

# Statistical matching of European Union statistics on income and living conditions (EU-SILC) and the household budget survey

P. SERAFINO AND R. TONKIN

2017 edition





**Statistical matching of  
European Union statistics on  
income and living conditions  
(EU-SILC) and  
the household budget survey**

**P. SERAFINO AND R. TONKIN**

**2017 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**

(\* ) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

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

Luxembourg: Publications Office of the European Union, 2017

ISBN 978-92-79-64144-2

ISSN 2315-0807

doi: 10.2785/933460

Cat. No: KS-TC-16-026-EN-N

**Theme: Population and social conditions**

**Collection: Statistical working papers**

© European Union, 2017

Reproduction is authorised provided the source is acknowledged.

For more information, please consult: <http://ec.europa.eu/eurostat/about/policies/copyright>

Copyright for the photograph of the cover: ©Shutterstock.

For reproduction or use of this photo, permission must be sought directly from the copyright holder.

The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

## Preface

Eurostat is the Statistical Office of the European Union (EU). Its mission is to provide 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) (quasi-)joblessness.

In this context, the Second Network for the Analysis of EU-SILC (Net-SILC2) is bringing together 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.

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://ec.europa.eu/eurostat/en/web/products-statistical-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.



## Abstract<sup>(1)</sup>

The Europe 2020 social inclusion target is 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. So, this paper compares 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, the paper describes methodological work conducted to statistically match expenditure from the Household Budget Survey with income and material deprivation contained within EU-SILC using data for six EU countries. 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.

<sup>(1)</sup> Richard Tonkin and Paola Serafino 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: hie@ons.gsi.gov.uk. The authors would like to thank David Gordon, Tony Atkinson and Eric Marlier for their helpful comments and discussions.





# Table of contents

<b>1. Introduction</b>	<b>9</b>
<b>2. Statistical matching</b>	<b>10</b>
2.1 Overview of statistical matching	10
2.2 Reconciliation of the data sources	10
2.2.1 Household	11
2.2.2 Household reference person	11
2.2.3 Population and Sampling Frame	11
2.2.4 Reference Period	11
2.3 Harmonization of variables	12
2.4 Choosing the matching variables	12
2.4.1 Coherence of distributions	12
2.4.2 Explanatory power of the variables	14
2.5 Matching methods	15
<b>3. Results of statistical matching</b>	<b>16</b>
3.1 Comparison of mean expenditure: EU-SILC imputed versus HBS observed	16
3.2 Comparison of expenditure by matching variables - EU-SILC imputed versus HBS observed	18
3.3 Comparison of expenditure by matching variables – observed versus imputed HBS	19
3.4 Comparison of expenditure by variables not used in statistical matching	20
<b>4. Conditional independence assumption</b>	<b>22</b>
<b>5. Conclusions</b>	<b>23</b>
<b>6. References</b>	<b>24</b>



# 1. Introduction

Most evidence-based 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<sup>(2)</sup>.

It is the consumption of goods and services, along with other inputs such as time that ultimately satisfies a household's wants. Because of this, it 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) as well as the OECD Framework for Statistics on the Distribution of Household Income, Consumption and Wealth (2013).

This Net-SILC2 work package therefore aimed to compare people's exposure to poverty using four different measures: income, expenditure, material deprivation and low work intensity, 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, the first stage of this project involved statistically matching expenditure from the Household Budget Survey (HBS) with income and material deprivation contained within EU Statistics on Income and Living Conditions (SILC). Preliminary work was carried out to develop the methodology using 2005 UK data (see Webber & Tonkin, 2013). This paper builds on that work by presenting the results of statistical matching of HBS and EU-SILC data for a number of countries using the 2010 wave of the HBS. The selection of countries was constrained by both restrictions on access to HBS microdata and the suitability of the two data sources for statistical matching. As a result, the matching was carried out for six EU countries: Belgium; Germany; Spain; Austria; Finland; and the UK.

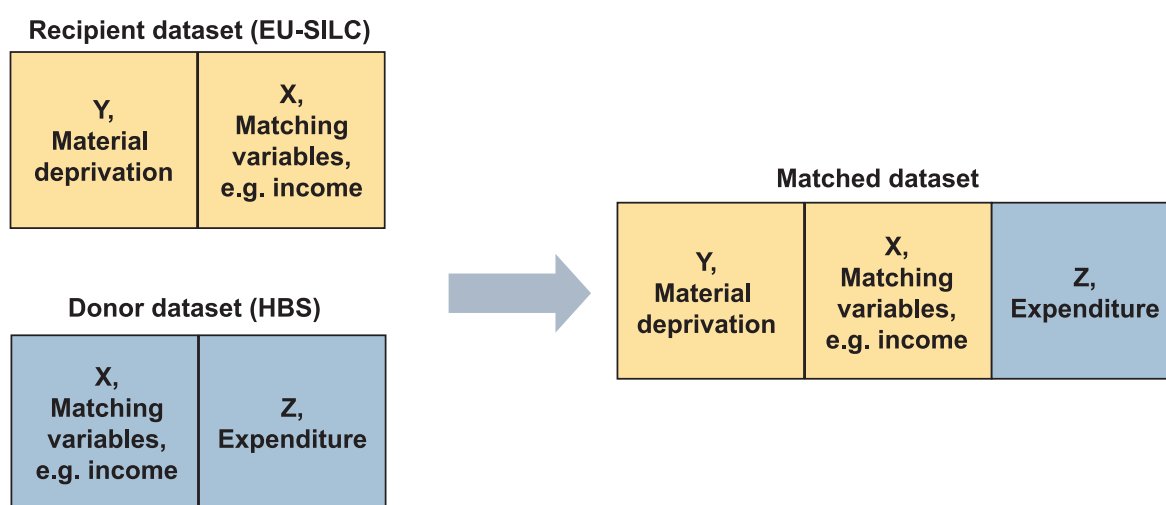
Joint analysis of income and expenditure based poverty and severe material deprivation carried out using the statistically matched datasets presented in this paper is can be found in the chapter by Serafino & Tonkin (2016) in the forthcoming book 'Monitoring social Europe' by Atkinson, Guio and Marlier (Eds).

<sup>(2)</sup> 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 et al. (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 is linked to EU-SILC, which contains information on income, material deprivation and work intensity.



### 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:

- Harmonization of the definition of units.
- Harmonization of reference periods.
- Completion of population.
- Harmonization of variables.
- Harmonization of classifications.
- Adjustment for measurement errors (accuracy).
- Adjustment for missing data.
- Derivation of variables.

Before carrying out statistical matching it is necessary to ensure that the key concepts are defined in a comparable way in the donor and recipient, in this case, the definitions of household, household reference person, population and income reference period.

### 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.

### 2.2.2 Household reference person

In HBS the household reference person (HRP) is clearly defined and identified. 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.

In EU-SILC, there is no household reference person as such. However, there are identifiers for up to two people responsible for the household accommodation. These identifiers are defined in a similar way to the HRP on HBS, except that in the case of joint householders, the default is to report the oldest householder, with no consideration of income.

Since 2001/02 the concept of household reference person (HRP) has been adopted on all UK Government sponsored surveys. Therefore, for the UK, the definition of the person responsible for the accommodation is the same on EU-SILC as the HRP on the HBS. For the remaining countries, the alignment of these variables was tested (see 'Harmonisation of variables'). In common with the UK, for Belgium, Germany and Spain the person responsible for the accommodation aligned with the HRP so was used to allocate a HRP in EU-SILC. However, for Austria and Finland, in order to achieve alignment with the HBS, a HRP was identified by applying its definition directly to the EU-SILC data.

### 2.2.3 Population and Sampling Frame

In all countries studied, both sources cover the same population (private households, excluding collective establishments).

In all countries except Germany both sources also use the same sampling frame. In Belgium, Spain, Austria and Finland national population registers are used as the sampling frame, while in the UK, the Postcode Address File (a list of addresses provided by the UK Post Office) is used.

In Germany, the sampling frames are different. EU-SILC uses a random sample of households who have responded to the German microcensus and have agreed to participate in further voluntary surveys. The sample for the German Household Budget Survey is largely selected from respondents to the sample survey of household income and expenditure (EVS).

### 2.2.4 Reference Period

For the UK in 2010, both the EU-SILC dataset and the HBS dataset measure current annual income in 2010. For the remaining countries, the income reference period for SILC data is the previous calendar year (so 2009 in the case of 2010 data) while data collection periods and income reference periods for the HBS vary. For example, data for the 2010 wave of the HBS in Austria was collected predominantly in 2009 and in Finland was collected in 2012.

For countries where the income reference periods in the HBS did not align exactly with a single year of EU-SILC, some judgement needed to be exercised in determining which dataset to use for statistical matching and in some cases this involved performing the matching more than once to determine which provided the best match using the diagnostics described in this paper. In some instances, advice was also sought from the relevant NSI. This resulted in matching data from the 2010 wave of the HBS with 2009 EU-SILC data for Austria and 2012 EU-SILC data for Finland.

## 2.3 Harmonization of variables

Annex 1 contains the full list of variables common to both EU-SILC and HBS in this analysis. The original statistical matching methodology set out in Webber & Tonkin (2013) was developed using the 2005 HBS, which included a number of variables related to ownership of material goods that were dropped from the 2010 survey. In the UK, these data were still collected on the Living Costs and Food Survey (LCF), the survey that is used to derive the HBS variables. This meant that it was possible to merge these variables onto the 2010 HBS for the UK, and allowed them to be included in the matching process for this country. This was not possible for the other countries examined. Annex 2 contains details of the variables taken from the LCF for the UK and their comparable variables in EU-SILC.

The variables common to both datasets needed to be harmonized across the two sources 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 HBS variable that defines activity status, for instance, is more detailed than the corresponding EU-SILC variable. The detail in HBS therefore needed 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. The table in Annex 3 shows the coding of the derived variables.

Once the variables had 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. Where missing information would have resulted in the loss of too many cases, variables were excluded from further analysis.

## 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. There is no fixed rule regarding what degree of similarity is suitable for statistical matching purposes, though Leulescu & Agafitei (2013) suggest that a 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 Hellinger Distances for the common variables for each of the countries in the analysis. Missing values generally reflect that the variable(s) required were not available on one of the datasets.

Where the HD was found to exceed approximately 5% for the potential matching variables, various options were explored. Two outcomes could be coded to a single one to overcome large discrepancies in the original proportions of the outcomes. This can reduce the HD thereby ensuring that it is suitable as a potential matching variable. However, by limiting the possible outcome responses in the variable reduces its variation, thereby making it potentially less likely to be useful in explaining variations in material deprivation or expenditure. For example, as the HD for DV\_AGE2 was relatively high for Belgium (9.1) an alternative age variable including fewer categories was created (DV\_AGE3).

**Table 1:** Hellinger distances (HD) of EU-SILC and HBS variables (%)

Variable	Belgium	Germany	Spain	Austria	Finland	UK
<b>DV_SEX</b>	1.8	0.8	8.9	3.2	0.4	0.4
<b>DV_AGE2</b>	9.1	5.6	1.2	4.7	2.4	2.1
<b>DV_AGE3</b>	5.1	3.5	0.7	4.3	2.1	2.0
<b>DV_REGION</b>	0.1	-	0.3	0.0	0.3	3.4
<b>DV_URBAN</b>	2.5	3.9	1.1	0.7	32.5	7.4
<b>DV_HHSIZE</b>	1.6	0.7	4.5	0.2	0.7	2.2
<b>DV_HHTYPE1 (Household type)</b>	5.2	1.9	4.2	1.4	8.5	1.5
<b>DV_HHTYPE2 (Household type)</b>	-	-	-	-	0.7	-
<b>DV_DWELL</b>	-	-	-	-	-	6.6
<b>DV_TENURE</b>	-	-	-	-	-	1.7
<b>DV_MARSTA (Marital status)</b>	5.6	3.3	3.4	3.5	3.8	6.8
<b>DV_CONUNI (Consensual union)</b>	35.1	38.3	45.3	22.3	28.0	0.2
<b>DV_MAXEDU (Educational attainment)</b>	19.7	4.0	6.0	1.5	12.2	-
<b>DV_MAXEDU2 (Educational attainment)</b>	18.8	4.0	2.9	2.1	12.2	-
<b>DV_LABOUR</b>	2.1	1.1	-	3.1	1.5	14.7
<b>DV_ACTSTAT (Activity status)</b>	9.3	4.8	12.0	46.5	6.4	4.6
<b>DV_ACTSTAT2 (Activity status)</b>	6.8	4.3	9.9	2.4	8.9	16.9
<b>DV_OCC</b>	14.4	-	8.2	-	10.3	5.6
<b>DV_CAR</b>	-	-	-	-	-	1.9
<b>DV_TV</b>	-	-	-	-	-	1.7
<b>DV_PC</b>	-	-	-	-	-	0.2
<b>DV_WASH</b>	-	-	-	-	-	0.0
<b>DV_PHONE</b>	-	-	-	-	-	3.4
<b>INC_BAND</b>	4.4	3.4	9.5	7.4	1.9	4.1
<b>INC_BAND2</b>	-	-	6.3	7.0	1.6	-
<b>INC_BAND3</b>	-	-	8.1	6.7	1.9	-

Source: EU-SILC 2009 (Austria), 2010 and 2012 (Finland): EU-SILC Users' database; HBS 2010: Eurostat/ONS.

Consideration was also given to recoding outcomes with a high divergence as missing observations to be excluded from the analysis. Although this can reduce the HD to an acceptable level, it can remove an unacceptable number of observations.

Having explored the potential recoding options described, where the HD remained in excess of approximately 5%, the variables were generally dropped. There were exceptions to this. The income band variable with the lowest HD was retained for all countries, regardless of the level of the HD. This meant retaining income band variables in excess of 5% for Spain (6.3%) and Austria (6.7%).

In addition, where visual inspection of the weighted frequencies indicated a high level of similarity, variables with an HD marginally in excess of 5% were retained. This applied in the case of the variable DV\_OCC (5.6%) for the UK. In contrast, the derived variable for marital status, DV\_MARSTA, had an HD of 6.9% for the UK, and comparison of the weighted frequency distributions for this variable revealed large differences in 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. As a result, this variable was dropped. This highlights the importance for effective statistical matching of ensuring that definitions for common variables are harmonised across data sources. Table 2 shows the pool of potential matching variables retained for each country following harmonisation.

**Table 2:** Potential matching variables; those retained for matching are indicated by a solid circle

Variable	Belgium	Germany	Spain	Austria	Finland	UK
DV_SEX	○	○		○	○	○
DV_AGE2			●	●	●	●
DV_AGE3	●	●				
DV_REGION	○		●	○	○	●
DV_URBAN	○	○	○	●		
DV_HHSIZE	●	●	●	●	●	○
DV_HHTYPE (Household type)	●	○	●	○		○
DV_HHTYPE2 (Household type)					○	
DV_TENURE						●
DV_MARSTA (Marital status)	●	●	●	●	●	
DV_CONUNI (Consensual union)						●
DV_LABOUR	○	●		○	○	
DV_MAXEDU		●		●		
DV_MAXEDU2			●			
DV_ACTSTAT (Activity status)					●	●
DV_ACTSTAT2 (Activity status)		●		●		
DV_OCC						●
DV_CAR						●
DV_TV						○
DV_PC						●
DV_WASH						○
DV_PHONE						○
INC_BAND	●	●			●	●
INC_BAND2			●			
INC_BAND3				●		

Source: EU-SILC 2009 (Austria), 2010 and 2012 (Finland): EU-SILC Users' database; HBS 2010: Eurostat/ONS.

## 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:

Let  $\psi_A$  consist of all the common variables such that  $\psi_A$  is independent of  $Y$  given the other common variables in the recipient data set.

Let  $\psi_B$  consist of all the common variables such that  $\psi_B$  is independent of  $Z$  given the other common variables in the donor data set.

Let  $\psi = \psi_A \psi_B$ ; then the other common variables define  $X$ , the matching variables.

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



Material deprivation was defined as a binomial variable, taking a value of 1 if the respondent was materially deprived and 0 otherwise. A logistic regression was fitted to model deprivation using the variables shown in Table 2 for each country.

Next, an expenditure model was estimated on HBS data. As expenditure is highly positively skewed, the stepwise regression model for each country was estimated on the logarithm of expenditure, using the same variables as before.

As stated above, the variables that should be selected for matching are those which are significant in explaining material deprivation and expenditure. Relatively few of the initial pool of matching variables for the Finnish data fulfilled this criterion. As a result, the analysis was repeated for Finland including the derived variable for activity status, despite it having an HD of 6.4%. Table 2 shows which of the initial pool of matching variables were used in the final matching process.

## 2.5 Matching methods

Three different matching methods were used in this analysis, covering the three broad categories of approaches typically used in statistical matching:

- Non-parametric methods
- Parametric methods
- Mixed methods

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 method used, this process was 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 to 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 in the HBS. This was done in three ways:

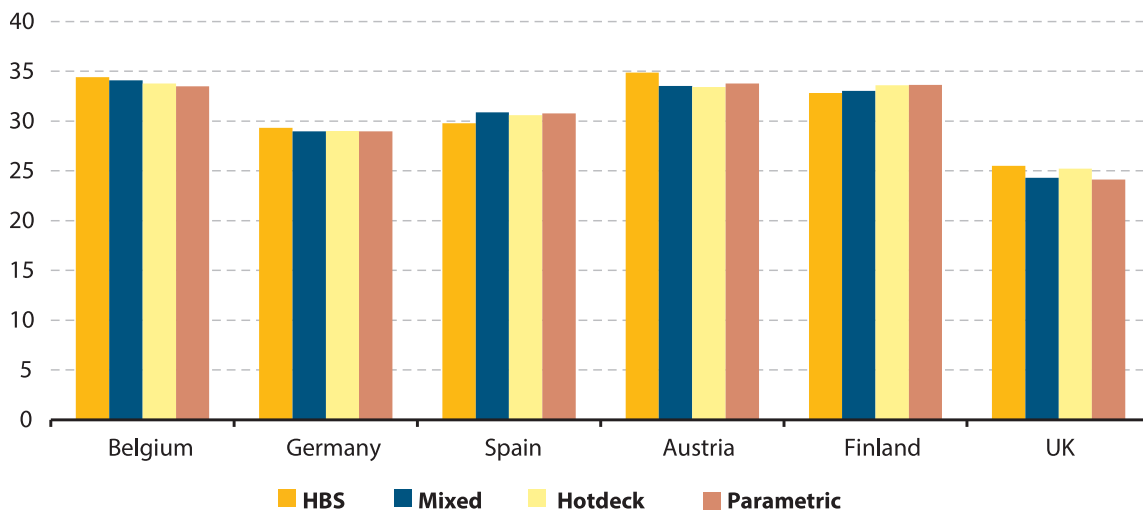
- By comparing mean expenditure by equalised expenditure decile to analyse the consistency of the overall expenditure distribution for each method.
- By comparing the consistency of mean expenditure by variables used in the statistical matching for observed and imputed expenditure.
- By comparing the relationship between expenditure and variables in both datasets but not included in the model.

The following section provides results of some of the main comparisons that were carried out, for all countries studied. Further details of the comparisons not reported here due to space constraints are available on request.

### 3.1 Comparison of mean expenditure: EU-SILC imputed versus HBS observed

Figure 1 compares the mean total expenditure in the HBS with each of the matched datasets. The matching method which most closely replicates mean expenditure in the HBS varies for each country, but overall, the differences between the methods within each country are relatively marginal, particularly for Germany and Spain.

**Figure 1: Mean expenditure for HBS and each of the matching methods**  
(thousands € per annum)

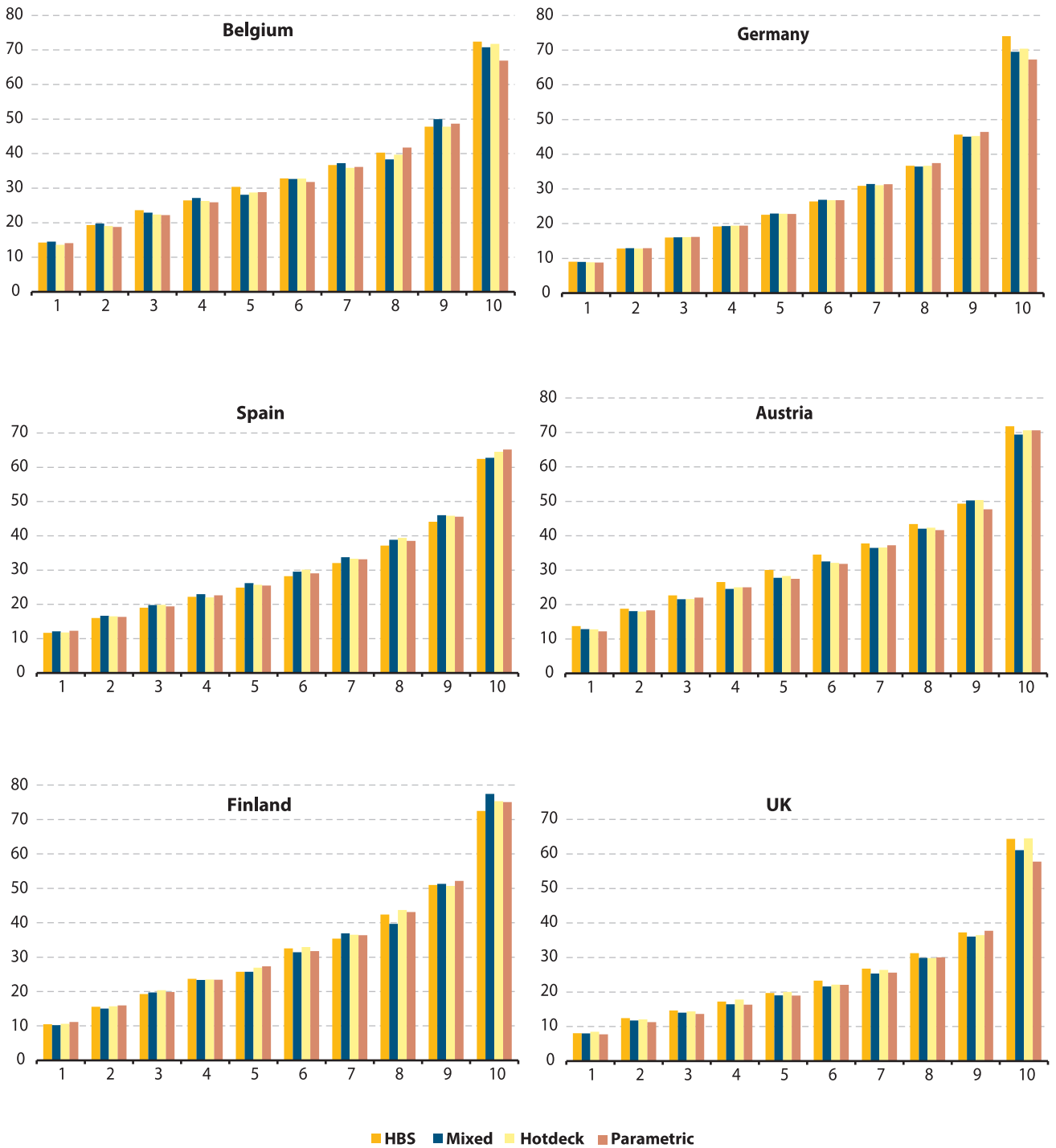


Source: EU-SILC 2009 (Austria), 2010 and 2012 (Finland): EU-SILC Users' database; HBS 2010: Eurostat/ONS.

Looking at the performance of the different matching methods across the expenditure distribution (Figure 2), all three methods appear to be relatively effective in replicating mean expenditure by expenditure deciles. No single method was consistently better across the entire expenditure range for any of the countries in the analysis.

The largest divergence between the values observed and those imputed onto EU-SILC through statistical matching was in the top expenditure decile. However, there is no clear pattern in these differences, with the mean expenditure for the top decile from the statistical matching generally lower than in the HBS in Belgium, Germany and the UK, but higher in Spain and Finland.

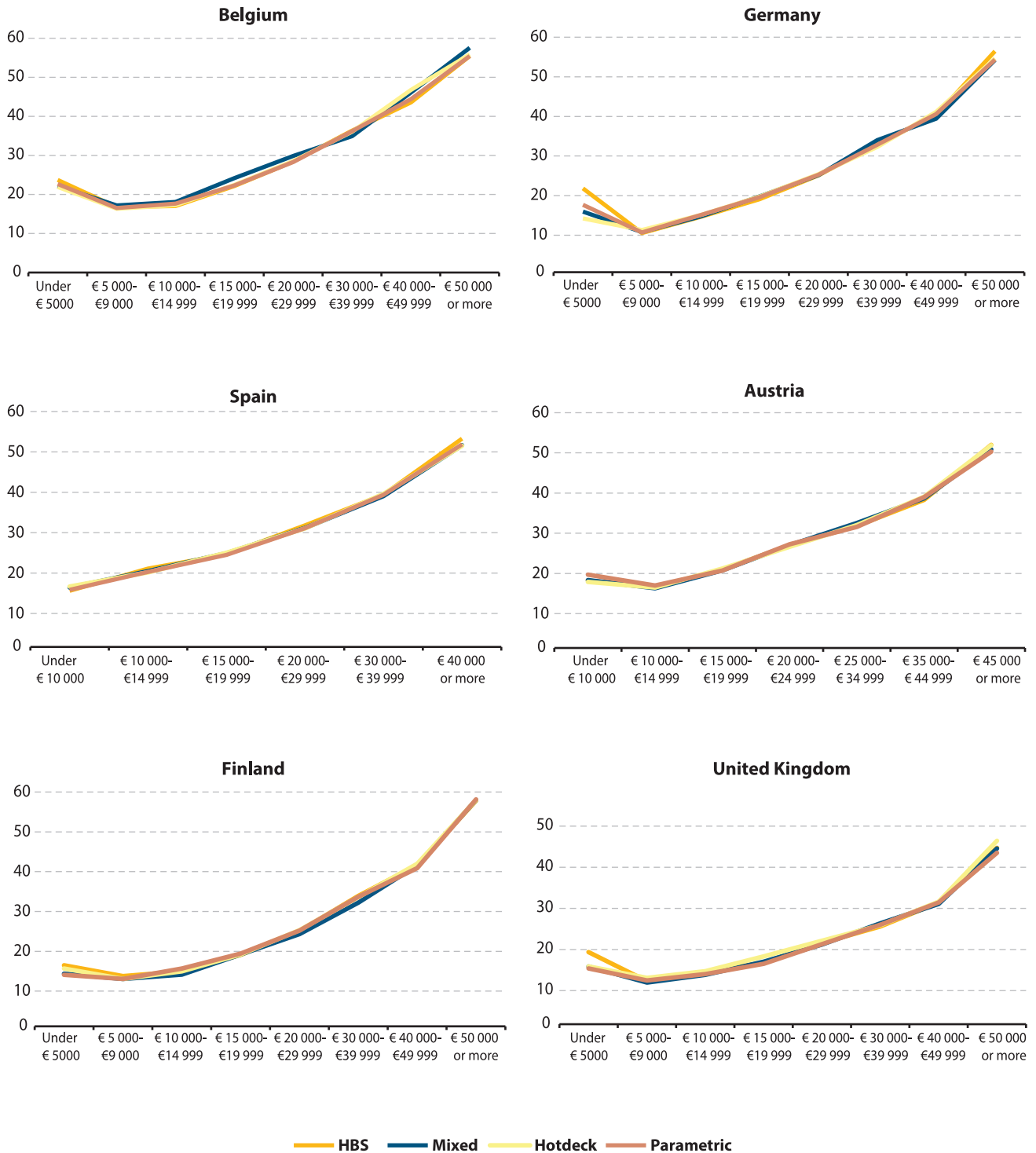
**Figure 2: Mean expenditure by equivalised household expenditure decile for HBS and different matching methods (thousands € per annum)**



Source: EU-SILC 2009 (Austria), 2010 and 2012 (Finland); EU-SILC Users' database; HBS 2010: Eurostat/ONS.

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

**Figure 3:** Mean total household expenditure by income band for HBS and matching methods (thousands € per annum)



Source: EU-SILC 2009 (Austria), 2010 and 2012 (Finland); EU-SILC Users' database; HBS 2010: Eurostat/ONS.

Figure 3 shows the distribution of actual total household expenditure in the HBS and expenditure derived from the matching methods across the income distribution.

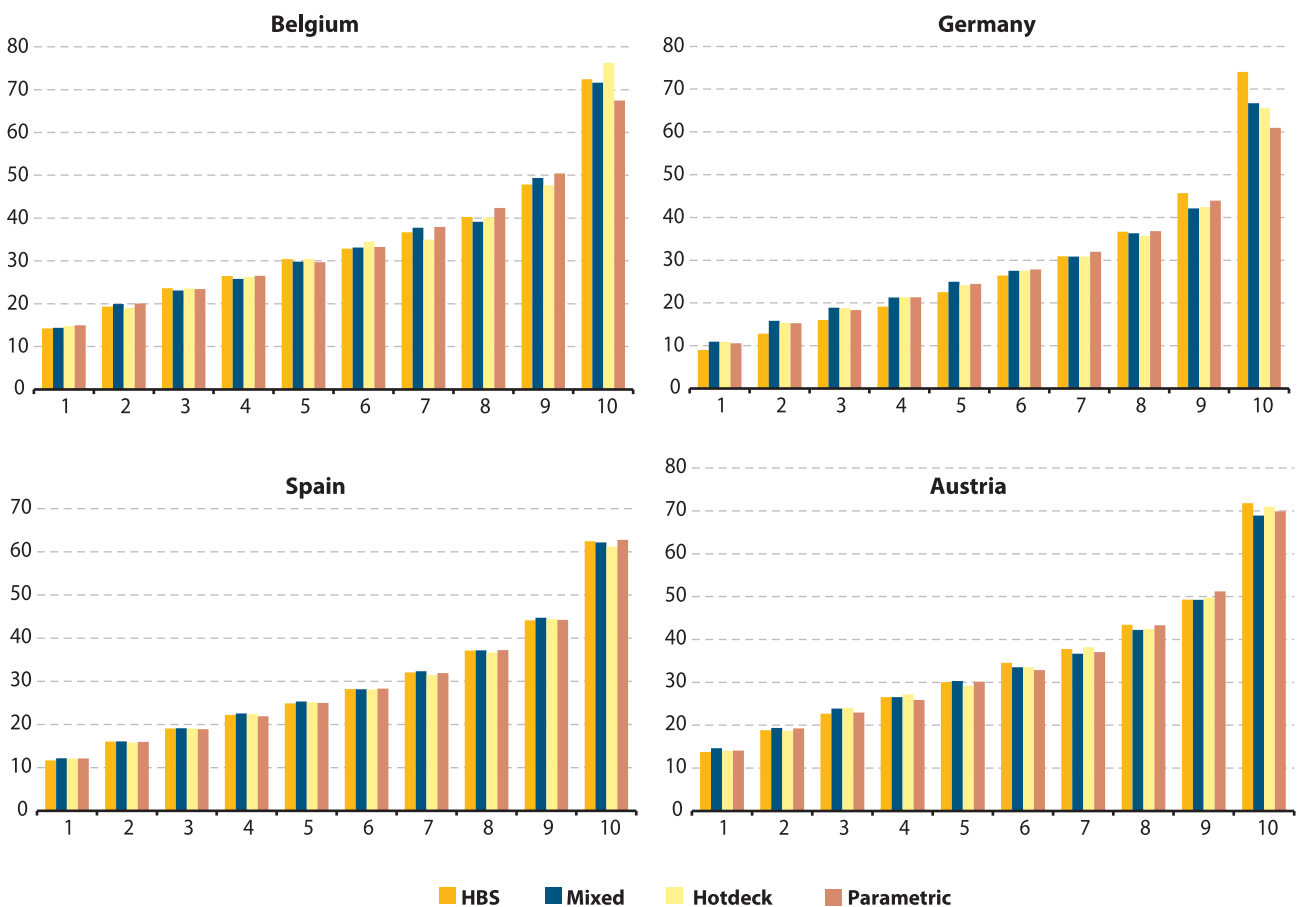
All three methods appear to perform well in general. With the exception of Spain, for all the countries in the analysis, at the low end of the income distribution we see the typical expenditure ‘tick’– higher average expenditure for the bottom income group than households in the second income group. The extent to which this is evident varies across the countries. For Finland, the ‘tick’ is almost negligible, while for Germany and the UK, it is quite considerable, though for these countries, all three matching methods under-estimate its extent. Across all the countries examined, none of the methods appears consistently better than the others at matching across the income distribution.

### 3.3 Comparison of expenditure by matching variables – observed versus imputed HBS

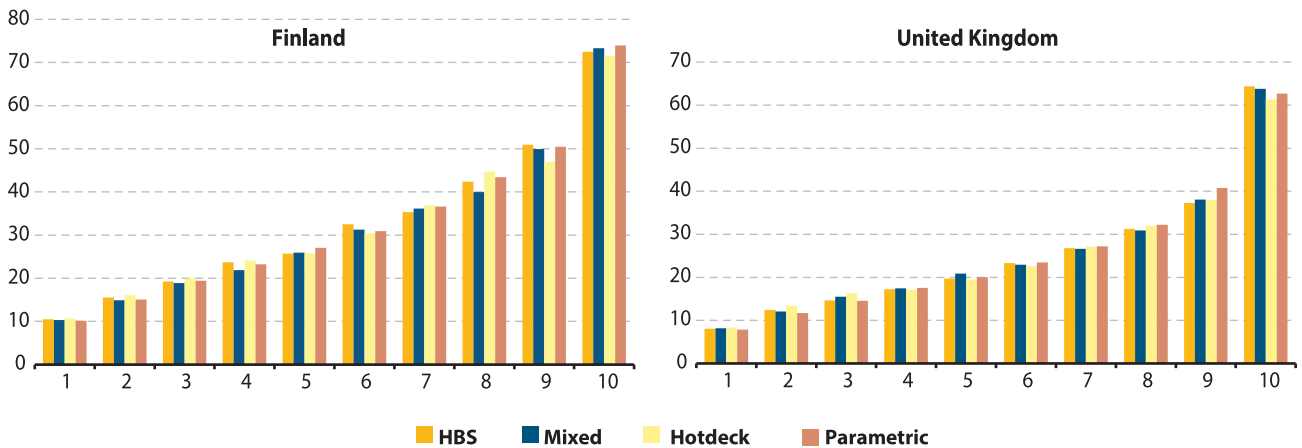
Another way of assessing the quality of the matching processes is to artificially remove expenditure from a random selection of half the HBS sample and then impute expenditure back on using each of the three methods. Figure 4 shows the distribution of mean expenditure by equivalised expenditure decile using this approach.

Again, all three methods appear relatively effective at replicating the expenditure distribution in the HBS, though with some underestimation of the higher deciles for the German data. Overall the mixed methods approach provides the closest match across the distribution as a whole for all six countries.

**Figure 4: Mean household expenditure by equivalised expenditure decile for HBS observed and HBS imputed (thousands € per annum)**



Source: EU-SILC 2009 (Austria), 2010 and 2012 (Finland); EU-SILC Users' database; HBS 2010: Eurostat/ONS

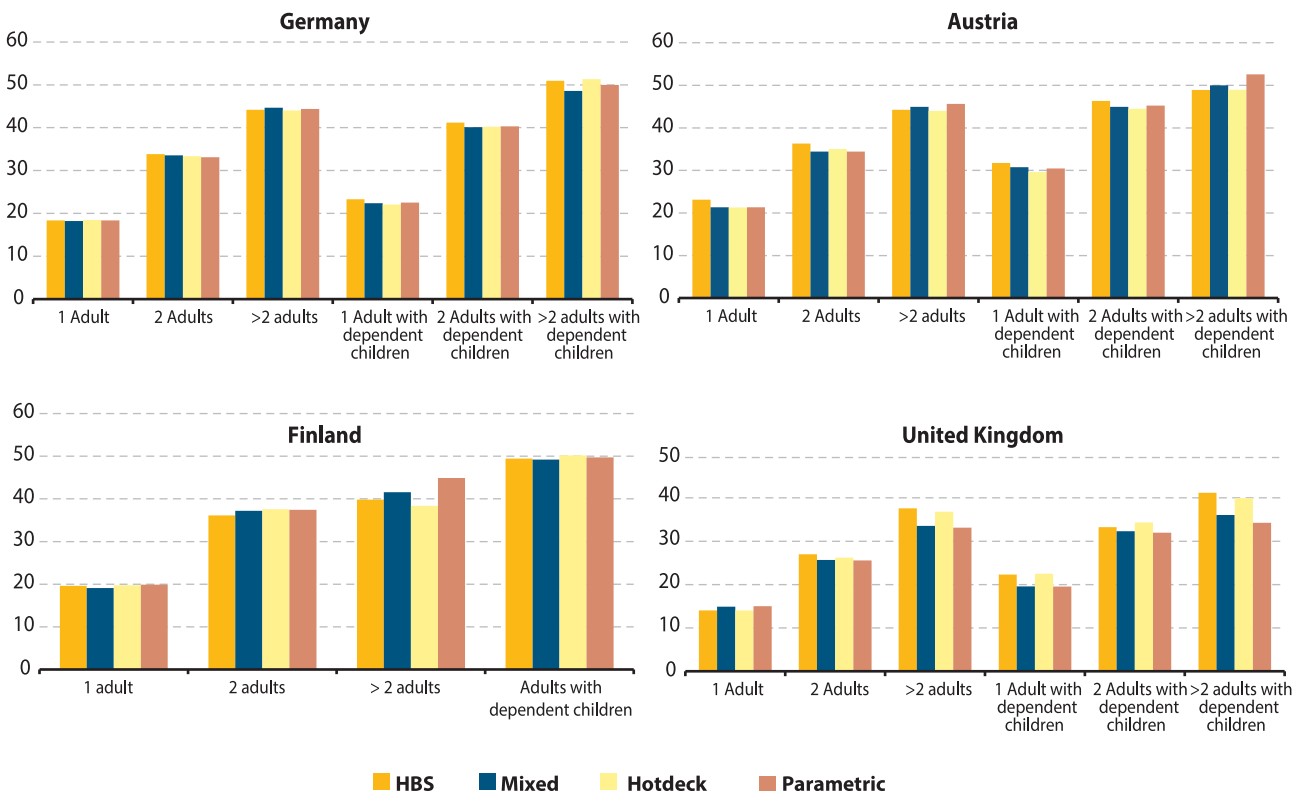


Source: EU-SILC 2009 (Austria), 2010 and 2012 (Finland): EU-SILC Users' database; HBS 2010: Eurostat/ONS

### 3.4 Comparison of expenditure by variables not used in statistical matching

Figures 5, 6 and 7 show the relative performances of the matching methods at estimating expenditure across a variable not used in the matching process. Figure 5 presents expenditure by household type for Germany, Austria, Finland and the UK. Since household type was a matching variable for Belgium and Spain, Figure 6 shows mean expenditure by employment contract for Belgium and Figure 7 shows mean expenditure by activity status for Spain.

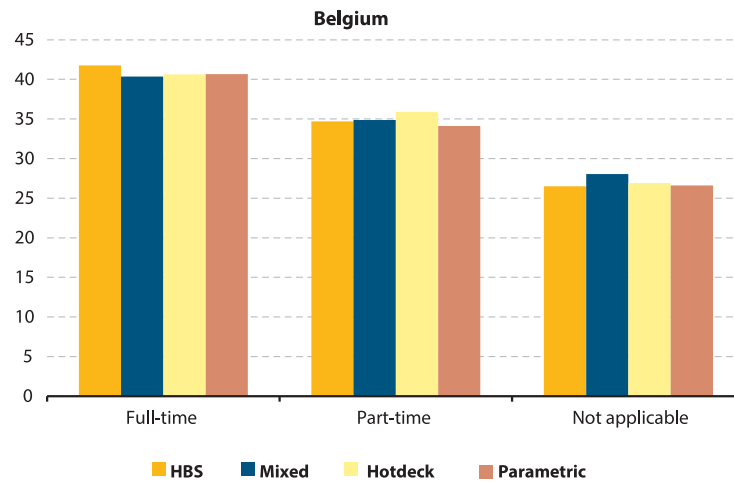
**Figure 5: Mean household expenditure by household type for HBS and matching methods, (thousands € per annum)**



Source: EU-SILC 2009 (Austria), 2010 and 2012 (Finland): EU-SILC Users' database; HBS 2010: Eurostat/ONS.

From Figure 5, all three methods perform reasonably well in replicating mean expenditure for different household types, particularly for the German and Finnish data. For Austria and the UK there is some over/under-estimation of expenditure for certain types of household for all methods. In particular, expenditure appears to be underestimated for single adult households and households comprising two adults. For the UK, expenditure is overestimated for single adult households, but underestimated for households with more than two adults.

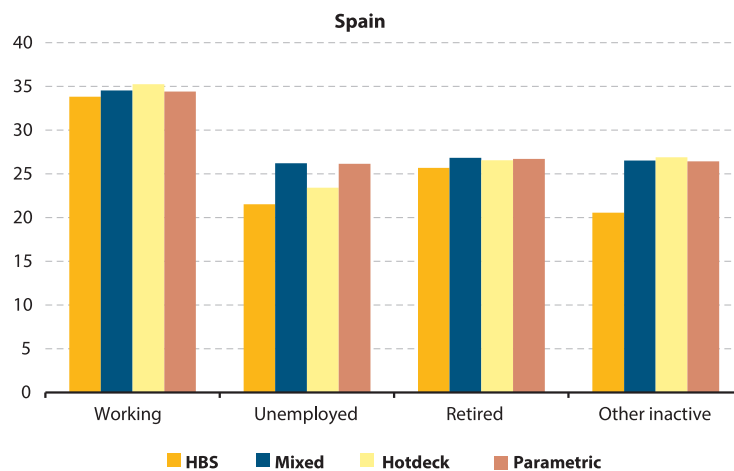
**Figure 6: Mean household expenditure by employment status for HBS and matching methods, Belgium, 2009**  
(thousands € per annum)



Source: EU-SILC 2009 (Austria), 2010 and 2012 (Finland): EU-SILC Users' database; HBS 2010: Eurostat/ONS.

All three methods also perform well in replicating mean expenditure by type of employment contract for the Belgian data (Figure 6) and mean expenditure by activity status for the Spanish data (Figure 7), though for the latter all three methods over-estimate expenditure to differing degrees.

**Figure 7: Mean household expenditure by activity status for HBS and matching methods, Spain, 2010**  
(thousands € per annum)



Source: EU-SILC 2009 (Austria), 2010 and 2012 (Finland): EU-SILC Users' database; HBS 2010: Eurostat/ONS.

## 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 in all 6 countries. 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.

The calculation of Fréchet bounds for this data was explored by Webber and Tonkin (2013). However, as it is necessary to first to harmonise the joint distribution of the matching variables (Renssen, 1998), something which is extremely difficult to carry out successfully with a large number of matching variables, it was only possible to use two matching variables in this process: `inc_band` and `DV_HHSIZE`. This limited the usefulness of the analysis as, while the uncertainty space was relatively large; it is likely that the use of a greater number of matching variables would have reduced this range of plausible values. Due to these limitations, equivalent analysis has not been presented in the current paper.



## 5. Conclusions

For the countries included in this paper, the results of the statistical matching are encouraging. Analysis of the joint distributions of the matching variables with imputed and actual expenditure indicates that the matching has been broadly effective, and that all three methods tested provide relatively good results, though the mixed methods and hotdeck approaches performed marginally better overall. For those countries for which imperfect alignment with the income reference period justified matching with more than one EU-SILC dataset, the choice of year made relatively little difference to the success of the matching, perhaps indicating that the relationship between expenditure and matching variables had changed relatively little between years.

It should be noted that the countries examined in this analysis are present mainly based on the ability to reconcile the EU-SILC and HBS datasets to a sufficient level to make matching viable, as well as data availability. Attempts to harmonise the available datasets ex-post to a sufficient degree for some other countries were unsuccessful. This highlights important issues regarding both the number of common variables and degree of harmonisation between them in EU-SILC and the HBS.

The number of potential matching variables available between the HBS and EU-SILC was actually lower in 2010 than 2005. Variables such as tenure status, dwelling type and number of rooms were not included on the 2010 HBS. This reduction in the pool of core variables that can potentially be used for matching could reduce the quality of matching or prevent it altogether for some countries. If there is enthusiasm for facilitating joint analysis, it is recommended that such variables are reintroduced to future waves of the HBS.

The definitions used for the common socio-demographic and related variables are equally important. For example, although the definition of the household reference person in HBS appears to align fairly well with the definition of the person responsible for the accommodation in EU-SILC, the critical difference in how to identify the relevant person where more than one person fills that role, using income in HBS and age in EU-SILC, may be significant in some countries. Certainly, based on the 2010 data examined so far for other countries, this lack of alignment is proving a serious obstacle to carrying out ex-post matching of this kind.

More generally, while statistical matching offers the opportunity to enhance the analytical possibilities of existing data collection exercises, particularly at an international level, until there is better harmonisation of definitions and outputs across all surveys, the opportunities are limited. The lack of harmonisation across different surveys presents the greatest barrier to this goal. The current modernisation programme for EU-SILC and ESS Social Surveys as a whole under the proposed Integrated European Social Statistics (IESS) regulation provides the opportunity to ensure greater comparability and consistency in the variables collected to allow these statistical techniques to be used.

Consideration should also be given to how cooperation in this area between the ESS and ESCB can be enhanced, given their role in running the Household Finance and Consumption Survey (HFCS), the main source of microdata on wealth in many EU countries. Improved methodological cooperation could significantly increase the possibility of producing integrated statistics on income, consumption and wealth at a European level, in line with the recommendations of OECD (2013).

In this context, it is welcome that Eurostat are currently developing EU-SILC module on consumption, wealth and over-indebtedness, which could be used as 'hooks' to improve the potential quality of matching between SILC and both the HBS and HFCS. These 'hook' variables will need to be carefully selected in order to fit well amongst other variables on the survey, not be burdensome on respondents or NSIs and have a strong relationship between the variables of interest in both sources. Ideally, such variables should also have standalone value in order to ensure the module is useful even where circumstances limit the opportunities for statistical matching. A number of countries will be testing these new variables in the 2017 EU-SILC operation.

## 6. References

- Atkinson, T., Cantillon, B., Marlier, E. and Nolan, B. (2002), *Social Indicators: The EU and Social Inclusion*, Oxford University Press, Oxford.
- Brewer, M., and O’Dea, C. (2012), “Measuring living standards with income and consumption: evidence from the UK”, *ISER Working Paper Series* n°2012-05, Institute for Social and Economic Research (ISER), Essex.
- Cutler, D., and Katz, L. (1991), “Macroeconomic performance and the disadvantaged”, *Brookings Papers on Economic Activity*, 2: 1-74.
- D’Orazio, M., Di Zio, M. and Scanu, M. (2006), *Statistical Matching: Theory and Practice*, John Wiley & Sons Ltd, Chichester.
- Jorgenson, D. and Slesnick, D. (1987), “Aggregate consumer behavior and household equivalence scales”, *Journal of Business and Economic Statistics*, 5(2): 219-232.
- Leulescu, A. and Agafitei, M. (2013), *Statistical matching: a model based approach for data integration*, Eurostat methodologies and working paper, Eurostat, Luxembourg.
- Meyer, B. and 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*. GESIS - Leibniz Institute for the Social Sciences, Mannheim. Available at: [http://www.gesis.org/fileadmin/upload/institut/wiss\\_arbeitsbereiche/soz\\_indikatoren/Publikationen/Household-Expenditures-Research-Report.pdf](http://www.gesis.org/fileadmin/upload/institut/wiss_arbeitsbereiche/soz_indikatoren/Publikationen/Household-Expenditures-Research-Report.pdf) [Accessed: 25 April 2016]
- OECD (2013), *OECD framework for statistics on the distribution of income, consumption and wealth*. OECD Publishing.
- Sabelhaus, J., Johnson, D., Ash, S., Swanson, D., Garner, T., Greenlees, J. and Henderseon, S. (2011), “Is the Consumer Expenditure Survey representative by income?”, in *Conference on Improving the Measurement of Consumer Expenditures* sponsored by Conference on Research in Income and Wealth and the NBER, with support from the Centre for Microdata Methods and Practice. Available at: <http://conference.nber.org/confer/2011/CRIWf11/CRIWf11prg.html> [Accessed: 25 April 2016]
- Serafino, P. and Tonkin, R.P. (forthcoming) “Comparing poverty estimates using income, expenditure and material deprivation”, in A.B. Atkinson, A.-C. Guio and E. Marlier (editors), *Monitoring social Europe*, Publishing Office of the European Union, Luxembourg.
- Stiglitz, J. E., Sen, A., and Fitoussi, J.-P. (2009), *Report by the Commission on the Measurement of Economic Performance and social progress*, French Government and the National statistics agency (INSEE), France.
- Webber, D. and Tonkin, R.P. (2013), *Statistical matching of EU-SILC and the Household Budget Survey to compare poverty estimates using income, expenditure and material deprivation (2013 edition)*, Eurostat Methodologies and working papers, Publications office of the European Union, Luxembourg.

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

EU-SILC		HBS	
Coding	Description	Coding	Description
<b>DB040</b>	<b>Region NUTS 2</b>	<b>HA08</b>	<b>Region NUTS2</b>
Specific to each country			
<b>DB100</b>	<b>Degree of Urbanisation</b>	<b>HA09</b>	<b>Population density domain</b>
1	Densely populated (at least 500 inhabitants/km <sup>2</sup> )	1	Densely populated (at least 500 inhabitants/ km <sup>2</sup> )
2	Intermediate (between 100 and 499 inhabitants/km <sup>2</sup> )	2	Intermediate (between 100 and 499 inhabitants/km <sup>2</sup> )
3	Sparsely populated (less than 100 inhabitants/km <sup>2</sup> )	3	Sparsely populated (less than 100 inhabitants/ km <sup>2</sup> )
<b>HX040</b>	<b>Household size</b>	<b>HB05</b>	<b>Household Size</b>
0+	Number of people in household	0+	Number of people in household
<b>HX060</b>	<b>Household type (age limit for children is 16 years old)</b>	<b>HB07.4</b>	<b>Type of household - 1 (age limit for children is 16 years old)</b>
5	One person household	1	1 adult
6	2 adults, no dependent children, both adults under 65 years	2	2 adults
7	2 adults, no dependent children, at least one adult under 65 years	3	More than 2 adults
8	Other households without dependent children	4	1 adult with dependent children
9	Single parent household, one or more dependent children	5	2 adults with dependent children
10	2 adults, one dependent child	6	More than 2 adults with dependent children
11	2 adults, two dependent children	9	Other
12	2 adults, three or more dependent children		
13	Other households with dependent children		
16	Other (these households are excluded from Laeken indicators calculation)		
<b>RB090</b>	<b>Sex</b>	<b>MB02</b>	<b>Sex of Reference Person</b>
1	Male	1	Male
2	Female	2	Female
<b>PX020</b>	<b>Age</b>	<b>MB03</b>	<b>Age in completed years of reference person</b>
00-120	Age in years	00-98	98 Years and older
		99	Not Specified
<b>PB190</b>	<b>Marital Status</b>	<b>MB04</b>	<b>Marital Status of Reference Person</b>
1	Never Married	1	Never Married
2	Married	2	Married or in a registered partnership
3	Separated	3	Widowed or with registered partnership that ended with death of partner (not remarried or in new registered partnership)
4	Widowed	4	Divorced or with registered partnership that was legally dissolved (not remarried or in new registered partnership)
5	Divorced	9	Not specified
<b>PB200</b>	<b>Consensual Union</b>	<b>MB04.2</b>	<b>Consensual Union of reference person</b>
1	Yes, on a legal basis	1	Person living in consensual union
2	Yes, without a legal basis	2	Person not living in consensual union
3	No	9	Not specified
<b>RB210</b>	<b>Main activity status during the income reference period</b>	<b>ME01</b>	<b>Current Activity status of household member</b>
1	At work	1	Working, including with employment but temporarily absent
2	Unemployed	2	Unemployed
3	In retirement or early retirement or has given up business	3	In retirement or early retirement or has given up business
4	Other inactive person	4	Pupil, student, further training, unpaid work experience

EU-SILC		HBS	
Coding	Description	Coding	Description
		5	Fulfilling domestic tasks
		6	Permanently disabled
		7	In compulsory military or community service
		8	Not applicable (legal age to work unfulfilled)
		9	Not specified
<b>PL030</b>	<b>Self-defined current economic status</b>	<b>ME02</b>	<b>Hours worked</b>
1	Employee working full-time	1	Full time
2	Employee working part-time	2	Part time
3	Self-employed working full time (including family worker)	8	Not applicable (do not work)
4	Self-employed working part time (including family worker)	9	Not specified
5	Unemployed		
6	Pupil, student, further training, unpaid work experience		
7	In retirement or in early retirement or has given up business		
8	Permanently disabled or/and unfit to work		
9	In compulsory military community or service		
10	Fulfilling domestic tasks and care responsibilities		
11	Other inactive person		
<b>PL140</b>	<b>Type of contract</b>	<b>ME03</b>	<b>Type of work contract for household member</b>
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 (does not work)
		9	Not specified
<b>PL040</b>	<b>Status in employment</b>	<b>ME12</b>	<b>Status in employment of household member</b>
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
<b>PL050</b>	<b>Occupation</b>	<b>ME0988</b>	<b>Occupation of household member (ISCO 1988 (COM))</b>
11	Legislators, Senior officials, and managers	01	Legislators, senior officials and managers
12	Corporate managers	02	Professionals
13	Managers of small enterprises	03	Technicians and associate professionals
21	Physical, mathematical, and engineering science professionals	04	Clerks
22	Life science and health professionals	05	Service Workers and shop and market sales workers
23	Teaching professionals	06	Skilled agricultural and fishery workers
24	Other professionals	07	Craft and related trades workers
31	Physical and engineering science associate professionals	08	Plant and Machine operators and assemblers
32	Life science and health associate professionals	09	Elementary occupations
33	Teaching associate professionals	00	Armed Forces
34	Other associate professionals	88	Not applicable (legal age to work unfulfilled)
41	Office clerks	99	Not Specified

EU-SILC		HBS	
Coding	Description	Coding	Description
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, handcraft, 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:** List of variables from UK Living Costs and Food Survey (LCF) in common with EU-SILC, 2010

EU-SILC Coding	Description	LCF	Description
<b>HH010</b>	<b>Dwelling type</b>		<b>Dwelling type</b>
1	Detached house	1	Whole house/bungalow detached
2	Semi-detached or terraced house	2	Whole house/bungalow semi-detached
3	Apartment or flat in a building with less than 10 dwellings	3	Whole house/bungalow terrace
4	Apartment or flat in a building with more than 10 dwellings	4	Purpose built flat maisonette
5	Some other kind of accommodation	5	Part of house converted flat
		6	Others
<b>HH020</b>	<b>Tenure status</b>		<b>Tenure status</b>
1	Owner	1	Local authority (Furnished/Unfurnished)
2	Tenant or subtenant paying rent at prevailing or market rate	2	Housing Association
3	Accommodation is rented at a reduced rate (lower price than the market price)	3	Other rented unfurnished
4	Accommodation is provided free	4	Rented furnished
		5	Owned with mortgage
		6	Owned by rental purchase
		7	Owned outright
		8	Rent free
<b>HH030</b>	<b>Number of rooms available to the household</b>		<b>Number of rooms used solely by the household</b>
1-9	1-9	0--100	0--100
10	10 or more		
<b>HS070</b>	<b>Do you have a telephone (including mobile phone)?</b>		<b>Phone?</b>
1	Yes	1	Fixed telephone
2	No - cannot afford	2	Mobile telephone
3	No - other reason	3	Fixed and mobile telephone
		4	No telephone present
<b>HS080</b>	<b>Do you have a colour TV?</b>		<b>TV?</b>
1	Yes	1	Television present
2	No - cannot afford	2	No television present
3	No - other reason		
<b>HS090</b>	<b>Do you have a computer?</b>		<b>Computer?</b>
1	Yes	1	Home computer present
2	No - cannot afford	2	No home computer
3	No - other reason		
<b>HS100</b>	<b>Do you have a washing machine?</b>		<b>Washing machine?</b>
1	Yes	1	Washing machine
2	No - cannot afford	2	No washing machine present
3	No - other reason		
<b>HS110</b>	<b>Do you have a car?</b>		<b>Car?</b>
1	Yes	1	Yes
2	No - cannot afford	2	No
3	No - other reason		

**Annex 3: Complete list of derived variables**

Harmonised Coding	Description
<b>DV_SEX</b>	<b>Sex of reference person</b>
1	Male
2	Female
<b>DV_AGE2</b>	<b>Age in completed years of reference person</b>
1	16-25
2	26-35
3	36-45
4	46-55
5	56-65
6	66-75
7	76 +
<b>DV_AGE3</b>	<b>Age in completed years of reference person</b>
1	16-30
2	31-40
3	41-50
4	51-60
5	61-70
6	71 +
<b>DV_REGION</b>	<b>Region of household</b>
1-....	Dependent on country
<b>DV_URBAN</b>	<b>Population density domain</b>
1	Densely populated (at least 500 inhabitants/km <sup>2</sup> )
2	Intermediate (between 100 and 499 inhabitants/km <sup>2</sup> )
3	Sparsely populated (less than 100 inhabitants/km <sup>2</sup> )
<b>DV_HHSIZE</b>	<b>Household Size</b>
0-5	Number of people in household
6	More than 5 people in household
<b>DV_HHTYPE</b>	<b>Household type</b>
1	One adult
2	Two adults
3	More than two adults
4	One adult with dependent children
5	Two adults with dependent children
6	More than two adults with dependent children
<b>DV_HHTYPE2</b>	<b>Household type</b>
1	One adult
2	Two adults
3	More than two adults
4	Adults with dependent children
<b>DV_DWELL</b>	<b>Dwelling Type</b>
1	Detached House
2	Semi detached or terraced house
3	Apartment
4	Other

Harmonised Coding	Description
<b>DV_TENURE</b>	<b>Tenure Status</b>
1	Owner
2	Renting
3	Accommodation is provided free
<b>DV_MARSTA</b>	<b>Marital Status of Reference Person</b>
1	Never Married
2	Married
3	Widowed
4	Divorced or separated
<b>DV_CONUNI</b>	<b>Consensual Union</b>
0	No
1	Yes
<b>DV_LABOUR</b>	<b>Labour status</b>
1	Full time
2	Part time
3	Not applicable
<b>DV_MAXEDU</b>	<b>Highest level of education achieved</b>
1	ISCED levels 0-1
2	ISCED level 2
3	ISCED levels 3-4
4	ISCED levels 5-6
<b>DV_MAXEDU2</b>	<b>Highest level of education achieved</b>
1	ISCED levels 0-2
2	ISCED levels 3-4
3	ISCED levels 5-6
<b>DV_ACTSTAT</b>	<b>Activity Status</b>
1	Working full time
2	Working part time
3	Unemployed
4	Student
5	Retired
6	Disabled
8	Other inactive
<b>DV_ACTSTAT2</b>	<b>Activity Status</b>
1	Working
2	Unemployed
3	Retired
4	Other inactive
<b>DV_OCC</b>	<b>Occupation</b>
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
11	Economically inactive



Harmonised Coding	Description
<b>DV_CAR</b>	<b>Do you have a car?</b>
0	No
1	Yes
<b>DV_TV</b>	<b>Do you have a TV?</b>
0	No
1	Yes
<b>DV_PC</b>	<b>Do you have a PC?</b>
0	No
1	Yes
<b>DV_WASH</b>	<b>Do you have a washing machine?</b>
0	No
1	Yes
<b>DV_PHONE</b>	<b>Do you have a telephone?</b>
0	No
1	Yes
<b>INC_BAND</b>	<b>Income band</b>
1	Under €5 000
2	€5 000-€9 999
3	€10 000-€14 999
4	€15 000-€19 999
5	€20 000-€29 999
6	€30 000-€39 999
7	€40 000-€49 999
8	€50 000+
<b>INC_BAND2</b>	<b>Income band</b>
1	Under €10 000
2	€10 000-€14 999
3	€15 000-€19 999
4	€20 000-€29 999
5	€30 000-€39 999
6	€40 000+
<b>INC_BAND3</b>	<b>Income band</b>
1	Under €10 000
2	€10 000-€14 999
3	€15 000-€19 999
4	€20 000-€24 999
5	€25 000-€34 999
6	€35 000-€44 999
7	€45 000+

## HOW TO OBTAIN EU PUBLICATIONS

### Free publications:

- one copy:  
via EU Bookshop (<http://bookshop.europa.eu>);
- more than one copy or posters/maps:  
from the European Union's representations ([http://ec.europa.eu/represent\\_en.htm](http://ec.europa.eu/represent_en.htm));  
from the delegations in non-EU countries ([http://eeas.europa.eu/delegations/index\\_en.htm](http://eeas.europa.eu/delegations/index_en.htm));  
by contacting the Europe Direct service ([http://europa.eu/europedirect/index\\_en.htm](http://europa.eu/europedirect/index_en.htm)) or  
calling 00 800 6 7 8 9 10 11 (freephone number from anywhere in the EU) (\*).

(\* ) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

### Priced publications:

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

# Statistical matching of European Union statistics on income and living conditions (EU-SILC) and the household budget survey

The Europe 2020 social inclusion target is 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. So, this paper compares 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, the paper describes methodological work conducted to statistically match expenditure from the Household Budget Survey with income and material deprivation contained within EU-SILC using data for six EU countries. 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.

---

**For more information**

<http://ec.europa.eu/eurostat/>

