Slide 1

ESRI

An Introduction to the ECHP for New Users - Day 2

Dorothy Watson, Bertrand Maître, Bernadette Ryan

EPUNet EuroPanel Users Network
Connecting ECHP Users

Slide 2

Day 2 Outline

- *Harmonisation and comparability*

- *Units of Analysis*
  - *Individuals*
  - *Households*
  - *Couples*
  - *Families*

Slide 3

Harmonisation and Comparability

- ECHP goal: Input harmonisation of
  - concepts
  - questionnaires
  - fieldwork procedures
  - editing, weighting, imputation
  - Some more flexibility with respect to sampling

Slide 4

Questionnaires

- *The EU Harmonised questionnaires are with the UDB documentation*
- *See C:\echp\Documentation*
  - *Documents showing question wording, variable list and routing for each questionnaire*

Slide 5

Questionnaires (2)

- *Have a look in C:\ECHP\Documentation*
  - *W1pan015-94Varlist.pdf*
  - *W2pan030-95Varlist.pdf*
  - *W3pan065-96VarList.pdf*
  - *W4pan081-97VarList.pdf*
  - *W5pan097-98VarList.pdf*
  - *W6pan112-99VarList.pdf*
  - *W7pan151-00VarList.pdf*
  - *W8pan159-00VarList.pdf*

Slide 6

Sampling (1)

- Probability samples
- Generally, stratified two-stage sampling designs
  - weighted to compensate for differences in selection probabilities and response rates, and calibrated to external control distributions.
- In some countries, direct (single stage) sampling of households or persons has been used

Sampling (2)

- *What is the difference between a simple random sample and a two-stage sample?*

- *What effect does this have on estimates?*

Sampling (4)

- *UDB data identify the PSU from which the household was originally selected and the order of selection of PSUs*
  - *HG005, HG006, HG007*
- *In Denmark, the Netherlands and Luxembourg, the sample is effectively a simple random sample of households*

Effect of Sample Design

- Generally, more clusters --> more efficient sample (e.g. Ireland 258, UK 249)

- Doc PAN 128/2000 examines Sampling Errors for Wave 2 data for a range of proportions, means etc.

- Design effects by variable and country

Effect of Sample Design (2)

Design effects (deft): ratio of the actual standard errors to that assuming simple random sampling

DE (ECHP)	1.20	IR	1.32
DK	1.06	IT	1.70
NL	1.10	GR	1.34
BE	1.20	ES.	1.29
LU (ECHP)	1.08	PT	1.87
FR	1.15	AT	1.34
UK (ECHP	1.17	All	1.29

Departures from Harmonisation

- Timing (countries beginning to participate later)
- Differences in mode of data collection
- Cloning (use of national sources)
- Differences in concepts

Different Start Dates

Countries	Full ECHP Data Format	ECHP Data Format derived from National Surveys
Belgium*, Denmark, France, Greece, Ireland, Italy, the Netherlands*, Spain, Portugal	1994-2001	-
Austria	1995-2001	-
Finland	1996-2001	-
Germany	1994-1996	1994-2001 (SOEP)
Luxembourg	1994-1996	1995-2001 (PSELL)
United Kingdom	1994-1996	1994-2001 (BHPS)
Sweden	-	1997-2001 (SLCS) (Cross-sectional only)

Differences in Data Collection

- *Mode of Interview*
  - *Particularly face to face, telephone or self-completion*
- *Proxy Interviews*
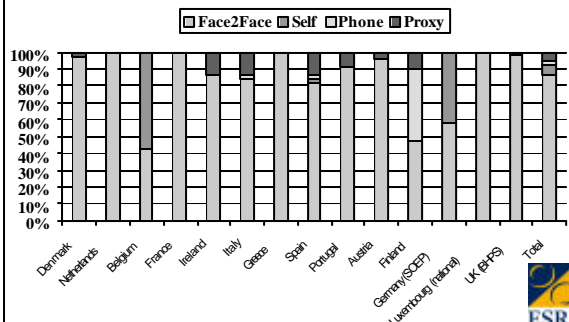 - *information obtained from another household member*

Question ...

- *What difference would you expect by mode of interview?*
  - *(face-to-face, telephone, self-administered)*
- *What difference would you expect between data from a proxy interview and data from a personal interview?*

Differences in Mode of Personal Interview

Cloning - Use of National Sources

- *Desire to use 'Best National Source'*
- *Switch from input-harmonisation to output harmonisation from 1996 -*
  - *Germany,*
  - *Luxembourg,*
  - *UK*
- *Finland - use of income data from registers*

Implications for Data: Cloning

- See paper by *Roland Günther for CHINTEX project*
  - *http://www.destatis.de/chintex/proj_des/wp_1.htm*
- *Main differences due to scale of measurement, different concepts, definition of variables*
- *Minor: definition of population, wording (but important for some variables), respondent, mode, correction for non-response, weighting*

Check Data Dictionary for comparability of information

- *PAN166-200312_Description.pdf*

- *Important codes:*
  - *1 =   Information not available*
  - *C =   confidential (not available)*
  - *TC=   Top Coding (e.g. age 86 = 86 years +)*
  - *RC=   Recoded (Confidentialised)*
  - *g=   Gross of tax*
- *E.g. PT023, PK001, PI110*

Missing Information

- *Germany - some variables missing ...*
  - *household durables*
  - *household financial situation*
  - *health*
  - *personal satisfaction*
  - *training and education*
  - *Caring*

Missing Information

- *UK - some variables missing in these groups ...*
  - *Household Financial Situation*
  - *Employment variables*
  - *Health*
  - *Caring*
  - *Satisfaction*

Missing Information

- *Luxembourg - some variables missing in these groups ...*
  - *Household durables*
  - *Employment/unemployment variables*
  - *Current education/training*
  - *Health*
  - *Caring*
  - *Migration*
  - *Satisfaction*

Examples of Differences (1)

- Relationships (GSOEP and BHPS record relationship to household manager only)
- Status 'at work' includes people on maternity etc. leave in ECHP, but not in GSOEP

Cloning and Income

- *Differences between ECHP and Cloned data in terms of income distribution, poverty rate etc*
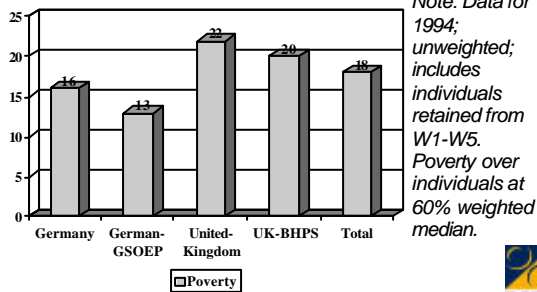- *Due to differences in concepts and measurement*

Examples of Differences - Income

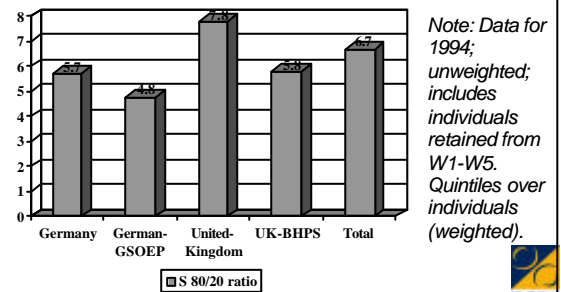- SOEP core concept is gross income (ECHP is net) – program for estimating gross/net ratio.
- BHPS lump sum income of employees or income from secondary or casual jobs not covered
- Reference period (ECHP)= previous calendar year
  - BHPS - social transfers reference period= Sept. of previous year to date of survey
  - GSOEP asks some varables for current year (e.g. overtime payments)
- GSOEP different bands for capital income.

Impact of Income Differences

Note: Data for 1994; unweighted; includes individuals retained from W1-W5. Poverty over individuals at 60% weighted median.

Germany 16, German-GSOEP 13, United-Kingdom 22, UK-BHPS 20, Total 18

Poverty

Impact of Income Differences (2)

Note: Data for 1994; unweighted; includes individuals retained from W1-W5. Quintiles over individuals (weighted).

Germany 5.7, German-GSOEP 4.8, United-Kingdom 7.6, UK-BHPS 5.8, Total 6.7

S 80/20 ratio

Levels of Ex-Post Harmonisation

- Gunther (CHINTEX WP 19)
  - 4. insert harmonised value into PDB (questionnaire variable)
  - 3. insert harmonised value into UDB (analysis variable)
  - 2. insert non harmonised value into PDB/UDB – document
  - 1. Insert no value; document; adjust estimation of population characteristic
  - 0. insert no value, document difference

Levels of Ex-Post Harmonisation (2)

- Goal was to harmonise at level 4
- In practice, some harmonisation at level 3 (e.g. income components could not be separated at PDB, but aggregated UDB variable was provided)
- Also, some harmonisation at level 2 (special codes in UDB and documented).

Critique of Conversion methodology

- (CHINTEX WP 19, *Roland Günther*)
  - Most mehods focus on producing variables at level of PDB – limits possible methods
  - Estimation techniques, making use of knowledge of empirical distributions and statistical relationships to predict harmonised values are almost never used
  - Information external to source of survey (distributions of target or explanatory variables) not used

---

The Unit of Analysis

Unit of Analysis

- *The Unit of analysis can make a difference for estimates of poverty*

  - *What is the Income Sharing Unit (household or family; rarely individual)*

  - *What is the base used to calculate median and percentage under median*

Possible Units of Analysis

- *Household*
- *Family Unit*
- *Economic Family Unit*
- *Individual*
- *In longitudinal work*
  - *Individual observed at particular time*

Income Sharing Unit (1)

- *Assumptions regarding income-sharing*
  - *The household (all members) - most commonly used*
  - *The family unit (persons related by marriage/cohabiting or by parent/child relationship: e.g. parents and all children)*
  - *The Economic family unit (ESU) - couples and dependent children (adult children in household form separate ESU)*
  - *The individual (rarely used) - useful for adult only analyses*

Implications of Choice of Unit

- *Income sharing unit affects results*
  - *e.g. adult children living at home tend to look poorer if they do not 'share' in total household income*
  - *Same for older adults*
  - *Children - do they have a share in income of all household members or just the income of parents?*

Other unit of analysis issues

- *Even if household is taken as income sharing unit,*
  - *is (equivalised) median income (an poverty) calculated over individuals or households?*
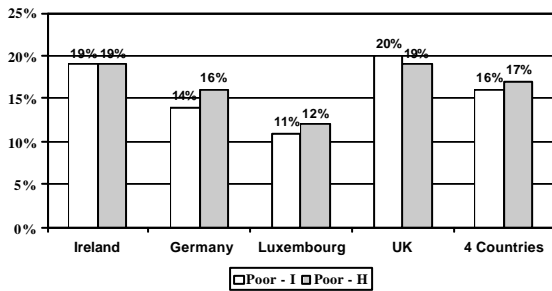 - *This makes a difference ...*

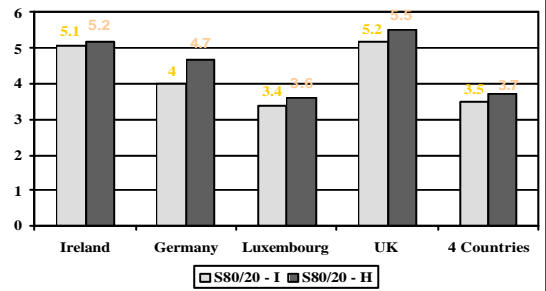## Reminder - Equivalised Income

- *Equivalised income is income per adult equivalent*
- *= (Total income) / (equivalisation factor)*
- *Eurostat tends to use modified OECD equivalence scale*
  - *first adult   = 1.0 (same as OECD)*
  - *other adults   = 0.5 (vs 0.7 in OECD)*
  - *child (under 14)   = 0.3 (vs 0.5 in OECD)*

Slide 1

Poverty over Individual vs Household - 1996

Bar chart (Poor - I, Poor - H):
- Ireland: 19%, 19%
- Germany: 14%, 16%
- Luxembourg: 11%, 12%
- UK: 20%, 19%
- 4 Countries: 16%, 17%

Legend: □ Poor - I ■ Poor - H

Slide 2

S80/20 Ratio over Individual vs Household - 1996

Bar chart (S80/20 - I, S80/20 - H):
- Ireland: 5.1, 5.2
- Germany: 4, 4.7
- Luxembourg: 3.4, 3.6
- UK: 5.2, 5.5
- 4 Countries: 3.5, 3.7

Legend: □ S80/20 - I ■ S80/20 - H

Slide 3

Question ...

- *What might account for these differences?*
  - *Greater inequality over individuals than over households (even when income shared among household members)*
  - *Slightly higher poverty risk over households in Germany, Luxembourg; opposite in UK*

ESRI

Slide 4

EPUNet Course participants

Slide 5

ESRI

Working with the UDB :
Matching within waves

Lab Session Day 2

EPUNet EuroPanel Users Network
Connecting ECHP Users

Slide 6

Identifiers within Files

Country File
- Country

Personal/Register File
- Country
- HID
- PID

Household File
- Country
- HID
- PID

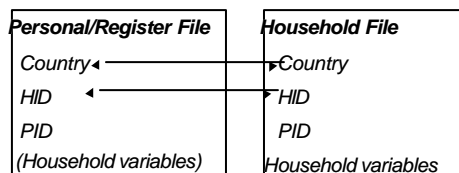**Relationship File**
- Country
- HID
- PID1
- Relation
- PID2
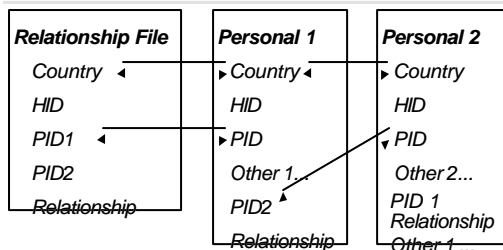
ESRI
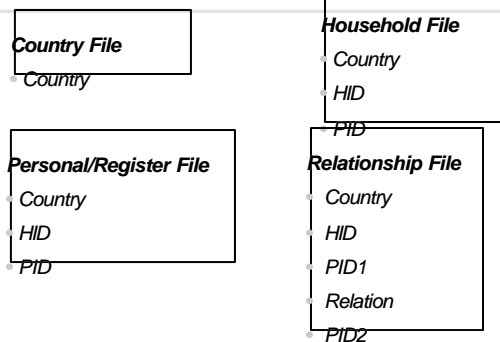
Matching Within Waves

- *Matching Households to Individuals*

- *Matching Individuals to Households*

- *Matching Individuals to Individuals*
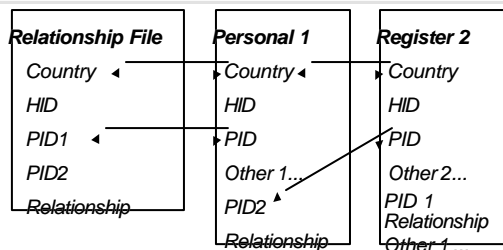
ESR

Matching Household Data to Individuals

Personal/Register File	Household File
Country	Country
HID	HID
PID	PID
(Household variables)	Household variables

ESR

Matching Spouses

Relationship File	Personal 1	Personal 2
Country	Country	Country
HID	HID	HID
PID1	PID	PID
PID2	Other 1...	Other 2...
Relationship	PID2	PID 1 Relationship
	Relationship	Other 1 ...

Rename PID for match ...

ESR

Matching Children to Parents

Relationship File	Personal 1	Register 2
Country	Country	Country
HID	HID	HID
PID1	PID	PID
PID2	Other 1...	Other 2...
Relationship	PID2	PID 1 Relationship
	Relationship	Other 1 ...

Rename PID for match ...

ESR

Matching Household Characteristics to Individuals

- *Importance of sorting (country, hid,pid)*

- *Characteristics : Household size, Household total income, Type of accomodation…*

ESR

Matching: Exercise 1

- *Match household size to each individual and run frequencies of household size for Ireland, wave 1*

- *Unit of analysis is individual*

ESR

Syntax for Exercise 1

```
GET FILE=w1regsav.
sort cases by country hid pid.
match files file=*/in=inr
  /table=w1hsav
    /in=inh
    /keep=country hid hd001
  /by country hid.
exe.
fre var=inr inh.
```

Check: result for Ireland

HD001 Household Size

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	532	3.6	3.6	3.6
	2	1746	12.0	12.0	15.6
	3	1932	13.2	13.2	28.9
	4	3060	21.0	21.0	49.8
	5	3135	21.5	21.5	71.3
	6	1950	13.4	13.4	84.7
	7	966	6.6	6.6	91.3
	8	680	4.7	4.7	96.0
	9	270	1.9	1.9	97.8
	10	130	.9	.9	98.7
	11	121	.8	.8	99.6
	12	24	.2	.2	99.7
	13	39	.3	.3	100.0
	Total	14585	100.0	100.0	

Matching Exercise 2

- *Match Ownership (type of dwelling) to each individuals interviewed and run frequencies for all countries, wave 1*

- *Unit of analysis is individual*

Syntax for Exercise 2

```
GET FILE=w1psav.
sort cases by country hid pid.
match files file=* /in=inp
  /table=w1hsav
  /in=inh
  /keep=country hid ha023
  /map
  /by country hid.
exe.
```

Example 2: Results

Ownership of Dwelling by Country

		Country Code		
		8 Ireland	51 Germany-National source	57 UK-National source
		%	%	%
Own/Rent Dwelling	1 owner	87.9%	39.6%	71.6%
	2 tenant/subtenant/paying rent	11.0%	58.0%	26.6%
	3 accommodation is rent-free	1.1%	2.4%	1.7%

Matching Individual Characteristics to Households

- *Importance of sorting (country, hid,pid)*

- *Characteristics of a specific person (Ref) : Age, Education level, Health…*

Matching Exercise 3

- *Match Principal Economic Status of the Reference person to the household, run freq for UK, wave 1*

- *Unit of analysis is household*

Exercise 3 Syntax

```
GET FILE=w1hsav /keep=country hid
  hg001.
sort cases by country hid.
match files /file=* /in=inh
  /table=w1psav /in=inp
  /rename (pid pe001=hg001 PesRef)
  /keep=country hid hg001 PesRef
  /by country hid hg001.
execute.
```

Exercise 3 Results

Main Activity Status - Self-Defined

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	paid employment 15+hours	2062	40.2	43.2	43.2
	paid apprenticeship 15+hours	1	.0	.0	43.2
	self-employment 15+hours	387	7.5	8.1	51.4
	education,training	103	2.0	2.2	53.5
	unemployed	240	4.7	5.0	58.5
	retired	1202	23.4	25.2	83.7
	housework	385	7.5	8.1	91.8
	other economically inactive	391	7.6	8.2	100.0
	Total	4771	93.1	100.0	
Missing	missing	7	.1		
	System	348	6.8		
	Total	355	6.9		
Total		5126	100.0		

Problem with match?

- *Notice that some households have no match to the individual*

- *Why might this happen?*

- *What is a possible solution?*

Matching Individuals to Individuals
Matching Spouses - Exercise 4

- *Importance of sorting (country, hid,pid)*

- *Exercise 4*
  Match partners and respective ages, produce mean ages and age differences by country, wave 1

- *Unit of analysis is Couple*

Example 4 Syntax

```
GET FILE=w1relsav.
sort cases by country hid pid1.
select if relation eq 1.
match files /file=*/in=inrel
  /table=w1regsav/in=inr
  /rename (pid rd003=pid1 age1)
  /keep=country hid pid1 age1
  pid2
  /by country hid pid1.
execute.
```

Example 4 Syntax (continued)

```
sort cases by country hid pid2.

match files /file=*/in=inrel2
  /table=w1regsav/in=inr2
  /rename (pid rd003=pid2 age2)
  /keep=country hid pid1 age1
  pid2 age2
  /by country hid pid2.
execute.
```

Exercise 4 Results

Average Age and Age Difference of Spouses by Country

	8 Ireland	51 Germany-National source	57 UK-National source
Age of spouse 1	50	46	47
Age of spouse 2	47	44	45
Age difference	3.56	3.80	3.80