Universität Trier

Variance Estimation for Measures of Change

S.A.M.P.L.E. CONFERENCE

Stefan Zins and Ralf Münnich

University of Trier, Faculty IV, VWL
Economic and Social Statistics

Siena, Wednesday, 06. October 2010

Universität Trier

Evolution of Measures of Poverty and Income Inequality Indicators in the European Union

Advanced Methodology for European Laeken Indicators

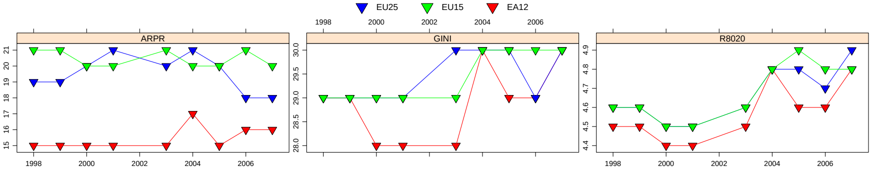
- Project is funded by the European Commission within the seventh Framework Programme
- Social Sciences and Humanities
  Area 6.2 – Developing better indicators for policy
- DG RTD in cooperation with DG ESTAT
- Project officer: Dr. Ian Perry
- EC contribution 1.089 M€

Co-ordinator: Ralf Münnich (muennich@uni-trier.de)
Homepage: http://ameli.surveystatistics.net

Universität Trier

Aim of EU-SILC

To monitor the process towards agreed policy goals we are interested in the evolution of social indicators.



Source: Eurostat, 2009, http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database

- Reading naively point estimator tables may lead to over-interpret the data.
- Was the change (in time) of an indicator value significant or not?
- How to *measure* significant changes of ARPR, GINI, and QSR?

Universität Trier

Two Problems arise:

1. The statistics in question (the Laeken indicators) are highly non-linear.
   - Basic variance estimation formulas cannot be applied directly.
2. The Surveys used to estimate the indicator values (EU-SILC) are often time dependent.
   - The correlation through time between indicators has to be taken into account.

Dell and d'Haultfoeuille (2007)

Goga, Deville, and Ruiz-Gazen (2009, Biometrika)

Universität Trier

Two Problems arise:

1. The statistics in question (the Laeken indicators) are highly non-linear.
   - Basic variance estimation formulas cannot be applied directly.
2. The Surveys used to estimate the indicator values (EU-SILC) are often time dependent.
   - The correlation through time between indicators has to be taken into account.

Dell and d'Haultfoeuille (2007)

Goga, Deville, and Ruiz-Gazen (2009, Biometrika)

Universität Trier

Two Problems arise:

1. The statistics in question (the Laeken indicators) are highly non-linear.
   - Basic variance estimation formulas cannot be applied directly.
2. The Surveys used to estimate the indicator values (EU-SILC) are often time dependent.
   - The correlation through time between indicators has to be taken into account.

Dell and d'Haultfoeuille (2007)

Goga, Deville, and Ruiz-Gazen (2009, Biometrika)

Universität Trier

Two Problems arise:

1. The statistics in question (the Laeken indicators) are highly non-linear.
   - Basic variance estimation formulas cannot be applied directly.
2. The Surveys used to estimate the indicator values (EU-SILC) are often time dependent.
   - The correlation through time between indicators has to be taken into account.

Dell and d'Haultfoeuille (2007)

Goga, Deville, and Ruiz-Gazen (2009, Biometrika)

Universität Trier

Two Problems arise:

1. The statistics in question (the Laeken indicators) are highly non-linear.
    - Basic variance estimation formulas cannot be applied directly.
2. The Surveys used to estimate the indicator values (EU-SILC) are often time dependent.
    - The correlation through time between indicators has to be taken into account.

Dell and d'Haultfoeuille (2007)

Goga, Deville, and Ruiz-Gazen (2009, Biometrika)

Universität Trier

Two Problems arise:

1. The statistics in question (the Laeken indicators) are highly non-linear.
   - Basic variance estimation formulas cannot be applied directly.
2. The Surveys used to estimate the indicator values (EU-SILC) are often time dependent.
   - The correlation through time between indicators has to be taken into account.

Dell and d'Haultfoeuille (2007)

Goga, Deville, and Ruiz-Gazen (2009, Biometrika)

Variance estimation for non-linear statistics

- Resampling methods
  Kovačević and Yung (1997)
  - Balanced repeated replication
  - Jackknife
  - Bootstrap
- Linearization methods
  - Taylor's method
  - Woodruff linearization
    Woodruff (1971) or Andersson and Nordberg (1994)
  - Estimating equations
    Kovačević and Binder (1997)
  - Influence functions
    Deville (1999)

Application to poverty and inequality indicators

Using the linearized values for the statistics ARPR, GINI, and QSR to approximated there variance:

$$\mathsf{V}(\widehat{\mathcal{I}}) \approx \mathsf{V}\Big(\sum_i \frac{1}{\pi_i} \cdot u_i\Big)$$

If the weights used in estimating $\mathcal{I}$ are obtained by a calibration of design weights, u_i are the residuals of the regression of the linearized values on the auxiliary variables used in the calibration, (cf. Deville, 1999).

Indicator $\mathcal{I}$	Source
ARPR:	Deville (1999), Osier (2009)
GINI:	Kovačević and Binder (1997)
QSR:	Hulliger and Münnich (2007)

Universität Trier

Ratio in time

$$\widehat{V}(\widehat{\mathcal{R}}_{\mathcal{I}}) = \widehat{V}(\widehat{\mathcal{I}}_{t_1}/\widehat{\mathcal{I}}_{t_0})$$
$$= \frac{1}{\widehat{\mathcal{I}}_{t_0}^2} \cdot \left(\widehat{\mathcal{R}}_{\mathcal{I}}^2 \cdot \widehat{V}(\widehat{\mathcal{I}}_{t_0}) + \widehat{V}(\widehat{\mathcal{I}}_{t_1}) - 2 \cdot \widehat{\mathcal{R}}_{\mathcal{I}} \cdot \widehat{Cov}(\widehat{\mathcal{I}}_{t_0}, \widehat{\mathcal{I}}_{t_1})\right)$$

Covariance estimation for non-linear statistics

$$\widehat{Cov}(\widehat{\mathcal{I}}_{t_0}, \widehat{\mathcal{I}}_{t_1}) = \widehat{Cov}\Big(\sum_{i \in S_{t_0}} \frac{u_i}{\pi_i}, \sum_{j \in S_{t_1}} \frac{u_j}{\pi_j}\Big)$$
$$= \sum_{i \in S_{t_0}} \sum_{j \in S_{t_1}} \left(1 - \frac{\pi_i \cdot \pi_j}{\pi_{ij}^*}\right) \cdot \frac{u_i}{\pi_i} \cdot \frac{u_j}{\pi_j}$$

$\pi_i = P(i \in S_{t.})$; $\quad \pi_{ij}^* = P(i \in S_{t_0}, j \in S_{t_1})$

Rotational samples in EU-SILC

Survey Year					
Y – 1	S_4^{Y-4}	S_3^{Y-3}	S_2^{Y-2}	S_1^{Y-1}	
Y – 0		S_4^{Y-3}	S_3^{Y-2}	S_2^{Y-1}	S_1^{Y-0}
	Y – 4	Y – 3	Y – 2	Y – 1	Y – 0

Selection Year

Actual Sampling Plan The population is partitioned into a rotational panel with 4 rotation groups (quarters).

- S_{CS}^{Y-1} A stratified sample drawn independently from groups U^{Y-4}, U^{Y-3}, U^{Y-2}, and U^{Y-1}.
- S_{CS}^{Y-0} A stratified sample drawn from U^{Y-4} plus the units in S_{CS}^{Y-1} without the units in S_4^{Y-4}, (assumes a static population).

Households as PSUs

Rotational samples in EU-SILC

Survey Year					
Y – 1	S_4^{Y-4}	S_3^{Y-3}	S_2^{Y-2}	S_1^{Y-1}	
Y – 0		S_4^{Y-3}	S_3^{Y-2}	S_2^{Y-1}	S_1^{Y-0}
	Y – 4	Y – 3	Y – 2	Y – 1	Y – 0

Selection Year

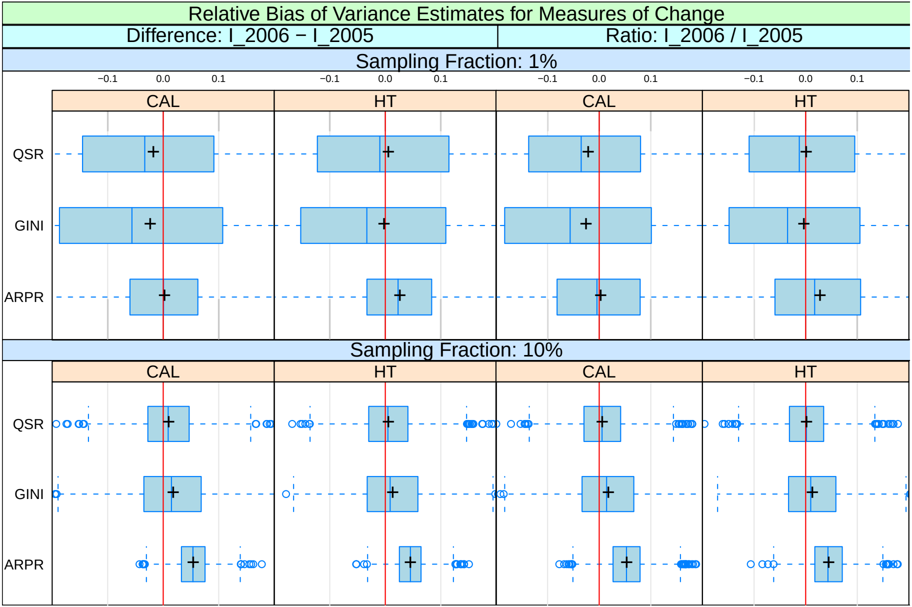
Actual Sampling Plan The population is partitioned into a rotational panel with 4 rotation groups (quarters).

S_{CS}^{Y-1} A stratified sample drawn independently from groups U^{Y-4}, U^{Y-3}, U^{Y-2}, and U^{Y-1}.

S_{CS}^{Y-0} A stratified sample drawn from U^{Y-4} plus the units in S_{CS}^{Y-1} without the units in S_4^{Y-4}, (assumes a static population).

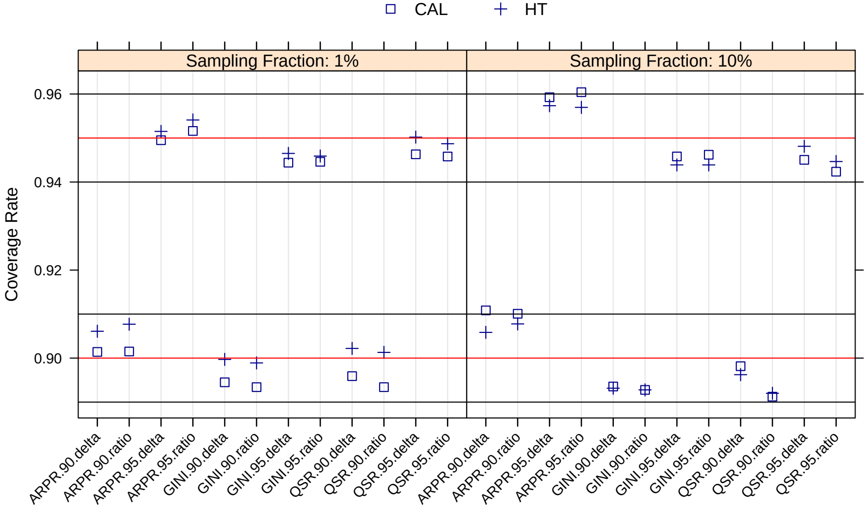
Households as PSUs

Universität Trier

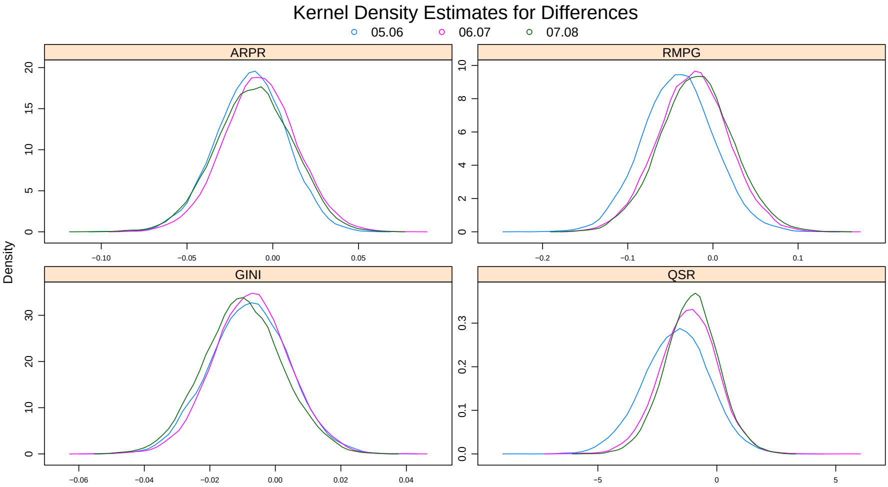


Universität Trier

Coverage Rates for the Different Sampling Fractions

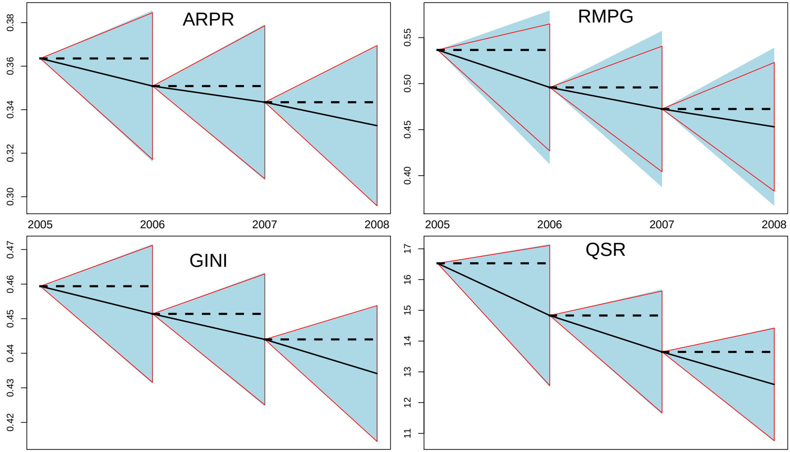


Universität Trier



Universität Trier

Test of Significant Change: $H_0 : \Delta = \Delta_0 = 0$



Hulliger (2005), Displays of indicators and of their accuracy, Conference on Visualising and Presenting Indicator Systems

Summary and outlook

- ARPR is less sensitive towards skewed distributions but more tends to be biased (density estimtaion)
- GINI and QSR are relatively non-robust against very skewed distributions
- Next steps
  - Non-linear calibration (on GINI or quantiles)
  - Estimation of the covariance between estimated totals in more complex dependent sampling surveys (Berger, 2004)
  - Introduction of robust methods for GINI and QSR