Robust estimation of the quintile share ratio with bias reduction

Beat Hulliger and Tobias Schoch
University of Applied Sciences Northwestern Switzerland
FHNW

Swiss Statistics Meeting, 30 October 2009, Geneva

Advanced Methodology for European Laeken Indicators

SEVENTH FRAMEWORK PROGRAMME

Contents

Laeken indicators

- European statistical **indicators on poverty and social exclusion**
- Lisbon Strategy: coordination of European social policies
- European Statistics on Income and Living Condition (EU-SILC): Coordinatet surveys across European Countries
- Survey methodology not main concern in development of the Laeken indicators: AMELI

Lisbon strategy (2000): "to become the most competitive and dynamic knowledge-based economy in the world capable of sustainable economic growth with more and better jobs and greater social cohesion"

AMELI

- Advanced Methodology for European Laeken Indicators
- Reserach project of 2008-2011 EU FP7 SSH programme.
- Methodology: robustness, small area estimation, variance estimation, visualisation.
- Development and evaluation by simulation.
- 10 Partners (4 Universities, 6 NSI)
- April 2008 - March 2011
- 1.4 M€

Quintile Share Ratio

Inequality of income distribution - Income quintile share ratio

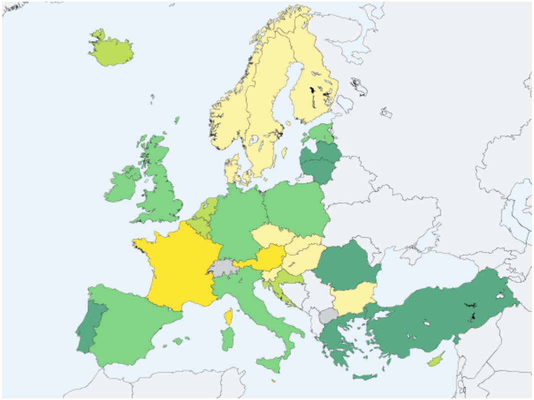
Short description: The ratio of total income received by the 20% of the population with the highest income (top quintile) to that received by the 20% of the population with the lowest income (lowest quintile). Income must be understood as equivalised disposable income.

QSR is much simpler to explain than Gini!

Official figures from Eurostat

Inequality of income distribution

Income quintile share ratio



Legend (Data 2007)

3.3 - 3.7
3.7 - 3.8
3.8 - 4.6
4.6 - 5.5
5.5 - 9.9
N/A

Exceptions: HR, TR(2003)

Minimum value:3.3 Maximum value:9.9 eu25:4.8 eu15:4.9

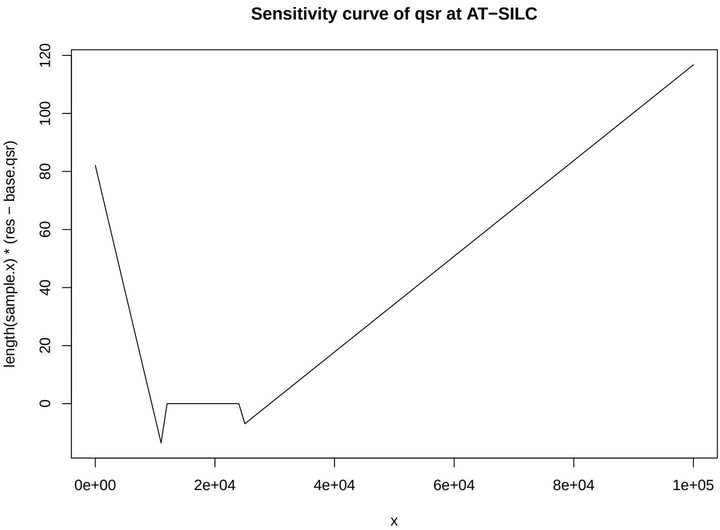
QSR

Estimand

$$\eta(F_U) = \frac{\sum_{i \in U} y_i \, \mathbf{1}\{Q_{0.8}(F_U) \leq y_i \leq Q_1(F_U)\}}{\sum_{i \in U} y_i \, \mathbf{1}\{Q_0(F_U) \leq y_i \leq Q_{0.2}(F_U)\}}$$

Estimator: $\hat{\eta} = \eta(F_S)$

- QSR has a breakdown point of 0: Moving a single observation any desired positive value can be produced.
- The influence function of the QSR is unbounded above.
- The variability of QSR is high.


Sensitivity curve of qsr at AT−SILC
length(sample.x) * (res − base.qsr)
0
20
40
60
80
100
120
0e+00
2e+04
4e+04
6e+04
8e+04
1e+05
x

i.i.d. observations: some Notation

Let $\alpha, \beta \in [0,1]$, $\alpha < \beta$ and $n\alpha, n\beta \in \{1, \ldots, n\}$
Trimmed mean

$$T_{\alpha;\,\beta}(F_n) := \frac{1}{n(\beta-\alpha)} \sum_{i=n\alpha+1}^{n\beta} X_{(i)}$$

The trimmed mean is an L-functional:

$$T_{\alpha;\,\beta}(F) = \int_0^1 F^{-1}(y) \frac{\mathbf{1}\{\alpha \le y \le \beta\}}{\beta-\alpha} dy$$

Robustification

- Parametric modelling and classical robustification (e.g. M-estimation)
- Parametric modeling of the tail above a threshold
- Trimming of extreme observations (Cowell and Victoria-Feser)
- AMELI: Nonparametric Robustification of QSR under complex sampling (Hulliger and Schoch 2009)

Trimming of extremes

Extreme observations are (symmetrically) trimmed from both tails. Upper (α_u) and lower (α_l) trimming proportions are equal.

$$\hat{\eta}_{TQSR}(\alpha_l, \alpha_u) = \frac{T_{0.8;\,1-\alpha_u}(F)}{T_{\alpha_l;\,0.2}(F)}$$

Both trimmings move TQSR down: High Bias!

Bias compensation

Trimming both quintiles above

$$\hat{\eta}_{BQSR}(\alpha_l, \alpha_u) = \frac{T_{0.8;\,1-\alpha_u}(F)}{T_{0;\,0.2-\alpha_l}(F)}$$

The trimming of numerator and denominator compensate at least partialley.

For a given (small) upper trimming proportion α_u there is a α_l such that BQSR is unbiased:

$$\hat{\eta}_{BQSR}\left(\alpha_l(\alpha_u), \alpha_u\right) = \hat{\eta}$$

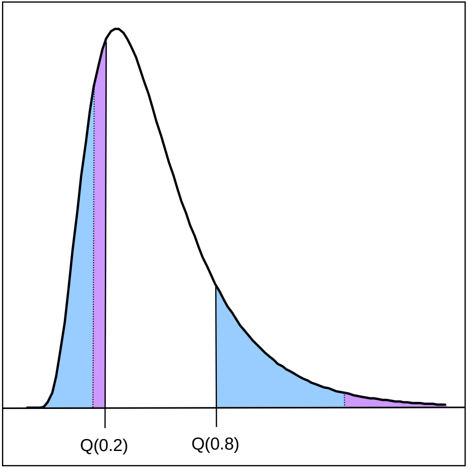
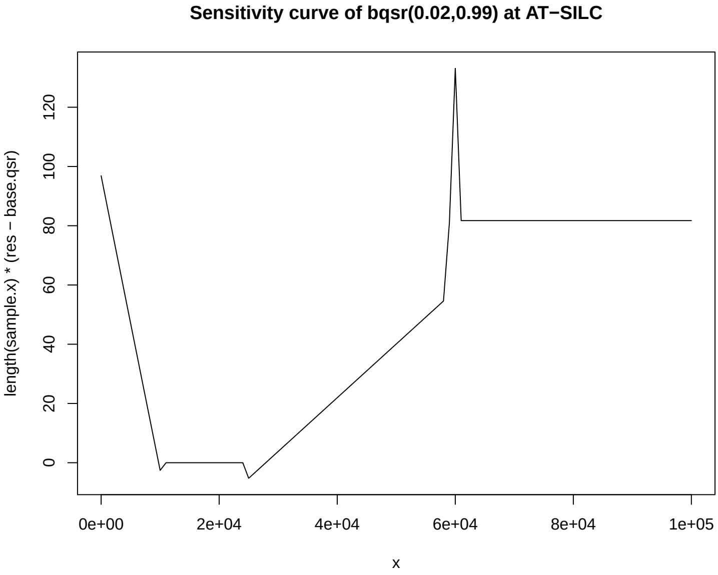


Figure: Idealized density function


Sensitivity curve of bqsr(0.02,0.99) at AT−SILC
length(sample.x) * (res − base.qsr)
0
20
40
60
80
100
120
0e+00
2e+04
4e+04
6e+04
8e+04
1e+05
x

Correction term

- Trim only numerator of QSR: $T_{0.8;\,1-\alpha_u}(F)$
- Use approximate correction term $\frac{T_{0.8;\,1-\alpha_u}(F)}{T_{0.8;\,1-2\alpha_u}(F)}$ by considering "double trimming".

$$\hat{\eta}_{CQSR}(\alpha_u) = \frac{T_{0.8;\,1-\alpha_u}(F)}{T_{0;\,0.2}(F)} \cdot \frac{T_{0.8;\,1-\alpha_u}(F)}{T_{0.8;\,1-2\alpha_u}(F)}$$

Sampling adaptations

- The functionals in $T_{\alpha;\beta}(F)$ have to be estimated from the sample, taking the sample design into account.
- Plug in an estimate of $F_U(t)$:
  $F_S(t) = \sum_S w_i \mathbf{1}\{x_i \leq t\} / \sum_S w_i$:

$$\hat{T} = T(F_S)$$

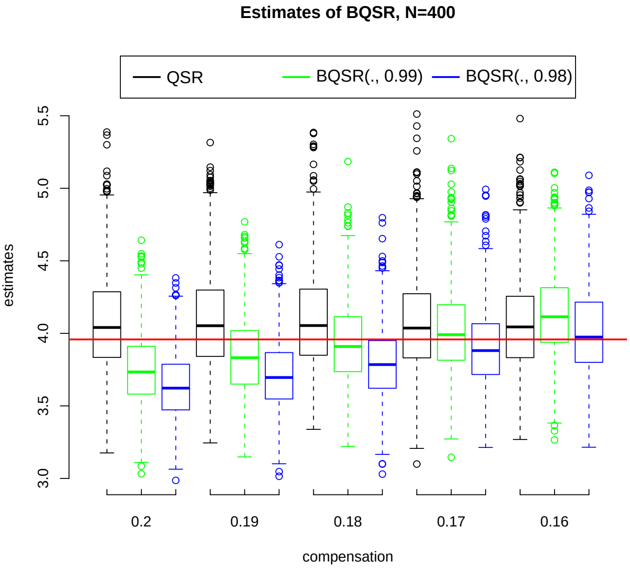
- Weighted trimmed means are straightforward
  $\Rightarrow$ Shao, Annals of Statistics 1994

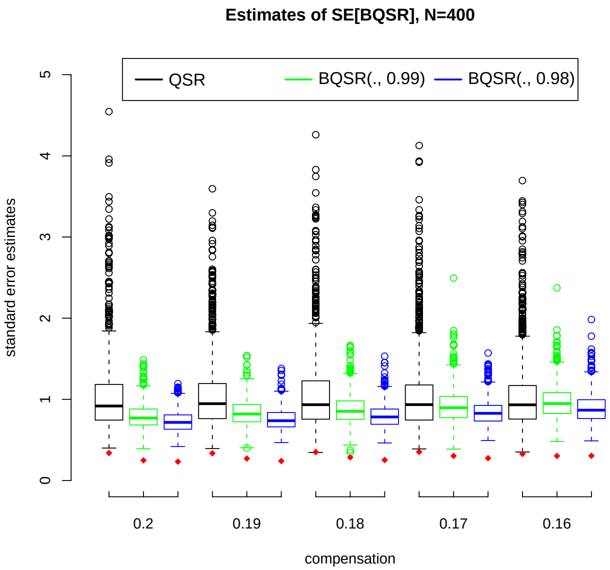
Variance estimation under sampling

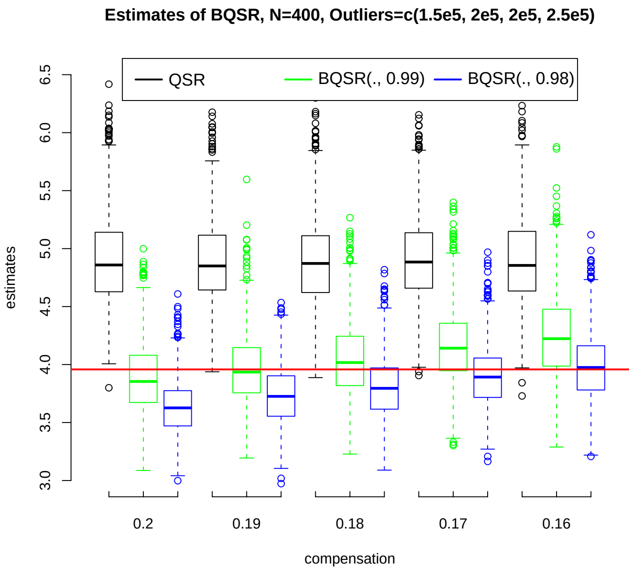
1. Estimate the variance covariance matrix of the influence function of the involved estimators taking into account the sample design.
2. Apply the Delta method.
3. If $T_j(F_n) = \sum_i^n w_i x_{ij}$ linearize first and take directional derivaitive (Woodruff 1971)(Anderson and Nordberg 1994)(Münnich 2008)

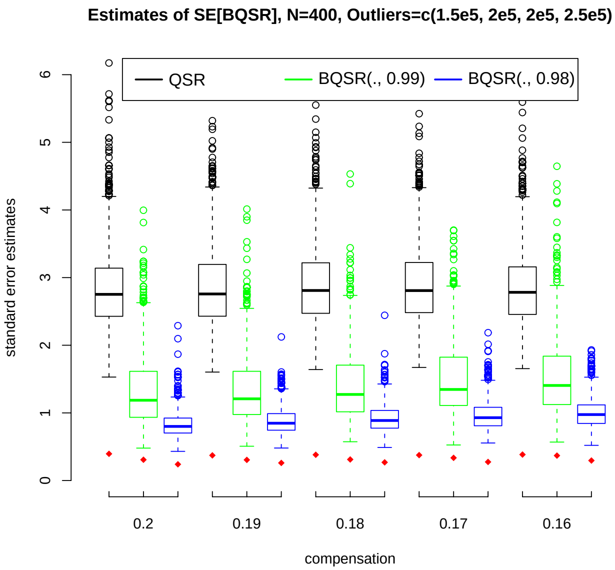
Some results with AT-SILC and BQSR

- Public Use Sample of AT-SILC 2004 (Sample for Research)
- Sample design: Stratified sample of households, proportional allocation. Rotational design. Complex non-response adjustment.
- Unit: persons ($n = 4626, p = 240$)
- Simulation with simple-random samples of size $n = 400$ and $r = 1000$ replicates.


Estimates of BQSR, N=400
QSR
BQSR(., 0.99)
BQSR(., 0.98)
estimates
3.0
3.5
4.0
4.5
5.0
5.5
0.2
0.19
0.18
0.17
0.16
compensation


Estimates of SE[BQSR], N=400
QSR
BQSR(., 0.99)
BQSR(., 0.98)
standard error estimates
0
1
2
3
4
5
0.2
0.19
0.18
0.17
0.16
compensation


Estimates of BQSR, N=400, Outliers=c(1.5e5, 2e5, 2e5, 2.5e5)
QSR
BQSR(., 0.99)
BQSR(., 0.98)
estimates
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
0.2
0.19
0.18
0.17
0.16
compensation


Estimates of SE[BQSR], N=400, Outliers=c(1.5e5, 2e5, 2e5, 2.5e5)
QSR
BQSR(., 0.99)
BQSR(., 0.98)
standard error estimates
0
1
2
3
4
5
6
0.2
0.19
0.18
0.17
0.16
compensation

Final remarks

- Work in progress!
- Variance estimator to be enhanced.
- We do not change the parameter η to estimate but the estimator.
- Variance estimation will be extensively tested in AMELI-simulations.
- The problem of choosing a tuning constant remains.

`http://ameli.surveystatistics.net`