An Introduction to the ECHP for New Users - Day 3

Dorothy Watson, Bertrand Maître, Bernadette Ryan

EPUNet European Panel Users Network
Connecting ECHP Users

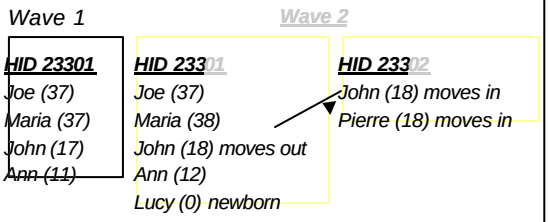
Day 3 Outline

- *Longitudinal Issues - Tracing Rules*
- *Attrition*
- *Cases available for panel analysis Presentation by Richard Layte*
- *More on Comparability of Data*

Sample Persons and Tracing

- *Sample person: someone who was a member of an ECHP household in the first wave*
- *Or a child born since the first wave to a mother who is a sample person*
- *Sample persons who move are traced*
  - *all members of a sample household are interviewed*
  - *even non-sample persons - to get total household income*
  - *So new people can be added to the panel if they move into a sample household or a sample person moves in with them*

Tracing and Newly-Generated Households



Who is followed in panel?

- *Sample person*
- *Moving to private or collective household*
- *In EU*
  - *In practice, not if move to another country*
- *Not traced if move outside EU, to institution, or is non-sample person*

Attrition

- Loss of sample persons over time in a panel survey
- Some inevitable loss (death, illness)
- Some loss due to fatigue (refusing in later waves)
- Some loss due to difficulties in tracing people who move

Attrition: Extent 1994-1998 (1)

- Focus on loss of individuals due to Household Non-response
  - Some individuals lost because of non-response to personal interview
- Ignore people out of scope
  - died, moved outside EU, moved to institution
- Notice that extent of attrition increases as number of waves increases

Attrition: Extent 1994-1998

	N waves	N cases	% retained
Belgium*	5	8,938	71
Denmark	5	7,465	58
Germany	3	12,270	89
Greece	5	15,765	69
Spain	5	22,481	64
France	5	18,642	68
Ireland	5	14,075	57
Italy	5	21,359	76
Luxembourg	3	2,793	89
The Netherlands*	5	13,029	72
Austria	4	9,399	74
Portugal	5	14,156	82
Finland	2	11,180	92
United-Kingdom	3	14,143	61
German-National *	5	15,848	79
Uk-National source*	5	12,595	81
Total	5	214,138	73

Patterns of Attrition (1)

- In panel survey, we have much more information on individuals lost through attrition than on initial non-respondents

Patterns of Attrition (3)

- Associated with
  - Changing address
  - Missing information in earlier wave
  - Change of interviewer
- Smaller effects:
  - Younger, single adults
  - In Northern countries: lower income, lower education
  - In Southern countries and Ireland: higher income, higher education

Patterns of Attrition (4)

- Most attrition is random with respect to observables
- Based on examining characteristics of individual in previous wave(s)

Impact of Attrition (1)

- This is the important question
- Obvious impact: Reduces number of cases available for analysis
- Less obvious: If non-random, **can** bias results in later waves
  - attrition on observables: e.g. sex, characteristics in earlier wave, change in circumstances between waves (measured)
  - attrition on non-observables: e.g. change in circumstances between waves (unmeasured)

Impact of Attrition (2)

- *In general, not as large as might expect*
- Model of w1-w5 attrition (Watson 2003) including
  - *Country and Wave*
  - *Individual characteristics (age, sex, marital status, education, economic status, socio-economic group)*
  - *Household Characteristics (household size, type,tenure, stability of residence, main income source, household income decile, poverty status)*

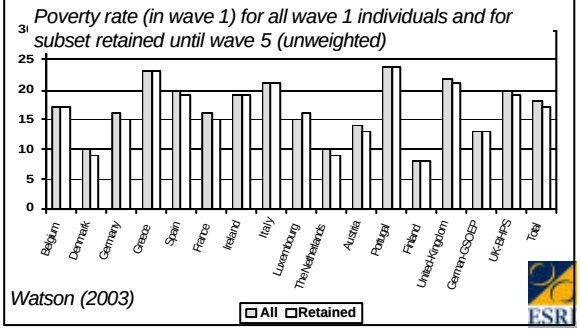
Impact of Attrition (3)

- The McKelvey-Zaviona pseudo-R^2 statistic shows all of the variables included in the final model account for only 11 per cent of the variance in attrition.
- Of this total, almost half (4.9 per cent) is due to differences in the level of attrition between countries and across waves.
- Most attrition, then, is not associated with the large number of substantive independent variables in the model.

Impact of Attrition (4)

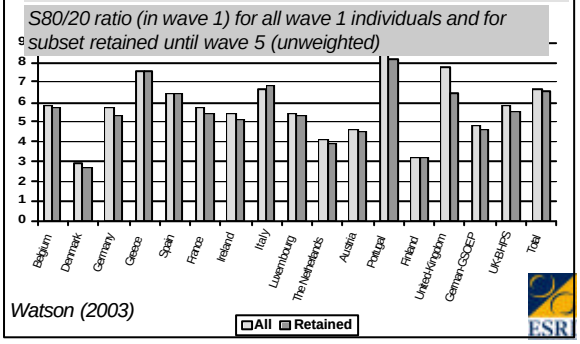
- *In general, not as large as might expect*
  - *even though attrition is patterned, impact on sample structure is small*
  - *most attrition is random and effects decline over life of panel*
  - *impact on sample estimates and coefficients from multivariate models also tend to be small*

Attrition and Poverty Rate (illustration)



Watson (2003)

Attrition and S80 / 20 ratio (illustration)



Watson (2003)

Cases available for panel analysis

	N. Persons (all ages)	N. Personal Interviews
Wave 1-2	179,464	132,220
Wave 2-3	187,573	139,594
Waves 1 to 7	99,516	70,966
Waves 1 to 8	92,350	65,622

Other Issues affecting comparability

- *Income in France and Finland*
- *Survey and Register data*
- *Pensions*
- *Sweden: cross-sectional data only*

Income: Gross and Net

- *In France and Finland, Income components are given as Gross amounts*
  - *But total personal income and total household income given as net amounts*
  - *Net/Gross conversion factor used*
- Sweden lacks some detail on components.

Gross components -- Total Net

- *In most countries:*
  - *HI100=HI110+HI120+HI130+HI140*
  - *HI100= SUM(PI100) + HI140*
- *In France:*
- *HI100=HI020*(HI110+HI120+HI130+HI140)*
- *In Finland (Hi140 is already net)*
- *HI100=HI020*(HI110+HI120+HI130)+HI140*

Pensions

- *Difficulty in harmonising according to original Eurostat categories*
- *See Pan 166 on PI1321 for details*
- *Not always possible to distinguish Old-age related and survivor's pensions*

So the lesson is ...

- *Carefully check the Data Dictionary for notes on the comparability and availability of variables you are interested in.*
- *Doc PAN 166*

Lab Session: Matching across Waves

- *8 waves*
- *Matching Individuals across waves*
- *Matching Households across waves*

Identifiers within Files

Country File
- Country

Household File
- Country
- HID
- PID

Personal/Register File
- Country
- HID
- PID

Link File
- Country
- PID
- HID1…HID8

Matching Across Waves

- All individuals ever in panel are in Link File
- Link File: shows situation of person in each wave
  - Person info
    - Fixed information: dob, gender, pid, sample person
    - Wave specific variables: resid status in Wi, pers elegible for interw in Wi, whether interviewed…
    - Wave-specific information on person's household: HID in Wi, sample status in Wi, location in Wi, household size in Wi,…
    - Movement in and out: new entrants, leavers,

Matching Across Waves: Exercise 1

- Identify individuals interviewed in all waves from 2 to 5,
- Match to individual file to find economic status in each wave
- Examine change in activity status by country

Exercise 1 Syntax (1)

"Match across waves.SPS"

```
get file=linksav.

* identify people interviewed all waves 2 to 5.
compute in2_5=0.
if pfnres2 eq 11 and pfnres3 eq 11
  and pfnres4 eq 11 and pfnres5 eq 11 in2_5=1.

fre var=in2_5.

select if in2_5 eq 1.
```

Exercise 1 Syntax (2)

```
* match to personal file for each wave to find
  main activity status i = wave number.

* wave i.
compute hid=hidi.
sort cases by country hid pid.
match files/file=*/in=inlinki
 /table=wipsav/in=inwip
 /rename (pe001=mainacti)
 /keep=country hid pid birthyy to pfnres8
  mainacti mainact… …./by=country hid pid.

fre var=inlinki inwip mainacti.
```

Exercise 1 Syntax (3)

```
* Construct variable for change in activity status.
compute chact=-9.
missing values chact (-9).
variable labels chact 'Change in activity, W2-5'.
value labels chact
 1 'Always working'
 2 'Never working'
 3 'Sometimes working'.
count nwork = mainact2 mainact3 mainact4 mainact5 (1,2,4,5,12).
count nother= mainact2 mainact3 mainact4 mainact5
   (3,6,7,8,9,10,11).
count nval =  mainact2 mainact3 mainact4 mainact5 (1 thru 12).
fre var=nwork nother nval.
if nwork eq nval chact=1.
if nother eq nval chact=2.
if nwork ge 1 and nother ge 1 chact=3.
fre var=chact.
```

Exercise 1 Results

Change in Activity Status, W2-W5

	8 Ireland	51 Germany	55 Luxembourg	57 UK
1.00 Always working	38%	43%	45%	47%
2.00 Never working	41%	33%	45%	29%
3.00 Sometimes working	21%	24%	11%	24%

Exercise 2: Changes in Household Characteristics

- *Identify individuals (any age) present in households interviewed in all waves from 2 to 5,*
- *Match to household file to find level of economic strain (HF002) in each wave*
- *Examine change level of economic strain by country*

Exercise 2 Syntax (1)"Match across waves.SPS"

```
get file=linksav.
* identify people interviewed in all
  waves from 2 to 5.
compute in2_5=0.
if hfnres2 eq 11 and hfnres3 eq 11 and
  hfnres4 eq 11 and hfnres5 eq 11
 in2_5=1.

fre var=in2_5.

select if in2_5 eq 1.
```

Exercise 2 Syntax (2)

```
* match to household file for each wave to find level of
    economic strain .
* wave i.
compute hid=hidi.
sort cases by country hid .
match files/file=*/in=inlinki
 /table=wihsav/in=inwih
 /rename (hf002=straini)
 /keep=country hid pid birthyy to pfnres8 straini (other
    strain variables) /by=country hid.
* check match.
fre var=inlinki inwih straini.
```

Exercise 2 Syntax (3)

- *Construct variable for change in economic strain (see chstrain in syntax file)*
- *Tabulate this variable by country*

Exercise 2: Results

Change in Economic Strain (Difficulty making ends meet), W2-W5

	8 Ireland	55 Luxembourg	57 UK
1.00 Always difficulty	7%	1%	2%
2.00 Never difficulty	57%	81%	79%
3.00 Sometimes difficulty	36%	17%	20%

Lab: More Within-Wave Matching

- Matching Parents to Children
  - Unit=Child
- Matching Children to Parents
  - Unit= Parent

Matching Individuals to Individuals Matching Parents to Children

- *Exercise 3:*
  Matching parents to children
  find mean age of children and parents, wave 1

  Unit of Analysis = Child

  Method:
  - *Begin with relationship file*
  - *identify parent/child pairs*
  - *aggregate to child level*
  - *Match ages from register (X 3)*

Exercise 3 Syntax (1)

```
* Step 1: get relationship file; select parent-child pairs.
GET FILE=w1relsav.
select if relation eq 2.
if relation eq 2 child eq pid2.
if relation eq 2 Parent=pid1.
* step 2: aggregate to level of child - keep id of both parents.
AGGREGATE
  /OUTFILE=*  /BREAK=country hid child
  /ncases=N
  /Parent1 = FIRST(parent) /Parent2 = LAST(parent).
if parent1 eq parent2 parent2=$sysmis.
descriptives var=all.
```

Exercise 3 Syntax (2)

```
* step 3: sort and match on age of child.
sort cases by country hid child.

match files /file=*/in=inc
   /table=w1regsav/in=inreg1
   /rename (pid rd003=child agec)
   /keep=country hid child ageC parent1 parent2
   /by country hid child.
execute.
fre var=inc inreg1.
```

Exercise 3 Syntax (3)

```
* step 4: Sort and match age of first parent.
sort cases by country hid parent1.
match files /file=*/in=inc2
   /table=w1regsav/in=inreg2
   /rename (pid rd003=parent1 agep1)
   /keep=country hid child ageC parent1 agep1 parent2
   /by country hid parent1.
execute.
fre var=inc2 inreg2.
```

Exercise 3 Syntax (4)

```
*step 5:  Sort and match age of second parent.
sort cases by country hid parent2.
match files /file=*/in=inc3
   /table=w1regsav/in=inreg3
   /rename (pid rd003=parent2 agep2)
   /keep=country hid child ageC parent1 agep1 parent2
   agep2
   /by country hid parent2.
execute.
fre var=inc3 inreg3.
variable labels agep1 'Age parent 1' agep2 'Age parent 2'
   agec 'Age of child'.
```

Exercise 3 Syntax (5)

```
* step 6: Table of ages of children and
  parents.
TABLES
  /FORMAT BLANK MISSING('.')
  /OBSERVATION= agep1 agep2 agec
  /GBASE=CASES
  /TABLE=agep1 + agep2 + agec  BY country
  /STATISTICS
  mean( '')
 /title='Average Ages of Parents and
  Children by Country'.
```

Exercise 3 Results

Average Ages of Parents and Children by Country

	8 Ireland	51 Germany-National source	57 UK- National source
Age parent 1	48	43	42
Age parent 2	44	40	39
Age of child	16	14	13

Matching Children to Parents

- *Exercise 4:*

  Find parents and get number of children aged under 5,aged under 15, aged under 18. Use wave 7

  Unit of analysis = parent

Exercise 4 Syntax (1)

* step 1: Go to relationship file and select parent/child pairs.

```
get file=w7relsav.

select if any(relation,2,3).
fre var=relation country.
```

Exercise 4 Syntax (2)

```
*step 2 match on age of parent - parent will be pid1.

sort cases by country hid pid1.

match files/file=*
 /rename (pid1=pid)/in=inrel
 /table=w7regsav/in=inr
 /rename (rd003    rd004    rd005
       =  rd003p1 rd004p1 rd005p1)
 /map
/by=country hid pid.
fre var=inrel inr.

rename variables (pid = pid1 ). /* put parent id back to pid1.
```

Exercise 4 Syntax (2)

* step 3: match on age of child - child will be pid2.

```
sort cases by country hid pid2.

match files/file=*
 /rename (pid2=pid)/in=inrel2
 /table=w7regsav/in=inr2
 /rename (rd003 rd004 rd005 =  rd003p2 rd004p2 rd005p2)
 /map   /by=country hid pid
 /keep= country hid pid1 pid rd003p1 rd004p1 rd005p1
    rd003p2 rd004p2 rd005p2.

fre var=inrel2 inr2.
descriptives var= rd003p1 rd003p2.
rename variables (pid = pid2 ). /* put child id back to pid2.
```

Exercise 4 Syntax (4)

* step 4: sort and aggregate to level of pid1 (the parent).

```
sort cases by country hid pid1.

aggregate outfile=*/presorted/break= country hid pid1
 /ncases=N
 /agynkid 'Age f youngest child' = min(rd003p2)
 /agolkid 'Age of oldest child' =max(rd003p2)
 /ownage 'Own age' =first(rd003p1)
 /punder18=pin(rd003p2,0,17)
 /punder5=pin(rd003p2,0,4)
 /punder15=pin(rd003p2,0,14).
```

Exercise 4 Syntax (5)

* convert percentages in each age group (from aggregate) to numbers .

```
do repeat x= nunder18 nunder15 nunder5
            /y=punder18 punder15 punder5.
compute x=ncases*y/100.
end repeat.

var labels nunder18 'N children under 18'
 /nunder15 'N children under 15'
 /nunder5 'N children under 5.

fre var=nunder18 nunder15 nunder5.
```

Exercise 4 Results

N children under 18

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	3957	29.6	29.7	29.7
	1.00	4226	31.7	31.7	61.3
	2.00	3531	26.5	26.5	87.8
	3.00	1244	9.3	9.3	97.1
	4.00	288	2.2	2.2	99.3
	5.00	66	.5	.5	99.8
	6.00	27	.2	.2	100.0
	7.00	4	.0	.0	100.0
	9.00	2	.0	.0	100.0
	Total	13345	100.0	100.0	
Missing	System	2	.0		
Total		13347	100.0		

N children under 15

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	5258	39.4	39.4	39.4
	1.00	3877	29.0	29.1	68.5
	2.00	3013	22.6	22.6	91.0
	3.00	944	7.1	7.1	98.1
	4.00	201	1.5	1.5	99.6
	5.00	42	.3	.3	99.9
	6.00	8	.1	.1	100.0
	7.00	2	.0	.0	100.0
	Total	13345	100.0	100.0	
Missing	System	2	.0		
Total		13347	100.0		

N children under 5

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	9855	73.8	73.8	73.8
	1.00	2868	21.5	21.5	95.3
	2.00	587	4.4	4.4	99.7
	2.00	2	.0	.0	99.8
	3.00	33	.2	.2	100.0
	Total	13345	100.0	100.0	
Missing	System	2	.0		
Total		13347	100.0		