Universität Trier

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

Universität Trier

Accuracy of poverty indicators derived from EU-SILC

Universität Trier

Accuracy of poverty indicators derived from EU-SILC

Universität Trier

Accuracy of poverty indicators derived from EU-SILC

Universität Trier

Accuracy of poverty indicators derived from EU-SILC

Universität Trier

Accuracy of poverty indicators derived from EU-SILC

Universität Trier

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)

Universität Trier

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)

Universität Trier

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)

Universität Trier

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)

Universität Trier

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)

Universität Trier

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

Universität Trier

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

Universität Trier

Poverty measures for EU-SILC (1)

At risk of poverty rate $0{,}6 \cdot z_{y_{\text{disp}}}$: At-risk-poverty rate

$$\text{ARPR} = \frac{\sum_i \mathbb{1}(y_{i,\text{disp}} < 0{,}6 \cdot z_{y_{\text{disp}}})}{N_{\text{pers}}}$$

many by gender or age and gender

Universität Trier

Poverty measures for EU-SILC (2)

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

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

- GINI coefficient (Lorenz curve)

$$\mathrm{GINI} = 2 \cdot F$$

Universität Trier

Poverty measures for EU-SILC (2)

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

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

- GINI coefficient (Lorenz curve)

$$\mathrm{GINI} = 2 \cdot F$$

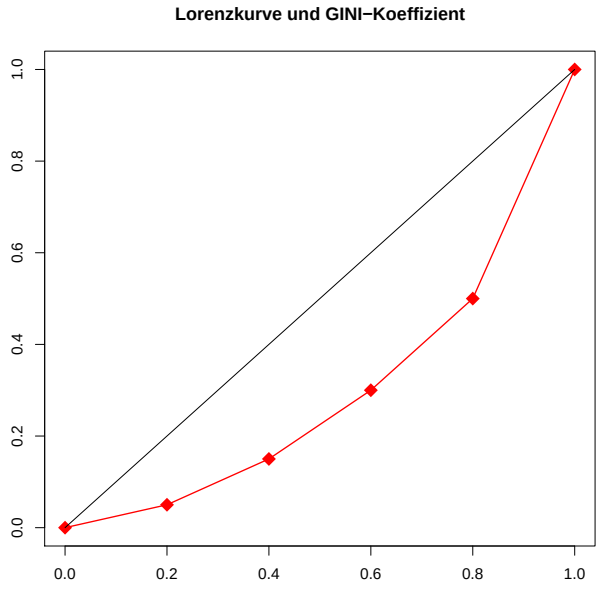
Universität Trier

Poverty measures for EU-SILC (2)

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

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

- GINI coefficient (Lorenz curve)



$$\mathrm{GINI} = 2 \cdot F$$

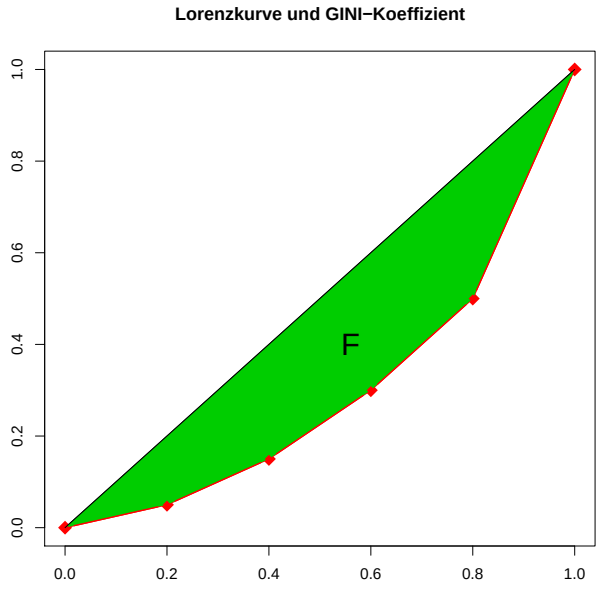
Universität Trier

Poverty measures for EU-SILC (2)

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

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

- GINI coefficient (Lorenz curve)



$$\text{GINI} = 2 \cdot F$$

Universität Trier

Accuracy of poverty indicators derived from EU-SILC

Universität Trier

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)

Universität Trier

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

Universität Trier

Balanced Repeated Replication

- Basic model with two observations per stratum
- The units in each stratum are divided randomly into two groups
- $\widehat{\theta}_r$ 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:

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

$$\widehat{V}_{\text{BRR}}\left(\widehat{\theta}\right) = \frac{1}{R} \sum_{r=1}^{R} \left(\widehat{\theta}_r - \widehat{\theta}\right)^2 \qquad .$$

Universität Trier

Jackknife variance estimation

- Delete 1 jackknife:
- $\widehat{\theta}_{-i}$ is the i-th estimation, where i-th element of the sample is deleted

$$\widehat{\mathrm{V}}_{\mathrm{d1JK}}\left(\widehat{\theta}\right) = \frac{n-1}{n} \sum_{i \in \mathcal{S}} \left(\widehat{\theta}_{-i} - \frac{1}{n} \sum_{j \in \mathcal{S}} \widehat{\theta}_{-j} \right)^2 \quad .$$

- Delete d jackknife:
- $\widehat{\theta}_{-i,d}$ is the i-th estimation, where d elements of the sample are deleted

$$\widehat{\mathrm{V}}_{\mathrm{ddJK}}\left(\widehat{\theta}\right) = \frac{n-d}{d \cdot m} \sum_{i=1}^{m} \left(\widehat{\theta}_{-i,d} - \frac{1}{m} \sum_{j=1}^{m} \widehat{\theta}_{-j,d} \right)^2 \quad .$$

Universität Trier

Bootstrap methods

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

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

- Sophisticated weighting for without replacement sampling
- Adjustment for complex designs

Universität Trier

Accuracy of poverty indicators derived from EU-SILC

Universität Trier

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

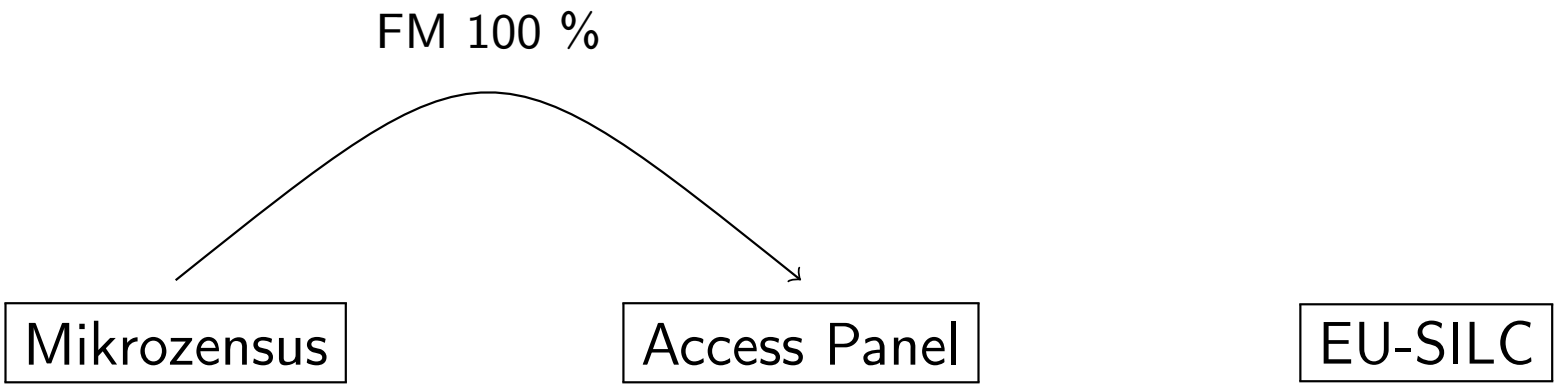
Universität Trier

Mikrozensus | Access Panel | EU-SILC

For further details see:

Münnich, Knobelspieß and Ohly (2006): Weiterführenden stichprobenmethodische Schätzungen im Rahmen des Aufbaus einer Dauerstichprobe befragungsbereiter Haushalte.

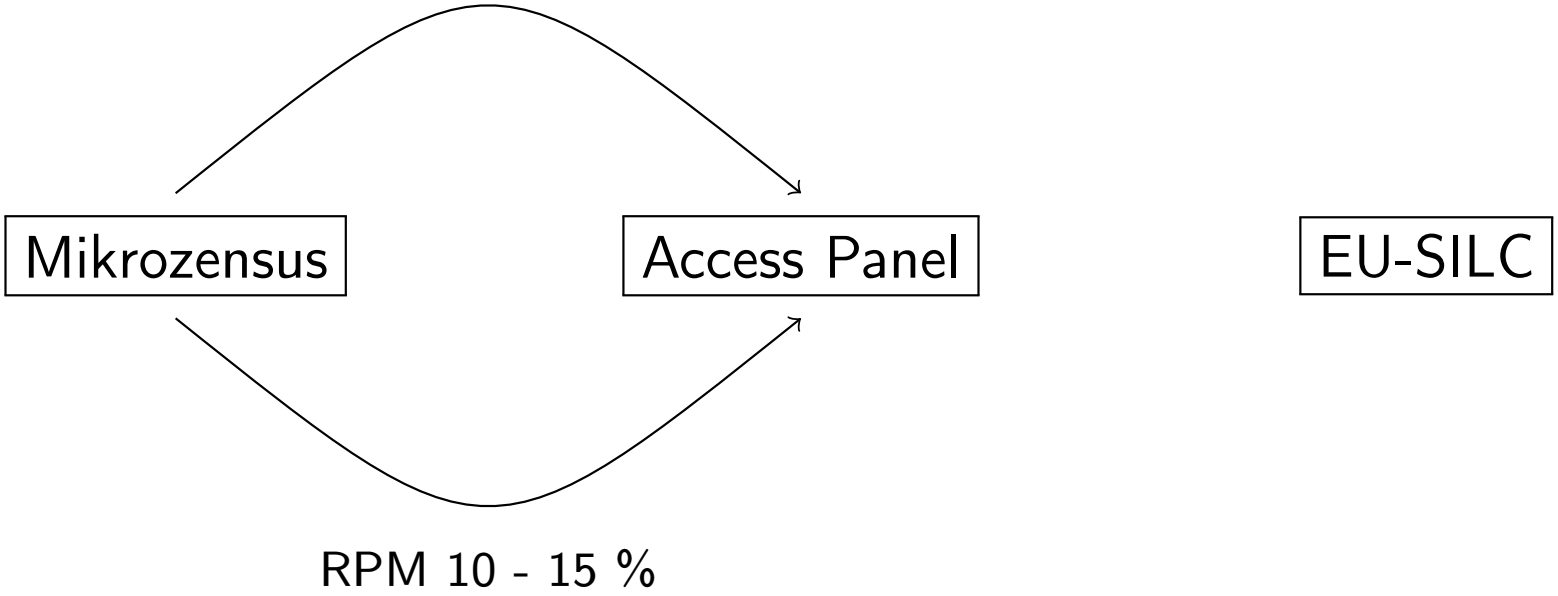
Universität Trier



For further details see:

Münnich, Knobelspieß and Ohly (2006): Weiterführenden stichprobenmethodische Schätzungen im Rahmen des Aufbaus einer Dauerstichprobe befragungsbereiter Haushalte.

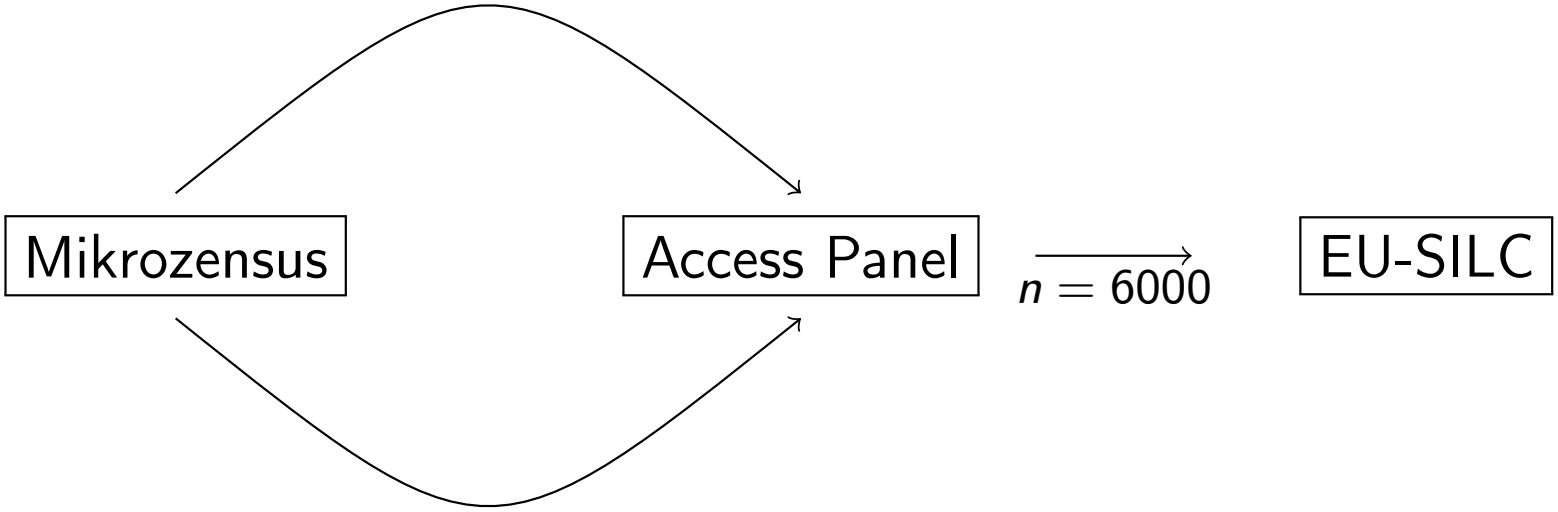
Universität Trier



For further details see:

Münnich, Knobelspieß and Ohly (2006): Weiterführenden stichprobenmethodische Schätzungen im Rahmen des Aufbaus einer Dauerstichprobe befragungsbereiter Haushalte.

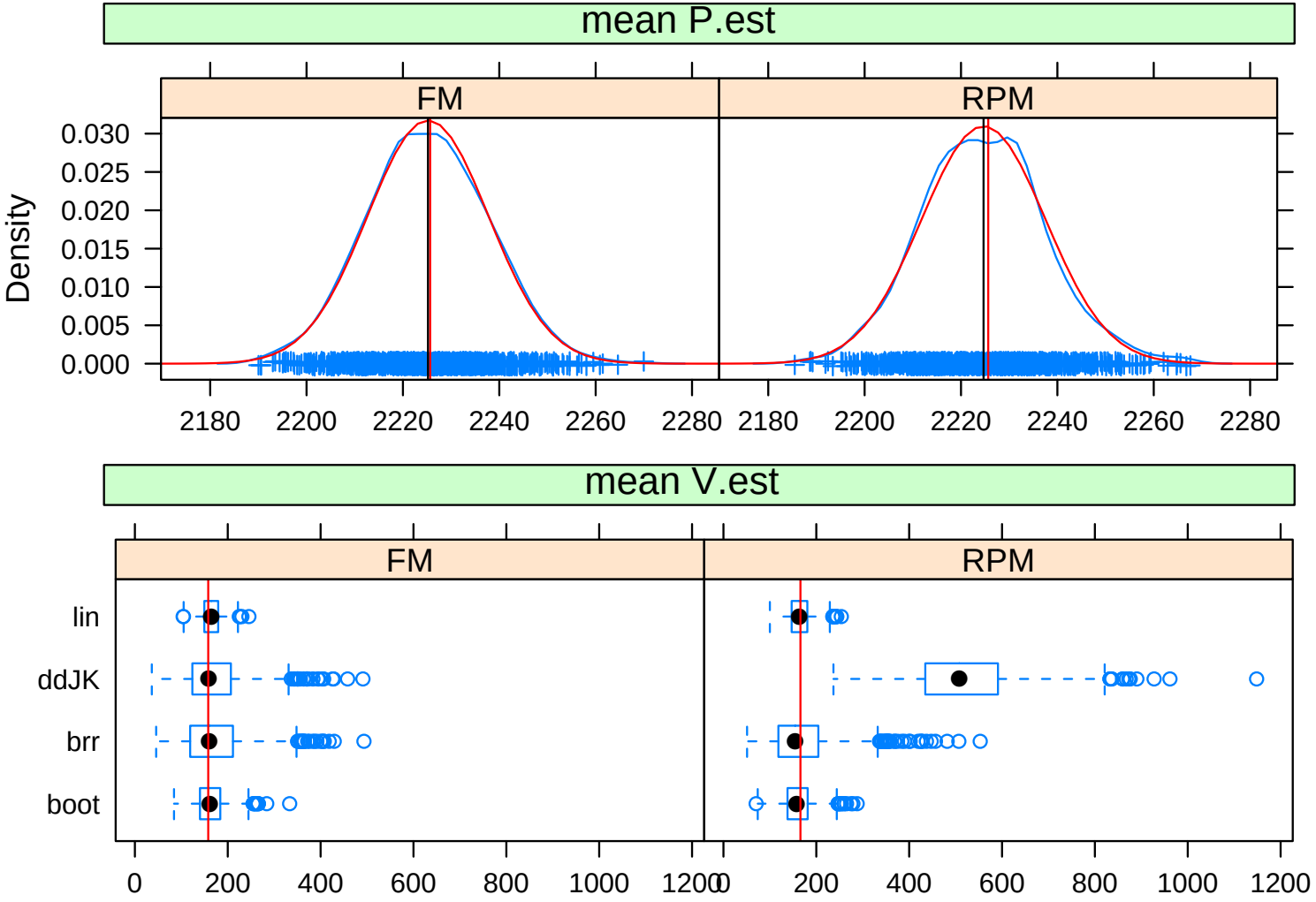
Universität Trier



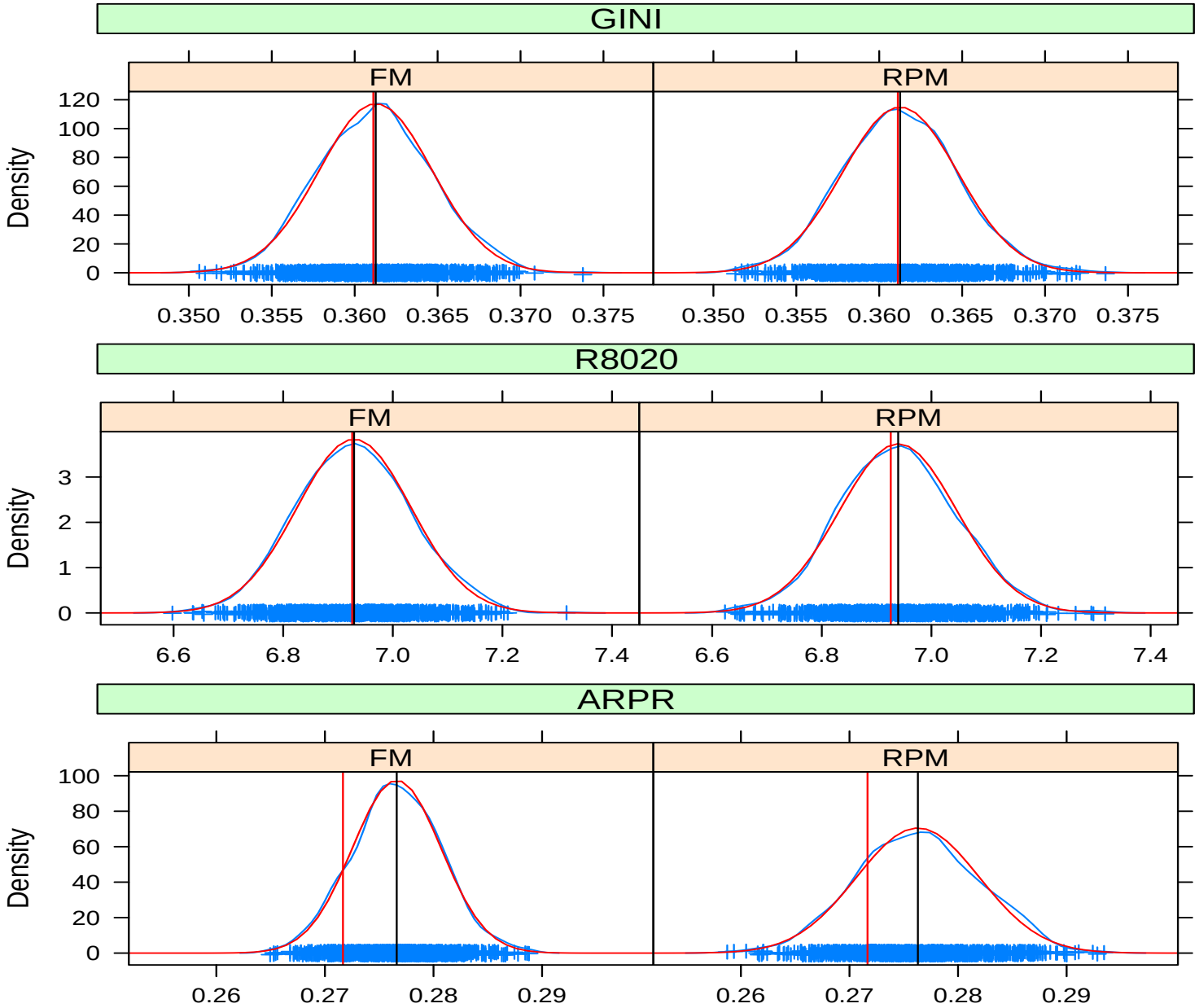
For further details see:

Münnich, Knobelspieß and Ohly (2006): Weiterführenden stichprobenmethodische Schätzungen im Rahmen des Aufbaus einer Dauerstichprobe befragungsbereiter Haushalte.

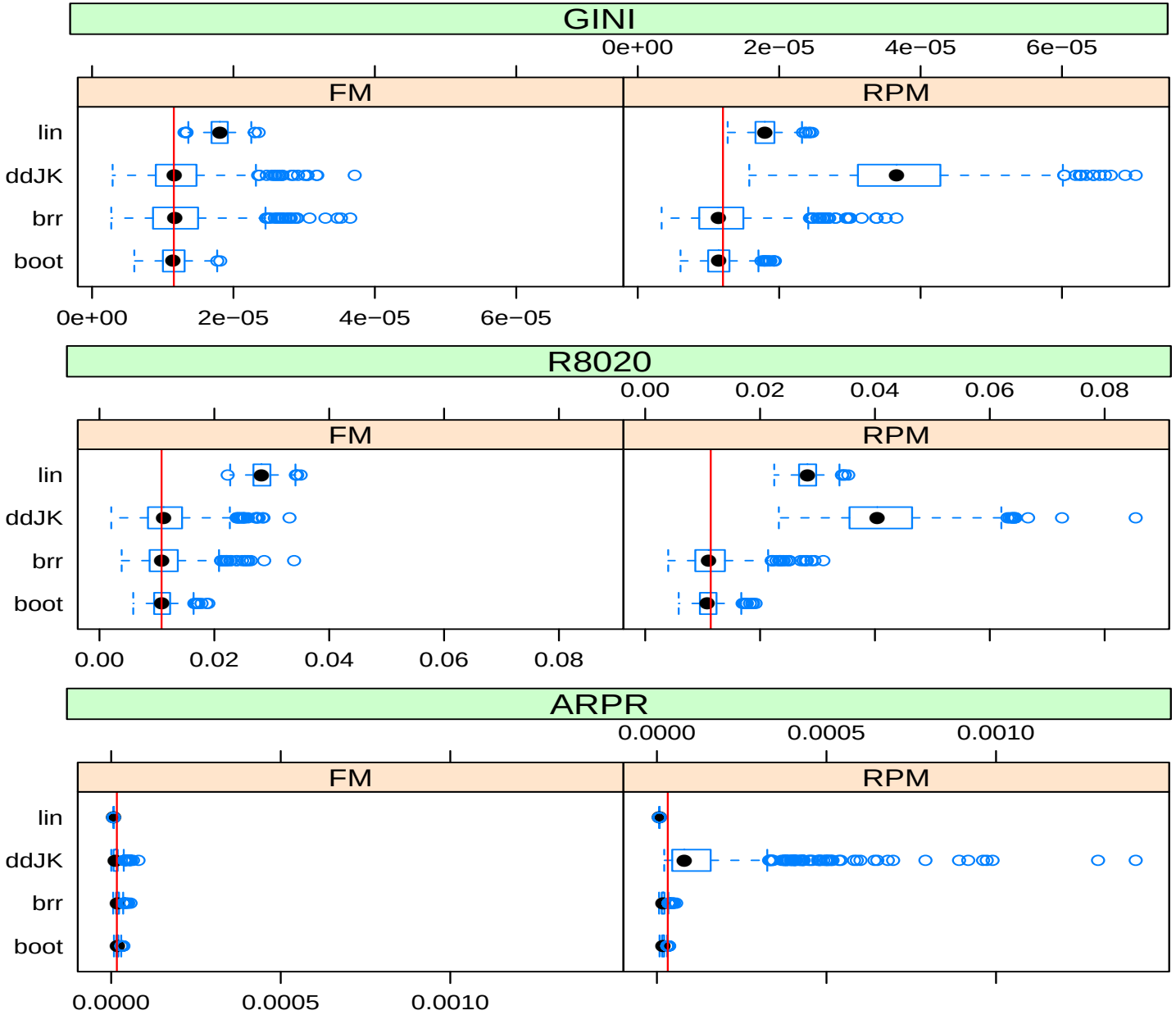
Universität Trier



Universität Trier



Universität Trier



Universität Trier

Accuracy of poverty indicators derived from EU-SILC

Universität Trier

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

Universität Trier

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

Universität Trier

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

Universität Trier

AMELI
Advanced Methodology for European Laeken Indicators

http://ameli.surveystatistics.net
DFG Priority Programme 1292 Survey Methodology

RHINE RIVER CRUISE 09
Recent Advances in Small Area Estimation

RRC09 Recent Advances in Small Area Estimation
Colloquium on small area statistics in Germany
at the Upper Middle Rhine World Heritage
22nd and 23rd June 2009

http://rrc09.surveystatistics.net
http://www.statistik.uni-trier.de

Universität Trier

References

- Andersson, Clæs; Nordberg, Lennart (1994): A Method for Variance Estimation of Non-Linear Functions of Totals in Surveys – Theory and Sorftware Implementation. In: Journal of Official Statistics, 10 (4), p. 395 – 405.
- Berger, Yves; Skinner, Christopher (2003): Variance estimation for a low income proportion. Applied Statistics, 52 (4), p. 457 – 468.
- Binder, David A.; Kovačević, M. S. (1995): Estimatin some measures of income inequality from survey data: An application of the estimating equation approach. In: Survey Methodology, 21, p. 137 – 145.
- Dennis, Ian; Guio, Anne-Catherine (2004): Poverty and social exlusion in the EU, Statistics in Focus 16/2004, Eurostat, Luxembourg.
- Deville, Jean-Claude (1999): Variance Estimation for Complex Statistics and Estimators: Linearization and Residual Techniques. In: Survey Methodology, 25 (2), p. 193 – 203.
- Horneffer, Birgit; Kuchler, Birgit (2008): Drei Jahre Panelerhebung EU-SILC: Erfahrungen und methodische Weiterentwicklungen. In: Wirtschaft und Statistik, 8/2008, p. 650 – 661.