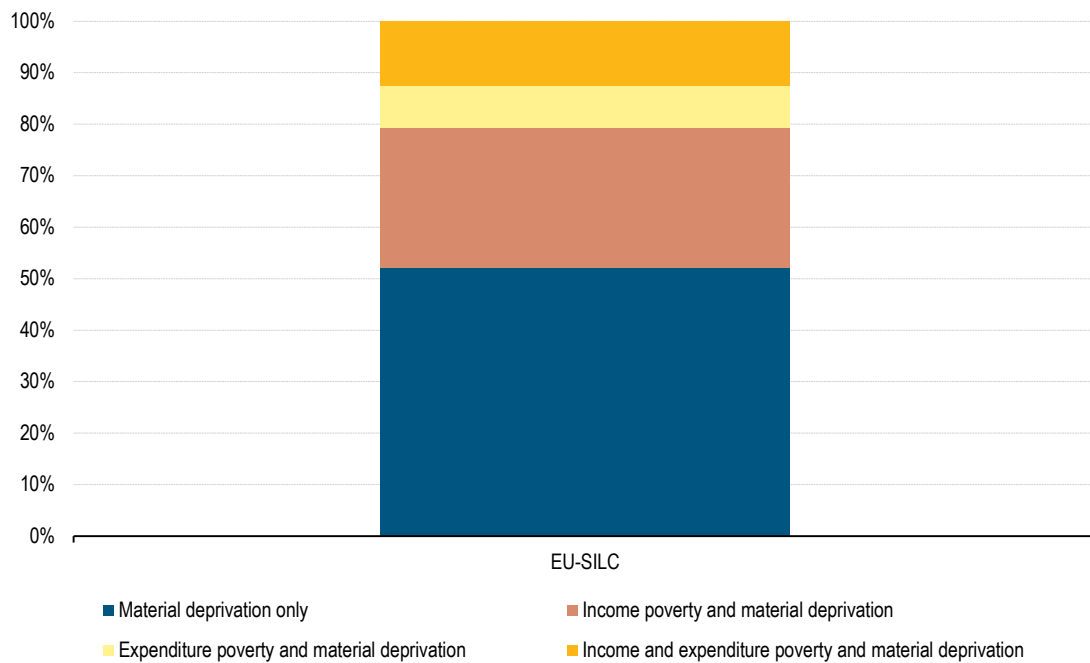


Despite this difference, the matched EU-SILC data and HBS show a consistent relationship between income and expenditure, with the proportion of expenditure poor people who are also income poor higher than the proportion of income poor who are also expenditure poor. This is consistent with other research which shows that those with the lowest expenditure also tend to have the lowest cash income, but that the inverse is not always true (Brewer & O’Dea, 2012).

The degree of overlap between expenditure poverty and material deprivation appears to be less than between income poverty and material deprivation. Just 14% of those who were expenditure poor also had an enforced lack of 3 or more material deprivation items (compared with 12% of the general population).

The apparent low degree of overlap between these two measures is also highlighted in Figure 12. This shows that, of those who were materially deprived on this measure, 21% were also expenditure poor, compared with 40% who were also income poor. Further analysis is needed to investigate this further, including an examination of different countries and time periods, and whether there is a stronger relationship between material deprivation and certain forms of expenditure (for example, non-durable goods).

Figure 12: Percentage of materially deprived individuals experiencing other forms of poverty, UK 2005 (%)



6. Conclusions and next steps

The outcomes of this initial statistical matching of EU-SILC and the HBS are encouraging. Analysis of the joint distributions of the matching variables with imputed and actual expenditure indicates that the matching has been broadly effective, with the mixed methods approach appearing slightly better than the others overall.

However, the statistical matching carried out so far has necessarily been limited by the availability of suitable matching variables for 2005 EU-SILC and HBS. This is because data for the 2005 dataset is only at the household level, limiting the information that is available for deriving potential matching variables.

For the UK data, harmonisation of concepts such as Household Reference Person (HRP) and generally consistent methods for collecting income data mean that it was possible to identify a potential set of common variables which broadly covered information about the household size and structure, other key characteristics relating to the household such as information regarding the type and size of dwelling, as well as socio-demographic characteristics of the HRP and income. However, even in this case, the number of potential matching variables which had sufficiently similar distributions in the two datasets was relatively limited. Having a smaller number of variables increases the risk that any model fitted to the data will be misspecified and the results of the matching will not be reliable.

In the case of data from other countries, if the definition of HRP is not standardised or income is not collected through similar methods in both surveys, there is limited value in attempting to carry out statistical matching with the 2005 datasets. The next stage of this project will therefore move to using 2010 HBS and EU-SILC data in order to make use of the additional variables available in the 2010 HBS. It may be that, with additional matching variables available, the parametric or mixed methods approaches become more effective in producing reliable matching results. This statistical matching will be carried out for a number of EU countries, selected on analytical and pragmatic grounds. The possibility of extending the statistical matching in order to match not just overall expenditure onto EU-SILC, but individual components of expenditure (such as expenditure on food) will also be explored.

6.1 Recommendations for future data collection

The Wiesbaden Memorandum, which was endorsed by the European Statistical System Committee (ESSC) in September 2011, called for a process of modernisation in social statistics. As part of this modernisation process, a review of the legal basis of EU-SILC is being undertaken, which shall report in September 2013. The findings of this paper highlight a number of important lessons which should be carefully considered as part of this modernisation agenda in order to maximise both the relevance and the efficiency of collecting these statistics.

Statistical matching allows the possibility of increasing the analytical potential of the data collated at a European level. However, the opportunities for fully utilising matching are limited due to the current extent of harmonisation of EU-SILC and other ESS surveys. It is therefore recommended that steps are taken through the EU-SILC Legal Basis Task Force and the overall modernisation programme to ensure that all variables measuring comparable concepts (not just the 'core social variables') are better harmonised across sources.

Furthermore, the EU-SILC Task Force should consider the inclusion within the EU-SILC nucleus of variables which could be used as 'hooks' to improve the potential quality of matching between SILC and other sources such as the HBS and HFCS (Household Finance and Consumption Survey of the ECB), thereby potentially allowing the production of integrated statistics on income, consumption and wealth. These 'hook' variables will need to be carefully selected in order to fit well amongst other variables in the nucleus, not be burdensome on respondents or NSIs and have a strong relationship between the variables of interest in both sources.

7. References

- Atkinson, T., Cantillon, B., Marlier, E., & Nolan, B. (2002). *Social Indicators: The EU and Social Inclusion*. Oxford: Oxford University Press.
- Brewer, M., & O'Dea, C. (2012). *Measuring living standards with income and consumption: evidence from the UK*. Retrieved June 23, 2012, from Institute for Social & Economic Research (ISER) - University of Essex: <https://www.iser.essex.ac.uk/publications/working-papers/iser/2012-05.pdf>
- Cutler, D., & Katz, L. (1991). Macroeconomic Performance and the Disadvantaged. *Brookings Papers on Economic Activity*, 2.
- D'Orzaio, M., Di Zio, M., & Scanu, M. (2006). *Statistical Matching: Theory and Practice*. Chichester: John Wiley & Sons Ltd.
- Jorgenson, D., & Slesnick, D. (1987). Aggregate Consumer Behavior and Household. *Journal of Business and Economic Statistics*, 5 (2), 219-232.
- Meyer, B., & Sullivan, J. (2011). Further Results on Measuring the Well-Being of the Poor Using Income and Consumption. *Canadian Journal of Economics*, 44 (1), 52-87.
- Noll, H-H. (2007). *Household Consumption, Household Incomes and Living Standards - A review of related recent research activities*. Retrieved 6 12, 23, from GESIS – Leibniz Institute for the Social Sciences: http://www.gesis.org/fileadmin/upload/institut/wiss_arbeitsbereiche/soz_indikatoren/Publikationen/Household-Expenditures-Research-Report.pdf
- Renssen, R.H. (1998) Use of statistical matching techniques in calibration estimation. *Survey Methodology*, 24, 171-183.
- Sabelhaus, J., Johnson, D., Ash, S., Swanson, D., Garner, T., Greenlees, J., et al. (2011). *Is the Consumer Expenditure Survey Representative by Income?* Retrieved August 28, 2012, from CRIW/NBER Conference: <http://conference.nber.org/confer/2011/CRIWf11/CRIWf11prg.html>
- Stiglitz, J. E., Sen, A., & Fitoussi, J. P. (2009). *Report by the Commission on the Measurement of Economic Performance and Social Progress*. http://www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf.

8. Annexes

Annex 1: Complete list of common variables EU-SILC and HBS, 2005

EU-SILC		HBS	
Codification	Description	Codification	Description
DB040	Region NUTS 2	HA08	Region NUTS1
UKC1	Tees Valley and Durham	1	North East
UKC2	Northumberland and Tyne and Wear	2	North West
UKD1	Cumbria	3	Yorkshire and the Humber
UKD2	Cheshire	4	East Midlands
UKD3	Greater Manchester	5	West Midlands
UKD3	Greater Manchester	6	Eastern
UKD4	Lancashire	7	London
UKD5	Merseyside	8	South East
UKE1	East Riding and North Lincolnshire	9	South West
UKE2	North Yorkshire	10	Wales
UKE3	South Yorkshire	11	Scotland
UKE4	West Yorkshire	12	Northern Ireland
UKF1	Derbyshire and Nottinghamshire		
UKF2	Leicestershire, Rutland and Northamptonshire		
UKF3	Lincolnshire		
UKG1	Herefordshire, Worcestershire and Warwickshire		
UKG2	Shropshire and Staffordshire		
UKG3	West Midlands		
UKH1	East Anglia		
UKH2	Bedfordshire and Hertfordshire		
UKH3	Essex		
UKI1	Inner London		
UKI2	Outer London		
UKJ1	Berkshire, Buckinghamshire and Oxfordshire		
UKJ2	Surrey, East and West Sussex		
UKJ3	Hampshire and Isle of Wight		
UKJ4	Kent		
UKK1	Gloucestershire, Wiltshire and North Somerset		
UKK2	Dorset and Somerset		
UKK3	Cornwall and Isles of Scilly		
UKK4	Devon		
UKL1	West Wales and The Valleys		
UKL2	East Wales		
UKM1	North Eastern Scotland		
UKM2	Eastern Scotland		
UKM3	South Western Scotland		
UKM4	Highlands and Islands		
UKN0	Northern Ireland		
UKZZ	Extra-Regio		
DB100	Degree of Urbanisation	HA09	Population density domain
1	Densely populated (at least 500 inhabitants/sq km)	1	Densely populated (at least 500 inhabitants/sq km)
2	Intermediate (between 100 and 499 inhabitants/sq km)	2	Intermediate (between 100 and 499 inhabitants/sq km)
3	Sparsely populated (less than 100 inhabitants/sq km)	3	Sparsely populated (less than 100 inhabitants/sq km)

		inhabitants/sq km)	
EU-SILC		HBS	
Codification	Description	Codification	Description
HH020	Tenure Status	HD01	Title (principle resident)
1	Owner	1	Owned, no outstanding loan or mortgage
2	Tenant or subtenant paying rent at prevailing or market price	2	Owned, with outstanding loan or mortgage
3	Accommodation is rented at reduced rate (lower than market price)	3	Rental (full rental)
4	Accommodation is provided free	4	Reduced rental
		5	Rent free
		9	Not specified
HH010	Dwelling Type	HD03	Type of dwelling
1	Detached House	1	Detached House
2	Semi detached or terraced house	2	Semi detached or terraced house
3	Apartment or flat in a building with less than 10 dwellings	3	apartment
4	Apartment or flat in a building with more than 10 dwellings	4	Other
5	Some other kind of accommodation	10	Not Specified
HH030	Number of rooms available to household	HD06	Number of rooms
1-9	Number of rooms	1-7	Number of rooms
10	10 or more rooms	8	8 rooms or more
		9	Not specified
HS110	Do you have a car?	HD14.02	Number of cars
1	Yes	0-3	Number of cars
2	No - cannot afford	4	4 or more
3	No - other reason	9	Not specified
HS080	Do you have a colour TV?	HD14.04	Number of televisions
1	Yes	0-3	Number of televisions
2	No - cannot afford	4	4 or more
3	No - other reason	9	Not specified
HS090	Do you have a computer?	HD14.08	Computer (PC)
1	Yes	0	No
2	No - cannot afford	1	Yes, without internet connection
3	No - other reason	2	Yes, with internet connection
		9	Not specified
HS100	Do you have a washing machine?	HD14.09	Washing machine at place of residence
1	Yes	0	No
2	No - cannot afford	1	Yes
3	No - other reason	9	Not specified
HS070	Do you have a telephone (including mobile phone?)	HD14.14	Number of mobile phones in use(with access to a cellular network)
1	Yes	0-3	Number of mobile phones
2	No - cannot afford	4+	4 or more mobiles
3	No - other reason	9	Not specified

EU-SILC		HBS	
Codification	Description	Codification	Description
Derived	Household type (age limit for children is 16 years old)	HB07_1	Type of household - 1 (age limit for children is 16 years old)
	ONE ADULT HOUSEHOLD		ONE ADULT HOUSEHOLD
1	One person, aged 65 years or more	1	One person, aged 65 years or more
2	One person, aged 30 to 64 years	2	One person, aged 30 to 64 years
3	One person, under 30 years	3	One person, under 30 years
4	One person with children up to 16 years old (exclusive)	4	One person with children up to 16 years old (exclusive)
	COUPLE WITHOUT CHILDREN		COUPLE WITHOUT CHILDREN
5	Couple without children, older member aged 65 or more	5	Couple without children, older member aged 65 or more
6	Couple without children, older member under 65	6	Couple without children, older member under 65
	COUPLE WITH CHILDREN AGED UP TO 16 YEARS OLD (EXCLUSIVE)		COUPLE WITH CHILDREN AGED UP TO 16 YEARS OLD (EXCLUSIVE)
7	One child	7	One child
8	Two children	8	Two children
9	Three or more children	9	Three or more children
	OTHER		OTHER
10	Single parent or couple with at least one child of 16 years old or older	10	Single parent or couple with at least one child of 16 years old or older
11	All other households	11	Other households with all members related
		12	Other households with one or more members without being related.
Derived	Household size	HB05	Household Size
0+	Number of people in household	0+	Number of people in household
RB090	Sex	HC03	Sex of Reference Person
1	Male	1	Male
2	Female	2	Female
PX020	Age	HC04	Age in completed years of reference person
00-120	Age in years	00-98	98 Years and older
		99	Not Specified
PB190	Marital Status	HC05	Marital Status of Reference Person
1	Never Married	0	Never Married
2	Married	1	Married or in cohabitation
3	Seperated	3	Widowed
4	Widowed	4	Divorced or seperated
5	Divorced	9	Not specified
PB200	Consensual Union	HC05.1	Consensual Union of reference person
1	Yes, on a legal basis	0	No
2	Yes, without a legal basis	1	Yes on a legal basis
3	No	2	Yes without legal basis
		9	Not specified

EU-SILC		HBS	
Codification	Description	Codification	Description
RB210	Main activity status during the income reference period	HC12	Current Activity status of reference person
1	At work		Economically Active
2	Unemployed	1	Working
3	In retirement or early retirement or has given up business	2	With employment but temporary absent
4	Other inactive person	3	Unemployed
			Economically Inactive
		4	Retired
		5	Student or in national service
		6	Non economic activity, housewife
		7	Unable to work
		8	Not applicable (legal age to work unfulfilled)
		9	Not specified
PL030	Self-defined current economic status	HC13	Hours worked by reference person
1	Working full time	1	Full time
2	Working part time	2	Part time
3	Unemployed	8	Not applicable (do not work)
4	Pupil, student, further training, unpaid work experience	9	Not specified
5	In retirement or early retirement or has given up business		
6	Permanently disabled or/and unfit to work		
7	In compulsory military community or service		
8	Fulfilling domestic tasks and care responsibilities		
9	Other inactive person		
PL140	Type of contract	HC14	Type of work contract for reference person
1	Permanent job/work contract of unlimited duration	1	Permanent job/work contract of unlimited duration
2	Temporary job/work contract of limited duration	2	Temporary job/work contract of limited duration
		8	Not applicable (do not work)
		9	Not specified
PL040	Status in employment	HC21	Status in employment of reference person
1	self-employed with employees	1	Employer
2	self-employed without employees	2	Self-employed person
3	employee	3	Employee
4	family worker	4	Unpaid family worker
		5	Apprentice
		6	Person not classified by status
		8	Not applicable (legal age to work unfulfilled)
		9	Not specified

EU-SILC		HBS	
Codification	Description	Codification	Description
PL050	Occupation	HC18	Occupation of reference person (ISCO 1988 (COM))
11	Legislators, Senior officials, and managers	1	Legislators, Senior officials, and managers
12	Corporate managers	2	Professionals
13	Managers of small enterprises	3	Technicians and associate professionals
21	Physical, mathematical, and engineering science professionals	4	Clerks
22	Life science and health professionals	5	Service Workers and shop and market sales workers
23	Teaching professionals	6	Skilled agricultural and fishery workers
24	Other professionals	7	Craft and related trades workers
31	Physical and engineering science associate professionals	8	Plant and Machine operators and assemblers
32	Life science and health associate professionals	9	Elementary occupations
33	Teaching associate professionals	0	Armed Forces
34	Other associate professionals	88	Not applicable (legal age to work unfulfilled)
41	Office clerks	99	Not Specified
42	Customer service clerks		
51	Personal and protective services workers		
52	Models, salespersons, and demonstrators		
61	Skilled agriculture and fishery workers		
71	Extraction and building trades workers		
72	Metal, machinery, and related trades workers		
73	Precision, handicraft, craft printing and related trades workers		
74	Other craft and related trades workers		
81	Stationary-plant and related operators		
82	Machine operators and assemblers		
83	Drivers and mobile plant operators		
91	Sales and services elementary occupations		
92	Agricultural, fishery and related labourers		
93	Labourers in mining, construction, manufacturing and output		
01	Armed Forces		

EU-SILC		HBS	
Codification	Description	Codification	Description
HH020	Tenure Status	HD01	Title (principle resident)
1	Owner	1	Owned, no outstanding loan or mortgage
2	Tenant or subtenant paying rent at prevailing or market price	2	Owned, with outstanding loan or mortgage
3	Accommodation is rented at reduced rate (lower than market price)	3	Rental (full rental)
4	Accommodation is provided free	4	Reduced rental
		5	Rent free
		9	Not specified
HH010	Dwelling Type	HD03	Type of dwelling
1	Detached House	1	Detached House
2	Semi detached or terraced house	2	Semi detached or terraced house
3	Apartment or flat in a building with less than 10 dwellings	3	apartment
4	Apartment or flat in a building with more than 10 dwellings	4	Other
5	Some other kind of accommodation	10	Not Specified
HH030	Number of rooms available to household	HD06	Number of rooms
1-9	Number of rooms	1-7	Number of rooms
10	10 or more rooms	8	8 rooms or more
		9	Not specified
HS110	Do you have a car?	HD14.02	Number of cars
1	Yes	0-3	Number of cars
2	No - cannot afford	4	4 or more
3	No - other reason	9	Not specified
HS080	Do you have a colour TV?	HD14.04	Number of televisions
1	Yes	0-3	Number of televisions
2	No - cannot afford	4	4 or more
3	No - other reason	9	Not specified
HS090	Do you have a computer?	HD14.08	Computer (PC)
1	Yes	0	No
2	No - cannot afford	1	Yes, without internet connection
3	No - other reason	2	Yes, with internet connection
		9	Not specified
HS100	Do you have a washing machine?	HD14.09	Washing machine at place of residence
1	Yes	0	No
2	No - cannot afford	1	Yes
3	No - other reason	9	Not specified
HS070	Do you have a telephone (including mobile phone?)	HD14.14	Number of mobile phones in use(with access to a cellular network)
1	Yes	0-3	Number of mobile phones
2	No - cannot afford	4+	4 or more mobiles
3	No - other reason	9	Not specified

EU-SILC		HBS	
Codification	Description	Codification	Description
Derived	Household type (age limit for children is 16 years old)	HB07_1	Type of household - 1 (age limit for children is 16 years old)
	ONE ADULT HOUSEHOLD		ONE ADULT HOUSEHOLD
1	One person, aged 65 years or more	1	One person, aged 65 years or more
2	One person, aged 30 to 64 years	2	One person, aged 30 to 64 years
3	One person, under 30 years	3	One person, under 30 years
4	One person with children up to 16 years old (exclusive)	4	One person with children up to 16 years old (exclusive)
	COUPLE WITHOUT CHILDREN		COUPLE WITHOUT CHILDREN
5	Couple without children, older member aged 65 or more	5	Couple without children, older member aged 65 or more
6	Couple without children, older member under 65	6	Couple without children, older member under 65
	COUPLE WITH CHILDREN AGED UP TO 16 YEARS OLD (EXCLUSIVE)		COUPLE WITH CHILDREN AGED UP TO 16 YEARS OLD (EXCLUSIVE)
7	One child	7	One child
8	Two children	8	Two children
9	Three or more children	9	Three or more children
	OTHER		OTHER
10	Single parent or couple with at least one child of 16 years old or older	10	Single parent or couple with at least one child of 16 years old or older
11	All other households	11	Other households with all members related
		12	Other households with one or more members without being related.
Derived	Household size	HB05	Household Size
0+	Number of people in household	0+	Number of people in household
RB090	Sex	HC03	Sex of Reference Person
1	Male	1	Male
2	Female	2	Female
PX020	Age	HC04	Age in completed years of reference person
00-120	Age in years	00-98	98 Years and older
		99	Not Specified
PB190	Marital Status	HC05	Marital Status of Reference Person
1	Never Married	0	Never Married
2	Married	1	Married or in cohabitation
3	Separated	3	Widowed
4	Widowed	4	Divorced or separated
5	Divorced	9	Not specified
PB200	Consensual Union	HC05.1	Consensual Union of reference person
1	Yes, on a legal basis	0	No
2	Yes, without a legal basis	1	Yes on a legal basis
3	No	2	Yes without legal basis
		9	Not specified

EU-SILC		HBS	
Codification	Description	Codification	Description
RB210	Main activity status during the income reference period	HC12	Current Activity status of reference person
1	At work		Economically Active
2	Unemployed	1	Working
3	In retirement or early retirement or has given up business	2	With employment but temporary absent
4	Other inactive person	3	Unemployed
			Economically Inactive
		4	Retired
		5	Student or in national service
		6	Non economic activity, housewife
		7	Unable to work
		8	Not applicable (legal age to work unfulfilled)
		9	Not specified
PL030	Self-defined current economic status	HC13	Hours worked by reference person
1	Working full time	1	Full time
2	Working part time	2	Part time
3	Unemployed	8	Not applicable (do not work)
4	Pupil, student, further training, unpaid work experience	9	Not specified
5	In retirement or early retirement or has given up business		
6	Permanently disabled or/and unfit to work		
7	In compulsory military community or service		
8	Fulfilling domestic tasks and care responsibilities		
9	Other inactive person		
PL140	Type of contract	HC14	Type of work contract for reference person
1	Permanent job/work contract of unlimited duration	1	Permanent job/work contract of unlimited duration
2	Temporary job/work contract of limited duration	2	Temporary job/work contract of limited duration
		8	Not applicable (do not work)
		9	Not specified
PL040	Status in employment	HC21	Status in employment of reference person
1	self-employed with employees	1	Employer
2	self-employed without employees	2	Self-employed person
3	employee	3	Employee
4	family worker	4	Unpaid family worker
		5	Apprentice
		6	Person not classified by status
		8	Not applicable (legal age to work unfulfilled)
		9	Not specified

EU-SILC		HBS	
Codification	Description	Codification	Description
PL050	Occupation	HC18	Occupation of reference person (ISCO 1988 (COM))
11	Legislators, Senior officials, and managers	1	Legislators, Senior officials, and managers
12	Corporate managers	2	Professionals
13	Managers of small enterprises	3	Technicians and associate professionals
21	Physical, mathematical, and engineering science professionals	4	Clerks
22	Life science and health professionals	5	Service Workers and shop and market sales workers
23	Teaching professionals	6	Skilled agricultural and fishery workers
24	Other professionals	7	Craft and related trades workers
31	Physical and engineering science associate professionals	8	Plant and Machine operators and assemblers
32	Life science and health associate professionals	9	Elementary occupations
33	Teaching associate professionals	0	Armed Forces
34	Other associate professionals	88	Not applicable (legal age to work unfulfilled)
41	Office clerks	99	Not Specified
42	Customer service clerks		
51	Personal and protective services workers		
52	Models, salespersons, and demonstrators		
61	Skilled agriculture and fishery workers		
71	Extraction and building trades workers		
72	Metal, machinery, and related trades workers		
73	Precision, handicraft, craft printing and related trades workers		
74	Other craft and related trades workers		
81	Stationary-plant and related operators		
82	Machine operators and assemblers		
83	Drivers and mobile plant operators		
91	Sales and services elementary occupations		
92	Agricultural, fishery and related labourers		
93	Labourers in mining, construction, manufacturing and output		
01	Armed Forces		

Source: Eurostat

Annex 2: Complete list of derived variables

Codification	Harmonised Description
DV_REGION	Region Nuts 1
1	North East
2	North West
3	Yorkshire and the Humber
4	East Midlands
5	West Midlands
6	Eastern
7	London
8	South East
9	South West
10	Wales
11	Scotland
12	Northern Ireland
DV_URBAN	Population density domain
1	Densely populate (at least 500 inhabitants/sq km)
2	Intermediate (between 100 and 499 inhabitants/sq km)
3	Sparsely populated (less than 100 inhabitants/sq km)
DV_TENURE	Tenure Status
1	Owner
2	Tenant or subtenant paying rent at prevailing or market price
3	Accommodation is rented at reduced rate
4	Accommodation is provided free
DV_DWELL	Dwelling Type
1	Detached House
2	Semi detached or terraced house
3	Apartment
4	Other
DV_ROOMS	Number of rooms
1-7	Number of rooms
8	8 rooms or more
DV_CARS	Do you have a car?
1	Yes
2	No
DV_TV	Do you have a TV?
1	Yes
2	No
DV_PC	Do you have a PC?
1	Yes
2	No
DV_WASH	Do you have a washing machine?
1	Yes
2	No

Codification	Harmonised Description
DV_PHONE	Do you have a telephone?
1	Yes
2	No
DV_HHTYPE	Household type
	ONE ADULT HOUSEHOLD
1	One person, aged 65 years or more
2	One person, aged 30 to 64 years
3	One person, under 30 years
4	One person with children up to 16 years old (exclusive)
	COUPLE WITHOUT CHILDREN
5	Couple without children, older member aged 65 or more
6	Couple without children, older member under 65
	COUPLE WITH CHILDREN AGED UP TO 16 YEARS OLD (EXCLUSIVE)
7	One child
8	Two children
9	Three or more children
	OTHER
10	Single parent or couple with at least one child of 16 years old or older
11	All other households
DV_HHSIZE	Household Size
0-5	Number of people in household
5+	More than 5 people in household
DV_SEX	Sex of reference person
1	Male
2	Female
DV_AGE	Age in completed years of reference person
1	16-25
2	26-35
3	36-45
4	46-55
5	56-65
6	66-75
7	75 plus
DV_MARSTA	Marital Status of Reference Person
1	Never Married
2	Married
3	Widowed
4	Divorced or separated
DV_CONUNI	Consensual Union
1	Yes, on a legal basis
2	Yes, without a legal basis
3	No
DV_ACTSTA	Activity Status
1	At work
2	Unemployed
3	In retirement or early retirement or has given up business
4	Other inactive person

Codification	Harmonised Description
DV_LABOUR	Full-time or part-time working
1	Full time
2	Part time
DV_PERM	Type of contract
1	Permanent job/work contract of unlimited duration
2	Temporary job/work contract of limited duration
3	Not applicable (not an employee)
4	Not specified
DV_OCCUP	Occupation
1	Legislators, Senior officials, and managers
2	Professionals
3	Technicians and associate professionals
4	Clerks
5	Service Workers and shop and market sales workers
6	Skilled agricultural and fishery workers
7	Craft and related trades workers
8	Plant and Machine operators and assemblers
9	Elementary occupations
10	Armed Forces

Annex 3: Output for material deprivation logistic regression model, EU-SILC 2005

	Estimate	Standard Error	Z value	Pr (> z)	Significance
(Intercept)	-3.69	0.32	-11.70	< 2e-16	***
DV_REGION2	-0.14	0.19	-0.76	0.447	
DV_REGION3	-0.33	0.20	-1.65	0.099	.
DV_REGION4	0.01	0.21	0.04	0.969	
DV_REGION5	0.34	0.19	1.78	0.075	.
DV_REGION6	-0.30	0.21	-1.41	0.157	
DV_REGION7	0.31	0.18	1.69	0.091	.
DV_REGION8	-0.06	0.19	-0.33	0.743	
DV_REGION9	0.21	0.20	1.07	0.283	
DV_REGION10	-0.10	0.23	-0.44	0.663	
DV_REGION11	-0.10	0.20	-0.49	0.622	
DV_REGION12	-0.29	0.32	-0.90	0.368	
DV_AGE2	0.20	0.15	1.34	0.180	
DV_AGE3	0.25	0.15	1.68	0.093	.
DV_AGE4	0.12	0.16	0.73	0.465	
DV_AGE5	-0.05	0.17	-0.30	0.763	
DV_AGE6	-0.42	0.23	-1.81	0.070	.
DV_AGE7	-1.20	0.25	-4.79	0.000	***
DV_ACTSTAT2	0.88	0.19	4.63	0.000	***
DV_ACTSTAT3	-0.36	0.18	-2.01	0.044	*
DV_ACTSTAT4	0.40	0.10	4.06	0.000	***
DV_CONUNI2	0.41	0.14	2.88	0.004	**
DV_CONUNI3	0.61	0.12	4.87	0.000	***
DV_TENURE2	1.21	0.09	13.18	< 2e-16	***
DV_TENURE3	0.45	0.31	1.47	0.141	
DV_DWELL2	0.42	0.15	2.83	0.005	**
DV_DWELL3	0.46	0.17	2.76	0.006	**
DV_DWELL4	-11.33	167.22	-0.07	0.946	
DV_SEX1	0.00	0.08	-0.06	0.953	
DV_HHSIZE2	0.37	0.12	3.03	0.002	**
DV_HHSIZE3	0.86	0.15	5.92	0.000	***
DV_HHSIZE4	0.98	0.17	5.85	0.000	***
DV_HHSIZE5	1.31	0.21	6.17	0.000	***
DV_HHSIZE6	1.65	0.26	6.33	0.000	***
DV_CAR	1.09	0.09	12.06	< 2e-16	***
DV_TV1	0.21	0.25	0.85	0.394	
DV_PC1	0.12	0.09	1.41	0.158	
inc_band2	0.01	0.17	0.07	0.944	
inc_band3	-0.34	0.18	-1.97	0.049	*
inc_band4	-0.71	0.19	-3.72	0.000	***
inc_band5	-0.96	0.19	-4.95	0.000	***
inc_band6	-1.27	0.23	-5.49	0.000	***
inc_band7	-1.80	0.31	-5.89	0.000	***
inc_band8	-2.16	0.33	-6.51	0.000	***

Source: Eurostat

Significance codes

***	0	*	0.05
**	0.01	.	0.1

Annex 4: Output for expenditure regression model, HBS 2005

	Estimate	Standard Error	Z value	Pr (> z)	Significance
(Intercept)	5.24	0.06	93.21	< 2e-16	***
DV_REGION2	0.03	0.03	0.95	0.342	
DV_REGION3	0.04	0.03	1.11	0.267	
DV_REGION4	0.02	0.03	0.55	0.582	
DV_REGION5	-0.01	0.03	-0.43	0.670	
DV_REGION6	0.08	0.03	2.37	0.018	*
DV_REGION7	0.10	0.03	2.86	0.004	**
DV_REGION8	0.07	0.03	2.15	0.032	*
DV_REGION9	0.05	0.03	1.50	0.133	
DV_REGION10	0.02	0.04	0.47	0.642	
DV_REGION11	0.05	0.03	1.54	0.123	
DV_REGION12	0.07	0.03	2.03	0.042	*
DV_AGE2	-0.06	0.03	-2.10	0.036	*
DV_AGE3	-0.06	0.03	-1.96	0.050	*
DV_AGE4	-0.02	0.03	-0.75	0.456	
DV_AGE5	-0.05	0.03	-1.58	0.114	
DV_AGE6	-0.12	0.04	-2.79	0.005	**
DV_AGE7	-0.30	0.04	-6.78	0.000	***
DV_ACTSTAT2	-0.04	0.04	-0.97	0.333	
DV_ACTSTAT3	0.04	0.03	1.28	0.199	
DV_ACTSTAT4	0.03	0.02	1.53	0.126	
DV_CONUNI2	0.02	0.02	0.75	0.452	
DV_CONUNI3	-0.02	0.02	-0.96	0.340	
DV_TENURE2	0.24	0.02	15.23	< 2e-16	***
DV_TENURE3	-0.12	0.05	-2.39	0.017	*
DV_DWELL2	-0.14	0.01	-9.64	< 2e-16	***
DV_DWELL3	-0.11	0.02	-5.03	0.000	***
DV_DWELL4	-0.01	0.04	-0.34	0.732	
DV_HHSIZE2	0.24	0.02	11.68	< 2e-16	***
DV_HHSIZE3	0.33	0.02	13.42	< 2e-16	***
DV_HHSIZE4	0.39	0.03	13.87	< 2e-16	***
DV_HHSIZE5	0.49	0.03	14.12	< 2e-16	***
DV_HHSIZE6	0.43	0.04	9.66	< 2e-16	***
DV_CAR1	-0.20	0.02	-12.49	< 2e-16	***
DV_SEX	0.02	0.01	1.50	0.135	
DV_TV1	-0.16	0.06	-2.90	0.004	**
DV_PC1	-0.13	0.01	-9.41	< 2e-16	***
inc_band2	0.00	0.03	-0.10	0.920	
inc_band3	0.18	0.03	5.25	0.000	***
inc_band4	0.27	0.04	7.69	0.000	***
inc_band5	0.44	0.04	12.26	< 2e-16	***
inc_band6	0.62	0.04	16.31	< 2e-16	***
inc_band7	0.77	0.04	18.86	< 2e-16	***
inc_band8	1.02	0.04	25.15	< 2e-16	***

Source: ONS/ Eurostat

Significance codes

***	0	*	0.05
**	0.01	.	0.1

European Commission

Statistical matching of EU-SILC and the Household Budget Survey to compare poverty estimates using income, expenditures and material deprivation

Luxembourg: Publications Office of the European Union

2013 — 52 pp. — 21 x 29.7 cm

ISBN 978-92-79-22844-5

ISSN 1977-0375

doi:10.2785/4151

Cat. No KS-RA-13-007-EN-N

Theme: Populations and social conditions

Collection: Methodologies & Working papers

HOW TO OBTAIN EU PUBLICATIONS

Free publications:

- via EU Bookshop (<http://bookshop.europa.eu>);
- at the European Union's representations or delegations. You can obtain their contact details on the Internet (<http://ec.europa.eu>) or by sending a fax to +352 2929-42758.

Priced publications:

- via EU Bookshop (<http://bookshop.europa.eu>).

Priced subscriptions (e.g. annual series of the *Official Journal of the European Union* and reports of cases before the Court of Justice of the European Union):

- via one of the sales agents of the Publications Office of the European Union (http://publications.europa.eu/others/agents/index_en.htm)



Publications Office

ISBN 978-92-79-22844-5



9 789279 228445