Assessment of socio-demographic sample composition in the European Social Survey

Achim Koch
GESIS Leibniz Institute for the Social Sciences
Mannheim, Germany

ITACOSM 2017
5th Italian Conference on Survey Methodology,
Bologna, Italy, 14-16 June 2017
Introduction I

European Social Survey (ESS)

- Since 2002 every 2\textsuperscript{nd} year
- 30+ countries

- Standards set by Central committee (CST)
  - Strict probability sampling
  - Minimum (effective) sample size
  - Face-to-face interviewing only
  - High target response rates
  - Briefing, workload, call attempts of interviewers
  - …
Introduction II

- Expectation: Following procedures + achieving high response rate => Sample of high quality, i.e., a sample which reflects the target population

- Empirical question: How “good” are the ESS samples actually?

- Errors of representation:
  - Sampling error
  - Coverage error
  - Nonresponse error
Evaluating sample quality by assessing demographic sample composition in ESS

Two approaches:

- Using an internal benchmark (derived from ESS samples itself)
- Comparison with external benchmark data
Two questions:

1. What is the **level + pattern of misrepresentation** of demographic groups in ESS samples?

2. Is the level of misrepresentation **related to two basic survey features** (response rate + type of sample)?
Part 1:

Level and patterns of misrepresentation of demographic groups in ESS samples
Internal benchmark criterion


- Criterion measures gender misrepresentation among a subset of respondents
  (namely respondents who live with a partner of the opposite gender in the same household)

- Can be estimated for surveys using samples in which 1 person per household is interviewed
Internal measure of misrepresentation

- In HHs with a gender heterogeneous couple: Male + female partner: same chance of being interviewed
  
  => 50% of the respondents from such couples should be women

- Misrepresentation = respondents gender distribution deviates from 50/50 female/male split

- Internal measure  = \[ \frac{(\% female - 50)}{\sqrt{(50*50) \over n}} \]

  Statistic follows a normal distribution; Values > |1.96| = misrepresentation due to differential nonresponse/undercoverage by gender.
Beyond rejection line: 14 out of 29 countries

Prevailing trend: Females overrepresented (11 countries)

Internal measure of misrepresentation, ESS 6

Internal measure = (% female – 50) / sqrt [(50 * 50) / n]

Subsample: HHs with a heterosexual couple
External benchmark data

- Comparing ESS estimates with data from official statistics
- Majority of ESS countries also participates in European Union Labour Force Survey (LFS)

Two requirements:

- ESS and LFS data have to be comparable
  - Measurement instruments
  - Date of measurement
  - Target population
- LFS data should be of high quality
External benchmark data

6 categorical variables:

- Gender (M/F)
- Age (10-year age categories)
- Marital status (Married, Y/N)
- Work status (In paid work at least 1 hour, Y/N)
- National of country (Y/N)
- HH-size (1 / 2 / 3 / 4 / 5+)
## Differences ESS vs. LFS estimates, ESS 6 (24 c.)

<table>
<thead>
<tr>
<th>Female</th>
<th>15-24 y.</th>
<th>25-34 y.</th>
<th>35-44 y.</th>
<th>Age</th>
<th>45-54 y.</th>
<th>55-64 y.</th>
<th>65-74 y.</th>
<th>75+ y.</th>
<th>Married</th>
<th>In paid work</th>
<th>Non-national</th>
<th>1p-hh</th>
<th>2p-hh</th>
<th>3p-hh</th>
<th>4p-hh</th>
<th>5+p-hh</th>
</tr>
</thead>
<tbody>
<tr>
<td>BE</td>
<td>-0.1</td>
<td>0.7</td>
<td>-1.5</td>
<td>-0.2</td>
<td>1.0</td>
<td>0.6</td>
<td>0.3</td>
<td>-0.8</td>
<td>0.3</td>
<td>3.4</td>
<td>-2.1</td>
<td>2.2</td>
<td>-1.3</td>
<td>2.2</td>
<td>-1.3</td>
<td>1.0</td>
</tr>
<tr>
<td>BG</td>
<td>5.1</td>
<td>-6.7</td>
<td>-4.9</td>
<td>-0.7</td>
<td>2.9</td>
<td>4.2</td>
<td>5.5</td>
<td>-0.1</td>
<td>3.3</td>
<td>-1.5</td>
<td>0.1</td>
<td>0.0</td>
<td>1.8</td>
<td>-1.9</td>
<td>-1.2</td>
<td>1.3</td>
</tr>
<tr>
<td>CH</td>
<td>-1.1</td>
<td>1.1</td>
<td>-2.3</td>
<td>-0.8</td>
<td>0.2</td>
<td>1.5</td>
<td>1.3</td>
<td>-1.0</td>
<td>3.5</td>
<td>2.3</td>
<td>-6.1</td>
<td>-1.7</td>
<td>0.4</td>
<td>0.1</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>CY</td>
<td>3.8</td>
<td>-1.0</td>
<td>-4.7</td>
<td>-0.2</td>
<td>2.8</td>
<td>2.0</td>
<td>1.5</td>
<td>-0.4</td>
<td>1.9</td>
<td>-6.1</td>
<td>-11.9</td>
<td>-1.7</td>
<td>0.4</td>
<td>0.1</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>CZ</td>
<td>-2.3</td>
<td>3.0</td>
<td>-5.2</td>
<td>1.8</td>
<td>5.0</td>
<td>0.4</td>
<td>-0.4</td>
<td>-4.6</td>
<td>4.9</td>
<td>3.0</td>
<td>-0.7</td>
<td>-5.3</td>
<td>-4.5</td>
<td>0.8</td>
<td>8.2</td>
<td>0.9</td>
</tr>
<tr>
<td>DE</td>
<td>-1.4</td>
<td>1.7</td>
<td>-0.9</td>
<td>-1.3</td>
<td>1.7</td>
<td>1.0</td>
<td>1.1</td>
<td>-3.4</td>
<td>1.8</td>
<td>2.8</td>
<td>-4.5</td>
<td>-5.3</td>
<td>-0.8</td>
<td>1.0</td>
<td>2.8</td>
<td>2.2</td>
</tr>
<tr>
<td>DK</td>
<td>-1.2</td>
<td>0.3</td>
<td>-3.5</td>
<td>-1.9</td>
<td>0.8</td>
<td>2.9</td>
<td>1.8</td>
<td>-0.3</td>
<td>4.8</td>
<td>2.3</td>
<td>-3.3</td>
<td>-1.1</td>
<td>0.0</td>
<td>3.3</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>EE</td>
<td>3.6</td>
<td>-1.0</td>
<td>-3.3</td>
<td>-1.1</td>
<td>-0.2</td>
<td>1.5</td>
<td>2.1</td>
<td>-2.2</td>
<td>2.0</td>
<td>-0.9</td>
<td>-0.7</td>
<td>5.9</td>
<td>-0.2</td>
<td>-1.1</td>
<td>-2.3</td>
<td>-2.3</td>
</tr>
<tr>
<td>ES</td>
<td>0.5</td>
<td>-0.8</td>
<td>-1.0</td>
<td>0.9</td>
<td>1.0</td>
<td>1.1</td>
<td>0.4</td>
<td>-1.4</td>
<td>2.1</td>
<td>1.8</td>
<td>-2.9</td>
<td>-2.9</td>
<td>1.4</td>
<td>1.7</td>
<td>0.0</td>
<td>-0.1</td>
</tr>
<tr>
<td>FI</td>
<td>-0.4</td>
<td>-2.6</td>
<td>-1.3</td>
<td>-0.5</td>
<td>0.5</td>
<td>1.1</td>
<td>2.6</td>
<td>0.3</td>
<td>2.4</td>
<td>-1.1</td>
<td>-0.1</td>
<td>1.6</td>
<td>0.9</td>
<td>-1.1</td>
<td>1.5</td>
<td>1.9</td>
</tr>
<tr>
<td>FR</td>
<td>1.9</td>
<td>-3.0</td>
<td>-3.7</td>
<td>1.3</td>
<td>3.4</td>
<td>1.6</td>
<td>1.1</td>
<td>-0.6</td>
<td>9.3</td>
<td>1.6</td>
<td>-1.9</td>
<td>-3.1</td>
<td>0.9</td>
<td>-1.1</td>
<td>1.5</td>
<td>1.9</td>
</tr>
<tr>
<td>HU</td>
<td>1.9</td>
<td>-0.8</td>
<td>-2.6</td>
<td>1.3</td>
<td>1.2</td>
<td>1.2</td>
<td>0.4</td>
<td>-0.8</td>
<td>3.3</td>
<td>2.5</td>
<td>-0.6</td>
<td>11.7</td>
<td>2.3</td>
<td>-1.5</td>
<td>7.1</td>
<td>-5.4</td>
</tr>
<tr>
<td>IE</td>
<td>0.9</td>
<td>1.0</td>
<td>-3.0</td>
<td>-1.1</td>
<td>0.7</td>
<td>3.0</td>
<td>1.3</td>
<td>-1.7</td>
<td>3.5</td>
<td>-7.0</td>
<td>-5.0</td>
<td>-1.3</td>
<td>-1.0</td>
<td>0.0</td>
<td>2.3</td>
<td>0.0</td>
</tr>
<tr>
<td>IS</td>
<td>0.9</td>
<td>3.1</td>
<td>-2.5</td>
<td>-1.3</td>
<td>-0.8</td>
<td>0.7</td>
<td>0.8</td>
<td>-4.3</td>
<td>-3.4</td>
<td>-3.6</td>
<td>-3.6</td>
<td>1.8</td>
<td>1.8</td>
<td>-3.6</td>
<td>1.8</td>
<td>-3.6</td>
</tr>
<tr>
<td>IT</td>
<td>0.3</td>
<td>2.2</td>
<td>1.1</td>
<td>-1.4</td>
<td>1.1</td>
<td>1.1</td>
<td>0.0</td>
<td>-4.2</td>
<td>-1.3</td>
<td>9.2</td>
<td>-4.3</td>
<td>-2.6</td>
<td>1.4</td>
<td>-2.3</td>
<td>1.6</td>
<td>1.8</td>
</tr>
<tr>
<td>LT</td>
<td>1.6</td>
<td>-2.5</td>
<td>-1.2</td>
<td>1.7</td>
<td>0.7</td>
<td>0.0</td>
<td>2.2</td>
<td>-5.9</td>
<td>2.1</td>
<td>2.4</td>
<td>-0.2</td>
<td>-6.6</td>
<td>2.0</td>
<td>5.8</td>
<td>2.0</td>
<td>-3.1</td>
</tr>
<tr>
<td>NL</td>
<td>2.4</td>
<td>-3.2</td>
<td>-3.2</td>
<td>0.7</td>
<td>2.0</td>
<td>1.9</td>
<td>1.9</td>
<td>-0.1</td>
<td>7.8</td>
<td>-0.6</td>
<td>-1.9</td>
<td>-3.6</td>
<td>0.2</td>
<td>2.6</td>
<td>0.3</td>
<td>0.6</td>
</tr>
<tr>
<td>NO</td>
<td>-2.5</td>
<td>-0.5</td>
<td>-2.7</td>
<td>-0.6</td>
<td>1.7</td>
<td>2.6</td>
<td>-0.4</td>
<td>3.8</td>
<td>4.2</td>
<td>1.8</td>
<td>-3.4</td>
<td>5.0</td>
<td>2.0</td>
<td>3.0</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>PL</td>
<td>0.0</td>
<td>0.5</td>
<td>-0.6</td>
<td>-1.2</td>
<td>-0.2</td>
<td>0.5</td>
<td>1.0</td>
<td>0.0</td>
<td>-0.4</td>
<td>1.9</td>
<td>-0.1</td>
<td>2.0</td>
<td>0.2</td>
<td>0.6</td>
<td>-0.4</td>
<td>-2.4</td>
</tr>
<tr>
<td>PT</td>
<td>7.0</td>
<td>-0.1</td>
<td>-2.6</td>
<td>-2.7</td>
<td>1.0</td>
<td>2.5</td>
<td>1.8</td>
<td>0.0</td>
<td>0.9</td>
<td>-7.7</td>
<td>-0.3</td>
<td>2.1</td>
<td>3.3</td>
<td>-0.4</td>
<td>-3.9</td>
<td>-1.0</td>
</tr>
<tr>
<td>SE</td>
<td>-0.3</td>
<td>-1.4</td>
<td>-0.9</td>
<td>-1.7</td>
<td>0.4</td>
<td>0.6</td>
<td>3.0</td>
<td>3.8</td>
<td>-0.6</td>
<td>-3.0</td>
<td>-3.0</td>
<td>8.6</td>
<td>-0.2</td>
<td>-3.8</td>
<td>1.6</td>
<td>-1.9</td>
</tr>
<tr>
<td>SI</td>
<td>3.2</td>
<td>0.7</td>
<td>-3.4</td>
<td>0.4</td>
<td>-1.1</td>
<td>2.0</td>
<td>0.8</td>
<td>0.7</td>
<td>0.7</td>
<td>-8.6</td>
<td>-0.2</td>
<td>-3.8</td>
<td>1.6</td>
<td>-1.9</td>
<td>0.6</td>
<td>3.6</td>
</tr>
<tr>
<td>SK</td>
<td>5.0</td>
<td>-7.1</td>
<td>-2.5</td>
<td>3.0</td>
<td>3.1</td>
<td>3.9</td>
<td>2.1</td>
<td>-2.4</td>
<td>10.1</td>
<td>7.8</td>
<td>0.4</td>
<td>0.1</td>
<td>5.3</td>
<td>1.2</td>
<td>-0.7</td>
<td>-5.9</td>
</tr>
<tr>
<td>UK</td>
<td>5.5</td>
<td>-2.1</td>
<td>-3.8</td>
<td>-1.2</td>
<td>0.7</td>
<td>2.6</td>
<td>2.7</td>
<td>1.2</td>
<td>6.9</td>
<td>-6.8</td>
<td>-3.7</td>
<td>-2.1</td>
<td>0.5</td>
<td>-0.1</td>
<td>-0.1</td>
<td>1.8</td>
</tr>
</tbody>
</table>

# sign. diff.

8+ / 1- 4+ / 6- 0+ / 15- 1+ / 2- 6+ / 0- 9+ / 0- 10+ / 0- 1+ / 6- 12+ / 2- 7+ / 6- 2+ / 16- 5+ / 10- 3+ / 1- 2+ / 1- 3+ / 3- 2+ / 5-

green = overrepresentation
red = underrepresentation
green/red dark colour = LFS estimate outside 95% CI of ESS estimate

12
External measure of misrepresentation

Index of dissimilarity for each variable

- \[ D = \frac{1}{2} \sum_{i}^{n} |ESS_i - LFS_i| \]

  (n = number of categories, 
  ESS_i = percentage in cat i of ESS, 
  LFS_i = percentage in cat i of LFS)

- Interpretation:
  % of persons who would have to change categories
  in order to achieve the same distribution in ESS and LFS

=> Average value of D across 6 variables
Index of dissimilarity (average across 6 variables)

ESS 6 (24 countries)

5 countries not included (no LFS data available)
Summary: Level/pattern of misrepresentation

Benchmarking with internal/external data:

- **Indications of misrepresentation** for a number of variables
- Misrepresentation is **not erratic**, certain **patterns** are discernible
- Level of misrepresentation **varies between countries**
- Question: **Correlates** of misrepresentation?
Part 2:

Correlates of misrepresentation of demographic groups in ESS samples

- Response rate
- Type of sample
Response rates and sample quality

- **‘Traditional’ view:** High RRs signal good sample quality
- **More recent research:** RRs and sample quality uncorrelated (Keeter et al. 2000; Groves/Peytcheva 2008)

![Response rates ESS 6 (in %, 24 countries)](chart)

Source: Beullens et al. 2014: Draft quality matrix for ESS 6
Type of sample and sample quality

Sampling methods using interviewers to select households/persons:

=> data sets that are less representative of the population (Kohler 2007, Sodeur 2007, Menold 2014)

=> ESS prefers samples of named individuals from a register

Sampling methods, acc. to interviewer involvement in sel. process

<table>
<thead>
<tr>
<th></th>
<th>Selection of household</th>
<th>Selection of person</th>
<th># of countries in ESS 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual register</td>
<td>No</td>
<td>No</td>
<td>14</td>
</tr>
<tr>
<td>Household register</td>
<td>No</td>
<td>Yes</td>
<td>7</td>
</tr>
<tr>
<td>Household walk</td>
<td>Yes</td>
<td>Yes</td>
<td>3</td>
</tr>
</tbody>
</table>
‘Dependent’ variables

- **Internal** measure:
  - Absolute value of \( \frac{\text{(%female} - 50)}{\sqrt{\frac{50 \times 50}{n}}} \)

- **External** measure:
  - Average index of dissimilarity across 6 variables,
    with \( D = \frac{1}{2} \sum^n \left| ESS_i - LFS_i \right| \)
Misrepresentation by response rates, ESS 6 (24 c.)

**Internal measure**

Pearson’s $r = .55$

**External measure**

Pearson’s $r = .07$

(gender only: $r = .36$)
Misrepresentation by type of sample, ESS 6

Internal measure

<table>
<thead>
<tr>
<th>Type of sample</th>
<th>Average misrepresentation (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual register</td>
<td>1.3 (14)</td>
</tr>
<tr>
<td>Household register</td>
<td>2.2 (7)</td>
</tr>
<tr>
<td>Household walk</td>
<td>3.3 (3)</td>
</tr>
</tbody>
</table>

External measure

<table>
<thead>
<tr>
<th>Type of sample</th>
<th>Average misrepresentation (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual register</td>
<td>3.0 (14)</td>
</tr>
<tr>
<td>Household register</td>
<td>4.6 (7)</td>
</tr>
<tr>
<td>Household walk</td>
<td>5.2 (3)</td>
</tr>
</tbody>
</table>
Misrepr. by response rate + sample type, ESS 1-6

n = 136 cases; 28 countries participating in 2 or more rounds of ESS

Source: Eckman/Koch: The relationship between response rates, sampling method and data quality: Evidence from the European Social Survey (under review)
Summary: Correlates of misrepresentation

- Sample quality lower when interviewers were involved in sample selection

- Response rate: Not a good indicator of sample quality
  - Sampling methods with high interviewer involvement: High response rates seem to indicate lower sample quality
  - Explanation: “Undocumented substitution“ (?)
    - Interviewers replace ‘difficult‘ sample units by more accessible/cooperative sample units
    => Response rates inflated
    => Larger misrepresentation of demographic groups
Conclusions / open questions

- Effect of bias in demographic variables on ‘key survey estimates’?

- Implications for fieldwork in future ESS rounds:
  - Targeted efforts for specific groups desirable/feasible?
  - Changes in interviewer training/back-checks recommendable?

- Use of post-stratification weights as a remedy?

- …
The end

(achim.koch@gesis.org)