Variance Reduction Methods for Parametric Bootstrap MSE-Estimation

Session: Different Inferential Issues in Area Level Models

Jan Pablo Burgard

Wirtschafts- und Sozialstatistik
Universität Trier, FB IV, VWL

04.09.2013

Statistical Challenge

- For reporting small area estimates precision measures are necessary.
- For some small area models analytical approximation to the MSE exist.
- Other models require resampling methods.

Statistical Challenge

- For reporting small area estimates precision measures are necessary.
- For some small area models analytical approximation to the MSE exist.
- Other models require resampling methods.
- One possible resampling method is the Parametric Bootstrap.

Statistical Challenge

- For reporting small area estimates precision measures are necessary.
- For some small area models analytical approximation to the MSE exist.
- Other models require resampling methods.
- One possible resampling method is the Parametric Bootstrap.
- For complex models computational expensive

Statistical Challenge

- For reporting small area estimates precision measures are necessary.
- For some small area models analytical approximation to the MSE exist.
- Other models require resampling methods.
- One possible resampling method is the Parametric Bootstrap.
- For complex models computational expensive
- **Challange** Is there a way to reduce the computational burden for PB MSE estimation?

PB MSE Estimator I

Recalling the parametric bootstrap method for estimating the MSE of a small area estimate

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

where ψ^*_d is the true value for one realisation of the superpopulation model defined by the used model, and $\widehat{\psi}^*_d$ being the estimate given the same realisation. Now the right hands side is written in function of the distribution of $y|X, Z$.

$$\mathrm{MSE}^*_{d,\mathrm{EST}} = \int\limits_{-\infty}^{\infty} \dots \int\limits_{-\infty}^{\infty} (\psi_d - \widehat{\psi}_{d,\mathrm{EST}})^2 f_{y|X,Z}(u_1, \dots, u_D, e_1 \dots, e_D)$$

$$du_1 \dots du_D \, de_1 \dots de_D \quad .$$

PB MSE Estimator II

Beautifying 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

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

- E.g. 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.

MSE Estimator III

- Two possible approaches
  - Numerical approximation (*curse of dimensionality* 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.

Variance Reduction Methods I

- The Monte-Carlo approximation of an integral often is not efficent
- Variance reduction methods try to
  - reduce the variance of the resulting estimate
  - whilst obtaining the **same** estimate as in plain Monte-Carlo

Variance Reduction Methods I

- The Monte-Carlo approximation of an integral often is not efficent
- Variance reduction methods try to
  - reduce the variance of the resulting estimate
  - whilst obtaining the **same** estimate as in plain Monte-Carlo
- If the variance is reduced it follows, that for a given precision less resamples are nedded.

$\mapsto$ Reduction of the computational burden.

Variance Reduction Methods II

- Latin Hypercube-Sampling

$\mapsto$ Did not show to improve the variance in the simulations performed

- Control Variables
- Variance reduction in bootstraps is presented by Hesterberg [1996].
- Here translated for the PB-MSE estimation

Control Variables I

Let $h(u, e)$ be the random variable produced within the parametric bootstrap. Then a function $g(u, e)$ is defined with known mean $\overline{g}$. Instead of now calculating the expectation of h via

$$\mathrm{E}\left[h(u, e)\right] = \frac{1}{R} \sum_{r=1}^{R} h(u^{(r)}, e^{(r)}) \quad ,$$

the control variate is introduced as a correction term

$$\mathrm{E}\left[h(u, e)\right]_{\mathrm{CV}} = \frac{1}{R} \sum_{r=1}^{R} h(u^{(r)}, e^{(r)}) + c\left(g(u^{(r)}, e^{(r)}) - \overline{g}\right) \quad . \quad (1)$$

As $\mathrm{E}\left[g(u^{(r)}, e^{(r)})\right] = \overline{g}$ and c is a constant it follows that $\mathrm{E}\left[c\left(g(u^{(r)}, e^{(r)}) - \overline{g}\right)\right] = 0$ and therefore $\mathrm{E}\left[h(u, e)\right]_{\mathrm{CV}} = \mathrm{E}\left[h(u, e)\right]$.

Control Variables II

The optimal constant c is given by

$$c = \frac{\mathrm{COV}\left[h(u,e),\ g(u,e)\right]}{\mathrm{V}\left[g(u,e)\right]} \tag{2}$$

Reduction of the variance by the rate of $\mathrm{COR}\left[h(u,e),\ g(u,e)\right]^2$. In practice, both $\mathrm{COV}\left[h(u,e),\ g(u,e)\right]$ and $\mathrm{V}\left[h(u,e)\right]$ are not known. Following Hesterberg [1996] these terms may be computed from the bootstrap resamples.

$$\widehat{c} = \frac{\widehat{\mathrm{COV}}\left[h(u,e),\ g(u,e)\right]}{\widehat{\mathrm{V}}\left[g(u,e)\right]} \tag{3}$$

The estimation induces a bias of order $\mathcal{O}(\frac{1}{R})$.

Control Variables III

- The central issue in order to apply this method is to define a function $g(u, e)$,
  - which has a known mean
  - and preferably a strong correlation with $h(u, e)$.
- Proof of concept a control variate for the PB-MSE estimate for the FH is derived

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

$$\widehat{\mu}_{d,\mathrm{FH}} = \overline{X}_d \widehat{\beta} + \widehat{u}_d \quad , \tag{4}$$

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

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

Control Variables for the FH I

$h(u, e)$ in the case for the estimation of a mean with the FH is given by

$$h(u,e)_{d,\mathrm{FH}} = (\widehat{\mu}^*_{d,\mathrm{FH}}(\overline{X}\widehat{\beta}, u^*, e^*) - \mu^*_d(\overline{X}\widehat{\beta}, u^*, e^*))^2 \tag{5}$$
$$= \left[\left(\overline{X}_d\widehat{\beta}^* + \gamma^*_d((\overline{X}\widehat{\beta} + u^*_d + e^*_d) - \overline{X}\widehat{\beta}^*)\right) - \overline{X}_d\widehat{\beta} + u^*_d\right]^2$$

and assuming that

$$\widehat{\beta} \approx \widehat{\beta}^*$$

Control Variables for the FH II

this may be approximated by

$$h(u,e)_{d,\mathrm{FH}} \approx \dot{h}(u,e)_{d,\mathrm{FH}} = \left(\gamma_d^*\left(u_d^* + e_d^*\right) - u_d^*\right)^2 \qquad (6)$$
$$= \left((\gamma_d^* - 1)u_d^* + \gamma_d^* e_d^*\right)^2 \quad ,$$

and by further assuming that

$$(\widehat{\sigma}_u, \widehat{\sigma}_{e,d}) \approx (\widehat{\sigma}_u^*, \widehat{\sigma}_{e,d}^*) \qquad (7)$$
$$\ddot{h}(u,e)_{d,\mathrm{FH}} = \left((\gamma_d - 1)u_d^* + \gamma_d e_d^*\right)^2 \quad ,$$

where u^* and e^* for area d are independently normally distributed with mean 0 and variances $\widehat{\sigma}_u^2$ and $\widehat{\sigma}_{e,d}^2$.

Control Variables for the FH III

Four choices for $g(u, e)$ then may be

$$g_d^{(1)}(u,e) = (u+e)^2 \qquad \bar{g}_d^{(1)} = \sigma_u^2 + \sigma_{e,d}^2 \quad , \tag{8}$$

$$g_d^{(2)}(u,e) = ((\gamma_d - 1)u + \gamma_d e)^2 \quad \bar{g}_d^{(2)} = (\gamma_d - 1)^2\sigma_u^2 + \gamma_d^2\sigma_{e,d}^2 \quad , \tag{9}$$

$$g_d^{(3)}(u,e) = (u)^2 \qquad \bar{g}_d^{(2)} = \sigma_u^2 \quad , \tag{10}$$

$$g_d^{(4)}(u,e) = (e)^2 \qquad \bar{g}_d^{(3)} = \sigma_{e,d}^2 \quad . \tag{11}$$

Control Variables for the FH IV

The correlations of these four functions with the approximation $\ddot{h}$ of h are

$$\mathrm{COR}\left[\ddot{h}(u,e)_{d,\mathrm{FH}},\, g_d^{(1)}(u,e)\right] = 0 \quad , \tag{12}$$

$$\mathrm{COR}\left[\ddot{h}_{d,\mathrm{FH}},\, g_d^{(2)}(u,e)\right] = 1 \quad , \tag{13}$$

$$\mathrm{COR}\left[\ddot{h}_{d,\mathrm{FH}},\, g_d^{(3)}(u,e)\right] = \frac{\sigma_{e,d}^2}{2(\sigma_{e,d}^2 + \sigma_u^2)} \quad , \tag{14}$$

and

$$\mathrm{COR}\left[\ddot{h}_{d,\mathrm{FH}},\, g_d^{(4)}(u,e)\right] = \frac{\sigma_u^2}{2(\sigma_{e,d}^2 + \sigma_u^2)} \quad . \tag{15}$$

Setup of the Monte-Carlo Simulation I

$$y_d \sim \mathrm{N}(x_d\beta + u_d, \sigma^2_{e,d})$$

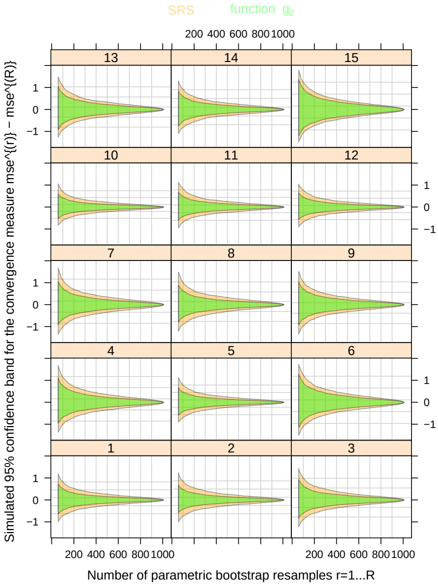
$$x_d \sim \mathrm{MVN}\left((20, 10), \begin{pmatrix} 5 & 0 \\ 0 & 3 \end{pmatrix}\right)$$

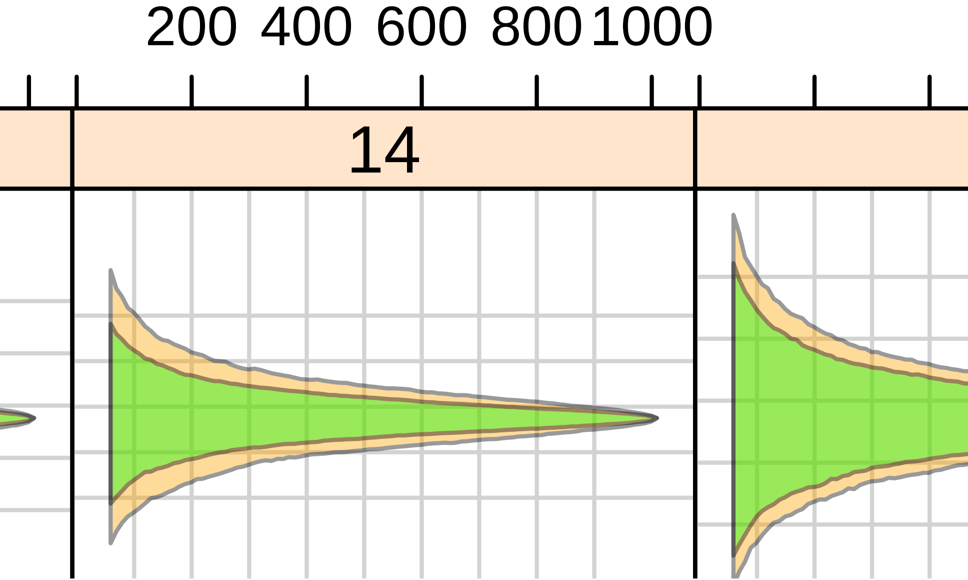
$$u_d \sim \mathrm{N}(0, \sigma^2_u)$$

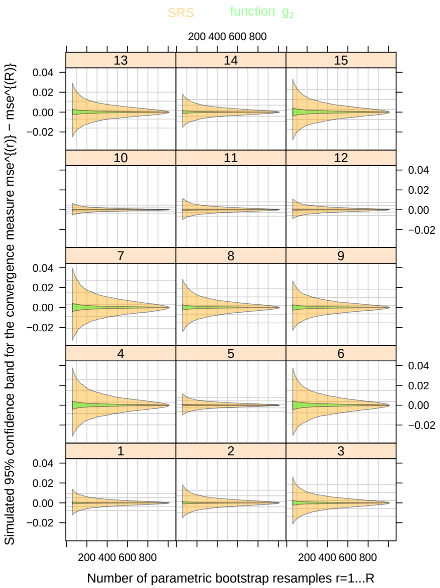
The x_d, u_d are generated only once, while the $y_d = x_d\beta + u_d + e_d$ are generated for every run randomly by drawing the e_d from a multivariate normal distribution with means zero and variance covariance matrix $(\sigma^2_{e,1}, .., \sigma^2_{e,D})I_{(D)}$.

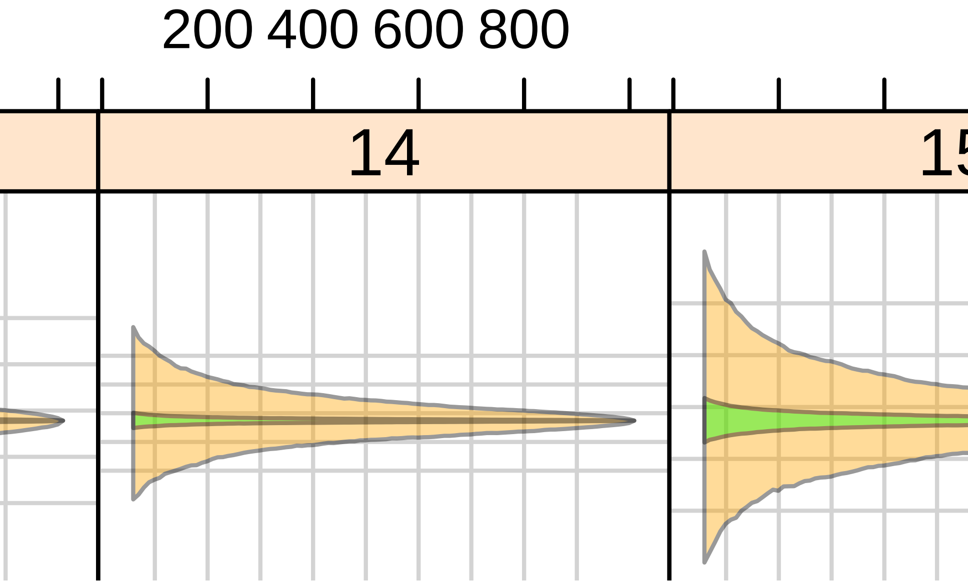
Setup of the Monte-Carlo Simulation II

population	D	$\sigma^2_{e,d}$	σ^2_u
1	15	$U(3,7)$	5
2	40	$U(3,7)$	5
3	100	$U(3,7)$	5
4	15	$U(0.01,0.1)$	15
5	40	$U(0.01,0.1)$	15
6	100	$U(0.01,0.1)$	15
7	15	$U(3,7)$	0.1
8	40	$U(3,7)$	0.1
9	100	$U(3,7)$	0.1
10	15	$U(.1,7)$	5
11	40	$U(.1,7)$	5
12	100	$U(.1,7)$	5


SRS
function g_2
13
14
15
10
11
12
7
8
9
4
5
6
1
2
3
200 400 600 800 1000
1
0
−1
Simulated 95% confidence band for the convergence measure mse^{(r)} − mse^{(R)}
Number of parametric bootstrap resamples r=1...R


200 400 600 800 1000
14


SRS
function g_2
200 400 600 800
13
14
15
10
11
12
7
8
9
4
5
6
1
2
3
0.04
0.02
0.00
−0.02
Simulated 95% confidence band for the convergence measure mse^{(r)} − mse^{(R)}
Number of parametric bootstrap resamples r=1...R


200 400 600 800
14
15

Summary and Outlook I

- The need to reduce computational burden when using parametric bootstrap MSE estimates is apparent.
- Many small area estimators require a lot of computation time for computing a single estimate.
- The use of control variates has been shown to be a computational easy implementable and reliable method.
- In some populations, the reduction of the needed resamples for a certain variability of the MSE estimate could be reduced by over 90%.
- This truly enables almost real-time computations of the parametric bootstrap MSE estimate.

Summary and Outlook II

- Only when σ_u^2 is very small, caution must be laid on the variance estimation method.
- Use generalized and adjusted maximum likelihood methods as proposed by Lahiri and Li [2009], Li and Lahiri [2007, 2010], and Yoshimori and Lahiri [2012].

D. L. Donoho. High-dimensional data analysis: The curses and blessings of dimensionality. *AMS Math Challenges Lecture*, pages 1–32, 2000.

R. E. Fay and R. A. Herriot. Estimates of income for small places: An application of james-stein procedures to census data. *Journal of the American Statistical Association*, Vol. 74, No. 366: 269–277, 1979.

Tim Hesterberg. Control variates and importance sampling for efficient bootstrap simulations. *Statistics and Computing*, 6(2):147–157, 1996. ISSN 0960-3174. doi: 10.1007/BF00162526.

P. Lahiri and H. Li. Generalized maximum likelihood method in linear mixed models with an application in small-area estimation. In *Proceedings of the Federal Committee on Statistical Methodology Research Conference*, 2009.

Huilin Li and P. Lahiri. An adjusted maximum likelihood method for solving small area estimation problems. *Journal of Multivariate Analysis*, 101:882–892, 2010. ISSN 0047-259X.

Yan Li and P Lahiri. Robust model-based and model-assisted predictors of the finite population total. *Journal of the American Statistical Association*, 102(478):664–673, 2007.

Masayo Yoshimori and Partha Lahiri. A New Adjusted Residual Likelihood Method for the Fay-Herriot Small Area Model. In *Section on Survey Research Methods at the Joint Statistical Meeting 2012*, 2012.