Correct Estimation
of Between-Country Differences
in the Context of Different Sample Designs
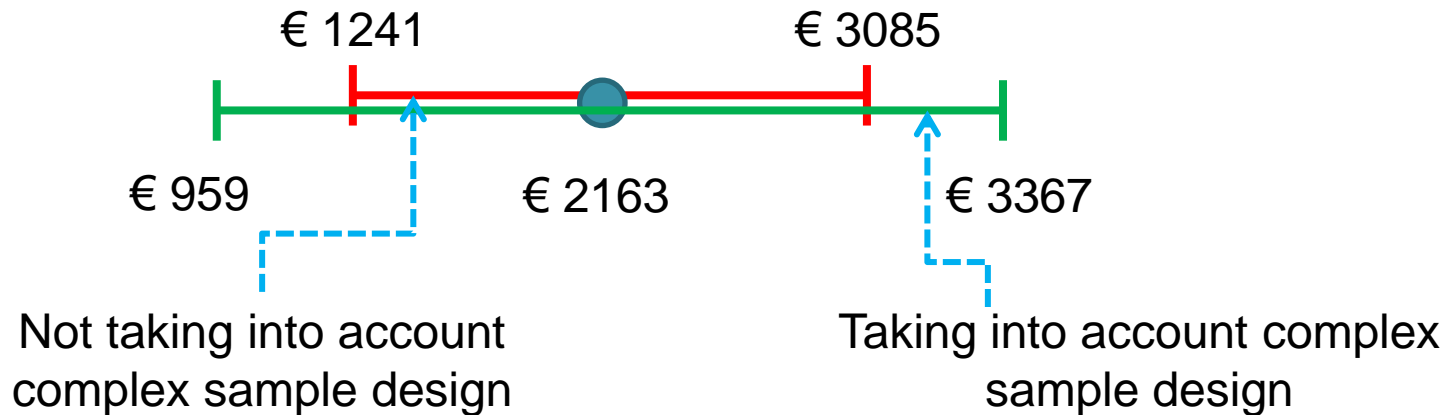
Olena Kaminska and Peter Lynn
ISER, University of Essex, UK

Between-country difference: estimation
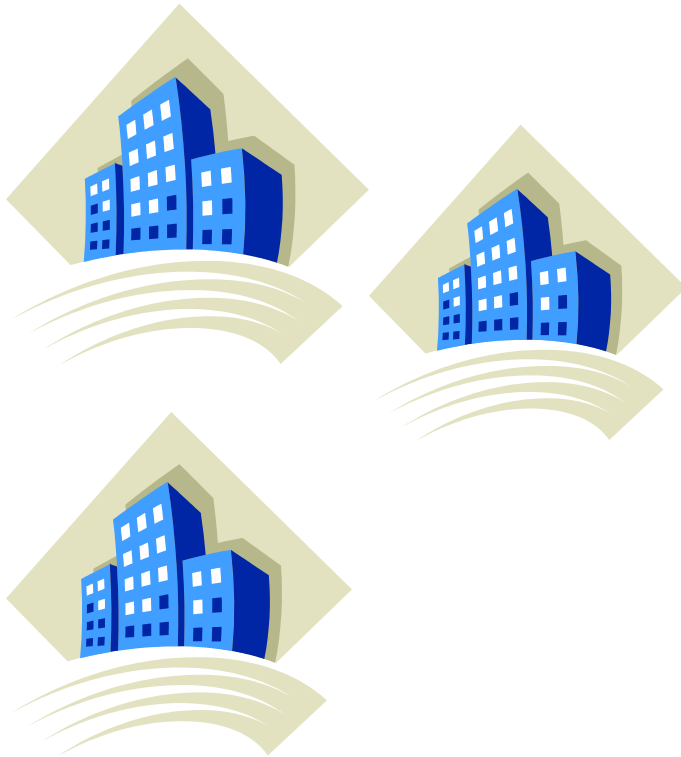# e.g. difference in total disposable HH income

- Country 1 (AT): mean(1)= € 33,914

- Country2 (BE): mean(2)= € 31,751

- Difference in disposable HH income: € 33,914 - € 31,751= € 2,163

€ 1241   € 3085

€ 959   € 2163   € 3367

Not taking into account
complex sample design

Taking into account complex
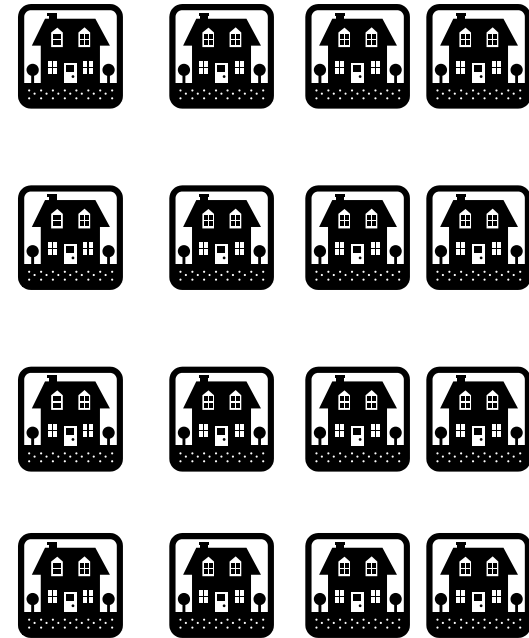sample design

Countries with clustered and unclustered designs
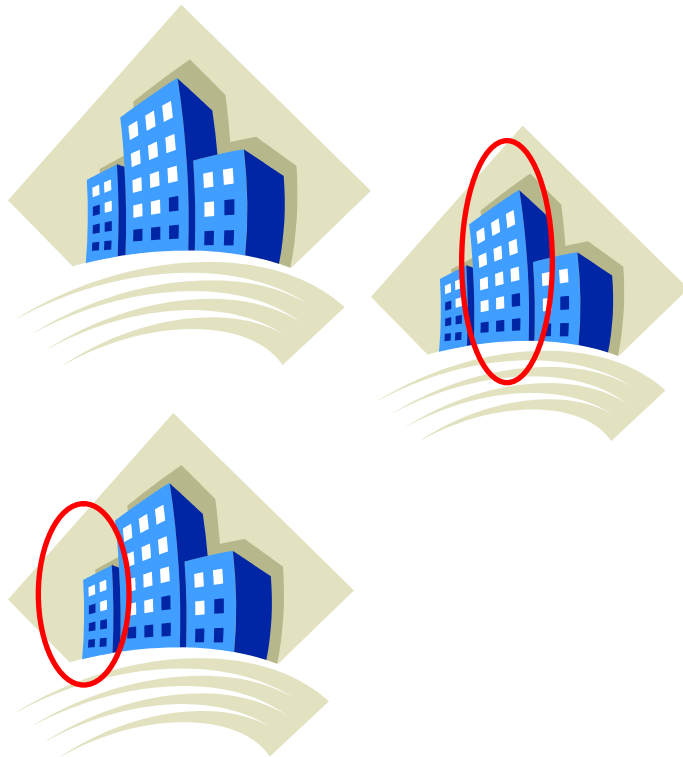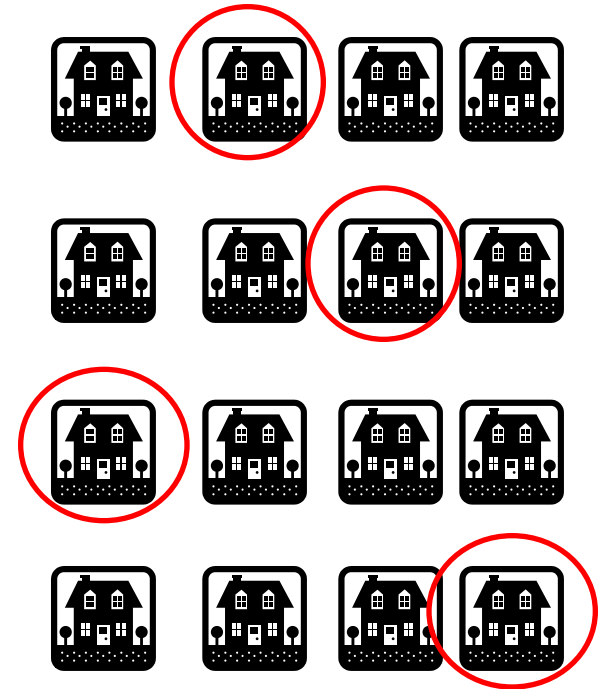
Country 1

Country 2

Countries with clustered and unclustered designs

Country 1 – Clustered Design

Country 2 – Unclustered Design

Comparability at sampling design stage

- Optimal sample design varies across countries

- Sample design does not have to be the same in each country, as long as they represent comparable populations and the probability of selection is known

- Different sample designs should achieve same precision: e.g. in a country with clustered design (where we lose precision) we should plan to interview more respondents

Taking into account sample design at estimation stage

- Statistical software have pre-programmed estimation for complex sample design

- But only for one-population situation. For cross-country study it is correct only when each country has the same sample design

- Our study develops method to estimate difference between means in the situation when one country has clustered and another country has unclustered design

What do we do?

- Two countries are two strata
- In the country with no weights, weights are equal to 1 for everyone (as long as the other country has weights scaled to 1)
- In the country with no clusters, households (or individuals themselves) are treated as clusters

Data: **EU-SILC – EU study of income, social inclusion and living conditions**

- 27 EU countries
- Information on everyone in the household 16+
- 2007 – cross-sectional part for this analysis

Analysis

We look at 90 between-country comparisons

- 10 with clustered design: Belgium, Czech Republic, France, Hungary, Italy, Latvia, Netherlands, Poland, Slovenia, United Kingdom

- 9 without clustered design: Austria, Cyprus, Estonia, Finland, Iceland, Lithuania, Luxemburg, Sweden, Slovakia

Variables:

- Total disposable household income

- Capacity afford holidays

- Capacity to afford meals

- Ability to make ends meet

- Number of household members

Comparison

- Difference in means: $\bar{y}_1 - \bar{y}_2$

- $\text{var}\left(\bar{y}_1 - \bar{y}_2\right)$ with "correct" estimation:

weights + clusters + strata (2 countries)

$\text{var}\left(\bar{y}_1 - \bar{y}_2\right)$ with incorrect estimation:

weights only

- Misspecification effect:

$$\text{meff} = \frac{incorrect\ \text{var}\left(\bar{y}_1 - \bar{y}_2\right)}{correct\ \text{var}\left(\bar{y}_1 - \bar{y}_2\right)}$$

Results for 5 variables: clustering effect on between-country difference estimation

Comparison of weight-only to weight+clusters+strata estimation

Over 90 comparisons	Income	Capacity to afford holidays	Capacity to afford meals	Ability to make ends meet	Number of household members
y1-y2, mean	19160.32	0.25	0.12	0.06	0.28
meff , mean	0.80	0.71	0.81	0.83	0.87
meff std	0.25	0.20	0.14	0.15	0.11
meff, min	0.07	0.33	0.54	0.43	0.55
meff, max	1.00	0.96	0.99	0.99	1.00
change in sig, number	3	2	2	1	1

Overall n=92,987 among 19 countries

Conclusions

- Misspecification effects (meffs) are quite substantial when clustering is ignored (e.g. for income, mean 0.802)

- Despite the substantial meff's, fairly few country comparisons are affected (9 out of 450)

Correct Estimation in Stata

```
svyset psu [pw=weight1], strata(strata1)
svy: mean var1 if cntry1==1 | cntry1==2, over(cntry1)
lincom [var1]1 - [var1]2
```

- Where psu is clusters for the country with clustered design and personal ID for the country with no clustered design

- Strata1 is strata for the country with stratification and country ID for the country without stratification

- Weight1 is weight for the country with weights (scaled to 1), and weight=1 for the country without weights

Thank you

Corresponding Author: Olena Kaminska
olena@essex.ac.uk