Accuracy of poverty indicators derived from EU-SILC

The 1st European User Conference organized by German Microdata Lab, GESIS, in cooperation with Eurostat

Stefan Zins, Jan-Philipp Kolb and Ralf Münnich

Fachbereich IV der Universität Trier
Wirtschafts- und Sozialstatistik

Mannheim, 5. March 2009
Accuracy of poverty indicators derived from EU-SILC

D-SILC and relevant Laeken Indicators

Accuracy measurement and variance estimation

The Monte-Carlo Study

Summary and outlook
Accuracy of poverty indicators derived from EU-SILC

D-SILC and relevant Laeken Indicators

Accuracy measurement and variance estimation

The Monte-Carlo Study

Summary and outlook
Accuracy of poverty indicators derived from EU-SILC

D-SILC and relevant Laeken Indicators

Accuracy measurement and variance estimation

The Monte-Carlo Study

Summary and outlook
Accuracy of poverty indicators derived from EU-SILC

D-SILC and relevant Laeken Indicators

Accuracy measurement and variance estimation

The Monte-Carlo Study

Summary and outlook
Accuracy of poverty indicators derived from EU-SILC

D-SILC and relevant Laeken Indicators

Accuracy measurement and variance estimation

The Monte-Carlo Study

Summary and outlook
EU-SILC and the D-SILC

Microcensus

- 1% sample of the population in Germany (dwellings)
- 4 rotation quarters

Access panel

- Recruitment from latest rotation quarter
- Very different recruitment rates

D-SILC

- Stratified sample from Access panel
- Panel attrition, item non-response and unit non-response

(cf. Horneffer / Kuchler, 2008)
EU-SILC and the D-SILC

Microcensus

- 1% sample of the population in Germany (dwellings)
- 4 rotation quarters

Access panel

- Recruitment from latest rotation quarter
- Very different recruitment rates

D-SILC

- Stratified sample from Access panel
- Panel attrition, item non-response and unit non-response

(cf. Horneffer / Kuchler, 2008)
EU-SILC and the D-SILC

Microcensus
- 1% sample of the population in Germany (dwellings)
- 4 rotation quarters

Access panel
- Recruitment from latest rotation quarter
- Very different recruitment rates

D-SILC
- Stratified sample from Access panel
- Panel attrition, item non-response and unit non-response

(cf. Horneffer / Kuchler, 2008)
EU-SILC and the D-SILC

Microcensus
- 1% sample of the population in Germany (dwellings)
- 4 rotation quarters

Access panel
- Recruitment from latest rotation quarter
- Very different recruitment rates

D-SILC
- Stratified sample from Access panel
- Panel attrition, item non-response and unit non-response

(cf. Horneffer / Kuchler, 2008)
EU-SILC and the D-SILC

Microcensus
- 1% sample of the population in Germany (dwellings)
- 4 rotation quarters

Access panel
- Recruitment from latest rotation quarter
- Very different recruitment rates

D-SILC
- Stratified sample from Access panel
- Panel attrition, item non-response and unit non-response

(cf. Horneffer / Kuchler, 2008)
Laeken-Indicators

1. At-risk-of-poverty rate (subgroups)
2. Inequality of income distribution S80/S20 income quintile share ratio
3. At-persistent-risk-of-poverty rate (60% median)
4. Relative median at-risk-of-poverty gap
5. Regional cohesion (dispersion of regional employment rates)
6. Long term unemployment rate
7. Persons living in jobless households
8. Early school leavers not in education or training
9. Life expectancy at birth
10. Self defined health status by income quintile
11. Dispersion around the at-risk-of-poverty threshold
12. Inequality of income distribution GINI coefficient
13. At-persistent-risk-of-poverty rate (50% median)
14. Long term unemployment share
15. Very long-term unemployment rate
16. Persons with low educational attainment

many by gender or age and gender
Laeken-Indicators

1. At-risk-of-poverty rate (subgroups)
2. Inequality of income distribution S80/S20 income quintile share ratio
3. At-persistent-risk-of-poverty rate (60% median)
4. Relative median at-risk-of-poverty gap
5. Regional cohesion (dispersion of regional employment rates)
6. Long term unemployment rate
7. Persons living in jobless households
8. Early school leavers not in education or training
9. Life expectancy at birth
10. Self defined health status by income quintile
11. Dispersion around the at-risk-of-poverty threshold
12. Inequality of income distribution GINI coefficient
13. At-persistent-risk-of-poverty rate (50% median)
14. Long term unemployment share
15. Very long-term unemployment rate
16. Persons with low educational attainment

many by gender or age and gender
Poverty measures for EU-SILC (1)

At risk of poverty rate $0.6 \cdot z_{y_{disp}}$: At-risk-poverty rate

$$ARPR = \frac{\sum_i \mathbb{1}(y_{i,disp} < 0.6 \cdot z_{y_{disp}})}{N_{pers}}$$

many by gender or age and gender
Poverty measures for EU-SILC (2)

- Inequality of income distribution S8020 income quintile share ratio $y_{0.2}$ and $y_{0.8}$:

$$QSR = \frac{\sum_i y_i,\text{disp} \cdot \mathbb{1}(y_i,\text{disp} > y_{0.8})}{\sum_i y_i,\text{disp} \cdot \mathbb{1}(y_i,\text{disp} \leq y_{0.2})} \cdot \frac{\sum_i \mathbb{1}(y_i,\text{disp} \leq y_{0.2})}{\sum_i \mathbb{1}(y_i,\text{disp} > y_{0.8})}$$

- GINI coefficient (Lorenz curve)

$$\text{GINI} = 2 \cdot F$$
Poverty measures for EU-SILC (2)

- Inequality of income distribution S8020 income quintile share ratio $y_{0,2}$ and $y_{0,8}$:

$$QSR = \frac{\sum_i y_{i,\text{disp}} \cdot 1(y_{i,\text{disp}} > y_{0,8})}{\sum_i y_{i,\text{disp}} \cdot 1(y_{i,\text{disp}} \leq y_{0,2})} \cdot \frac{\sum_i 1(y_{i,\text{disp}} \leq y_{0,2})}{\sum_i 1(y_{i,\text{disp}} > y_{0,8})}$$

- GINI coefficient (Lorenz curve)

$$\text{GINI} = 2 \cdot F$$
Poverty measures for EU-SILC (2)

- Inequality of income distribution S8020 income quintile share ratio $y_{0,2}$ and $y_{0,8}$:

$$QSR = \frac{\sum_i y_{i,\text{disp}} \cdot 1(y_{i,\text{disp}} > y_{0,8})}{\sum_i y_{i,\text{disp}} \cdot 1(y_{i,\text{disp}} \leq y_{0,2})} \cdot \frac{\sum_i 1(y_{i,\text{disp}} \leq y_{0,2})}{\sum_i 1(y_{i,\text{disp}} > y_{0,8})}$$

- GINI coefficient (Lorenz curve)

\[ \text{GINI} = 2 \cdot F \]
Poverty measures for EU-SILC (2)

- Inequality of income distribution S8020 income quintile share ratio $y_{0.2}$ and $y_{0.8}$:

$$QSR = \frac{\sum_i y_{i,\text{disp}} \cdot 1(y_{i,\text{disp}} > y_{0.8})}{\sum_i y_{i,\text{disp}} \cdot 1(y_{i,\text{disp}} \leq y_{0.2})} \cdot \frac{\sum_i 1(y_{i,\text{disp}} \leq y_{0.2})}{\sum_i 1(y_{i,\text{disp}} > y_{0.8})}$$

- GINI coefficient (Lorenz curve)

$$\text{GINI} = 2 \cdot F$$
Accuracy of poverty indicators derived from EU-SILC

D-SILC and relevant Laeken Indicators

Accuracy measurement and variance estimation

The Monte-Carlo Study

Summary and outlook
Accuracy Measurement

European Statistics Code of Practice (2005)

- Comparability of statistics
- Relevance of statistical concepts
- Completeness
- Accuracy of estimations
  - Sampling error: standard error, confidence interval, $n_{eff}$
  - Non sampling error: non-response, frame error

Variance estimation is an essential basis for accuracy measurement (asymptotic unbiased estimators)
Methods for Variance Estimation

- **Exact Methods**
  - Simple Random Sampling
  - Stratified Random Sampling

- **Approximation Methods**
  - Linearization of non-linear estimators
  - Approximation of second order IIP’s

- **Resampling-Methods**
  - Random Groups
  - Balanced Repeated Replication
  - Jackknife-Methods
  - Bootstrap-Methods
Balanced Repeated Replication

- Basic model with two observations per stratum
- The units in each stratum are divided randomly into two groups
- $\hat{\theta}_r$ is the $r$-th estimation of a sample, where one half sample is chosen in each $H$ stratum
- Balanced sample of $R \ll 2^H$
- BRR estimators:

$$\hat{\theta}_{BRR} = \frac{1}{R} \cdot \sum_{r=1}^{R} \hat{\tau}_r$$

$$\hat{V}_{BRR}(\hat{\theta}) = \frac{1}{R} \sum_{r=1}^{R} (\hat{\theta}_r - \hat{\theta})^2.$$
Jackknife variance estimation

- Delete 1 jackknife:
  - $\hat{\theta}_{-i}$ is the $i$-th estimation, where $i$-th element of the sample is deleted
  
  $$\hat{V}_{d1JK}(\hat{\theta}) = \frac{n-1}{n} \sum_{i \in S} \left( \hat{\theta}_{-i} - \frac{1}{n} \sum_{j \in S} \hat{\theta}_j \right)^2.$$

- Delete $d$ jackknife:
  - $\hat{\theta}_{-i,d}$ is the $i$-th estimation, where $d$ elements of the sample are deleted
  
  $$\hat{V}_{ddJK}(\hat{\theta}) = \frac{n-d}{d \cdot m} \sum_{i=1}^{m} \left( \hat{\theta}_{-i,d} - \frac{1}{m} \sum_{j=1}^{m} \hat{\theta}_{-j,d} \right)^2.$$

Mannheim, 5.-6. March 2009  Stefan Zins  Accuracy of poverty indicators 12 (23)
Bootstrap methods

- Theoretic bootstrap
- Monte Carlo bootstrap:
  Sample of size $n$ model with replacement

$$\hat{V}_{\text{Boot,MC}} = \frac{1}{B - 1} \sum_{i=1}^{B} \left( \hat{\theta}^*_{n,i} - \frac{1}{B} \sum_{j=1}^{B} \hat{\theta}^*_{n,j} \right)^2 .$$

- Sophisticated weighting for without replacement sampling
- Adjustment for complex designs
Accuracy of poverty indicators derived from EU-SILC

D-SILC and relevant Laeken Indicators

Accuracy measurement and variance estimation

The Monte-Carlo Study

Summary and outlook
The simulation study

- Synthetic universe with 16 federal states (Microcensus 2004)
- Additional stratification variables: Household income (continuous, categorized with 5 categories), social status head of the household.
- Comparison between a full response model (FM) and a model with response propensities (RPM)
- Response is generated by a logit model, with the covariates:
  - Federal state
  - Household type
  - Social status head of the household
  - Household net income
For further details see:

For further details see:

For further details see:

For further details see:

D-SILC and relevant Laeken Indicators
Accuracy measurement and variance estimation
The Monte-Carlo Study
Summary and outlook

Mannheim, 5.-6. March 2009  Stefan Zins  Accuracy of poverty indicators 17 (23)
D-SILC and relevant Laeken Indicators
Accuracy measurement and variance estimation
The Monte-Carlo Study
Summary and outlook

Mannheim, 5.-6. March 2009 Stefan Zins
Accuracy of poverty indicators 18 (23)
D-SILC and relevant Laeken Indicators
Accuracy measurement and variance estimation
The Monte-Carlo Study
Summary and outlook

GINI
boot
brr
ddJK
lin
0e+00 2e−05 4e−05 6e−05
●
●
●
●
●●
● ● ●● ●● ● ●● ●●●● ●● ●● ●●● ●● ●● ●●●● ●● ●● ●●●●
●●●●●●●●●

R8020

ARPR

Mannheim, 5.-6. March 2009 Stefan Zins

Accuracy of poverty indicators
Accuracy of poverty indicators derived from EU-SILC

D-SILC and relevant Laeken Indicators

Accuracy measurement and variance estimation

The Monte-Carlo Study

Summary and outlook
Summary and outlook

- Accuracy measurement
  - Approximate unbiased point estimators for GINI and R8020
  - Good variance estimates for GINI and R8020 with BRR and bootstrap routines
  - ARPR needs some further improvements

- End-user needs for variance estimation
  - Regional information in micro data in general not available
  - Inclusion of replicate weights into datasets

- Investigation of alternative methods
  - Parametric estimation
  - Robust methods
  - Small area estimation
Summary and outlook

- **Accuracy measurement**
  - Approximate unbiased point estimators for GINI and R8020
  - Good variance estimates for GINI and R8020 with BRR and bootstrap routines
  - ARPR needs some further improvements

- **End-user needs for variance estimation**
  - Regional information in micro data in general not available
  - Inclusion of replicate weights into datasets

- **Investigation of alternative methods**
  - Parametric estimation
  - Robust methods
  - Small area estimation
Summary and outlook

- Accuracy measurement
  - Approximate unbiased point estimators for GINI and R8020
  - Good variance estimates for GINI and R8020 with BRR and bootstrap routines
  - ARPR needs some further improvements

- End-user needs for variance estimation
  - Regional information in micro data in general not available
  - Inclusion of replicate weights into datasets

- Investigation of alternative methods
  - Parametric estimation
  - Robust methods
  - Small area estimation
References


