



An Introduction to the ECHP for New Users - Day 4

Dorothy Watson, Bertrand Maître,
Bernadette Ryan

EPUNet European
Users Network
Connecting ECHP Users

Day 4 Outline

- Sample Weights and Grossing Factors
- The Country File
- Chris Whelan on Using Latent Class Models
- Item Non-Response and imputation
- Unit Non-Response and imputation



Purpose of Weights

- Purpose of weights is to correct for any distortion or lack of representativeness in sample
- Could be due to
 - characteristics of sampling frame
 - sampling error
 - differential response rates
 - attrition



Weights in the UDB

- Two types of Weights
 - Crosssectional
 - Longitudinal (Base weight)



Weights available in UDB files

File type	Name of the weight variable	
	Base weight	Cross-sectional weight
Register	RG003	RG002
Personal	PG003	PG002
Household	-	HG004



Base Weights

- RG003 and PG003
- Sample persons assigned a base weight.
- The base weight of non-sample persons is 0.
- Base weights are scaled such that the average of the base weights of sample and non-sample persons is 1



Cross-Sectional Weights

- RG002, PG002 and HG004
- Cross-sectional weights for persons (RG002 and PG002) are derived from the base weights by sharing the weights of all persons in the same household.
- Scaled such that average weight is 1.
- The cross-sectional household weight (HG004) corresponds to the average weight of its members.
- Scaled such that average HG004 = 1.



When to Use

- *Base weights* - use for longitudinal analysis (must exclude non-sample persons)
- *Cross-sectional weights* - use for cross-sectional analyses (can include non-sample persons)



Cross-sectional analysis

- For cross-sectional analysis at country level, the normalised cross-sectional weights can be used.
 - PG002 for interviewed persons
 - RG002 for persons in interviewed households
 - HG004 for interviewed households



Country Level vs Multi-Country Analysis

- The weight variables for each country sum to the sample size of the country (mean=1)
- This is ok if we just want to conduct analyses country by country
- But, e.g., to get EU-level figures, we would generally want the weights for each country to sum to population of country



Grossing Weights to Population totals (1)

Country	Sample n Households (1995)	Population N Households (1995)
Ireland	3,584	1,150,451
Germany	6,336	36,228,000
Luxembourg	2,978	155,935
UK	5,032	24,244,000
Total	17,930	61,778,386

Ireland: 20% of sample; 2% of population
 What would be the result of ignoring population size?
 Luxembourg: 17% of sample; 0.3% of population



Cross-Sectional Multi-country analysis

- For multi-country analysis the weights for each country need to be 'inflated' to the population total
 - so that sum of weights = population total (rather than = sample size)
- This also applies if you want to get population aggregates, e.g. total number of persons in the UK who are working part-time



Grossing Weights to Population totals (2)

- This inflation factor = (N/n)
 - divide the country population (N) by the actual sample size (n).
- Population figures are in the 'country file'
- The inflation factors to be applied are
 - For PG002: "POP16P/ n for this country in the P-file"
 - For RG002: "POPTOT/ n for this country in the R-file"
 - For HG002: "POPHHD/ n for this country in the H-file"



Longitudinal Analysis at Country Level

- Longitudinal analysis over several waves confined to persons interviewed in all waves or to persons living in an interviewed household in all waves.
- For longitudinal analysis up to a wave i , the normalised base weights of wave i can be used. These are:
 - PG003 for interviewed sample persons
 - RG003 for sample persons in interviewed households



Multi-Country Longitudinal Analysis

- Apply an 'inflation factor' = (N/n) to the normalised weights.
 - divide the country population (N) by the actual sample size (n).
- Population figures are in the 'country file'
- The inflation factors to be applied are
 - For PG003: "POP16P/ n for this country in the P-file"
 - For RG003: "POPTOT/ n for this country in the R-file"



Comparing Income across Countries

- Purchasing Power Parity (ppp) - converts amounts in national currencies to a purchasing power standard (PPP)
 - Currency converter
 - Plus price converted
- 1 PPS buys same 'basket' of goods and services in each country



Using PPP

- For example, To compare total net income for 1998,
 - $HI100x = HI100 / PPP98$
- Note: Does not control for changes in prices across waves (need national CPI for this)



Non-Response and Imputation

- Types of Non-Response
 1. Initial Non-Response of Household
 2. Attrition (Non-Response in Later waves)
 3. Unit non-response to personal Interview
 4. Item Non-Response
- Eurostat strategies for each type
 - Weighting (1, 2 and 3)
 - HI140 - Income of non-interviewed adult(3)
 - Imputation (HI140 for 3; item imputation 4)



Eurostat Methodology

- See Pan Doc 165 for weighting
 - Two steps:
 - Adjust for sample design characteristics
 - Calibrate to population totals
- See Pan Doc 164 for imputation of income variables
 - Generally, imputation was conducted at the most detailed level (PDB variables)



Non-response to personal interview

- Within Household non-response
- Extent
- Patterns?



Unit Non-Response to Personal Interview

Within-Household Non-Response, ECHP 2001

	Personal interview	
	Interviewed	Not interviewed
Denmark	96%	4%
The Netherlands	92%	8%
Belgium	93%	7%
France	97%	3%
Ireland	95%	5%
Italy	99%	1%
Greece	100%	0%
Spain	98%	2%
Portugal	99%	1%
Austria	99%	1%
Finland	98%	2%
Germany, national	96%	4%
Luxembourg, national	99%	1%
UK, national	91%	9%
Total	97%	3%



The Income Estimate for Non-Respondents HI140

Use of HI100, Wave 8

	HI140	
	Not used	used
Denmark	95%	5%
The Netherlands	91%	9%
Belgium	93%	7%
France	99%	1%
Ireland	92%	8%
Italy	99%	1%
Greece	100%	0%
Spain	98%	2%
Portugal	98%	2%
Austria	99%	1%
Finland	97%	3%
Sweden	75%	25%
Germany, national	94%	6%
Luxembourg, national	99%	1%
UK, national	87%	13%
Total	94%	6%



Proxy Interviews

- Extent
- Patterns
- Implications



Mode of Interview

Mode of Interview, ECHP 2001

	Mode of Interview			
	Face-2-face	self-adm nistered	phone int	proxy interview
Denmark	97%	0%		3%
The Netherlands	100%			
Belgium	42%	57%		0%
France	100%			
Ireland	86%			14%
Italy	83%		3%	13%
Greece	99%	0%	0%	0%
Spain	83%	2%	2%	14%
Portugal	91%	0%	0%	9%
Austria	96%			4%
Finland	47%		43%	10%
Germany-National source	58%	42%	0%	0%
Luxembourg-National source	100%			
UK-National source	98%		2%	
Total	85%	6%	3%	5%



Subjective Variables and Proxy Interviews

Work Satisfaction by Mode of Interview, ECHP 2001

	Response to Satisfaction With Work		
	Answered, in person	Answered, proxy	Missing
Denmark	97%	2%	1%
The Netherlands	100%	0%	0%
Belgium	95%	0%	5%
France	99%	0%	1%
Ireland	84%	0%	15%
Italy	86%	13%	0%
Greece	98%	0%	2%
Spain	86%	12%	3%
Portugal	91%	9%	0%
Austria	96%	4%	0%
Finland	90%	0%	10%
UK, national	56%	0%	44%
Total	89%	5%	6%

*This variable
not available
for Germany
(national),
Luxembourg
(National),
Sweden*



Lab Session Day 4

*Weights, Grossing Factors and
using the PPP*

Lab Session: Exploring Weights (1)

- *Exercise 1: Use the cross-sectional household weight to get weighted household income in wave 8*



Exercise 1 (2)

- *Begin with Household File*
 - *What checks can we do on weights?*
 - *Why might we have weights of zero?*
 - *Is there a difference between weighted and unweighted household income?*
 - *Why might this difference arise?*



Exercise 1 (2)

- *get file=w8hsav.*
- ** check the weight variable hg004.*
- ** Table for mean and median income, unweighted.*
- *Weight by hg004.*
- ** Table for mean and median income, weighted.*



Exercise 2: Using PPP

- *Exercise 2: Use the PPS to compare income across countries in Wave 8*
 - *Begin with household file*
 - *Match on country file (sorted by cntrepubl)*
 - *Compute net income in PPS*
 - *Tabulate weighted PPS by country*



Exercise 2a: Personal Weights

- Take the personal file for wave 8 and examine weighted mean and median personal income
- Begin with personal file



Exercise 3(1): Multi-country analysis

- Get the weighted average income across households in the four countries (UK, Ireland, Germany, Luxembourg) in PPS units.
- Find the total number of people in these four countries with PPS income below 10,000 (population aggregate)



Exercise 3(2)

- To get weighted total across countries (and population aggregates), need to adjust weights so that they sum to total population within country
- For PG002: "POP16P/n for this country in the P-file"
- For RG002: "POPTOT/n for this country in the R-file"
- For HG004: "POPHTD/n for this country in the H-file"



Exercise 3 (3): Results

Average Net Household Income in Purchasing Power Standard(Mean and Median) by Country and Average Across Countries

	Household Net Income in PPS		
	Mean	Median	Count
8 Ireland	27,203	22,568	1,290,580
51 Germany-National source	27,955	24,096	37,711,001
55 Luxembourg-National source	45,017	38,579	171,982
57 UK-National source	27,873	23,315	25,564,001
TOTAL	27,953	23,873	64,737,564

Weighted



Exercise 3 (4): Results

Number of households with PPS Below 10,000 by Country

	PPS under 10,000	
	.00 Under 10,000	1.00 10,000 or above
8 Ireland	249,707	1,039,924
51 Germany-National source	4,044,771	33,568,352
55 Luxembourg-National source	953	171,029
57 UK-National source	3,771,442	21,566,777
TOTAL	8,066,874	56,346,082

Weighted



Exercise 4: Using Base Weights

- Select individuals who completed a personal interview in all waves, 1995 to 2001.
- Examine Changes in Self-Rated Health using the appropriate weight



Exercise 4 (2)

Change in Health Status 1995 to 1998 by Country

		51 German-Nati onal source	57 Uk-National source
1.00 Always good	66.3%	26.3%	49.0%
2.00 Mostly good	13.9%	14.9%	18.8%
3.00 Mixed	6.3%	12.3%	11.8%
4.00 Mostly bad	6.8%	14.2%	8.7%
5.00 Always bad	6.8%	32.3%	11.7%

Weighted



Exercise 5: Country aggregates

- Find the number of people in each country who always rate their health as good or very good
- Get the percentage of people across the four countries who always rate their health as very good



Exercise 5 (2)

- Use the appropriate variable from the country file to get country aggregates (numbers of cases)
- Also used to get percentages across countries (multi-country analysis)
- for pg003, use pop15p for the appropriate (latest) wave
 - divide pop16p by sample size.



Exercise 5 (3): Results

Change in Self-Rated Health, 1995-1998, by Country

	8 Ireland	51 German-Nati onal source	57 Uk-National source	TOTAL
1.00 Always good	66.3%	26.3%	49.0%	36.3%
2.00 Mostly good	13.9%	14.9%	18.8%	16.4%
3.00 Mixed	6.3%	12.3%	11.8%	11.9%
4.00 Mostly bad	6.8%	14.2%	8.7%	11.9%
5.00 Always bad	6.8%	32.3%	11.7%	23.5%
	189,218	21,758,747	5,405,858	27,353,824

Weighted

