Variance estimation of some EU-SILC based indicators at regional level

Jean Monet Lecture in Pisa – 2017

Ralf Münnich
Trier University, Faculty IV
Chair of Economic and Social Statistics

1. Introduction to variance estimation

2. Linearization methods

3. Resampling methods

4. Further topics in variance estimation

Unemployment in Saarland

Unemployed		14 – 24	25 – 44	45 – 64	65 +	$\sum$
Women	τ	2.387	7.248	4.686	128	14.449
Men	τ	4.172	9.504	10.588	0	24.264
$\sum$	τ	6.559	16.752	15.274	128	38.713

- *True* values in Saarland
- Estimates from the Microcensus
- Is the quality of the cell estimates identical?

Unemployment in Saarland

Unemployed		14 – 24	25 – 44	45 – 64	65 +	$\sum$
Women	τ	2.387	7.248	4.686	128	14.449
	$E\,\widehat{\tau}$	2.387	7.238	4.684	128	14.436
Men	τ	4.172	9.504	10.588	0	24.264
	$E\,\widehat{\tau}$	4.172	9.505	10.598	0	24.275
$\sum$	τ	6.559	16.752	15.274	128	38.713
	$E\,\widehat{\tau}$	6.558	16.743	15.282	128	38.711

- *True* values in Saarland
- Estimates from the Microcensus
- Is the quality of the cell estimates identical?

Unemployment in Saarland

Unemployed		14 – 24	25 – 44	45 – 64	65 +	$\sum$
Women	τ	2.387	7.248	4.686	128	14.449
	$E\,\widehat{\tau}$	2.387	7.238	4.684	128	14.436
Men	τ	4.172	9.504	10.588	0	24.264
	$E\,\widehat{\tau}$	4.172	9.505	10.598	0	24.275
$\sum$	τ	6.559	16.752	15.274	128	38.713
	$E\,\widehat{\tau}$	6.558	16.743	15.282	128	38.711

- *True* values in Saarland
- Estimates from the Microcensus
- Is the quality of the cell estimates identical?

How can we measure quality?

- Main emphasis is put on sample surveys
- Statistical items of interest:
  - Sampling distribution
    - Theoretical distribution
    - Approximate distribution
    - *Simulated distribution*
  - Focus on properties
    - Large sample properties
    - Small sample properties
- View of Official Statistics

Quality and the Code of Practice

European Statistics Code of Practice

15 principles on *institutional environment* (6), *statistical processes* (4), and **statistical output** (5):

Relevance of the statistical concept:
End-user, *user needs*, hierarchical structure and contents

Accuracy and reliability:

- Sampling errors: standard error, CI coverage
- Non-sampling errors: nonresponse, coverage error, measurement errors

Timeliness and punctuality: Time and duration from data acquisition until publication

Coherence and comparability: Preliminary and final statistics, annual and intermediate statistics (regions, domains, time)

Accessibility and clarity: Data, analysis and method reports

`http://ec.europa.eu/eurostat/quality`

Evaluation of samples and surveys

Practicability

Costs of a survey

Accuracy of results

- Standard errors
- Confidence interval coverage
- Disparity of sub-populations

Robustness of results

In order to adequately evaluate the estimates from samples, *appropriate* evaluation criteria have to be considered.

Why do we need variance estimation

Most *accuracy measures* are based on variances or variance estimates!

- Measures for point estimators
  - Bias, variance, MSE
  - CV, relative root MSE
  - Bias ratio, confidence interval coverage
  - Design effect, effective sample size
- Problems with measures:
  - *Theoretical* measures are problematic
  - Estimates from the sample (e.g. bias)
  - Availability in simulation study
  - Does large sample theory help much?
  - Small sample properties

Do we need special measures for variance estimators or variance estimates?

Why do we need variance estimation

Most *accuracy measures* are based on variances or variance estimates!

- Measures for point estimators
  - Bias, variance, MSE
  - CV, relative root MSE
  - Bias ratio, confidence interval coverage
  - Design effect, effective sample size
- Problems with measures:
  - *Theoretical* measures are problematic
  - Estimates from the sample (e.g. bias)
  - Availability in simulation study
  - Does large sample theory help much?
  - Small sample properties

Do we need special measures for variance estimators or variance estimates?

Why do we need variance estimation

Most *accuracy measures* are based on variances or variance estimates!

- Measures for point estimators
  - Bias, variance, MSE
  - CV, relative root MSE
  - Bias ratio, confidence interval coverage
  - Design effect, effective sample size
- Problems with measures:
  - *Theoretical* measures are problematic
  - Estimates from the sample (e.g. bias)
  - Availability in simulation study
  - Does large sample theory help much?
  - Small sample properties

Do we need special measures for variance estimators or variance estimates?

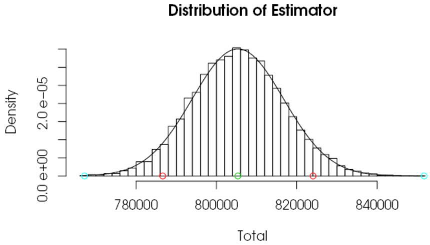
Why do we need variance estimation

Most *accuracy measures* are based on variances or variance estimates!

- Measures for point estimators
  - Bias, variance, MSE
  - CV, relative root MSE
  - Bias ratio, confidence interval coverage
  - Design effect, effective sample size
- Problems with measures:
  - *Theoretical* measures are problematic
  - Estimates from the sample (e.g. bias)
  - Availability in simulation study
  - Does large sample theory help much?
  - Small sample properties

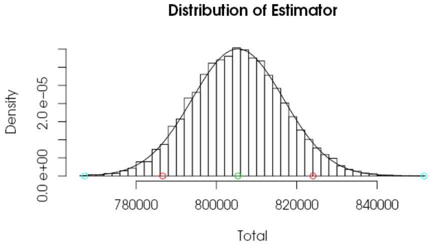
Do we need special measures for variance estimators or variance estimates?

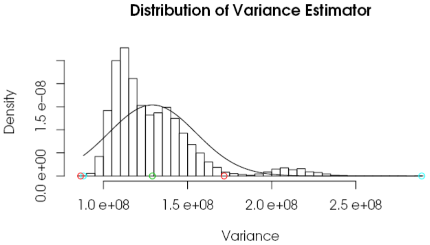
Example: Men in Hamburg



τ:	805258.00	N:	1669690				
$\hat{\tau}$:	805339.10	$V\hat{\tau}$:	1.29e+008	$E\,\hat{V}(\tau)$:	1.29e+008	$V\,\hat{V}(\tau)$:	6.72e+014
Bias Est:	81.10	MSE Est:	1.29e+008	Bias Var:	-3.78e+005	MSE Var:	6.72e+014
Skew Est:	0.0747	Curt Est:	3.0209	Skew Var:		Curt Var:	
CI (90%):				CI (95%):			

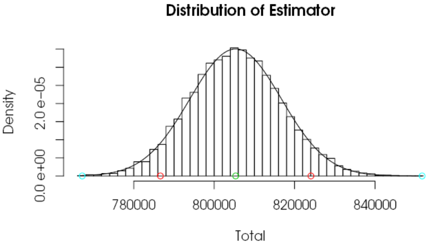
Example: Men in Hamburg

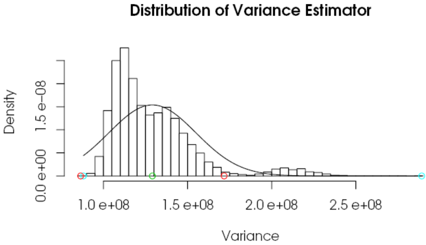




τ:	805258.00	N:	1669690				
$\hat{\tau}$:	805339.10	$V\hat{\tau}$:	1.29e+008	$E\,\widehat{V}(\tau)$:	1.29e+008	$V\,\widehat{V}(\tau)$:	6.72e+014
Bias Est:	81.10	MSE Est:	1.29e+008	Bias Var:	-3.78e+005	MSE Var:	6.72e+014
Skew Est:	0.0747	Curt Est:	3.0209	Skew Var:	1.8046	Curt Var:	6.9973
CI (90%):				CI (95%):			

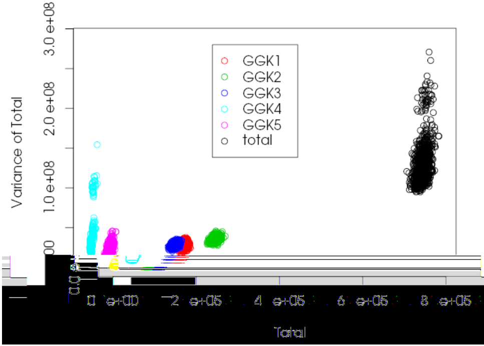
Example: Men in Hamburg

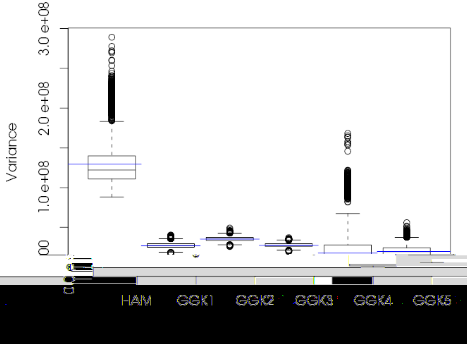




τ:	805258.00	N:	1669690				
$\hat{\tau}$:	805339.10	$V\hat{\tau}$:	1.29e+008	$E\,\hat{V}(\tau)$:	1.29e+008	$V\,\hat{V}(\tau)$:	6.72e+014
Bias Est:	81.10	MSE Est:	1.29e+008	Bias Var:	-3.78e+005	MSE Var:	6.72e+014
Skew Est:	0.0747	Curt Est:	3.0209	Skew Var:	1.8046	Curt Var:	6.9973
CI (90%):	90.16	(4.1;5.7)		CI (95%):	94.79	(2.0;3.2)	

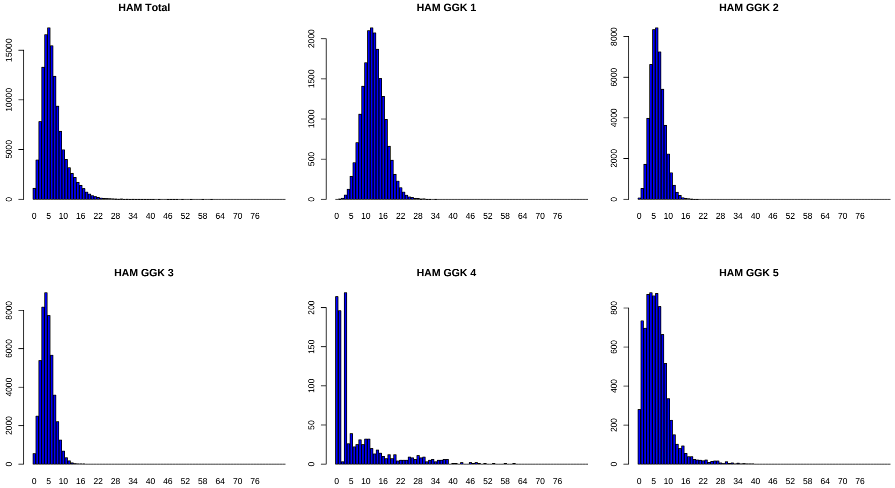
Total estimate separated by house size class (GGK)



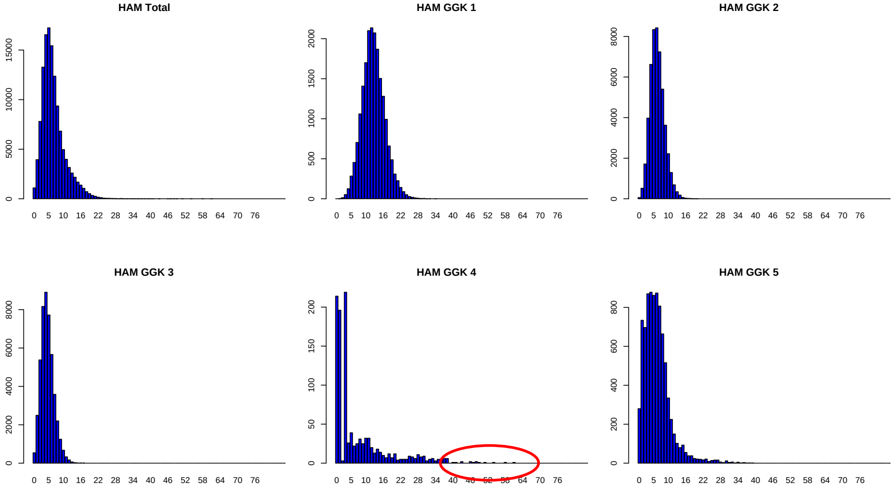


	GGK1	GGK2	GGK3	GGK4	GGK5	total
Persons	468293	651740	439745	9940	99970	1669690
Sampling units	173	446	414	10	75	1118

Distribution of men in HAM (per SU)

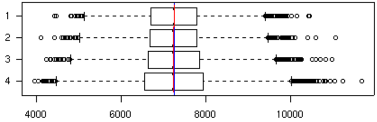


Distribution of men in HAM (per SU)

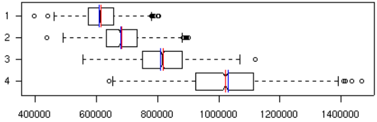


Unemployed women, 25 – 44

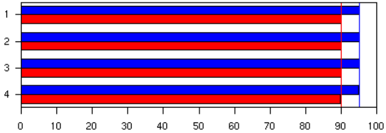
Raking estimator



variance estimator

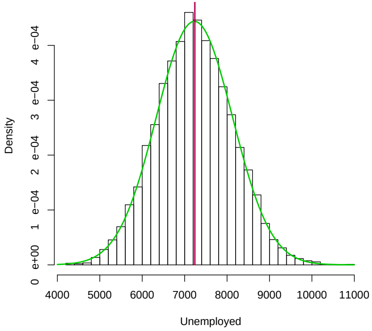


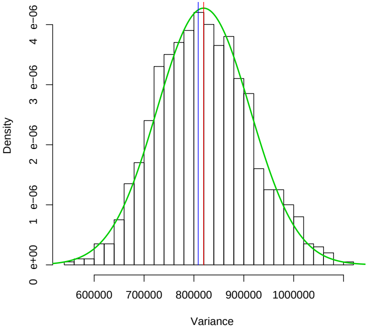
NR rates: 1: 5%, 2: 10%, 3: 25%, 4: 40%



95% 90%

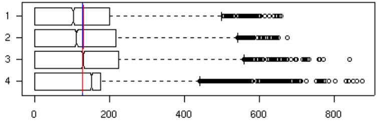
Unemployed women, 25 – 44, distribution of point and variance estimator (25% NR)



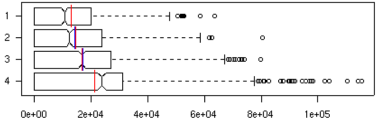


Unemployed women, 65 +

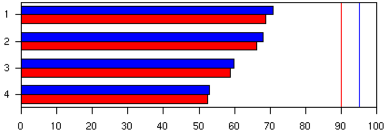
Raking estimator



variance estimator

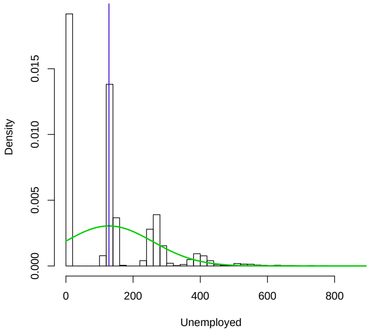


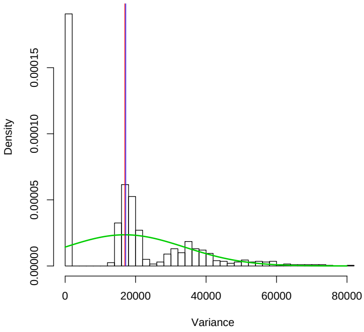
NR rates: 1: 5%, 2: 10%, 3: 25%, 4: 40%



95% 90%

Unemployed women, 65 +, distribution of point and variance estimator (25% NR)





Direct variance estimators for single stage sampling

Stratified random sampling Since $E(s_q^2) = \frac{N}{N-1} \cdot \sigma_q^2 = S_q^2$

$$\widehat{V}(\widehat{\mu}_{\text{StrRSWOR}}) = s_p^2 = \sum_{q=1}^{L} \gamma_q^2 \cdot \frac{s_q^2}{n_q} \cdot \frac{N_q - n_q}{N_q}$$

is an unbiased estimate for $V(\widehat{\mu}_{\text{StrRSWOR}})$.

Cluster sampling Similarly:

$$\widehat{V}(\widehat{\mu}_{\text{SIC}}) = \frac{L^2}{N^2} \cdot \frac{s_e^2}{l} \cdot \frac{L - l}{L}$$

where

$$s_e^2 = \frac{1}{l-1} \cdot \sum_{r=1}^{l} \left(N_r^{\text{sel}} \mu_r^{\text{sel}} - \frac{N \cdot \widehat{\mu}_{\text{SIC}}}{L} \right)^2 .$$

Direct variance estimator for two stage sampling

- *Direct variance estimator*:

$$\hat{\mathrm{V}}\left(\hat{\tau}_{2St}\right) = L^2 \cdot \left(\frac{L-l}{L}\right) \cdot \frac{s_e^2}{l} + \frac{L}{l} \sum_{q=1}^{l} \left(\frac{N_q - n_q}{N_q}\right) \cdot N_q^2 \cdot \frac{s_q^2}{n_q}$$

with $s_e^2 = \frac{1}{l-1} \sum_{q=1}^{l} \left(\hat{\tau}_q - \frac{\hat{\tau}}{L}\right)^2, s_q^2 = \frac{1}{n_q - 1} \cdot \sum_{i=1}^{n_q} \left(y_{qi} - \overline{y}_q\right)^2$

cf. Lohr (1999), p. 147.

- The estimator is unbiased, but the first and second term do not estimate the variance at the respective stage (cf. Särndal et al. 1992, p. 139 f., Lohr 1999, p. 210):

$$\mathrm{E}\left[L^2 \cdot \left(\frac{L-l}{L}\right) \cdot \frac{s_e^2}{l}\right] = L^2 \cdot \left(1 - \frac{l}{L}\right) \cdot \frac{\sigma_e^2}{l} + \frac{L}{l}\left(1 - \frac{l}{L}\right) \sum_{q=1}^{L} \mathrm{V}\left(\hat{\tau}_q\right)$$

Aim of the lecture

- Learn more about the statistic $\widehat{\pi}$
  - Distribution of the estimator
  - Derivation of more information from this
- How do $V(\widehat{\pi})$ and $\widehat{V}(\widehat{\pi})$ correspond to each other?
- Is the *characteristic equation for variance estimation* all we need to know?

$$E\left(\widehat{V}(\widehat{\pi})\right) = V\left(\widehat{\pi}\right) \tag{1}$$

- How do we *measure* the characteristic equation?
- What influences the reliability of the characteristic equation?
- Can we properly simulate this?
- By the way
  - Do we have more than one choice?
  - Can we apply this to more sophisticated estimators?
  - How does the *statistical production process* influence the findings, e.g w.r.t. non-sampling errors

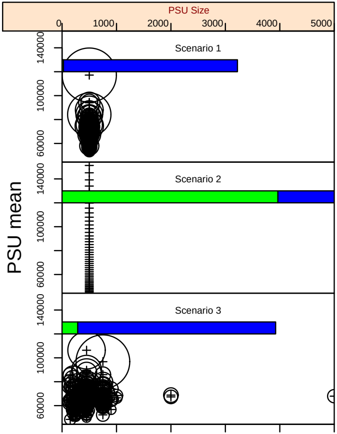
Experimental study: Sampling design

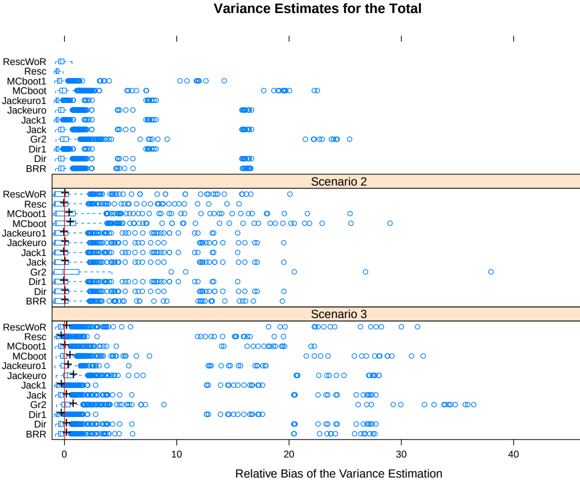
- Two stage sampling with stratification at the first stage, 25 strata
- 1. Stage: Drawing 4 PSU in each stratum (contains 8 PSU on average, altogether 200 PSU)
- 2. Stage: Proportional allocation of the sample size (1,000 USU) to the PSU (contains 500 USU on average, altogether 100,000 USU)

Experimental study: Scenarios

- *Scenario 1* : Units within PSU are heterogeneous with respect to the variable of interest $Y \sim LN(10, 1.5^2)$, PSU are of equal size
- *Scenario 2* : Units within PSU are homogeneous with respect to the variable of interest, PSU are of equal size
- *Scenario 3* : Units within PSU are heterogeneous with respect to the variable of interest $Y \sim LN(10, 1.5^2)$, PSU are of unequal size

Variance estimates for the total





Second order inclusion probabilities

In case of unequal probability sampling designs, we also need the second order inclusion probabilities for variance estimation:

Second order inclusion probability

The probability that both elements i and j are drawn in the sample is denoted by

$$\pi_{ij} = \sum$$

Horvitz-Thompson estimator

For estimating a total τ_Y of a variable of interest Y (πps, WOR) we take

$$\widehat{\tau}_{\mathrm{HT}} = \sum_{i \in \mathcal{S}} \frac{y_i}{\pi_i} = \sum_{i \in \mathcal{S}} d_i \cdot y_i \quad ,$$

where $d_i = 1/\pi_i$ denote the design weights (as reciprocal of the first order inclusion probabilities.

In order to estimate means, one should take the Hájek estimator

$$\widehat{\mu} = \frac{\sum\limits_{i \in \mathcal{S}} d_i \cdot y_i}{\sum\limits_{i \in \mathcal{S}} d_i}$$

even if the population total is known.

Properties of the HT estimator

- HT estimator: class of homogeneous, linear unbiased estimators
- The HT estimator is admissible;
  the HH estimator is not necessarily admissable
  (improved HH via Rao-Blackwellization)
- Negative variance estimates are possible in some designs!

Cassel, C.; Särndal, C.-E.; Wretman, J.H. (1977): Foundations of inference in survey sampling. Wiley.
Gabler, S. (1990): Minimax Solutions in Sampling from Finite Populations. Lecture Notes in Statistics, 64. Springer.
Hedayat, A.S.; Sinha, B.K. (1991); Design and Inference in Finite Population Sampling, Wiley.

Variance of the HT estimator

Assuming positive second order inclusion probabilities we obtain

$$V(\widehat{\tau}) = \sum_{i\in\mathcal{U}} \pi_i(1-\pi_i)\cdot\left(\frac{y_i}{\pi_i}\right)^2 + 2\cdot\sum\sum_{\substack{i,j\in\mathcal{U}\\ i<j}} (\pi_{ij}-\pi_i\cdot\pi_j)\cdot\frac{y_i}{\pi_i}\cdot\frac{y_j}{\pi_j}$$

as the variance of the Horvitz-Thompson estimator $\widehat{\tau}$.
An unbiased estimator of this variance is:

$$\widehat{V}_{\mathrm{HT}}(\widehat{\tau}) = \sum_{i\in\mathcal{S}} (1-\pi_i)\cdot\left(\frac{y_i}{\pi_i}\right)^2 + 2\cdot\sum\sum_{\substack{i,j\in\mathcal{S}\\ i<j}} \left(1-\frac{\pi_i\cdot\pi_j}{\pi_{ij}}\right)\cdot\frac{y_i}{\pi_i}\cdot\frac{y_j}{\pi_j}$$

Sen-Yates-Grundy variance estimator

Alternatively, for designs with fixed sample sizes, we can use the Sen-Yates-Grundy variance estimator:

$$V_{\mathrm{SYG}}(\widehat{\tau}) = -\frac{1}{2} \sum\sum_{\substack{i,j \in \mathcal{U} \\ i \neq j}} (\pi_{ij} - \pi_i \cdot \pi_j) \cdot \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2$$

$$= \sum\sum_{\substack{i,j \in \mathcal{U} \\ i < j}} (\pi_i \cdot \pi_j - \pi_{ij}) \cdot \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2$$

As unbiased estimator can be applied:

$$\widehat{V}_{\mathrm{SYG}}(\widehat{\tau}) = \sum\sum_{\substack{i,j \in \mathcal{S} \\ i < j}} \frac{\pi_i \cdot \pi_j - \pi_{ij}}{\pi_{ij}} \cdot \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2$$

Example for approximations

- In presence of a sampling design with maximum entropy the following general approximation of the variance results:

$$V_{approx}\left(\hat{\ }\right) = \sum_{i \in \mathcal{U}} \frac{b_i}{{}_i^2} \cdot \left(y_i - y_i^*\right)^2$$

$$y_i^* = \ {}_i \cdot \frac{\sum\limits_{j \in \mathcal{U}} b_i \cdot y_i = \ {}_j}{\sum\limits_{j \in \mathcal{U}} b_j}$$

- Hájek approximation:

$$b_i^{Hajek} = \frac{{}_i \cdot (1 - \ {}_i) \cdot N}{N-1}$$

Cf. Matei and Tillé (2005) or Hulliger et. al (2011)

Linearization of non-linear statistics

Suppose we have d different study variables and we want to estimate parameter θ of the finite population $\mathcal{U}$ of size N, which has the following form

$$\theta = f\left(\boldsymbol{\tau}\right) \ , \tag{2}$$

where $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_k, \ldots, \tau_d)$ and $\tau_k = \sum_{i \in \mathcal{U}} y_{ki}$, with y_{ki} as the observation of k-th study variable of the i-th element in $\mathcal{U}$.
Then we substitute the unknown parameter vector $\boldsymbol{\tau}$ in (2) by its estimate $\hat{\boldsymbol{\tau}} = (\hat{\tau}_1, \ldots, \hat{\tau}_k, \ldots, \hat{\tau}_d)$, which yields

$$\hat{\theta} = f\left(\hat{\boldsymbol{\tau}}\right) \ ,$$

with $\hat{\tau}_k = \sum_{i \in s} y_{ki} w_i$ as the estimated total of the k-th study variable and w_i is the survey weight of the i-th element in s.
Further, it is assumed that $\hat{\tau}_k$ is a consistent estimator of τ_k.

In case the function f is continuously differentiable up to order two at each point in the open set $\mathbb{S}$ containing $\boldsymbol{\tau}$ and $\hat{\boldsymbol{\tau}}$, we can use a Taylor expansion

$$\hat{\theta} - \theta = \sum_{k=1}^{d} \left[\frac{\partial f(p_1, \ldots, p_d)}{\partial \tau_k} \right]_{\mathbf{p}=\boldsymbol{\tau}} (\hat{\tau}_k - \tau_k) + R(\hat{\boldsymbol{\tau}}, \boldsymbol{\tau}) \,, \tag{3}$$

where

$$R(\hat{\boldsymbol{\tau}}, \boldsymbol{\tau}) = \frac{1}{2!} \sum_{k=1}^{d} \sum_{l=1}^{d} \left[\frac{\partial^2 f(p_1, \ldots, p_d)}{\partial p_k \partial p_l} \right]_{\mathbf{p}=\ddot{\boldsymbol{\tau}}} (\hat{\tau}_k - \tau_k)(\hat{\tau}_l - \tau_l)$$

and $\ddot{\boldsymbol{\tau}}$ is in the interior of line segment $L(\boldsymbol{\tau}, \hat{\boldsymbol{\tau}})$ joining $\boldsymbol{\tau}$ and $\hat{\boldsymbol{\tau}}$. For the remainder term R we have $R = O_p(r_n^2)$, where $r_n \to 0$ as $n \to \infty$. Further, we have $\hat{\theta} - \theta = O_p(r_n)$. Thus, in most applications it is common practice to regard R as negligible in (3) for sample sizes large enough.

This justifies the use of the following approximation:

$$\hat{\theta} - \theta \approx \sum_{k=1}^{d} \left[\frac{\partial f(p_1, \ldots, p_d)}{\partial \tau_k} \right]_{\mathbf{p}=\boldsymbol{\tau}} (\hat{\tau}_k - \tau_k) \ . \tag{4}$$

Note, that in expression (4) only the linear part of the Taylor series is kept. Now, we can use (4) to derive an approximation of the mean square error (MSE) of $\hat{\theta}$ which is given by

$$\begin{aligned} \text{MSE}\left(\hat{\theta}\right) &\approx V\left(\sum_{k=1}^{d} \left[\frac{\partial f(p_1, \ldots, p_d)}{\partial \tau_k} \right]_{\mathbf{p}=\boldsymbol{\tau}} \hat{\tau}_k \right) \qquad (5) \\ &= \sum_{k=1}^{d} a_k^2 V(\hat{\tau}_k) + 2 \sum_{k=1}^{d} \sum_{\substack{l=1 \\ k<l}}^{d} a_k a_l Cov(\hat{\tau}_k, \hat{\tau}_l) \ , \end{aligned}$$

where $a_k = [\frac{\partial f(p_1, \ldots, p_d)}{\partial \tau_k}]_{\mathbf{p}=\boldsymbol{\tau}}$.

Because $\text{MSE}(\hat{\theta}) = V(\hat{\theta}) + \text{Bias}(\hat{\theta})^2$, where $\text{Bias}(\hat{\theta}) = \hat{\theta} - \theta$, we can approximate the variance of $\hat{\theta}$ by $\text{MSE}(\hat{\theta})$ since $V(\hat{\theta})$ is of higher order than $\text{Bias}(\hat{\theta})^2$ for unbiased or at least consistent estimators. Thus, we can use (5) as an approximation of the design variance of $\hat{\theta}$.

The methodology holds for *smooth* functions. For non-differentiable functions (e.g. quantiles), influence functions could be applied (cf. Deville, 1999). Finally, estimating equations may also lead to linearized variables (cf. Binder and Patak, 1994).

Finally, *only* the linearized variables have to be derived in order to apply the linear methodology for non-linear statistics (e.g. poverty indicators).

To estimate (5), we could simply substitute the variances and covariances with their corresponding estimates. This, however, might become unpractical if d, the number of estimated totals in $\hat{\boldsymbol{\tau}}$, becomes large. To evade this problem, Woodruff (1971) suggested the following:

$$\begin{aligned}\mathrm{MSE}\left(\hat{\theta}\right) &\approx V\Big(\sum_{k=1}^{d} a_k \hat{\tau}_k\Big) \approx V\Big(\sum_{k=1}^{d} a_k \sum_{i=1}^{n} w_i y_{ki}\Big)\\ &\approx V\Big(\sum_{i=1}^{n} w_i \sum_{k=1}^{d} a_k y_{ki}\Big) \approx V\Big(\sum_{i=1}^{n} w_i z_i\Big) \qquad ,\end{aligned}$$

where

$$z_i = \sum_{k=1}^{d} a_k y_{ki} \qquad \text{with} \qquad \mathrm{MSE}\left(\hat{\theta}\right) \approx V\left(\sum_{i \in s} w_i z_i\right) \;. \qquad (6)$$

CLAN: Function of Totals

Andersson and Nordberg introduced easy to computer SAS macros in order to produce linearized values for functions of totals:

Let $\theta = \tau_1 \circ \tau_2$ a function of totals from $\circ \in \{+, -, \cdot, /\}$. Then

Operator	z transformation
$+$	$z_k = y_{1k} + y_{2k}$
$-$	$z_k = y_{1k} - y_{2k}$
$\cdot$	$z_k = \theta \cdot (y_{1k}/t_1 + y_{2k}/t_2)$
$/$	$z_k = \theta \cdot (y_{1k}/t_1 - y_{2k}/t_2)$

The proof follows from applying Woodruff's method. Now, any functions using the above operators of totals can be recursively developed, which can be integrated in software (cf. Andersson and Nordberg, 1994).

Evidence-based Policy Decision Based on Indicators

- Indicators are seen as *true* values
- In general, indicators are simply survey variables
- No modelling is used to
  - Improve quality and accuracy of indicators
  - Disaggregate values towards domains and areas
- Reading naively point estimator tables may lead to misinterpretations
  - Change (Münnich and Zins, 2011)
  - Benchmarking (change in European policy)

- How accurate are estimates for indicators (ARPR, RMPG, GINI, and QSR)?
- This leads to applying the adequate variance estimation methods

Linearization and Resampling Methods

The statistics in question (the Laeken indicators) are highly non-linear.

- 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)
  - Demnati and Rao (2004)

Application to poverty and inequality indicators

Using the linearized values for the statistics ARPR, GINI, and QSR to approximate their variances.

Calibrated weights w_i: z_i are 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)
GINI:	Kovačević and Binder (1997)
QSR:	Hulliger and Münnich (2007)
RMPG:	Osier (2009)

For CI estimation, empirical likelihood methods may be preferable (cf. Berger, De La Riva Torres, 2015).

Example: Quintile share ratio

Starting the quintiles of the groups R and P we get

$$\widehat{\mu}_{\mathrm{R}} = \sum_i w_i \cdot \big(y_i - y_i \cdot \mathbb{1}(y_i \leq \widehat{y}_{0,8})\big) \Big/ \sum_i w_i \cdot (1 - 0{,}8)$$

$$\widehat{\mu}_{\mathrm{P}} = \sum_i w_i \cdot y_i \cdot \mathbb{1}(y_i \leq \widehat{y}_{0,2}) \Big/ \sum_i w_i \cdot 0{,}2 \quad .$$

Next, we define the QSR as a function of four totals:

$$\widehat{\mathrm{QSR}} = \frac{\widehat{\mu}_{\mathrm{R}}}{\widehat{\mu}_{\mathrm{P}}} = \frac{\widehat{\tau}_1}{\widehat{\tau}_2} \Big/ \frac{\widehat{\tau}_3}{\widehat{\tau}_4}$$

In order to apply linearized variance estimation, we have to derive the linearized variables.

$$
\begin{aligned}
u_{1i} &= y_i - \left((y_i - y_{0,8}) \cdot \mathbb{1}(y_i \le \widehat{y}_{0,8}) + 0{,}8 \cdot y_{0,8} \right) \\
u_{2i} &= 0{,}2 \\
u_{3i} &= (y_i - y_{0,2}) \cdot \mathbb{1}(y_i \le \widehat{y}_{0,2}) + 0{,}2 \cdot y_{0,2} \\
u_{4i} &= 0{,}2 \\
u_{5i} &= \left(u_{1i} - \frac{\widehat{\tau}_1}{\widehat{\tau}_2} \cdot u_{2i}\right) \cdot \frac{1}{\widehat{N} \cdot 0{,}2} = \widehat{\mu}_\mathrm{R} \\
u_{6i} &= \left(u_{3i} - \frac{\widehat{\tau}_3}{\widehat{\tau}_4} \cdot u_{4i}\right) \cdot \frac{1}{\widehat{N} \cdot 0{,}2} = \widehat{\mu}_\mathrm{P}
\end{aligned}
$$

where $\widehat{\tau}_2 = \widehat{\tau}_4 = \widehat{N} \cdot 0{,}2$. Finally, we get

$$
z_i = \left(u_{5i} - \widehat{QSR} \cdot u_{6i}\right) \cdot \frac{\widehat{\tau}_4}{\widehat{\tau}_3}
$$

Note: The linearization of quintiles will subject to the methods of using *influence functions* (see Deville, 1999).

Resampling methods

- Idea: draw repeatedly (sub-)samples from the sample in order to build the sampling distribution of the statistic of interest
- Estimate the variance as variability of the estimates from the resamples
- Methods of interest
  - Random groups
  - Balanced repeated replication (balanced half samples)
  - Jackknife techniques
  - Bootstrap techniques
- Some remarks:
  - If it works, one doesn't need second order statistics for the estimate
  - May be computationally exceptional
  - What does influence the quality of these estimates

Random groups

- Mahalanobis (1939)
- Aim: estimate variance of statistic θ
- Random partition of sample into R groups (independently)
- $\widehat{\theta}_{(r)}$ denotes the estimate of θ on r-th subsample
- Random group points estimate:

$$\widehat{\theta}_{\mathrm{RG}} = \frac{1}{R} \cdot \sum_{r=1}^{R} \widehat{\theta}_{(r)}$$

- Random group variance estimate:

$$\widehat{V}(\widehat{\theta}_{\mathrm{RG}}) = \frac{1}{R} \cdot \frac{1}{R-1} \cdot \sum_{r=1}^{R} \left(\widehat{\theta}_{(r)} - \widehat{\theta}_{\mathrm{RG}}\right)^2$$

- Random selection versus random partition!

Balanced repeated replication

- Originally we have two observations per stratum
- Random partioning of observations into two groups
- $\widehat{\theta}_r$ is the estimate of the r-th selection using the H *half samples*
- Instead of recalling all possible $R \ll 2^H$ replications, we use a balanced selection via Hadamard matrices
- We obtain:

$$\widehat{\theta}_{\text{BRR}} = \frac{1}{R} \cdot \sum_{r=1}^{R} \widehat{\tau}_r \text{ and } \widehat{V}_{\text{BRR}}\big(\widehat{\theta}\big) = \frac{1}{R} \sum_{r=1}^{R} \big(\widehat{\theta}_r - \widehat{\theta}\big)^2 \quad .$$

- May lead to highly variable variance estimates, especially when H is small (cf. Davison and Sardy, 2004). Repetition of random grouping may be useful (cf. Rao and Shao, 1996)
- Use special weighting techniques for improvements

The Jackknife

Originally, the Jackknife method was introduced for estimating the bias of a statistic (Quenouille, 1949).
Let $\widehat{\theta}(Y_1, \ldots, Y_n)$ be the statistic of interest for estimating the parameter θ. Then,

$$\widehat{\theta}_{-i} = \widehat{\theta}(Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_n)$$

is the corresponding statistic omitting the observation Y_i which is therefore based on $n-1$ observations. Finally, the delete-1-Jackknife (d1JK) bias for θ is

$$\widehat{\mathrm{B}}_{\mathrm{d1JK}}(\widehat{\theta}) = (n-1) \cdot \left(\frac{1}{n} \sum_{i \in \mathcal{S}} \widehat{\theta}_{-i} - \widehat{\theta} \right)$$

(cf. Shao und Tu, 1995).

The jackknife (continued)

From the bias follows immediately the Jackknife point estimate

$$\begin{aligned}\widehat{\theta}_{\mathrm{d1JK}} &= \widehat{\theta} - \widehat{\mathrm{B}}_{\mathrm{d1JK}}\left(\widehat{\theta}\right) \\ &= n \cdot \widehat{\theta} - \frac{n-1}{n} \sum_{i \in \mathcal{S}} \widehat{\theta}_{-i}\end{aligned}$$

which is a delete-1-Jackknife bias corrected estimate. This estimator is under milde smoothness conditions of order n^{-2}.

Jackknife variance estimation

Tukey (1958) defined the so-called jackknife pseudo values $\widehat{\theta}_i^* := n \cdot \widehat{\theta} - (n-1) \cdot \widehat{\theta}_{-i}$ which yield under the assumption of stochastic independency and approximately equal variance of the $\widehat{\theta}_i^*$. Finally

$$\widehat{V}_{\mathrm{d1JK}}\left(\widehat{\theta}\right) = \frac{1}{n(n-1)} \cdot \sum_{i \in \mathcal{S}} \left(\widehat{\theta}_i^* - \overline{\widehat{\theta}}^*\right)^2$$

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

Problem: What is $\widehat{\theta}_i^*$ and $\widehat{V}_{\mathrm{d1JK}}\left(\widehat{\theta}\right)$ for $\widehat{\theta} = \overline{Y}$?

Advantages and disadvantages of the jackknife

- Very good for *smooth* statistics
- Biased for the estimation of the median
- Needs special weights in stratified random sampling (missing independency of jackknife resamples)

$$\widehat{V}_{\text{d1JK,strat}}(\widehat{\theta}) = \sum_{h=1}^{h} \frac{(1-f_h)\cdot(n_h-1)}{n_h} \cdot \sum_{i=1}^{n_h} \left(\widehat{\theta}_{h,-i} - \overline{\widehat{\theta}}_h\right)^2$$

where $-i$ indicates the unit i that is left out.
- Specialized procedures are needed for (really) complex designs (cf. Rao, Berger, and others)
- Huge effort in case of large samples sizes (n):
  - grouped jackknife (m groups; cf. Kott and R-package EVER)
  - delete-d-jackknife (m replicates with d sample observations eliminated simultaneously; $m \ll \binom{n}{d}$)

Bootstrap resampling

- Theoretical bootstrap
- Monte-Carlo bootstrap:
  Random selection of size n (SRS) yields

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

- Special adaptions are needed in complex surveys
- Insufficient estimates in WOR sampling and higher sample fractions

Monte-Carlo bootstrap

Efron (1982):

1. Estimate $\widehat{F}$ as the empirical distribution function (non-parametric maximum likelihood estimation);
2. Draw bootstrap samples from $\widehat{F}$, that is

$$X_1^*, \ldots, X_n^* \quad \overset{\text{IID}}{\sim} \quad \widehat{F}$$

   of size n;
3. Compute the bootstrap estimate $\widehat{\tau}^*_{n,i} = \widehat{\tau}\left(X_1^*, \ldots, X_n^*\right)$;
4. Repeat 1. to 3. B times (B arbitrarily large) and compute finally the variance

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

Properties of the Monte-Carlo bootstrap

The bootstrap variance estimates converge by the law of large numbers to the *true* (theoretical) bootstrap variance estimate (cf. Shao and Tu, 1995, S. 11)

$$\widehat{\mathrm{V}}_{\mathrm{Boot,MC}} \xrightarrow{\text{a.s.}} V_{\mathrm{Boot}} \qquad .$$

Analogously, one can derive the bootstrap bias of the estimator by

$$\widehat{\mathrm{B}}_{\mathrm{Boot,MC}} = \frac{1}{B}\sum_{i=1}^{B} \widehat{\tau}^{*}_{n,i} - \widehat{\tau} \qquad .$$

Bootstrap confidence intervals

- Via variance estimation

$$\left[\widehat{\tau} - \sqrt{\widehat{\mathrm{V}}_{\mathrm{Boot,MC}}\left(\widehat{\tau}\right)} \cdot z_{1-\alpha/2}; \widehat{\tau} - \sqrt{\widehat{\mathrm{V}}_{\mathrm{Boot,MC}}\left(\widehat{\tau}\right)} \cdot z_{\alpha/2}\right]$$

- Via bootstrap resamples:

$$z_1^* = \frac{\widehat{\tau}_1^* - \widehat{\tau}}{\sqrt{\widehat{\mathrm{V}}_{\mathrm{Boot,\ MC}}\left(\widehat{\tau}_1^*\right)}} \quad , \ldots , \quad z_B^* = \frac{\widehat{\tau}_B^* - \widehat{\tau}}{\sqrt{\widehat{\mathrm{V}}_{\mathrm{Boot,\ MC}}\left(\widehat{\tau}_B^*\right)}}$$

From this empirical distribution, one can calculate the $\alpha/2$- and $(1-\alpha/2)$ quantiles $z^*_{\alpha/2}$ and $z^*_{1-\alpha/2}$ respectively by

$$\left[\widehat{\tau} - \sqrt{\widehat{\mathrm{V}}_{\mathrm{Boot,MC}}\left(\widehat{\tau}\right)} \cdot z^*_{B(1-\alpha/2)}; \widehat{\tau} - \sqrt{\widehat{\mathrm{V}}_{\mathrm{Boot,MC}}\left(\widehat{\tau}\right)} \cdot z_{B\alpha/2}\right]$$

This is referred to as the *studentized* bootstrap confidence interval.

Rescaling bootstrap

- *Rescaling bootstrap*: In case of multistage sampling only the first stage is considered. I^* (must be chosen) instead of I PSU are drawn with replacement (see Rao, Wu and Yue, 1992, Rust, 1996) The weights are adjusted by:

$$w^*_{qi} = \left[\left(1 - \left(\frac{I^*}{I-1}\right)^{1/2}\right) + \left(\frac{I^*}{I-1}\right)^{1/2} \cdot \left(\frac{I}{I^*}\right) \cdot r_q\right] \cdot w_{qi}.$$

- *Rescaling bootstrap without replacement*: From the I units of the sample, $I^* = \lfloor I/2 \rfloor$ units are drawn without replacement (see Chipperfield and Preston, 2007). In case of single stage sampling, the weights are adjusted by:

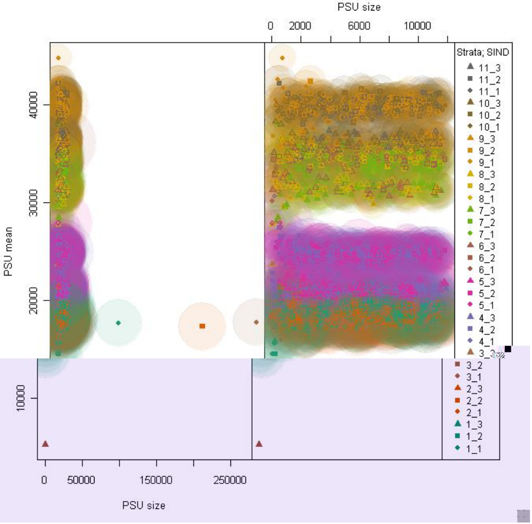
$$w^*_i = \left(1 - \lambda + \lambda \cdot \frac{n}{n^*} \cdot \delta_i\right) \cdot w_i, \text{with } \lambda = \sqrt{n^* \cdot \frac{(1-f)}{(n-n^*)}},$$

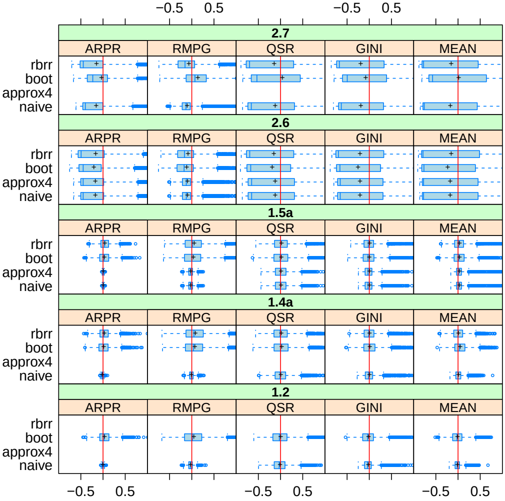
where δ_i is 1 when element i is chosen and 0 otherwise. For multistage designs (cf. Preston, 2009) the weights are adjusted at each stage by adding the term $-\lambda_G \cdot \left(\prod_{g=1}^{G-1} \sqrt{(n_g/n_g^*) \cdot \delta_g}\right) + \lambda_G \cdot \left(\prod_{g=1}^{G-1} \sqrt{(n_g/n_g^*) \cdot \delta_g}\right) \cdot (n_G/n_G^*) \cdot \delta_g$ at each stage G with $\lambda_G = \sqrt{n_G^* \left(\prod_{g=1}^{G-1} f_g\right) \cdot \frac{(1-f_G)}{n_G - n_G^*}}$.

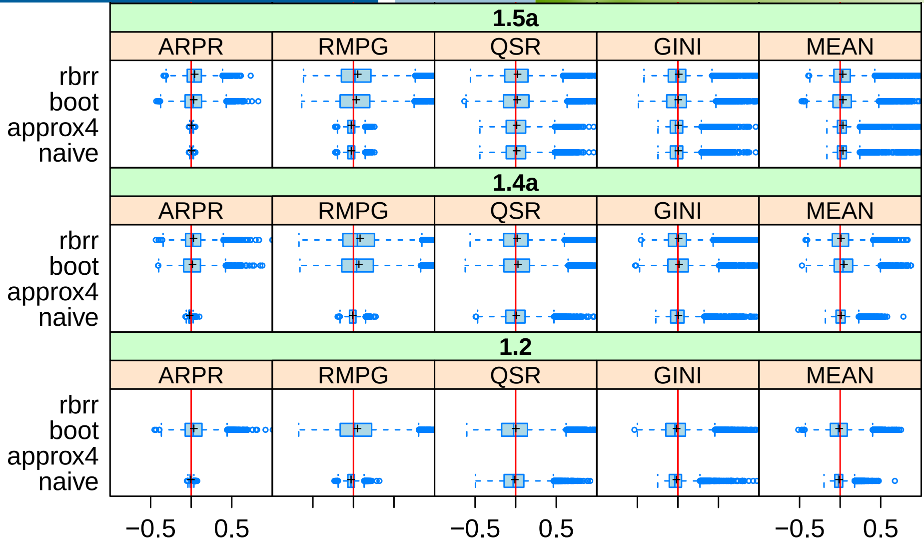
Comparison (cf. Bruch et al., 2011)

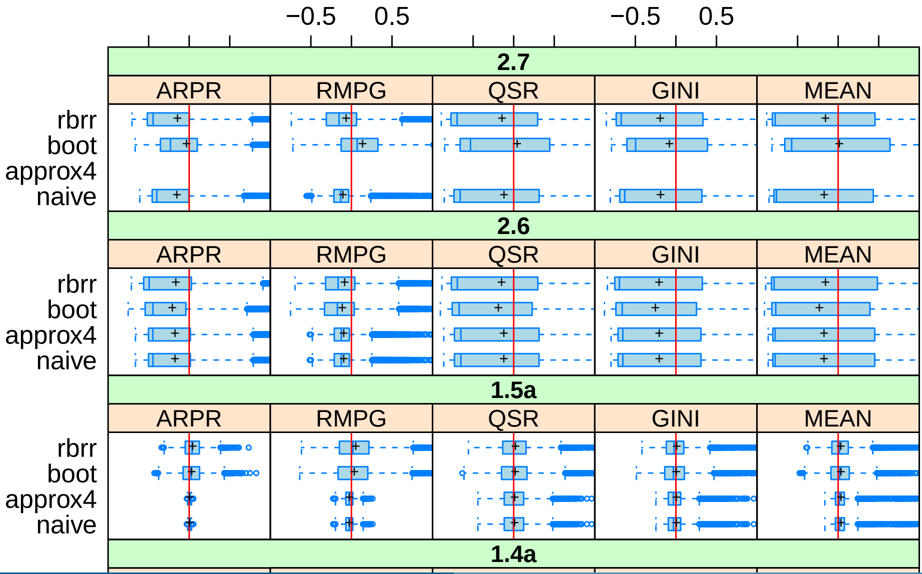
Method	BRR (Basic Model)	BRR (Group)	Delete-1 Jackknife	Delete-d Jackknife	Delete-a-Group Jackknife	Monte Carlo Bootstrap	Rescaling Bootstrap	Rescaling Bootstrap WoR
Statistic	Smooth and non-smooth	Smooth and non-smooth	Only for smooth statistics	Smooth and non-smooth	Smooth and non-smooth	Smooth and non-smooth	Smooth and non-smooth	Smooth and non-smooth
Stratification	Only when 2 elements per stratum	Required	Appropriate	Appropriate	Appropriate	Appropriate	Appropriate	Appropriate
Unequal Probability Sampling	Wolter (2007, p. 113)	Not considered	Berger (2007)	Not considered	Not considered	The ordinary Monte Carlo Bootstrap may lead to biased variance estimates	Not considered	Not considered
Sampling WR/WoR	WR	WoR	WR/WoR	WR/WoR	WR/WoR	WR	WR	WoR
FPC	Not considered	Considered	Possible	Possible	Possible	Not considered	Not considered	Considered

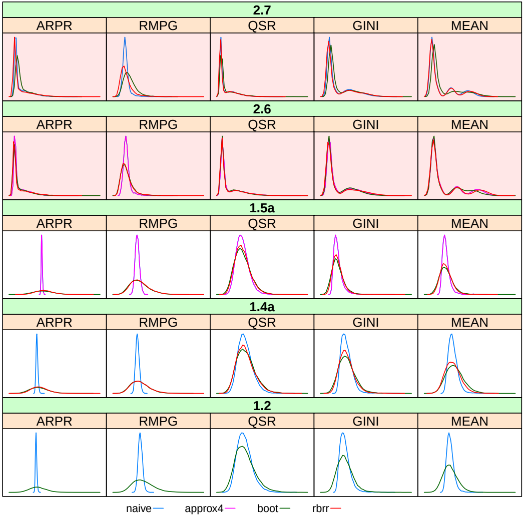
Characteristics of the AMELI universe

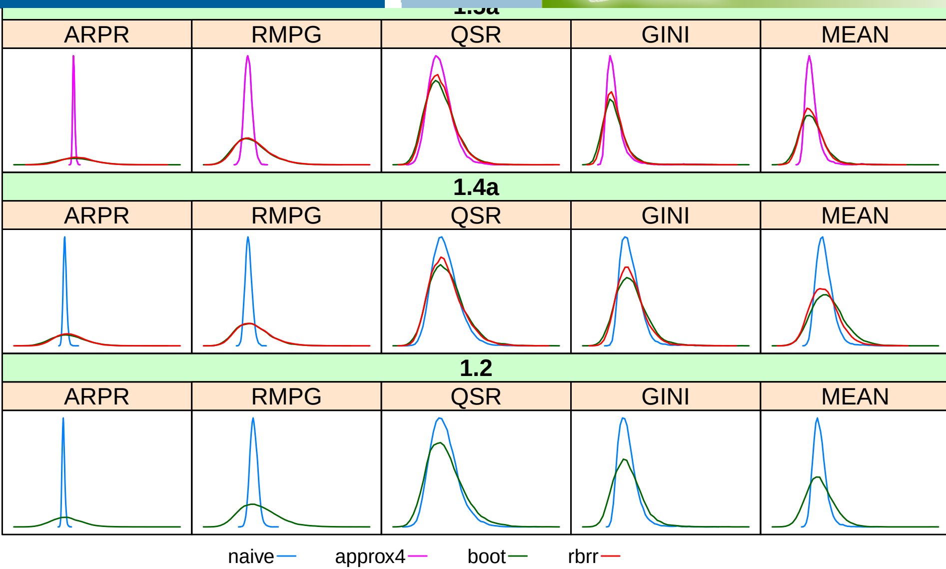


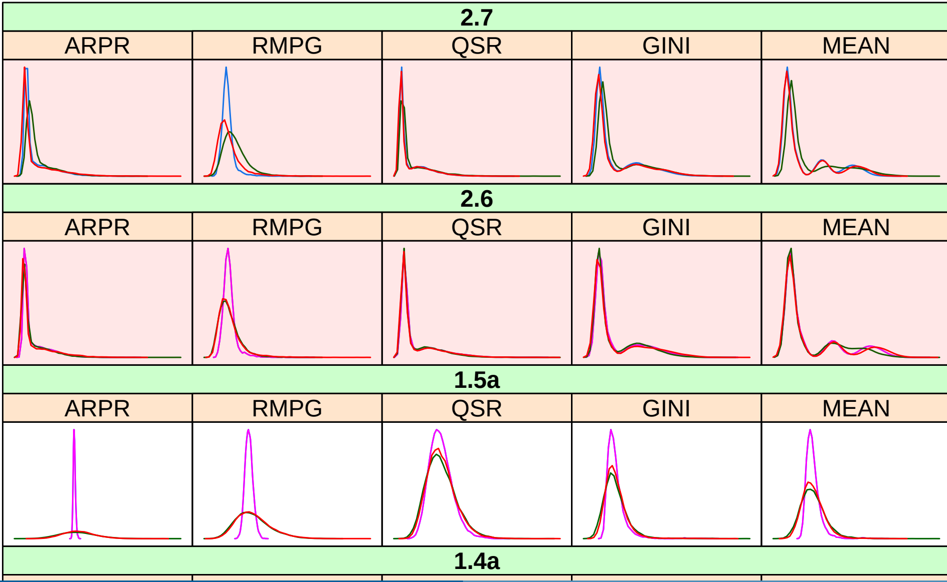

2.7
2.6
1.5a
1.4a
1.2
ARPR
RMPG
QSR
GINI
MEAN
rbrr
boot
approx4
naive
−0.5
0.5


1.5a
ARPR
RMPG
QSR
GINI
MEAN
rbrr
boot
approx4
naive
1.4a
1.2
−0.5
0.5


−0.5
0.5
−0.5
0.5
2.7
ARPR
RMPG
QSR
GINI
MEAN
rbrr
boot
approx4
naive
2.6
1.5a
1.4a


2.7
ARPR
RMPG
QSR
GINI
MEAN
2.6
ARPR
RMPG
QSR
GINI
MEAN
1.5a
ARPR
RMPG
QSR
GINI
MEAN
1.4a
ARPR
RMPG
QSR
GINI
MEAN
1.2
ARPR
RMPG
QSR
GINI
MEAN
naive
approx4
boot
rbrr


ARPR
RMPG
QSR
GINI
MEAN
1.4a
ARPR
RMPG
QSR
GINI
MEAN
1.2
ARPR
RMPG
QSR
GINI
MEAN
naive
approx4
boot
rbrr


2.7
ARPR
RMPG
QSR
GINI
MEAN
2.6
ARPR
RMPG
QSR
GINI
MEAN
1.5a
ARPR
RMPG
QSR
GINI
MEAN
1.4a

Coverage Rates (in %) of indicator estimates

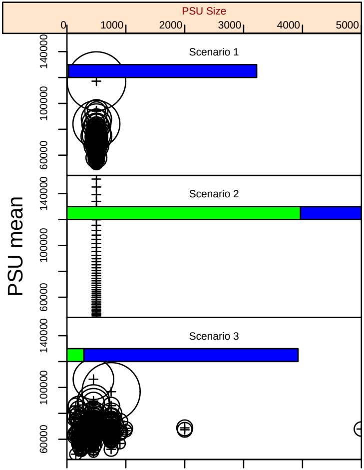
Direct/appr.	1.2	1.4a	1.5a	2.6	2.7
ARPR	95.070	94.700	94.950	89.340	90.640
RMPG	94.640	94.790	94.550	92.930	92.650
QSR	94.620	95.260	94.850	83.880	83.690
GINI	94.440	95.090	95.140	84.230	85.550
MEAN	94.850	95.070	95.320	78.720	79.960
Bootstrap	**1.2**	**1.4a**	**1.5a**	**2.6**	**2.7**
ARPR	95.100	94.910	94.810	87.850	93.070
RMPG	94.410	94.750	94.600	92.390	94.940
QSR	94.280	95.180	94.220	82.210	88.260
GINI	94.240	94.770	94.660	81.890	90.070
MEAN	94.620	95.260	95.090	77.630	90.340

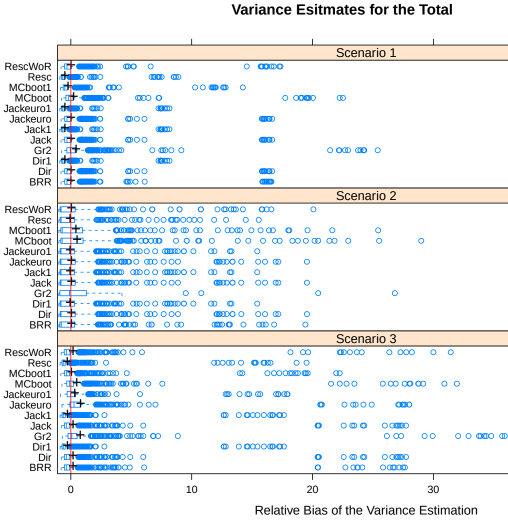
Experimental study: Sampling design

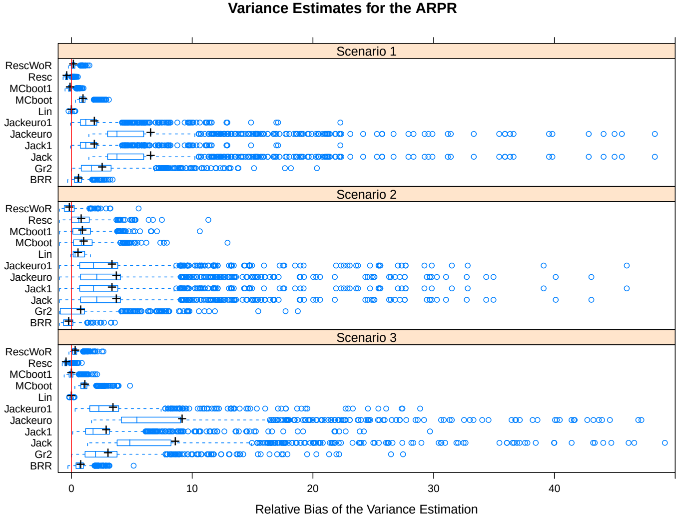
- Two stage sampling with stratification at the first stage, 25 strata
- 1. Stage: Drawing 4 PSU in each stratum (contains 8 PSU in average, altogether 200 PSU)
- 2. Stage: Proportional allocation of the sample size (1,000 USU) to the PSU (contains 500 USU in average, altogether 100,000 USU)

Experimental study: Scenarios

- *Scenario 1* : Units within PSU are heterogeneous with respect to the variable of interest $Y \sim LN(10, 1.5^2)$, PSU are of equal size
- *Scenario 2* : Units within PSU are homogeneous with respect to the variable of interest, PSU are of equal size
- *Scenario 3* : Units within PSU are heterogeneous with respect to the variable of interest $Y \sim LN(10, 1.5^2)$, PSU are of unequal size


PSU Size
0
1000
2000
3000
4000
5000
Scenario 1
Scenario 2
Scenario 3
PSU mean
60000
100000
140000


Variance Esitmates for the Total
Scenario 1
Scenario 2
Scenario 3
RescWoR
Resc
MCboot1
MCboot
Jackeuro1
Jackeuro
Jack1
Jack
Gr2
Dir1
Dir
BRR
0
10
20
30
Relative Bias of the Variance Estimation


Variance Estimates for the ARPR
Scenario 1
Scenario 2
Scenario 3
RescWoR
Resc
MCboot1
MCboot
Lin
Jackeuro1
Jackeuro
Jack1
Jack
Gr2
BRR
0
10
20
30
40
Relative Bias of the Variance Estimation

Replication weights

- Doing resampling methods by adjusting the weights
- Advantage: partial anonymization
  only the design weights are required (may not be fully true)
- BRR: Adjusting weights by

$$w_{h,i}^{(r)} := \begin{cases} w_{hi} \cdot \left[1 + \left\{ \frac{(n_h - m_h)\cdot(1-f_h)}{m_h} \right\}^{1/2} \right]; & {}_{rh} = 1; \\ w_{hi} \cdot \left[1 - \left\{ \frac{m_h \cdot (1-f_h)}{n_h - m_h} \right\}^{1/2} \right]; & {}_{rh} = -1; \end{cases}$$

  where ${}_{rh}$ indicates if the first or second group in stratum h in replication r is chosen and $m_h = \lfloor n_h{=}2 \rfloor$ (cf. Davison and Sardy, 2004)
- Delete-1-Jackknife: The weights of the deleted unit are 0, all others are computed by $\frac{n_h}{n_h - 1} \cdot w_{hi}$
- Monte-Carlo bootstrap: Computing weights by $w_{hi} \cdot c_{hi}$ where c_{hi} indicates how often unit i in stratum h is drawn with replacement

Jackknife method for EBP I

Jiang et al. (2998) and Chattopadhyay et al. (1999) propose to use the Jackknife for the estimation of the MSE of an empirical best predictor. Here reduced to an area-level EBP, they show that the MSE of $\widehat{\theta}^{\mathrm{EBP}}$ can be decomposed into:

$$MSE\left(\widehat{\theta}^{\mathrm{EBP}}\right) = MSE\left(\widehat{\theta}^{\mathrm{BP}}\right) + E\widehat{\theta}^{\mathrm{EBP}} - \widehat{\theta}^{\mathrm{BP}^2} \tag{7}$$

Jackknife method for EBP II

Further, they propose to estimate $MSE(\widehat{\theta}^{\mathrm{BP}})$ by

$$\widehat{\mathrm{MSE}}\left[\widehat{\theta}^{\mathrm{BP}}\right]_{Jack} = \widehat{\mathrm{MSE}}\left[\widehat{\theta}^{\mathrm{BP}}\right] - \frac{D-1}{D}\sum_{j=1}^{D}\left(\widehat{\mathrm{MSE}}\left[\widehat{\theta}_{-j}^{\mathrm{BP}}\right] - \widehat{\mathrm{MSE}}\left[\widehat{\theta}^{\mathrm{BP}}\right]\right) \tag{8}$$

and $E\widehat{\theta}^{\mathrm{EBP}} - \widehat{\theta}^{\mathrm{BP}^2}$ by

$$\widehat{\mathrm{E}}\left[\widehat{\theta}^{\mathrm{EBP}} - \widehat{\theta}^{\mathrm{BP}}\right]^2 = \frac{D-1}{D}\sum_{j=1}^{D}\left(\widehat{\theta}_{-j}^{\mathrm{EBP}} - \widehat{\theta}^{\mathrm{EBP}}\right)^2 \quad . \tag{9}$$

The $\widehat{\theta}_{-j}^{\mathrm{EBP}}$ is the $\widehat{\theta}^{\mathrm{EBP}}$ when omitting area j in the estimation of the model parameters of the underlying model Chattopadhyay et al. (1999), and $\widehat{\mathrm{MSE}}\left[\widehat{\theta}^{\mathrm{BP}}\right]$ depends on the BP at hand.

The Fay-Herriot estimator I

(Fay and Herriot, 1979) proposed the so called Fay-Herriot estimator (FH) for the estimation of the mean population income in a small area setting.

- Covariates only available at aggregate level.
- Covariates are true population parameters, e.g. population means $\overline{X}$.
- Direct estimates $\widehat{\mu}_{d,\text{direct}}$ are used as dependent variable.
  - Only one observation per area.
- The model they use may be expressed as

$$\widehat{\mu}_{d,\text{direct}} = \overline{X}\beta + u_d + e_d \quad .$$

$$u_d \sim \mathrm{N}(0, \sigma_u^2) \quad \text{and} \quad e_d \sim \mathrm{N}(0, \sigma_{e,d}^2)$$

The Fay-Herriot estimator II

The FH is the prediction from this mixed model and is given by

$$\widehat{\mu}_{d,\mathrm{FH}} = \overline{X}_d \widehat{\beta} + \widehat{u}_d \quad , \tag{10}$$

$$\widehat{u}_d = \frac{\widehat{\sigma}_u^2}{\widehat{\sigma}_u^2 + \sigma_{e,d}^2} (\widehat{\mu}_{d,\mathrm{direct}} - \overline{X}\widehat{\beta}) \qquad .$$

- $\widehat{\sigma}_u^2$ and $\widehat{\beta}$ are estimates
- $\sigma_{e,d}^2, d = 1..D$ are assumed to be known

The Fay-Herriot PB MSE Estimator I

There exist different analytical approximations to the MSE of the FH depending on the estimation method for σ_u^2 (cf. Datta et al., 2005).
Recalling the parametric bootstrap method for the FH

$$\mathrm{MSE}^*_{d,\mathrm{EST}} = \mathbb{E}^*\left[(\psi^*_d - \widehat{\psi}^*_{d,\mathrm{FH}})^2\right] \quad .$$

Now the right hands side is written in function of the distribution of $y|X, Z$.

$$\mathrm{MSE}^*_{d,\mathrm{EST}} = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (\psi_d - \widehat{\psi}_{d,\mathrm{EST}})^2 f_{y|X,Z}(u_1, \dots, u_D, e_1 \dots, e_D) \, du_1 \dots du_D \, de_1 \dots de_D \quad .$$

The Fay-Herriot PB MSE Estimator II

Simplifying the equation one can write $h(u) := (\psi_d - \widehat{\psi}_{d,\mathrm{FH}})^2$ and $f_{u,e} := f_{y|X,Z}$.
Then the MSE estimate obtains the form

$$\mathrm{MSE}^*_{d,\mathrm{EST}} = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h(u) f_{u,e}(u_1, \dots, u_D, e_1 \dots, e_D) \, du_1 \dots du_D \, de_1 \dots de_D .$$

- Multivariate normal probability distribution function $f_{u,e}$ does not have a closed form integral

$\mapsto$ The equation above generally will not be tractable analytically.

The Fay-Herriot PB MSE estimator III

- Two possible approaches
  - Numerical approximation (*curse of dimensionality*; cf. Donoho, 2000)
  - Monte-Carlo approximation (classical parametric bootstrap)
- It follows so far, that the parametric bootstrap may be written as a special case of a Monte-Carlo integration problem.
- Thus, methods to improve estimates gained by Monte-Carlo integration may be helpful in estimating the parametric bootstrap MSE estimate as well.

Cf. Burgard (2015) – he also proposed variance reduction techniques for resampling in small area.

Missing Data - *Everybody has them, nobody wants them*

Missingness may be either

- **MCAR** (missing completely at random),
- **MAR** (missing at random), or
- **MNAR** (missing not at random)

Rubin and Little (1987, 2002)

Methods to handle missing data

- Procedures based on the **available cases** only, i.e., only those cases that are completely recorded for the variables of interest
- **Weighting procedures** such as Horvitz-Thompson type estimators or raking estimators that adjust for nonresponse
- **Single imputation** and correction of the variance estimates to account for imputation uncertainty
- **Multiple imputation** (MI) according to Rubin (1978, 1987) and standard complete-case analysis
- **Model-based corrections** of parameter estimates such as the expectation-maximization (EM) algorithm

Resampling under non-response

- **Balanced Repeated Replication**: The method is applied as before but it comes to an adjustment of the imputed values in each resample:
  $y_{imp,i} + E\left(y_{imp,i}\right)^{*} - E\left(y_{imp,i}\right);$
  where $E\left(y_{imp,i}\right)$ is the expectation of the imputes value with respect to the original imputation procedure and $E\left(y_{imp,i}\right)^{*}$ indicates the same except that the imputation is applied in the resample (cf. Shao, Chen, und Chen, 1998).
- **Jackknife**: The same adjustments of imputed values $y_{imp,i}$ as for the BRR in each resample: $y_{imp,i} + E\left(y_{imp,i}\right)^{*} - E\left(y_{imp,i}\right)$
  (cf. Rao and Shao, 1992, Chen and Shao 2001, Shao and Tu 1995)
  Newer developments with specialised routines by Berger and Rao
  For jackknife in case of NN-Imputation cf. Chen and Shao (2001)
- **Bootstrap**: Same imputation as in the original data in each bootstrap sample
  (cf. Shao und Sitter, 1996)

Computational burdon may be very high!

Variance estimation under multiple imputation

- Multiple imputation (Rubin, 1987): $\widehat{\theta}^{(j)}$ and $\widehat{var}(\widehat{\theta}^{(j)})$
- Multiple imputation point estimate $\widehat{\theta}_{MI} = {}^{1}$

Kish's definition of the design effect

- The design effect measures the inflation of the variance of an estimator due to deviation from SRS:

$$deff_{\text{Kish}} = \frac{\frac{s_m^2}{m}}{\frac{s^2}{n}}$$

- More generally:

$$deff = \frac{V_{\text{cd}}\left(\hat{\theta}\right)}{V_{\text{SRS}}\left(\hat{\theta}\right)}$$

  where cd means complex design, i.e. the variance under the complex sampling design.

- In practice, the design effect may be used as prior information (design stage) or has to be estimated from the sample (posterior evaluation)

Estimating design effects

- The estimated design effect is:

$$\widehat{deff} = \frac{\widehat{V}_{\text{cd}}\left(\hat{\theta}\right)}{\widehat{V}_{\text{SRS}}\left(\hat{\theta}\right)}$$

- Find suitable estimators for both
  - the enumerator (classical variance estimation)
  - the denominator (non-trivial!)

Estimation of the denominator

- Problem how to compute $\widehat{V}_{\text{SRS}}\left(\hat{\theta}\right)$, because the sample is received by using the complex sampling design
- One possibility is to treat the data as if they had drawn using SRS (adjustment needed)
- In case of unequal probability sampling the denominator is computed by:

$$\widehat{V}_{\text{SRS}}\left(\hat{\theta}\right) = N^2 \frac{s^2}{n} \frac{N-n}{n}$$

$$\text{with } s^2 = \frac{1}{\sum\limits_{i \in \mathcal{S}} \pi_i^{-1} - 1} \sum_{i \in \mathcal{S}} \frac{1}{\pi_j} \left(y_i - \frac{1}{\sum\limits_{i \in \mathcal{S}} \pi_i^{-1}} \sum_{j \in \mathcal{S}} \frac{y_i}{\pi_j} \right)^2$$

- Alternatively: Make usage of parametric bootstrap (cf. Bruch, Münnich, and Seger, in submission)

Some further comments of design effects

- Use deff to determine the effective sample size:

$$n_{\text{eff}} = n \Big/ \mathit{deff}$$

- In practice, approximate priors for *deff* might be used
- The European Social Survey is a pretty good example for using design effects
  - prior determination of comparable sample sizes in a European sample
  - posterior evaluation of the effects
- Model-based design effect estimates are used for two- and multi-stage sampling (ESS context see above), cf. Ganninger (2010)
- In the meanwhile, different effects are separated, i.e. by clustering, by unequal probabilities, and by interviewers

Generalized Variance Functions (GVF)

- Instead of directly computing variance estimates, a model is evaluated at the survey estimates. The estimation of parameters is done from past data or a small subset of the items of interest.
- Useful in case of a large number of publication estimates
- Main reasons
  - Variance estimation can be computational intensive, especially when many basic estimates are considered. The same is for the publication of all variance estimates.
  - This becomes more intense when various combinations of results (e.g. ratios) are of interest.
  - When using GVF the variance is simultaneously estimated for groups of statistics rather than statistic-by-statistic. The question is if it leads to more stable estimates?

cf. Wolter(2007)

Models for variance functionals

A model has to be chosen which defines the relationship between the variance respectively the relative variance of the estimator of interest $\hat{X}$ and its expectation $X = E(X)$.

- The relative variance is defined by: $V^2 = V(\hat{X})/X^2$
- Examples of models for the relative variance:

1) $V^2 = \alpha + \beta/X$ (often used, e.g. Current Population Survey)
2) $V^2 = \alpha + \beta/X + \gamma/X^2$
3) $V^2 = (\alpha + \beta/X)^{-1}$
4) $V^2 = \left(\alpha + \beta/X + \gamma X^2\right)^{-1}$
5) $log\left(V^2\right) = \alpha - \beta log\left(X\right)$

where α, β and γ are unknown parameters that to be estimated (cf. Wolter, 2007).

Some final comments on GVF

- GVF are highly appreciated in statistical production, eg. in surveys with many variables
- Gershunskaya and Lahiri (2005) developed a small area statistics GVF for the monthly estimates of employment levels `http://www.bls.gov/osmr/pdf/st050260.pdf`
- Eltinge (2003) used GVF under non-response `http://www.amstat.org/sections/SRMS/Proceedings/y2003/Files/JSM2003-000787.pdf`

See also Valliant et al. (2013)

Final comments

- Due to the development of computer power, more and more resampling methods are applied
- The integration of different aspects into variance estimation becomes more and more attractive, e.g. non-response or statistical modeling
- GVF are still important for statistical production but need further development
- Variance estimation for change can be seen as ratio or differences of (dependent) estimates, but with changing universes
- Variance estimation (solely) does not necessarily yield best confidence intervals – always consider peculiarities in data!
- Variance estimation can be seen:
  - a priori (design-stage)
  - a posteriori (quality measurement)