Data Quality Issues in the EU-SILC Intergenerational Modules

Table of Contents

Overview of Data Quality Issues in the EU-SILC Intergenerational Module

In this document based on EU-SILC 2005 version three (June 2008) we report some finding reported at an EQUALSOC seminar held in Dublin at the ESRI Dublin on 26/27 March. These include

- An overview by Maître and Whelan relating to overall data quality issues.
- Schneider and Müeller provide a detailed analysis of the education variables that is based on analysis across a range of date sources
- A further discussion of EU-SILC data issues by Van Rie and Van den Bosch including those relating to the reference period employed, sampling and levels of aggregation of education variables.
- Watson provides an overview of the issues involved in coding ESeC with EU-SILC data
- The discussion by Vallet looks in detail at results from the French utilising both the national survey and the common EU-SILC data set.
- Barbieri and Carossa focus on a range of issues relating not to the intergenerational module as such but issue relating to the reliability of income information and cross-national comparability in EU-SILC.
- Finally Whelan, Nolan and Maître compare the effects of a set of intergenerational influences.

General Data Quality Issues in the EU-SILC 2005 Inter-generational Module

Bertrand Maitre and Christopher T. Whelan

The EU-SILC 2005 includes a special module on inter-generational transmission of poverty. The information was collected for all household members or selected respondents aged over 24 and less then 66 at the end of the reference period of the income. The reference period in relation to the module is when the interviewee was a young teenager, between the ages of 12 and 16.

The module includes 11 variables (excluding the flag variables) and Table 1 shows the percentage of missing values across countries for these variables. Table 1 presents the "true" missing values that are the missing values corrected for when there were no father/mother in the household or when in case of the occupation variable for example when the father/mother did not have any activity (mainly on home duties for the mothers) to have an occupational code.

Table 1: Percentage of missing values on the multigenerational variables across countries

	PM010	PM020	PM030	PM035	PM040	PM050	Pm060	PM070	PM080	PM090	PM100
	'Family composition when young'	'Year of birth father'	'Year of birth mother'	'Siblings'	'Education father'	'Education mother'	'Activity father'	'Occupation father'	'Activity mother'	'Occupation mother'	'Economic circumstances '.
AT	1	15	9	2	8	**33**	5	2	3	1	3
BE	2	10	7	2	7	6	4	19	4	4	2
CY	0	5	1	0	0	0	0	0	0	0	0
CZ	1	11	3	0	1	1	1	1	2	1	2
DE	2	9	4	3	14	14	4	11	3	4	**100**
DK	0	0	0	0	12	23	1	8	1	**38**	1
EE	0	27	9	0	1	1	0	1	0	1	1
ES	2	18	14	4	5	4	3	6	3	1	4
FI	3	18	8	3	5	5	5	4	4	16	6
FR	0	12	8	2	11	7	8	9	4	1	7
GR	0	5	2	0	0	0	0	0	0	1	**100**
HU	1	14	5	3	2	1	2	6	1	2	2
IE	2	10	6	0	7	4	6	10	2	0	2
IS	1	17	7	1	3	4	1	3	1	3	2
IT	0	13	31	1	0	0	0	0	0	0	1
LT	1	16	3	1	2	2	1	1	1	0	2
LU	0	7	5	0	6	6	1	2	1	2	2
LV	1	27	5	1	3	3	1	1	1	1	3
NL	2	16	9	3	15	12	1	4	1	1	8
NO	0	10	5	0	12	6	10	12	9	**39**	4
PL	0	16	12	0	7	5	6	9	4	4	3
PT	1	17	10	1	2	2	2	1	1	0	**100**
SE	3	6	6	**100**	5	4	5	**74**	3	**51**	6
SI	0	16	6	1	1	2	1	3	2	5	2
SK	0	6	3	2	3	3	4	8	2	2	1
UK	9	24	17	9	**39**	**34**	**20**	**58**	14	19	20

From this table we see than the main problems relate to the high percentages of missing values for the education variable of the father (PM050) for the UK (39%) as well as for the education of the mother in Austria (33%) and in the UK (34%). Regarding the occupation of the father (PM070) we see a very high percentage of missing values in Sweden (74%) and in the UK (58%). Looking at the same variable for the mother occupation (PM090) we find three countries with high percentage that are Denmark (38%), Norway (39%) and Sweden (51%).

Finally the “economic circumstances” variable (PM100) is completely missing for three countries, Germany and Portugal while the information was collected during the survey which was not the case in Greece.

A second issue relating to the data set is the distribution of the education variable of the father (PM040) and of the mother (PM050) as shown in Table 2 and Table 3.

The main concern on the education variable of the father is the large number of countries having no or almost no cases of father having primary or less than primary education level as can be seen in Table 2. Austria, the Czech Republic, Denmark, Germany and Norway have almost nobody less than lower secondary education level while in the UK 54% of father have less than primary education level. A the other end of the spectrum Germany and Norway have a significant high percentage of fathers with a high level of education (ISCED 5 & 6) at 32 % and 21 % respectively.

Table 2: Education of the Father across Countries

	Less than primary	Primary	Lower secondary	Upper secondary	Post secondary (non-tertiary)	First stage of tertiary education (ISCED 5 & 6)
AT	**0.0**	**0.2**	59.7	35.4	0.3	4.5
BE	17.1	32.7	15.7	17.5	2.1	14.9
CY	30.6	42.3	6.4	13.5	0.8	6.4
CZ	**0.0**	**0.7**	20.4	70.1	1.0	7.7
DE	**0.0**	**2.0**	12.9	51.1	1.6	**32.3**
DK	**0.0**	**0.0**	41.8	39.1	4.4	14.7
EE	1.9	22.3	27.4	29.9	5.4	13.0
ES	26.1	54.3	5.6	5.5	0.4	8.1
FI	7.6	24.6	40.0	13.7	0.8	13.3
FR	6.7	56.2	21.7	6.4	0.3	8.7
GR	31.9	46.0	9.0	5.3	2.3	5.6
HU	1.5	26.1	23.4	37.7	3.3	8.0
IE	3.1	68.9	11.1	7.0	2.6	7.3
IS	3.3	20.5	16.5	35.0	13.0	11.8
IT	18.8	51.2	16.2	10.8	0.0	3.1
LT	11.5	40.0	18.1	11.7	10.3	8.4
LU	6.2	47.9	4.4	24.0	5.6	11.9
LV	3.9	18.7	36.3	25.7	5.1	10.3
NL	**0.0**	33.0	31.9	18.0	0.0	17.0
NO	**0.0**	**0.0**	35.8	29.0	14.6	**20.6**
PL	17.3	41.2	0.7	35.7	0.5	4.6
PT	43.1	48.2	3.3	2.5	0.1	2.8
SE	**1.1**	**50.9**	22.5	9.2	2.7	13.6
SI	6.1	39.4	11.0	35.6	3.6	4.3
SK	**0.0**	**11.1**	29.4	51.1	0.0	8.4
UK	**54.4**	**0.0**	10.4	3.5	17.4	14.3

In Table 3 we look at the corresponding figures for the mother for the same countries and observe again show similar patterns as in the case of the education of the father. These countries are Austria, the Czech Republic, Germany, Denmark and Norway with less than 4% having less than secondary education level. In the UK almost 68% of mother have less than primary education level and 0% have a primary education level. As for the father's education variable, the Norway records the highest percentage having the highest education level (ISCED 5 & 6) at 25%.

Table 3: Education of the Mother across Countries

	Less than primary	Primary	Lower secondary	Upper secondary	Post secondary (non-tertiary)	First stage of tertiary education (ISCED 5 & 6)
AT	**0.0**	**3.4**	72.9	19.4	2.0	2.3
BE	18.5	37.8	17.0	15.2	2.0	9.4
CY	43.4	35.7	5.7	11.3	0.5	3.4
CZ	**0.0**	**1.5**	42.3	52.5	0.7	3.0
DE	**0.0**	**3.6**	37.4	46.8	2.1	10.2
DK	**0.0**	**0.1**	68.8	20.0	0.0	11.0
EE	2.6	23.7	28.0	27.1	5.8	12.8
ES	30.4	56.8	5.4	3.8	0.2	3.4
FI	6.5	26.7	40.9	16.0	0.5	9.3
FR	7.5	62.4	18.0	6.7	0.3	5.2
GR	38.6	45.1	6.7	4.7	2.0	2.9
HU	1.9	30.5	33.9	25.5	3.9	4.2
IE	2.4	66.4	13.0	9.8	2.7	5.7
IS	3.7	28.7	38.2	21.5	1.7	6.2
IT	23.5	54.0	13.3	8.0	0.0	1.3
LT	14.8	41.4	14.6	10.6	11.0	7.6
LU	8.5	61.4	7.1	15.1	0.7	7.2
LV	6.4	20.0	33.2	26.7	5.2	8.4
NL	**0.0**	39.3	41.8	12.1	0.0	6.8
NO	**0.0**	**0.0**	42.8	32.2	0.0	**25.0**
PL	19.3	46.5	0.7	29.3	1.1	3.0
PT	53.6	40.7	2.1	1.3	0.1	2.2
SE	1.9	53.5	23.0	8.6	3.3	9.7
SI	8.2	58.2	5.2	24.0	2.6	1.8
SK	**0.0**	**13.2**	42.6	40.2	0.0	4.0
UK	**67.6**	**0.0**	13.8	3.6	4.3	10.7

The final issue concerns the occupational variables for the father (PM070) and the mother (PM090) which allows us to construct a social class position. In the SILC module the occupational variable is a 2 digits ISCO-88 and we can derive a "rough" four category social class variable with the following classification:

ISCO-88	**Social class position**
11 to 34	Highly skilled non-manual
41 to 52	Lower skilled non-manual
61 to 83	Skilled manual
91 to 93	Elementary occupation

Table 4 presents the distribution of the social class of the father where we can see that four countries are quite distinctive in having a high percentage of fathers in the highly skilled non-manual class position these comprise the Netherlands (47%), Ireland (40%) and Norway and

Iceland both at 37%. For the UK we see also that 43% of fathers have a lower skilled non-manual class position while for the other countries it ranges only from 2 % to 16%, and 13% of the UK fathers are found into the skilled manual category versus an average of just above 50% for the other countries.

The corresponding table for the mothers produces similar results and patterns as for the fathers.

Table 4: Social Class of the Father across Countries

	Highly skilled non-manual	Lower skilled non-manual	Skilled manual	Elementary occupation	Not at work
AT	20	15	47	16	3
BE	28	16	43	13	0
CY	10	16	54	20	0
CZ	23	8	60	8	1
DE	33	11	50	6	1
DK	32	10	44	13	0
EE	22	2	59	15	1
ES	14	12	51	21	1
FI	26	5	63	3	2
FR	25	8	57	8	2
GR	16	9	66	9	1
HU	16	7	63	12	2
IE	**40**	11	29	20	0
IS	**37**	8	50	5	0
IT	17	8	51	12	10
LT	16	4	52	27	1
LU	**30**	8	56	4	1
LV	20	4	56	20	1
NL	**47**	11	37	5	0
NO	**37**	9	53	1	0
PL	12	5	73	9	0
PT	12	10	63	15	1
SE	25	11	59	2	2
SI	17	9	65	6	3
SK	22	7	51	19	0
UK	21	43	13	22	2

Conclusions

Overall is clear that the scale of missing valued for Sweden and the UK means that it is not possible to include these countries in any analysis focused on occupation/ class. Denmark. The Netherlands and the UK have particularly high missing values for mother’s occupation. The Czech Republic, Greece and Portugal have have 100% missing values.

Educational distributions vary across countries in ways that are very difficult to accept at face value. The discussion by Schneider and Müeller that follows explores issues relating to education in a good deal of further detail and confirms the view that results employing this variable must be interpreted with a great deal of caution.

The “class schema” employed here is of a very aggregated and rough nature, Although the variables available for the parents’ generation involve a level of detail that is well below the level for successful implementation of ESeC or EGP procedures, in principle a somewhat improved schema could be developed. However, given that class distributions vary across countries in a manner that is difficult to explain the question arises as to whether the investment of effort involved in following this route can be justified.

Overall our assessment would be that cross-national analysis involving the EU Intergenerational module must be interpreted with great care. This does not preclude the possibility of productive analysis involving particular countries and specific variables. However, even here one would ideally wish to be able to compare the results deriving from the EU-SILC combined data sets with those deriving from national data sets. Schneider and Müeller and Vallet in their discussions provide an illustration of what could be achieved

Measurement of Education in EU-SILC – Preliminary Evaluation of Measurement Quality[1].

Silke Schneider and Walter Müller

In EU-SILC education is measured along a simplified version of the ISCED 97-classification. The ISCED version used in the harmonized EU-SILC database distinguishes 6 levels of education. It neglects, however, any distinctions within levels of education according to the sub-dimensions programme orientation and programme destination, which are also available in the ISCED 97-classification (for details of the classification see UNESCO 2007 [1997], 1999; OECD 1999, Schneider/Kogan 2008). Roughly speaking, the former sub-dimension refers to the distinction between predominantly general/academic vs. predominantly vocational types of courses of study. This distinction has been shown to be quite crucial for a number of education related outcomes, such as for returns on the labour market, educational opportunities of children or various attitudinal or behavioural correlates of education. (Braun/Müller 1997; Schneider 2007,2008a,b; Müller/Klein 2008).

Neglecting the vocational/general distinction can lead to substantial heterogeneity within the aggregated levels of education in terms of the various correlates of education. Usually, therefore, the explanatory power of education (as measured e.g. by the variance explained) is considerably underestimated when vocational and general courses of study are not distinguished. Below, we will illustrate this phenomenon for one aspect of occupational outcomes.

In comparative research these issues lead to a further problems: The significance of the vocational / general distinction likely varies between countries, among others depending on

- the relative prevalence of either type of education,
- the differences in teaching contents, learning requirements (selectivity) and competence development in vocational in contrast to general courses of study;

[1] Paper prepared for the EQUALSOC EDUC/INCDIS Meeting on the Potential of the EU-SILC Interneretional Module. ESRI, Dublin, 26th-27th March 2009.

- the role the distinction plays in structuring labour market and other opportunities of individuals in a given country;
- the prestige and reputation attached to different courses of study;
- the costs of respective courses of study.

Depending on such social and institutional differences in a nation's educational setup, the explanatory power of educational attainment is likely to be attenuated in different ways and to a different extent in different countries. Thus, cross-national comparisons based on the simplified ISCED used in EU-SILC (and other cross-national surveys) are likely to be seriously misleading. The comparability of both bivariate associations, education effects in multivariate models and even effects of other variables (because educational attainment is insufficiently controlled for) are dubious.

In this paper we give a brief account of some first explorations of several aspects of the quality of education measurement in EU-SILC. At this stage, we essentially concentrate on the distributions of respondents along the ISCED categories distinguished. We do not enter into the issue whether national educational categories have been coded properly, that is in accordance with the coding rules established in the definition of ISCED, into the international classification, since the EU-SILC does not allow such checks. Examining the distributions, however, may already give hints to problematic coding. In an attempt to evaluate the stability of educational coding in EU-SILC, we compare the educational distributions over successive survey years. To check consistency of measurement in different databases, we also compare the educational distribution in EU-SILC with the corresponding distribution found in EU-LFS for identical survey years and corresponding population definitions. Finally, we will turn to the issue of explanatory power of education as measured in EU-SILC. We thus consider the following points:

1. The distribution of education along the ISCED 97 level categories
2. The stability of education measures over successive survey years in EU-SILC
3. Comparison of education measures in EU-SILC and EU-LFS

4. Explanatory power of education as measured in EU-SILC

In examining these issues with the SILC-, LFS- and ESS-data the paper finds a number of serious problems in the measurement available for education in the EU-SILC data.

As a consequence of using the simplified “only levels version” of the ISCED classification the resulting measurement of education comes out as a very rough classification with large proportions of the population coded into large highly heterogeneous categories. This is particularly true for ISCED level 3, but with increasing significance among the younger cohorts also for ISCED level 5. But even in countries, in which educational attainment is more evenly distributed among the various levels distinguished by ISCED, the lack of differentiation within heterogeneous categories can damage measurement quality.

In a number of countries we find high measurement instability from wave to wave drawing into doubt the interwave reliability of the measures. Even larger discrepancies in education distributions result from the comparison of education measures in EU-SILC and the education measures in EU-LFS in spite of the fact that both databases are produced by the same NSIs. In some of the countries, dissimilarities between the two sources are rather large.

Also in terms of criterion validity EU-SILC needs improvement. In a few countries, the correlations between ISCED and ISEI vary considerably between waves. Partly, there are also large differences in the size of the criterion correlation between EU-SILC and ESS, when in both data the same ISCED classification is used. Evidently, the latter differences can also be due to particular features of ESS. However, the correlations with the country specific education variables (CSEV) in ESS unmistakably support the view that EU-SILC tends to underestimate the role of educational attainment for occupational attainment. This seems to be the case for all countries, yet to a different extent. Needless to say, the varying loss of explanatory power for education when collapsing national educational categories to the ISCED level classification can partly also derive from inadequate differentiation in the CSEV available for the different countries.

These results are preliminary in various respects. But they are useful to direct the search for further clarification and the design of additional analyses. In terms of further analyses, several strategies could be pursued. The examination of criterion validity could be extended to include other outcomes of education such as unemployment risks, labor force participation, class position or income. Analyses done here with ESS and SILC should be compared with similar analyses done by the EQUALSOC EDUC team validating educational measurement in the EU-LFS. As at least for some of the countries, the LFS provides more detailed measures, the relevance of particular distinctions for the explanatory power of education could be explored.

As to internationally comparative studies concerning substantive issues related to education, the results found here do not suggest promoting at this stage EU-SILC as a promising data base. One of the first crucial steps to improve the situation would be to release to the research community better data documentation and the original national measures from which the internationally "harmonized" variables have been derived. From this one might be able to construct more adequate measures for European comparative research than so far available.

EU SILC Intergenerational module Data quality issues and recommendations

Tim van Rie and Karel van Den Bosch

The reference period

The reference period for the intergenerational module is defined as 'when the interviewee was a young teenager, between the ages of 12 and 16. If the respondent hesitates or asks for a specific age, the age of 14 should be used'. [2]

This 'three-fold' definition of the reference period has produced differences in implementation across Member States, as can be seen most clearly in the questionnaires for variable PM100[3].

While a large number of MS refer to the period of age 12 to 16 (CY, CZ, EE, FI, HU, IE, IS, IT, LU, LV, NO, SE, UK), some countries refer to the notion of young teenager (DK, EE, ES, FR, LT, UK). In two Member States (AT, BE) the question refers to age 14[4].

These differences in reference period are not trivial. Whereas some variables will remain relatively stable over the four year period (e.g. educational attainment of the parents), others are more likely to change over a four-year reference period (e.g. family composition, activity status and occupation of the parents). These differences in reference periods hamper cross-country comparability. For instance, 'unemployment' as main activity status in a one-year reference is not comparable to unemployment as main activity status for a four-year reference period.

In my view, if the module was to be replicated, it would be advisable to include a more clear definition of the reference period, preferably one year (in keeping with the annual reference period of SILC, and in view of difficulties of linking module results to the longitudinal dataset).

The sample

First, it would appear that a number of datasets are circulating, which is an issue in itself. Whereas most of my results for missing values are consistent with those that Bertrand Maître presented at the

[2] Official Journal of the European Union, Commission regulation (EC) No16/2004 of 6 January 2004
[3] http://circa.europa.eu/Public/irc/dsis/eusilc/library
[4] In DE, EL and PL no info on PM100 is provided. For NL and PL, it is not possible to deduct which version was used.

Dublin workshop, I find a much higher share of missing values for IE. Moreover, in my data, the missing values for IE are overrepresented among males.

Second, as some several participants at the workshop indicated, the number of cases for the ‘register countries’ is limited, since the data for the intergenerational module were only collected from the sample. Whereas in some Members States (e.g. Finland) this still results in a fairly large sample, for others (e.g. Sweden) the combination of ‘respondents not selected’ (coded -5) and ‘missing values’ (coded -1) leads to a relatively small sample (under 4000 cases). This should certainly be taken into account when disaggregating the data. In the joint Excel files, I provide the breakdown in more detail. Below I add a table showing the number of filled values per Member State and per variable of the Intergenerational module.

Education variables

For many purposes the ISCED categories used in the SILC are to rough, and too aggregated. Example: for the analysis of the transition into higher education (an important issue in intergenerational research, and in inequality of life-chances generallly) it is essential to know who is eligible for this transition, i.e. has the required diploma to enter higher education. In many countries only some forms of higher secondary eduction give access to higher education, while others do not. In the SILC education variables all these forms are lumped together in category ISCED-3. A disaggregation of ISCED levels is required.

Tim Van Rie
Karel Van den Bosch
Centre for Social Policy, Antwerp
tim.vanrie@ua.ac.be

A Note on constructing the European Socio-Economic Classification on EU-SILC Data

Dorothy Watson

Overview of ESeC

Socio-Economic Classifications are intended to capture the root causes of enduring inequalities. They draw on the Weberian tradition of class analysis (in their emphasis on inequalities in life chances) and are typically rooted in market position, especially the labour and capital markets. The European Socio-Economic Classification (ESeC) is a classification system developed and validated for use in comparative international research in Europe[5]. In 2004, a consortium of two National Statistical Institutes (ONS and INSEE) and researchers from six countries (Germany, Ireland, Italy, the Netherlands, Sweden and the UK) received EU Sixth Framework funding to develop and validate this scale. It is unique in that it has been developed using the international classification of occupations typically available in international datasets (ISCO88) and explicitly validated for use across a range of European countries using a variety of international datasets, including the European Social Survey (ESS) and the European Socio-Economic Household Panel Survey (ECHP).

The basic premise of ESeC, drawing on the work of Goldthorpe (2007), is that in market economies, market position (especially position in the occupational division of labour) is fundamental to social inequality. Classes are sets of structural positions defined by sets of social relationships with different levels of power in market. The key positions are employer, self-employed, employee and excluded groups and, among employees, positions differentiated by the ease of difficulty in monitoring the work and 'asset specificity' (the importance to the employer of the employee's technical and organisational knowledge). Where it is difficult to monitor the work

[5] See the ESeC website for details on the work of the international consortium who developed the scale, and the ESeC User Guide: http://www.iser.essex.ac.uk/research/esec.

and asset specificity is high, a service contract will be typical; labour contracts apply where labour is more easily replaceable in these terms.

At the lowest level (level 1), ten classes are distinguished. These are as follows:

1. Large employers, higher managerial & professionals (including senior managers and higher professionals such as doctors, lawyers, university lecturers);
2. Lower managerial & professional occupations (including junior managers and professions such as primary and secondary teachers, nurses and social workers);
3. Intermediate occupations (including occupations such as book-keepers, secretaries, clerical workers);
4. Small employers & own account workers (small employers and self-employed occupations outside of agriculture);
5. Employers & self-employed in agriculture (farmers);
6. Lower supervisory & lower technician occupations (including supervisors of those in classes 7 to 9, foremen, precision craft workers in metal);
7. Lower services & sales occupations (including occupations such as shop salesperson, cashier, teller);
8. Lower technical occupations (such as carpenters, mechanics, plumbers);
9. Routine occupations (including machine operator, driver, caretaker, cleaner);
10. Never worked & long term unemployed (those excluded from the labour market for whatever reason);

The classes are not intended to form a hierarchy, as the life chances associated with them may differ across political regimes and economies. However, among employees, on many measures of life chances there is a hierarchy running from class 1, 2, 3, 6, 7, 8 to 9.

Constructing ESeC on SILC data

The construction of ESeC ideally requires information on economic status (employer, self-employed, employee), occupation coded to three digits on ISCO88, supervisory status, and industrial sector (to distinguish farmers and other self-employed or small employers). However, in the absence of all the information required, there are suggested compromises that can be made in constructing ESeC.

The EU-SILC data from 2005 has a module on intergenerational mobility, and many researchers will be interested in constructing a measure of social class for both the respondent and the family of origin using this dataset. The table below shows the variables needed to construct the 'full' version of ESeC and the nearest equivalents available on the SILC dataset for the respondent and for the respondent's parents. A number of compromises will be needed in constructing ESeC on SILC data and the consequences of these are outlined below.

Figure 1: Variables required to construct ESeC classes

Ideal variable	*Ideal Codes*	*SILC 2005 -Respondent*	*SILC 2005 -Parent*
Employment Status	Employer Self-Employed Employee	PL050 Self-Employed Employee (Employer = self-employed and establishment size >1)	PM060/PM080 Self-employed Employee
Occupation	ISCO88 3-digit	PL050 ISCO88 – 2 digit	PM070/PM090 ISCO88 – 2 digit
Supervisory Status	Supervisor Not a supervisor	PL150 (supervisor/non-supervisor)	Not available
Size of business	Fewer than 10 More than 10	PL150 (<10, >10)	Not available
Sector	Agriculture/other	PL110 (agriculture/other)	Not available

Use of two-digit ISCO code for occupation

Table 1 shows the consequences of using a two-digit ISCO coding of occupations rather then the recommended three-digit version for constructing ESeC classes. As SILC has only the two-digit version, this compromise is necessary with this dataset. The table is based on an analysis of the ESS data and comes from the ESeC User Guide.

There are two shifts of over one fifth of the cases (in terms of frequency of the classes in the European population). About 28 per cent of ESeC Class 2 cases (lower managerial and professional occupations) are lost to Class 1 (Higher managerial and professional occupations). This occurs because the coding is not detailed enough to identify lower-level managers (production and operation managers) and certain professions (nursing and midwifery) which would otherwise be allocated to class 2. A second large shift involves the loss of 22 per cent of cases (including materials recording and transport clerks) from Class 7 (lower service and sales) to Class 3. .

Table 1: Correspondence between ESeC classes constructed from 2-digit and 3-digit ISCO Codes (row percentages)

		ESeC from 2-Digit ISCO									
		1	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***Total***
ESeC from 3-Digit ISCO	***1***	***91***	***9***								***100***
	2	***28***	***72***								***100***
	3			***100***							***100***
	4		***1***		***99***					***0***	***100***
	5					***100***					***100***
	6		***14***				***86***				***100***
	7			***22***				***77***		***1***	***100***
	8								***94***	***3***	***100***
	9			***3***				***12***		***85***	***100***

Source: Harrison and Rose, 2006, ESeC User Guide, 2006, Table 3.

The net effect of using a two-digit ISCO code is to increase the sizes of social classes 1 and 3, and reducing the sizes of classes 2 and 7. However, the ESeC based on two-digit ISCO code was used in analyses based on both ECHP (Watson et al 2009) and SILC data (Watson et al 2007), and worked very well.

Absence of indicator of supervisory status

Although an indicator of supervisory status is available for the respondent on EU-SILC, it is missing for the occupation of the mother and father. The main impact of this is on ESeC class 6 (lower supervisory and lower technical occupations). This class is mostly comprised of supervisors of manual occupations and when supervisors cannot be identified, the occupants of this class are assigned, based on occupation, to classes 7, 8 and 9 (nearly one third to each). As a result, ESeC class 6 is reduced to a very small number of precision craft workers. A second large impact is on class 2, which loses about 17 per cent of its occupants to class 3. Most of these are supervisors of clerical workers and technicians.

In constructing parent's social class, we would suggest combining the occupations in social class 6 with those in social class 7 (lower service and sales occupations). This combination was used in an earlier analysis of SILC data (Watson et al 2007), and an analysis of persistent poverty using the ECHP data (Watson et al 2009, forthcoming) suggested that these classes face similar levels of persistent poverty.

Absence of indicator of size of establishment and sector

No information is available on EU-SILC for size of establishment or economic sector for occupation of parents. However, the loss of this information is less consequential in terms of the proportion of occupations affected.

Size of establishment is used to distinguish small employers (employing 1-9), from large employers (employing 10 or more) and the self-employed (with no employees)[6]. The suggested compromise in ESeC (Harrison and Rose 2006) is to classify the self-employed into class 4 or 5. The number of cases mis-assigned using this procedure is less than one percent of all occupations.

Sector is used to distinguish social classes 4 (small employers and self-employed other than agriculture) and 5 (small employers and self-employed in agriculture). It should be possible to distinguish these groups based on occupation, however, if farmers are correctly coded into ISCO 61[7].

Conclusion

[6] Note that self-employed professionals should be classified into social class 1 or 2, as appropriate. In these cases, the professional skill of the occupant is considered to be of more consequence for life chances than the ownership of the business.

[7] If farmers are coded as ISCO 13 (managers of small enterprises), this will not be possible. Whether this is the practice in the country could be checked in the respondent data, where sector is available.

The European Socio-Economic Classification is a theoretically grounded social class measure that has been specifically developed for use in international comparative research. It has been validated for use across a wide range of European countries with respect to a range of dimensions of life chances, including poverty and deprivation, unemployment, health and education (Rose and Harrison, 2009, forthcoming). It remains relatively robust even when all of the data requirements are not met and compromises need to be made in constructing it. As such, it holds promise as the basis from which to begin exploring hypotheses around intergenerational transmission of inequality in Europe.

References:

Goldthorpe, J. H. 2007 'Social Class and Differentiation of Employment Contracts' in *On Sociology Second Edition, Volume Second Edition: Volume Two Illustration and Retrospect*, Stanford, Stanford University Press.

Harrison, E. and Rose, D. (2006) The European Socio-economic Classification (ESeC) User Guide (Revised Version). Institute for Social and Economic Research, University of Essex, Colchester, UK.
Available at http://www.iser.essex.ac.uk/research/esec.

Watson, D., Whelan C. and Maître, B (2007) "Social Class Variation in Income Poverty, Deprivation and Consistent Poverty: An Analysis of EU-SILC" in Comparative EU Statistics on Income and Living Conditions: Issues and Challenges - Proceedings of the EU-SILC Conference (Helsinki 6-8 November 2006) .
Available at:
http://epp.eurostat.ec.europa.eu/portal/page?_pageid=1073,46587259&_dad=portal&_schema=PORTAL&p_product_code=KS-RA-07-007

Watson, D., Whelan C. and Maître, B (2009, forthcoming) "Class and Poverty: Cross-sectional and dynamic analysis of income poverty and life-style deprivation" in Rose, D. and Harrison, E (Eds.) Social Class in Europe: An Introduction to the European Socio-Economic Classification. London: Routledge.

Whelan, C. and Maître, B. (2007) "Social Class Variation in Risk: A Comparative Analysis of the Dynamics of Economic Vulnerability ".Dublin: ESRI Working Paper number 202, June 2007.

Whelan, C. and Maître, B. (2008) "'New' and 'Old' Social Risks: Life Cycle and Social Class Perspectives on Social Exclusion in Ireland", Dublin: ESRI Workign Paper number 226, April 2008.

Educational and Occupational Coding in the French EU-SILC 2005 Intergenerational Module

Louis-André Vallet
CNRS & CREST, Paris

According to Commission Regulation (EC) 16/2004 of 6 January 2004 that includes the list of target secondary variables in the 2005 'Intergenerational transmission of poverty' module, the parental education variables (highest level of education attained by father (PM040) and by mother (PM050)) have to be coded using the 1-digit ISCED-97 classification, and the parental occupation variables (main occupation of father (PM070), main occupation of mother (PM080)) have to be coded using the ISCO-88(COM) classification with 2 digits. Having access to the French file of the 2005 Intergenerational Module, I therefore tried to examine the quality of the coding for these variables and how they relate to the same variables using 'national' classifications that were created after the administration of the questionnaire in France.

Parental Education Variables (father and mother)

The French version of the questionnaire asks for the highest degree possessed (NIVISCPERE/ NIVISCMERE with a classification in 10 categories), and, in case of 'no degree', 'don't know' or 'refusal', the highest stage of the educational system that was attained (SCOLPERE/SCOLMERE with a classification in 7 categories). As a consequence, if needed, it is certainly possible to do better than using 1-digit ISCED-97. In particular, as the French variables distinguish between general and vocational education, it would be straightforward to implement the CASMIN educational scheme (with 9 categories). N of valid cases is as follows: 11,008 (PM040), 11,709 (PM050), 9,853 (Niviscpere), 10,561 (Niviscmere).

Cross-classifications for Parental Education

I examined the four two-way classifications between a ‘French’ variable and the EU variable: Niviscpere*PM040 and Scolpere*PM040, Niviscmere*PM050 and Scolmere*PM050. Tables are approximately diagonal, so coding in ISCED-97 seems to be generally good. However, looking at the two three-way classifications (Niviscpere*Scolpere*PM040 and Niviscmere*Scolmere*PM050), there are 327 cases related to father and 518 cases related to mother for which the ISCED-97 variable is filled, but none of the ‘French’ variables, what is surprising (is it imputation? if so, on which basis?); and there are 57 cases related to father and 69 cases related to mother for which ‘studies in vocational secondary education that do not lead to *baccalauréat*’ is coded as ‘3’ in the ISCED-97 variable instead of ‘2’ (in my opinion). So, even in ISCED-97, some improvement is probably possible in the data.

Parental Occupation Variables (father and mother)

The French version of the questionnaire asks for the occupation with an open question (PROFESSPERE/ PROFESSMERE), the employment status in 4 categories (STATUTPERE/STATUTMERE), and the job qualification in 10 categories (CLASSIFPERE/CLASSIFMERE). The detailed-occupation variables are not included in the French file, but on the basis of occupation, employment status and job qualification, 2-digit *Catégorie socioprofessionnelle* variables are created (CS_PERE/CS_MERE) with 31 categories. As a consequence, if needed, it is possible to do something else than using 2-digit ISCO-88(COM). For instance, implementing the CASMIN class scheme with 10 categories is possible (classes IVa and IVb cannot be distinguished because ‘number of employees’ is unknown). N of valid cases is as follows: 11,329 (CS_PERE), 8,479 (CS_MERE), 11,324 (PM070), 8,462 (PM090).

How Are the ISCO-88(COM) Variables Created?

In French public statistics, coding in ISCO-88(COM) is mainly implemented in the Labour Force Survey (*Enquête Emploi*) on the basis of the 4-digit PCS (*Profession et Catégorie Socioprofessionnelle*) variable, the detailed classification of economic activity, and the number of employees in the economic establishment. My opinion is that the mapping is not extremely well maintained by the French Statistical Office. Three years ago, I examined it with a colleague, we proposed a number of detailed modifications and improvements but, as far as I know, they have not been fully examined as yet.

According to the 2005 EU-SILC INSEE Documentation

As regards the respondents, use of the mapping resulted in 2-digit ISCO-88(COM) occupational coding in 95% of the cases. Imputation was used in the remaining cases. The ISCO code was randomly assigned according to the distribution of the 2-digit ISCO codes per 2-digit PCS in the Labour Force Survey (second quarter 2005). As regards the parents whose occupation may be less well known or less precisely described by the respondents, coding in the French classification was generally less accurate, resulting more often in a 2-digit PCS code. As a consequence, transcoding in 2-digit ISCO-88(COM) was successful in 85% of the cases. Imputation was used in the remaining cases. The ISCO code was randomly assigned according to the distribution of the 2-digit ISCO codes per 2-digit PCS in the 2003 *Formation & Qualification Professionnelle* Survey.

Cross-classifications for Parental Occupation

For both father and mother, I finally examined the two-way classification between 2-digit PCS and 2-digit ISCO-88(COM) in detail: CS_PERE*PM070 and CS_MERE*PM090. I did not find any strong or obvious anomaly in the correspondence. However, that does not imply that coding in ISCO is optimal (see above). In particular, it is unclear whether the situation will be better or worse if father’s and mother’s occupation in 2-digit ISCO-88(COM) is combined with their employment status and job qualification to create an ESeC variable.

CONCLUSION

Especially in case of output harmonisation, this brief report illustrates to what extent it may be useful, in order to strengthen comparative research, to have access, in the same dataset, to both ‘national’ variables and EU-standardised variables.

Short notes on data quality and cross-national comparability in income measures using EU-SILC cross-sectional and panel data

Paolo Barbieri and Sabrina Carrossa
University of Trento

Through its unique system of measurement, income inequality metrics determine the dispersion of income among participants within a particular economy. Specifically, income has been most commonly used to examine inequality among either individuals or households. In the life course approach, (work-)income represents a source as well an outcome of intergenerational transmission of inequality.

These notes focus on fixed term versus permanent contract wage penalty, as they derive from another Equalsoc research project, namely the "Atypical employment and welfare regimes". Wage gaps between flexible forms of employment and standard work contracts, in fact, when they stratify in stable patterns of wage differentiation become true part of a wider phenomenon of LM inequality. The definition of income can be clearly defined through the use of income metrics and because of this fact, we have focused our attention on data quality and cross-sectional comparability in income measures.

The first part of the paper provides a formal analysis of few general problems concerning Eu-Silc data, specifically in regards to income information. The second section compares the ECHP and Eu-SILC data through wage penalty analysis. Both empirical analyses are based on a specific group of countries: Germany, Italy, Netherlands, Ireland and United Kingdom. Lastly, the conclusion discusses and summarizes the most important problems within these analyses and provides a list of recommendations for the next Eurostat Conference.

1. General problems with Eu-Silc data

The general problems with Eu-Silc data derive from three central issues:

(a) *differences of information across countries*,

(b) *differences among respondents* and *sampling methods,*

(c) *differences in the variables between cross-sectional and longitudinal databases*.

Respectively, these problems have three consequences: comparability among countries (how can we compare the income information among countries since they recorded this information in different way?), estimates accountability and reliability (how can we be confident about the estimates, especially for the so-called "register-countries"), and limitation in the object of the analysis (how can we analyse intergenerational transmission of inequality using the longitudinal database since there are no information about parents?).

1.1 Problem (a): different information across countries

The Eu-Silc survey collected different information on income, both at the individual and household level. This paper is specifically focused on the individual income information. Furthermore, the reference period is the "current reference period" (that is time t) or the "income reference period" (time t-1). The different types of income collected in the survey at the individual level are displayed as synthesis in Table 1 in comparison to the ECHP survey. Unfortunately, in the Eu-Silc data each of these variables have a problem.

In regards to the yearly income variables, they are not standardized and directly comparable across countries. Each country measures and collects data on yearly incomes in its own way: some countries collect *gross* income, others collect income *net of income taxes and social contributions* while other countries collect income *net of social contributions*. In order to solve this problem, we need to define a set of rules to convert gross to net incomes or vice versa, as in the EUROMOD model. This suggested systemise a micro-simulation tax-benefit model, that has been broadly tested with ECHP but not yet with EU-SILC data. Even with the use of this micro-simulation, however, the EU-SILC lacks relevant information to simulate the income conversion (ex: number and age of children is missing).

In Appendix A there are two cross-tabulations of the yearly gross/net income flag variable by countries, underlining the heterogeneous nature of the income information among countries.

Table 1: Comparison on income information at individual level between Eu-Silc and ECHP survey.

<table>
<tr><th>Reference period</th><th>Eu-Silc survey</th><th>ECHP survey</th></tr>
<tr><td>t-1</td><td>from variable y010g/n to variable py140g/n kinds of (gross and net) yearly incomes from different sources (job, social benefits…)</td><td>from variable pi001 to variable pi138a: kinds of (net) yearly incomes from different sources (job, social benefits…)</td></tr>
<tr><td rowspan="2">t</td><td>variable py200g: gross monthly earnings for dependent employees</td><td>variable pi211mg: gross monthly earnings for dependent employees</td></tr>
<tr><td></td><td>variable pi211m: net monthly earnings for dependent employees</td></tr>
</table>

Similar to the yearly information, the monthly income information also is flawed in the Eu-Silc database. Its central problem is that it has a lot of country-missing values, most specifically the “-2” values, that in the codebook is coded as: “*na (PL040 not = 3 or MS has other source to calculate gender pay gap or for analyse it)*”.

Furthermore, using the yearly income is a somewhat complicated. Indeed, the yearly income information is usable solely with the longitudinal database, especially in the case of wage penalty analysis. Due to the panel structure of the data, the yearly income information is about the income reference period (time t-1) while the information on occupation are about the current reference period (time t). In Appendix B there is a cross-tabulation of the monthly gross income flag variable by countries to underline the heterogeneity in the income information among countries.

1.2 Problem (b): different respondents and sampling methods

Another problem that arises is due to the sampling differences. In most of the countries (so-called “survey” countries), all the information is collected by interviewing all the household members.

Other countries (mostly the Nordic Countries), have population registers with income information. In these countries, there is a “selected respondent” in each household, and he/she answers to most of the income and non income questions. As a representative of all of the household members, the register provides information for income, education and housing. This method explains why in Denmark, Finland, Iceland, Norway, Slovenia, Sweden and The Netherlands there are a lot of cases with missing value (usually the value -3) for a number of variables, especially concerning occupation. This (we checked for the NDL) causes serious biases: the Dutch missing cases, in fact, were quite numerous (about **47%** among the selected cases) and well predictable by sex (female) and (young) age…

1.3 Problem (c): different variables between cross-sectional and longitudinal databases

A third problem is that some variables are present solely in the cross-sectional database and others only in the longitudinal one (i.e. the secondary variables – as well as additional occupational information - are lacking in the longitudinal database). This is a problem because it is not possible to improve some kind of analysis. Specifically in the case of income analyses, since there is no monthly income information disposable for the majority of countries, one has to use the longitudinal database and the yearly income variable, which is then limited in terms of control variables. In the Appendix C there is a list of variables lacking in the longitudinal database.

2. An empirical comparison between ECHP and EU-SILC

This second section is concerned with the specific problems arising with the analyses of the fixed term and permanent contract wage penalty. The focus of this empirical analysis is the comparison between ECHP and Eu-Silc data. Comparing the two models helps to determine the Eu-Silc efficiency and if it can be a reliable source for income comparisons (despite of the problems previously discussed).

Our empirical example refers to the "fixed term contract wage penalty". In these kind of analyses, the dependent variable is the logarithm of the hourly gross wage, usually computed via the monthly gross wage. But as in Eu-Silc this information is disposable just for a limited set of countries, we used the yearly gross income that is disposable for more - but not for all – countries (Problem a). This fact has a consequence in terms of countries selection: i.e. we need to give up to analyze FR and SWE, as these countries collect net income (and report it in the gross income variable, even if "flagged" ...).

Once made the choice to run the analyses on yearly gross income, one is obliged to use the longitudinal db because the information on income and occupation have to refer to the same time period. Furthermore, using the longitudinal db one meets additional problems such as a lack of variables (collected and recorded just in the cross-sectional db) and "*the non selected respondent*" issue (which by the way represents a problem for the longitudinal db as well for the cross-sectional one). Given these problems, the next paragraphs describes and justifies the case selection we adopted to analyze wage penalty.

2.1 What about the selection?

We analysed the wage penalty in Germany, Italy, The Netherlands, Ireland and United of Kingdom. Within this process we selected fixed term contract and permanent employees older than 25 years (because we can't control for apprenticeship). Unfortunately the information on the apprenticeship status is disposable just in the cross-sectional db (variable pe010, "current education activity"). The info on apprenticeship is quite relevant for the German case, as we know. But it is not just a problem of having a proper definition of our individuals "at risk", as it turns to be a substantial issue: in fact, inserting the under25 in the analyses, would produce a wage gap of about 40% in Germany – that is it would almost double the wage penalty for FTC in Germany…

In order to use the "yearly gross income" as "gross wage" we must be certain that this income information (at t-1) concerns one specific occupation (at t0), therefore one needs to check if there were job changes between the two waves. If there are no job changes between the two waves of the longitudinal database one can assume that the yearly gross income collected in the 2nd wave can be used as gross wage for the occupation collected in the 3rd wave.

Specifically, we used the yearly gross income and the longitudinal database for Germany, The Netherlands, Ireland and United Kingdom. For Italy with Eu-Silc database and ECHP, we used the monthly gross wage and the cross-sectional database.

2.2 Empirical results

Despite of the problems presented above, the comparison between the ECHP and Eu-Silc produces quite stable estimates and r-squared that gives us confidence in respect to the reliability of Eu-Silc data as a source of income comparison.

Overall, the OLS as well as the quintile regression estimates obtained with ECHP data are quite close to the Eu-Silc results. **With the only (relevant) exception of the Netherlands**, all coefficients displayed in Table 2 are quite similar between surveys and within each country. Interestingly enough, Germany highlights the higher penalty, also controlling for age. Noteworthy is the effect of the part-time contract: the dependent variable is the logarithm of hourly gross wage, so the dependent variable already controls for the number of hours worked. Nonetheless, the effect of the type of contract itself remains significant, therefore showing a specific contract effect. Other determinants that explain the wage penalty are: gender (female), age (young), lower education and having a manual occupations (Isco).

The quintile regressions show – between surveys and within each country - quite stable estimates for each quintile (Table 3), especially for the German and the Italian cases. Also for Ireland and United Kingdom, we find non statistically significant estimates in conformity with ECHP estimates and with our expectations about the existence of a low or null fixed term contract wage penalty in such flexible Lms. Again, **the Netherlands stands as a quite problematic case** because the results between the two databases are differ substantially. Of course, it could be interesting to assess how much of this variation between ECHP and EUSILC estimates for NL is due to a data problem (specified as problem b) and how much is due to a real change in the Dutch labour market situation...

Unfortunately, we have to search for better data for the NL to try to answer to this point…

Table 2. Fixed term contract wage penalty, OLS Regression. ECHP 2001 vs Eu-SILC 2005-2006, Yvar: Ln Hourly Gross Wage

	ECHP	Eu-Silc	ECHP	Eu-Silc	ECHP	Eu-Silc	ECHP	Eu-Silc	ECHP	Eu-Silc
	DE	*DE*	IE	*IE*	NL	*NL*	UK	*UK*	IT	*IT*
lnhwage	*beta*	*beta*	*beta*	*beta*	*beta*	*beta*	*beta*	*beta*	*beta*	*beta*
ref. c. permanent	-	-	-	-	-		-	-		-
temp-job	-0,22 *	-0,21 *	0,02	-0,07	-0,30 *	-0,18 *	-0,07	0,10	-0,07 *	-0,11 *
ref.c. male	-	-	-	-	-	-	-	-		-
female	-0,24 *	-0,21 *	-0,20 *	-0,14 *	-0,18 *	-0,16 *	-0,18 *	-0,21 *	-0,12 *	-0,13 *
age	0,04 *	0,06 *	0,04 *	0,05 *	0,04 *	0,05 *	0,05 *	0,04 *	0,04 *	0,04 *
age2	0,00 *	0,00 *	0,00 *	0,00 *	0,00 *	0,00 *	0,00 *	0,00 *	0,00 *	0,00 *
ref. c. usecondary	-	-	-	-	-	-	-	-		-
lsecondary	0,05 *	-0,09 *	-0,18 *	-0,09 *	0,05	-0,11 *	-0,05 *	-0,11 *	-0,11 *	-0,12 *
tertiary	0,04 *	0,09 *	0,23 *	0,19 *	0,07	0,17 *	0,12 *	0,13 *	0,15 *	0,10 *
ref. c. fulltime	-	-	-	-	-	-	-	-		-
parttime	0,00	-0,02	-0,09 *	-0,11 *	0,05 *	-0,07 *	-	-0,04 *	0,12 *	-0,03 *
ref. c. Isco5	-	-	-	-	-	-	-	-		-
isco1	0,49 *	0,48 *	0,31 *	0,45 *	0,39 *	0,36 *	0,50 *	0,48 *	0,56 *	0,33 *
isco2	0,52 *	0,48 *	0,58 *	0,60 *	0,42 *	0,36 *	0,51 *	0,51 *	0,38 *	0,35 *
isco3	0,34 *	0,32 *	0,26 *	0,31 *	0,25 *	0,24 *	0,44 *	0,35 *	0,17 *	0,15 *
isco4	0,27 *	0,25 *	0,16 *	0,12 *	0,15 *	0,12 *	0,19 *	0,21 *	0,09 *	0,04 *
isco6	-0,17 *	-0,12	-0,12	-0,32	0,11	-0,11 *	-0,06	-0,18	-0,21 *	
isco7	0,10 *	0,09 *	0,17 *	0,19 *	0,02	-0,01	0,19 *	0,13 *	-0,04 *	-0,07 *
isco8	0,05 *	0,00	0,09 *	0,03	-0,05	0,07 *	0,05	0,01	0,03	-0,02
isco9	-0,05	-0,10 *	-0,01	-0,07	-0,12 *	-0,05	-0,01	-0,06	-0,08 *	-0,14 *
_cons	2,33 *	1,19 *	1,33 *	1,62 *	2,33 *	1,58 *	0,86 *	1,80 *	1,90 *	1,44 *
adj r-sqr	*0,31*	*0,31*	*0,47*	*0,40*	*0,25*	*0,45*	*0,33*	*0,34*	*0,44*	*0,32*

Table 3. Fixed term contract wage penalty, Quintile regression. ECHP '01 vs Eu-Silc '05-'06 , Yvar: Ln Hourly Gross Wage

	q.20		q.40		q.60		q.80	
DE echp	-0,26	*	-0,25	*	-0,17	*	-0,15	*
DE eu-silc	-0,32	*	-0,21	*	-0,15	*	-0,10	*
NL echp	-0,31	*	-0,29	*	-0,22	*	-0,11	
NL eu-silc	-0,26	*	-0,13	*	-0,13	*	-0,11	*
IT echp	-0,11	*	-0,09	*	-0,05	*	-0,05	
IT eu-silc	-0,13	*	-0,13	*	-0,11	*	-0,09	*
IE echp	-0,02		-0,03		-0,01		0,01	
IE eu-silc	-0,14	*	-0,11		-0,13	*	-0,07	
UK echp	-0,08		-0,10	*	-0,14	*	-0,05	
UK eusilc	-0,06		-0,02		0,10		0,18	*

3. Conclusion

This notes deal with data quality and cross-national comparability in income measures. Our analysis shows that the Eu-Silc encounters problems with income information and other variables that damage/reduce comparability among countries, as well the type of the analysis and, in some cases, the estimates reliability.

In the first part of these notes we detailed the more general problems with Eu-Silc database in comparison with ECHP, while in the second part we underlined the specific problems arising with the analyses of the FTC-Permanent contract wage penalty. In general, the regression results on the wage penalty are quite stable between Eu-Silc and ECHP – with some exceptions (NL).

To conclude, paying the price of a reduction both in the number of countries and in the variables we can insert in our models, one can be relatively confident in Eu-Silc as a source for income comparisons. Nonetheless, some problems are still present.

In specific, some "adjustments" would definitely better the quality of the dataset:

- each yearly income variable should be made strictly comparable among countries (problem a)
- all the (time constant) cross-national variables should be reported in the longitudinal database;

A. Appendix – Cross-tabulation of gross/net yearly income variable by countries

* *(dep. employees in full or part-time job)*
keep if ((pl030==1)|(pl030==2))&(pl040==3)
* *country flag variable on gross yearly income*
tab pb020 py010g_f

country	py010g_f -5	0	1	3	4	5	Total
AT	0	91	0	0	4,799	0	4,890
BE	0	189	464	0	3,518	0	4,171
CY	0	125	0	0	4,016	0	4,141
CZ	0	49	1,804	0	1,998	0	3,851
DE	0	162	55	0	10,218	0	10,435
DK	0	169	0	0	6,606	0	6,775
EE	0	166	3,888	0	408	95	4,557
ES	11,683	0	0	0	0	0	11,683
FI	0	73	0	0	9,557	0	9,630
FR	0	223	0	8,532	0	0	8,755
GR	3,471	0	0	0	0	0	3,471
HU	0	402	0	0	5,604	0	6,006
IE	0	12	181	0	4,552	0	4,745
IS	0	75	0	0	4,103	0	4,178
IT	15,127	0	0	0	0	0	15,127
LT	0	144	3,191	0	907	0	4,242
LU	0	57	0	0	3,569	0	3,626
LV	3,564	0	0	0	0	0	3,564
NL	0	146	0	0	8,635	0	8,781
NO	0	92	0	0	6,640	0	6,732
PL	0	480	11,384	0	0	0	11,864
PT	4,295	0	0	0	0	0	4,295
SE	0	223	0	6,062	0	0	6,285
SI	0	89	0	0	9,605	0	9,694
SK	0	234	0	0	5,919	0	6,153
UK	0	209	0	0	7,842	0	8,051
Total	38,140	3,410	20,967	14,594	98,496	95	175,702

B. Appendix – Cross-tabulation of gross monthly income variable by countries

** dep. employees in full or part-time job*
keep if ((pl030==1)|(pl030==2))&(pl040==3)
** country flag variable on gross monthly income*
tab pb020 py200g_f

	py200g_f							
country	-2	-1	0	1	3	4	5	Total
AT	0	0	0	0	0	4,890	0	4,890
BE	0	308	0	0	0	3,863	0	4,171
CY	4,141	0	0	0	0	0	0	4,141
CZ	3,851	0	0	0	0	0	0	3,851
DE	10,435	0	0	0	0	0	0	10,435
DK	6,775	0	0	0	0	0	0	6,775
EE	4,557	0	0	0	0	0	0	4,557
ES	685	670	0	4,142	0	6,186	0	11,683
FI	9,630	0	0	0	0	0	0	9,630
FR	0	8,755	0	0	0	0	0	8,755
GR	0	0	0	0	0	3,471	0	3,471
HU	6,006	0	0	0	0	0	0	6,006
IE	17	0	10	0	0	4,718	0	4,745
IS	4,178	0	0	0	0	0	0	4,178
IT	52	0	0	0	0	15,075	0	15,127
LT	4,242	0	0	0	0	0	0	4,242
LV	3,564	0	0	0	0	0	0	3,564
NL	8,781	0	0	0	0	0	0	8,781
NO	10	6,722	0	0	0	0	0	6,732
PL	1	0	306	0	0	11,557	0	11,864
PT	0	0	0	312	140	3,644	199	4,295
SE	6,285	0	0	0	0	0	0	6,285
SI	9,694	0	0	0	0	0	0	9,694
SK	6,153	0	0	0	0	0	0	6,153
UK	487	0	1	0	0	7,563	0	8,051
Total	89,544	16,455	317	4,454	140	60,967	199	172,076

C. Appendix - List of variables lacking in the longitudinal personal database

code	label	Relevant for:
pb210	Country of birth	analysis on migrations
pb220A	Citizenship	
pe010	Current education activity	These variables are the only proxy of apprenticeship
pe020	ISCED level currently attended	
pe030	Year when highest level of education was attained	
pl100	Total number of hours usually worked in second, third… jobs	Fundamental info to construct the correct hourly wage from the yearly income variable.
pl110	NACE (REV 1.1)	Fundamental info to construct EGP class schema
pl130	Number of persons working at the local unit	
pl150	Managerial/supervisor position	
py200g	Gross monthly earnings for dep. employees	Fundamental info in the wage penalty panel analysis

Moreover, in the longitudinal database all the secondary variables are lacking. In the case of the Intergenerational Transmission of Poverty module, these time invariant variables could have been added to the longitudinal files, as they are of extreme relevance for the analysis of different forms of inequality.

Intergenerational Transmission of Economic Vulnerability and Poverty

Christopher T. Whelan, Brian Nolan and Bertrand Maître

Introduction

- Key independent variables – parental education and class, childhood economic circumstances and current education
- Key dependent variables at risk of poverty and economic vulnerability.
- Economic vulnerability comprises a multidimensional profile covering income poverty, consumption deprivation and economic stress involving inability to cope with unanticipated expenses.

Welfare Regimes

Whelan, Nolan and Maître conduct an analysis at welfare regime distinguishing the following regimes.

- Social Democratic
- Corporatist
- Liberal
- Southern European
- Post Socialist Corporatist
- Post Socialist Liberal

In conducting such analysis it is necessary to exclude countries at each stage where data is missing or of dubious quality. It is also necessary to conduct analysis at high levels of conclusion.

Variables

- Income poverty
- Economic vulnerability derived from a latent class analysis
- Aggregated 4 category parents’ education variable using information for mother and father and dominance approach
- Aggregated 4 category parent social class variable using information for mother and father and dominance approach
- Childhood economic circumstances

Economic Vulnerability

- Levels of economic vulnerability vary across welfare regimes in the expected fashion
- The economic vulnerability group are sharply differentiated from the non-vulnerable in every regime.

Intergenerational Influences on Poverty and Econmic Vulnerability

Certain broad conclusion emerge

- Generally the impact of parents’ education is somewhat stronger than parents’ social class.
- The cumulative impact of parents’ class, parents’ education and childhood economic circumstances is generally significant.
- The impact of these factors is substantially less than that relating to the individual’s education but of clear substantive importance.
- Intergenerational influences are weakest for the Social Democratic regime
- Parents’ education effects are particularly strong in the post socialist regimes.

- Parents' social class effects are effects are strongest in the liberal and corporatist regimes.
- Childhood economic circumstances have their weakest impact in the social democratic and post-socialist counties and their strongest influence in the liberal, corporatist and southern European regimes.
- Intergenerational influences have a consistently stronger influence of economic vulnerability than income poverty.

Caveat

While the broad pattern of results reported is probably fairly robust, concerns about the quality of the occupational and education variables means that it is very difficult to have confidence in the precision of the estimates.