Bootstrap test based on data collected according to a continuous sampling design

Janusz L. Wywiał, Grzegorz Sitek

Department of Statistics, Econometrics and Mathematics
University of Economics in Katowice, Poland,
janusz.wywial@ue.katowice.pl; grzegorz.sitek@ue.katowice.pl

The 4th Congress of Polish Statistics, July 2-4, 2024,
Warsaw, Poland

Contents:

- Horvitz-Thompson statistic from sample selected according to continuous sampling design
- Inclusion function of sampling design proportional to values of auxiliary variable
- Sampling schemes
- McKay's bivariate gamma distribution
- Algorithms of the hypothesis testing
- Assessing of the test power
- Conclusions

Continuous sampling design according to Cordy (1993)

Let the population $U \subset R$. The sample space, denoted by $\boldsymbol{S}_n = U^n$, is the set of ordered samples denoted by $\mathbf{x} = (x_1, ..., x_n)$, $x_k \in U$, $k = 1, ..., n$, where x_k is the outcome of the variable observed in the k-th draw. Let $\mathbf{x}$ be a value of the n-dimensional random variable $\mathbf{X} = (X_1, ..., X_n)$ with the sampling design density function $f(\mathbf{x}) = f(x_1, ..., x_n)$. Let $f_i(x)$ and $f_{i,j}(x, x')$, $x \in U$, $x' \in U$, be marginal density functions of X_i and (X_i, X_j), respectively, $j > i = 1, ..., n$. The inclusion func.:

$$\pi(x) = \sum_{i=1}^{n} f_i(x), \quad \pi(x, x') = \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} f_{i,j}(x, x'), \quad x \in U, x' \in U \tag{1}$$

and $\int_U \pi(x)dx = n$, $\int_U \int_U \pi(x, x')dxdx' = n(n-1)$.

$$f(x_n, ..., x_i, x_{i-1}, ..., x_1) = f(x_1) \prod_{i=2}^{n} f(x_i | x_{i-1}, x_{i-2}, ..., x_1) \tag{2}$$

Continuous version of Horvitz-Thompson statistic

The estimated parameter: $\theta = \int_U g(x)dx$.

$$T_{\mathbf{X}} = \sum_{i=1}^{n} \frac{g(X_i)}{\pi(X_i)} \qquad (3)$$

***Theorem* 1.** [Cordy (1993)] The statistic $T_{\mathbf{X}}$ is an unbiased estimator for θ, if the function $g(x)$ is either bounded or non-negative, and $\pi(x) > 0$ for each $x \in U$.

***Theorem* 2** [Cordy (1993)] If the function $g(x)$ is bounded, $\pi(x) > 0$ for each $x \in U$, and $\int_U (1/\pi(x))dx < \infty$, then

$$V(T_{\mathbf{X}}) = \int_U \frac{g^2(x)}{\pi(x)} dy + \int_U \int_U g(x)g(x') \frac{\pi(x, x') - \pi(x)\pi(x')}{\pi(x)\pi(x')} dx dx'.$$

Cordy (1993) also proposed the unbiased estimator of this variance.

Moreover, in the case of simple random sample $\bar{X} = \frac{1}{n} \sum_{k=1}^{n} X_k$ is particular case of $T_{\mathbf{X}}$.

Estimation using auxiliary variable

Let $h(x,y)$, $(x,y) \in U \subseteq R^2$, be the density function. The marginal densities are: $h_1(x)$ and $h_2(y)$.
$h(y|x) = h(x,y)/h_1(x)$. $\mu_y = E(Y) = \int_{-\infty}^{-\infty} yh_2(y)dy$,
$\mu_x = E(X) = \int_{-\infty}^{-\infty} xh_1(x)dx$, $E(Y|x) = \int_{-\infty}^{-\infty} yh(y|x)dy$,
$V(Y|x) = \int_{-\infty}^{-\infty}(y - E(Y|x))^2 h(y|x)dy$.
Our purpose is estimation of parameter θ where

$$\theta = \mu_y = \int_{-\infty}^{\infty} g(x)dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yh(y|x)h_1(x)dxdy \quad (4)$$

where:

$$g(x) = E(Y|x)h_1(x) = h_1(x)\int_{-\infty}^{\infty} yh(y|x)dy. \quad (5)$$

Estimation using auxiliary variable

We replace $g(X) = E(Y|X)h_1(X)$ with $Yh_1(X)$ in the definition of $T_{\mathbf{X}}$. This leads to the following estimator of μ_y:

$$T_{\mathbf{X},\mathbf{Y}} = \sum_{i=1}^{n} \frac{Y_i h_1(X_i)}{\pi(X_i)} \tag{6}$$

We set:

$$h(\boldsymbol{y}|\boldsymbol{x}) = h(y_1, ..., y_n|x_1, ..., x_n) = \prod_{i=1}^{n} h(y_i|x_i) \tag{7}$$

Estimation using auxiliary variable

***Theorem* 3**[*Wywial* (2020)] If $E(Y) < \infty$ and $\pi(x) > 0$ for all $(x, y) \in U$ and assumption (7) holds, then $E\left(T_{\mathbf{X},\mathbf{Y}}\right) = \mu_y$.

***Theorem* 4**[*Wywial* (2020)] If the function $E(Y)$ is bounded, $\pi(y) > 0$ for each $(x, y) \in U$, and $\int_U (1/\pi(y))dy < \infty$, then

$$V\left(T_{\mathbf{X},\mathbf{Y}}\right) = \int_U \frac{V(Y|x)h_1^2(x)}{\pi(x)}dx + \int_U \frac{E^2(Y|x)h_1^2(x)}{\pi(x)}dx + A \quad (8)$$

where

$$A = \int_U \int_U E(Y|x)h_1(x)E(Y|x')h_1(x')\frac{\pi(x,x') - \pi(x)\pi(x')}{\pi(x)\pi(x')}dxdx'$$

or

$$A = \int_U \int_U E(Y|x)h_1(x)E(Y|x')h_1(x')\frac{\pi(x,x')}{\pi(x)\pi(x')}dxdx' - E^2(Y).$$

The density function of the sampling design is known:

$$f(x_1, ..., x_n) = \prod_{i=1}^{n} f(x_i), \quad f(x_i) = \frac{x_i h_1(x_i)}{\mu_x}, \quad \pi(x) = nf(x_i). \tag{9}$$

The estimated density func. of the sampl. design is known:

$$\hat{f}(\boldsymbol{x}) = f(x_1, ..., x_n, \hat{\theta}_1 ... \hat{\theta}_r) = \prod_{i=1}^{n} \hat{f}(x_i), \quad \hat{f}(x) = \frac{x h_1(\hat{\theta}_1 ... \hat{\theta}_r)}{\bar{x}}. \tag{10}$$

This and equation (6) lead to:

$$T_{\mathbf{X,Y}} = \tilde{Y}_R = \frac{\bar{x}}{n} \sum_{i \in s} \frac{Y_i}{X_i}. \tag{11}$$

The uniform kernel estimator of $h_1(x)$ leads to the following estimator of $f(x)$ (Wywial (2020)):

$$\tilde{f}(x) = \sum_{i=1}^{N} w_i \tilde{f}_i(x, x_i, \Delta), \quad \tilde{f}_i(x, x_i, \Delta) = \begin{cases} \frac{x}{2x_i\Delta}, & x \in [x_i - \Delta; x_i + \Delta], \\ 0, & x \notin [x_i - \Delta; x_i + \Delta], \end{cases}$$

where $\Delta > 0$ is the bandwidth parameter and $w_i = \frac{x_i}{N\bar{x}}$.

$$\tilde{F}(x) = \int_{-\infty}^{x} \tilde{f}(t)dt = \sum_{i=1}^{N} w_i \tilde{F}_i(x, x_i, \Delta), \qquad (12)$$

$$x = \tilde{F}_i^{-1}(u) = \sqrt{4x_i\Delta u + (x_i - \Delta)^2}, \qquad z \in [0; 1] \qquad (13)$$

where u has uniform distribution on interval $[0; 1]$. This let us easy generate the pseudovalues on interval $[x_i - \Delta; x_i + \Delta]$.

Estimation using auxiliary variable

Sampling scheme. Algorithm A

- Purpose: selection of sample $\mathbf{x}_s = [x_1, ..., x_k, ..., x_n]$ from $\mathbf{d}_{x0} = [x_1, ..., x_k, ..., x_N]$;
- values of vector $\mathbf{x}'_s = [x'_1, ..., x'_n]$ are generated by means of quantile functions $x' = \hat{F}^{-1}(u)$, where u is the value of the uniformly distributed variable on [0; 1] and

$$\hat{F}(x) = \int_{-\infty}^{x} \hat{f}(t)dt, \quad \hat{f}(x) = \frac{xh_1(\hat{\theta}_1 ... \hat{\theta}_r)}{\bar{x}} \tag{14}$$

- elements of $\mathbf{x}_s$ are selected from $\mathbf{x}$ according to

$$x_k = arg \min_{j=1,...,N} |x_j - x'_k|. \tag{15}$$

Estimation using auxiliary variable

Sampling scheme. Algorithm B

- In Algorithm I, the function $\hat{f}(x)$, $\hat{F}(x)$ is replaced by $\tilde{f}(x)$,$\tilde{F}(x)$, respectively:

$$\tilde{f}(x) = \sum_{i=1}^{N} w_i \tilde{f}_i(x, x_i, \Delta), \ \tilde{f}_i(x, x_i, \Delta) = \begin{cases} \frac{x}{2x_i\Delta}, & x \in [x_i - \Delta; x_i + \Delta], \\ 0, & x \notin [x_i - \Delta; x_i + \Delta], \end{cases}$$

$$\tilde{F}(x) = \int_{-\infty}^{x} \tilde{f}(t)dt = \sum_{i=1}^{N} w_i \tilde{F}_i(x, x_i, \Delta), \quad \Delta > 0;$$

- Select x_i with probability $\propto w_i = \frac{x_i}{N\bar{x}}$;

  next: $$x_i' = \tilde{F}_i^{-1}(u) = \sqrt{4x_i\Delta u + (x_i - \Delta)^2}, \qquad u \in [0; 1]$$

  where u has uniform distribution on interval $[0; 1]$;
- the elements of $\mathbf{x}_s$ are selected from $\mathbf{d}_{x0}$ according to

$$x_k = arg \min_{j=1,\dots,N} |x_j - x_k'|.$$

The case of McKay's bivariate gamma distribution

Let U_i, $i = 0, 1$, have gamma distributions with densities:

$$l_i(u_i) = l_i(u_i, \theta_i, c) = \frac{c^{\theta_i}}{\Gamma(\theta_i)} u_i^{\theta_i - 1} e^{-cu_i} \tag{16}$$

The joint distribution of (X, Y) where $X = U_0 + U_1$, $Y = U_0$ has density function (see McKay (1934)):

$$l(x, y) = \frac{c^{\theta_0 + \theta_1}}{\Gamma(\theta_0)\Gamma(\theta_1)} y^{\theta_0 - 1}(x - y)^{\theta_1 - 1} e^{-cx}, \quad x > y > 0. \tag{17}$$

The case of McKay's bivariate gamma distribution

We estimate $\mu_y = \frac{\theta_0}{c}$ by means the sampling design:

$$f(\boldsymbol{x}) = \prod_{i=1}^{n} f(x_i), \quad f(x_i) = \frac{x_i h_1(x_i)}{\mu_x} = \frac{c^{\theta_0+\theta_1+1}}{\Gamma(\theta_0+\theta_1+1)} x_i^{\theta_0+\theta_1} e^{-cx_i}.$$

In this case (Wywial (2020)):

$$deff(\hat{Y}_R) = \frac{V(\hat{Y}_R)}{V(\bar{Y})} = \frac{\theta_1}{\theta_0+\theta_1+1} < 1.$$

$V(\hat{Y}_R)$ can be estimated by means of the following statistics:

$$\tilde{V}(\tilde{Y}_R, \hat{f}(x)) = \frac{1}{n}\tilde{Y}_R(\bar{x} - \tilde{Y}_R)\frac{\hat{\gamma}_x^2}{1+\hat{\gamma}_x^2}. \tag{18}$$

Bootstrap-type estimators of variance:

$$\tilde{V}\left(\tilde{Y}_R\right) = \frac{1}{B-1}\sum_{k=1}^{B}\left(\tilde{Y}_R^{(k)} - \tilde{Y}_R\right)^2, \quad \tilde{Y}_R^{(k)} = \frac{\bar{x}}{n}\sum_{k=1}^{n}\frac{Y_i^{(k)}}{X_i^{(k)}} \tag{19}$$

Hypothesis and test statistics

$$H_0 : \mu_y = \mu_{y0}, \quad H_1 : \mu_y = \mu_{y1} > \mu_{y0}. \tag{20}$$

test statistics based on studentized versions of mean value estimators:

$$T_S = \frac{\bar{Y}_S - \mu_0}{\sqrt{V_S(Y)}}\sqrt{n}, \quad \hat{T}_S = \frac{\tilde{Y}_S - \mu_0}{\sqrt{\hat{V}(\tilde{Y}_S)}}, \quad \tilde{T}_S = \frac{\tilde{Y}_S - \mu_0}{\sqrt{\tilde{V}(\tilde{Y}_S)}} \tag{21}$$

where $V_S(Y) = \sum_{i \in S}(Y_i - \bar{Y}_S)^2/(n-1)$, $\bar{Y}_S = \sum_{i \in S} Y_i/n$.

Testing strategies

- The pair (testing strategy) $(\hat{T}_S, \hat{f}(x))$ is based on the statistic $\hat{T}_S$ from sample selected by means of sampling scheme explained by Algorithm A where the estimator $\hat{f}(x)$ of the density function $f(x)$ is used;
- The strategy $(\tilde{T}_S, \tilde{f}(x))$ is based on the statistic $\tilde{T}_S$ from sample selected by mean of sampling scheme explained by Algorithm B where sampling the kernel estimator $\tilde{f}(x)$ of the density function $f(x)$ is used;
- $(T_S, h_2(y))$ is the simple random sample selected from distribution of variable under study Y which density function $h_2(y)$ is the marginal distribution of (X, Y) explained by density: $h(x, y)$.

Bootstr. significant test strategy $(\hat{T}_S, \hat{f}(x))$. Algorithm 1

$d_{x0} = \{x_1, ..., x_N\}$ observations of an auxiliary variable distributed according density function $h_1(x, \hat{\theta}_1, ..., \hat{\theta}_r)$;

1. Sample $s = ((x_1, y_1), ..., (x_n, y_n))$ is selected from d_{x0} according to Algorithm A;

2. Evaluate test statistic $\hat{t}_s$ given by (21);

4. Draw the bootstrap samples of size n denoted by $s_j = ((x_1, y_1)_j, ..., (x_N, y_n)_j)$ from s, $j = 1, ..., B \geq 1000$;

5. Calculate $\hat{t}_{s_j} = \frac{\tilde{y}_{s_j} - \bar{y}_s}{\sqrt{\hat{V}(\hat{Y}_{s_j})}}$, $j = 1, ..., B$, $\hat{V}(\hat{Y}_{s_j})$ is given by (18);

6. Evaluate the critical value: $\hat{t}_s(\alpha)$ as the sample quantile of order $1 - \alpha$ on the basis of sequence $\left(\hat{t}_{s_j} \leq \hat{t}_{s_{j+1}}, j = 1, ..., B\right)$;

7. Calculate the empirical $\hat{p}$-value according to

$$\hat{p} = \frac{1}{B} \sum_{j=1}^{B} I(\hat{t}_{s_j}) \text{ where } I(\hat{t}_{s_j}) = \begin{cases} 1, & if \quad \hat{t}_{s_j} \geq \hat{t}_s, \\ 0, & if \quad \hat{t}_{s_j} < \hat{t}_s; \end{cases}$$

Bootstr. significant test strategy $(\tilde{T}_S, \tilde{f}(x))$. Algorithm 2

1. Select sample s according to Algorithm B;
2. Draw the bootstrap samples of size n denoted by $s_j = ((x_1, y_1)_j, ..., (x_N, y_n)_j)$ from s, $j = 1, ..., B \geq 1000$;
3. Evaluate statistics: $\tilde{y}_{s_j}$ according to the expression (11), $j = 1, ..., B$;
4. Calculate $\tilde{t}_s$ according to (21) where $\tilde{V}\left(\tilde{Y}_S\right)$ is given by (19);
6. Select bootstr. sampl. S_{jk} of size n from S_j, $j, k = 1, ..., B$;
7. Evaluate: $\tilde{T}_{s_j} = \frac{\tilde{Y}_{s_j} - \tilde{y}_s}{\sqrt{\tilde{V}(\tilde{Y}_{S_j})}}$, $\tilde{V}(\tilde{Y}_{S_j}) = \frac{1}{B-1}\sum_{k=1}^{B}\left(\tilde{Y}_{S_{jk}} - \tilde{Y}_{S_j}\right)^2$, $\tilde{Y}_{S_j} = \frac{\bar{x}}{n}\sum_{i \in S_j}\frac{Y_i}{X_i}$, $\tilde{Y}_{S_{jk}} = \frac{\bar{x}}{n}\sum_{i \in S_{jk}}\frac{Y_i}{X_i}$;
8. Evaluate the critical value: $\tilde{t}_s(\alpha)$ as the sample quantile of order $1 - \alpha$ on the basis of sequence $(\tilde{t}_{s_j} \leq \tilde{t}_{s_{j+1}}, j = 1, ..., B)$;
9. Calculate the empirical $\tilde{p}$-value according to $\tilde{p} = \frac{1}{B}\sum_{j=1}^{B} I(\tilde{t}_{s_j})$ where $I(\tilde{t}_{s_j}) = \begin{cases} 1, & if \quad \tilde{t}_{s_j} \geq \tilde{t}_s, \\ 0, & if \quad \tilde{t}_{s_j} < \tilde{t}_s; \end{cases}$

Power simulation of $(T_S, h_2(y))$. Algorithm 3

Let $d_i = (y_1, ...y_N)$ be population when hypot. H_i is true, $i = 0, 1$;
1. Draw the simple random sample: $s = (y_1, ..., y_n)$ from d_0;
2. Evaluate test statistic t_s given by (21);
3. Select the bootstrap samples: $s_j = (y_1^{(j)}, ..., y_n^{(j)})$ from s;
4. Evaluate $t_{s_j} = \frac{\bar{y}_{s_j} - \bar{y}_s}{\sqrt{V_{s_j}(Y)}} \sqrt{n}$, $j = 1, ..., B$;
5. Let $t_s(\alpha)$ be the $(1 - \alpha)$ quantile from $\left(t_{s_j}, j = 1, ..., B\right)$;
6. Repeat 1-5 A-times for evaluate critical value:
$\bar{t}_s(\alpha) = \frac{1}{A} \sum_{k=1}^{A} t_s^{(k)}(\alpha)$;
7. Draw the simple $s = (y_1, ..., y_n)$ from the set d_1;
8. Select the bootstrap samples: $s_j = (y_1^{(j)}, ..., y_n^{(j)})$ from s;
10. Evaluate statistics: $t'_{s_j} = \frac{\bar{y}_{s_j} - \mu_0}{\sqrt{V_{s_j}(Y)}} \sqrt{n}$, $j = 1, ..., B$;
11. Asses the power according to $\hat{\beta} = \frac{1}{B} \sum_{j=1}^{B} I(t'_{s_j})$ where

$$I(t'_{s_j}) = \begin{cases} 1, if & t'_{s_j} \geq \bar{t}_s(\alpha), \\ 0, if & t'_{s_j} < \bar{t}_s(\alpha). \end{cases}$$

Power simulation of $(\hat{T}_S, \hat{f}(x))$. Algorithm 4

1. Generate population data $d_i = ((x_1, y_1), ..., (x_N, y_n))$, $i = 0, 1$ according to the density $h(x, y, \theta_0, ..., \theta_r)$;
2. Repeat A-times steps 1-6 of the algorithm 1 in order to evaluate the mean critical value: $\hat{\hat{t}}(\alpha) = \frac{1}{A}\sum_{k=1}^{A} \hat{t}_k(\alpha)$;
3. Select sample $s = ((x_1, y_1), ..., (x_n, y_n))$ from d_1 according to Algorithm A;
4. Draw samples $s_j = ((x_1, y_1)_j, ..., (x_N, y_n)_j)$ from s, $j = 1, ..., B$;
5. Calculate $\hat{t}_{s_j} = \frac{\tilde{y}_{s_j} - \mu_0}{\sqrt{\hat{V}(\hat{Y}_{s_j})}}$, $j = 1, ..., B$, $\hat{V}(\hat{Y}_{s_j})$ is given by (18);
6. Calculate: $\hat{\beta} = \frac{1}{B}\sum_{j=1}^{B} I(\hat{t}_{s_j})$ where

$$I(\hat{t}_{s_j}) = \begin{cases} 1, if & \hat{t}_{s_j} \geq \hat{\hat{t}}(\alpha), \\ 0, if & \hat{t}_{s_j} < \hat{\hat{t}}(\alpha). \end{cases}$$

7. Repeat A-times steps 3-6 and asses the mean power $\hat{\hat{\beta}} = \frac{1}{A}\sum_{k=1}^{A} \hat{\beta}_k$.

Power simulation of $(\tilde{T}_S, \tilde{f}(x))$. Algorithm 5

1. Generate population data $d_i = ((x_1, y_1), ..., (x_N, y_n))$, $i = 0, 1$ according to the density $h(x, y, \theta_0, ..., \theta_r)$;
2. Repeat A-times steps 1-8 of the algorithm 2 in order to evaluate the mean critical value: $\tilde{\tilde{t}}(\alpha) = \frac{1}{A}\sum_{k=1}^{A} \tilde{t}_k(\alpha)$;
3. Select sample $s = ((x_1, y_1), ..., (x_n, y_n))$ from d_1 according to Algorithm B;
4. Draw samples $s_j = ((x_1, y_1)_j, ..., (x_N, y_n)_j)$ from s, $j = 1, ..., B$;
5. Calculate $\tilde{t}_{s_j} = \frac{\tilde{y}_{s_j} - \mu_0}{\sqrt{\tilde{V}(\tilde{Y}_{s_j})}}$, $j = 1, ..., B$, $\tilde{V}(\tilde{Y}_{s_j})$ is given by (19);
6. Calculate: $\tilde{\beta} = \frac{1}{B}\sum_{j=1}^{B} I(\tilde{t}_{s_j})$ where

$$I(\tilde{t}_{s_j}) = \begin{cases} 1, if & \tilde{t}_{s_j} \geq \tilde{\tilde{t}}(\alpha), \\ 0, if & \tilde{t}_{s_j} < \tilde{\tilde{t}}(\alpha). \end{cases}$$

7. Repeat A-times steps 3-6 and asses the mean power $\tilde{\tilde{\beta}} = \frac{1}{A}\sum_{k=1}^{A} \tilde{\beta}_k$.

Results of the simulation analysis

$$H_0: \ \mu_y = \mu_{y0} = \frac{\theta_0}{c} = 10, \quad H_1: \ \mu_y = \mu_{y1} > \mu_{y0}, r = 0.95.$$

$$H_0^{'}: \ \mu_y = \mu_{y0} = \frac{\theta_0}{c} = 20, \quad H_1: \ \mu_y = \mu_{y1} > \mu_{y0},$$
$$r = 0.75, r = 0.85.$$

Results of the simulation analysis H_0

The significant level: $\alpha = 0.05$

H1	Alg_3			Alg_4 r=0.95		
	n=30	n=60	n=90	n=30	n=60	n=90
$1.025m_0$	0.23	0.23	0.27	0.58	0.51	0.68
$1.05m_0$	0.26	0.42	0.44	0.77	0.88	0.97
$1.075m_0$	0.38	0.4	0.62	0.93	0.99	1
$1.1m_0$	0.54	0.66	0.79	0.98	1	1
$1.15m_0$	0.7	0.91	0.97	0.99	1	1
$1.2m_0$	0.9	0.98	1	1	1	1

Table 1. Power of the bootstrap test.Source: own calculations.

Results of the simulation analysis H_0

The significant level: $\alpha = 0.1$

H1	Alg_3			Alg_4 r=0.95		
	n=30	n=60	n=90	n=30	n=60	n=90
$1.025m_0$	0.32	0.32	0.37	0.67	0.62	0.77
$1.05m_0$	0.35	0.52	0.54	0.84	0.93	0.97
$1.075m_0$	0.48	0.5	0.71	0.96	1	1
$1.1m_0$	0.74	0.75	0.86	0.99	1	1
$1.15m_0$	0.79	0.94	0.99	1	1	1
$1.2m_0$	0.94	0.99	1	1	1	1

Table 2. Power of the bootstrap test.Source: own calculations.

Results of the simulation analysis H_0

The significant level: $\alpha = 0.2$

H1	Alg_3			Alg_4 r=0.95		
	n=30	n=60	n=90	n=30	n=60	n=90
$1.025m_0$	0.44	0.43	0.49	0.78	0.73	0.86
$1.05m_0$	0.46	0.64	0.66	0.89	0.96	0.99
$1.075m_0$	0.6	0.62	0.81	0.98	1	1
$1.1m_0$	0.75	0.83	0.92	0.99	1	1
$1.15m_0$	0.86	0.97	0.99	1	1	1
$1.2m_0$	0.97	0.99	1	1	1	1

Table 3. Power of the bootstrap test.Source: own calculations.

Results of the simulation analysis H_0'

The significant level: $\alpha = 0.05$

H1	Alg_4 r=0.85			Alg_4 r=0.75		
	n=30	n=60	n=90	n=30	n=60	n=90
$1.025m_0$	0.32	0.4	0.27	0.24	0.24	0.43
$1.05m_0$	0.43	0.65	0.86	0.46	0.62	0.73
$1.075m_0$	0.69	0.94	0.99	0.57	0.84	0.87
$1.1m_0$	0.92	0.98	1	0.78	0.98	0.93
$1.15m_0$	0.97	1	1	0.93	1	1
$1.2m_0$	1	1	1	1	1	1

Table 4. Power of the bootstrap test.Source: own calculations.

Results of the simulation analysis H_0'

The significant level: $\alpha = 0.1$

H1	Alg_4 r=0.85			Alg_4 r=0.75		
	n=30	n=60	n=90	n=30	n=60	n=90
$1.025m_0$	0.42	0.5	0.35	0.32	0.32	0.53
$1.05m_0$	0.55	0.75	0.91	0.55	0.71	0.8
$1.075m_0$	0.77	0.96	0.99	0.67	0.89	0.92
$1.1m_0$	0.95	0.99	1	0.84	0.97	0.96
$1.15m_0$	0.98	0.94	1	0.97	1	1
$1.2m_0$	1	1	1	1	1	1

Table 5. Power of the bootstrap test.Source: own calculations.

Results of the simulation analysis H_0'

The significant level: $\alpha = 0.2$

H1	Alg_4 r=0.85			Alg_4 r=0.75		
	n=30	n=60	n=90	n=30	n=60	n=90
1.025m_0	0.54	0.62	0.46	0.43	0.43	0.65
1.05m_0	0.68	0.84	0.95	0.67	0.81	0.87
1.075m_0	0.85	0.97	1	0.77	0.94	0.95
1.1m_0	0.97	0.99	1	0.9	0.98	0.98
1.15m_0	0.99	0.97	1	0.98	1	1
1.2m_0	1	1	1	1	1	1

Table 6. Power of the bootstrap test. Source: own calculations.

Results of the simulation analysis H_0'

The significant level: $\alpha = 0.05$

H1	Alg_5 r=0.85			Alg_5 r=0.75		
	n=30	n=60	n=90	n=30	n=60	n=90
1.025m_0	0.36	0.5	0.38	0.29	0.4	0.29
1.05m_0	0.53	0.61	0.65	0.44	0.52	0.61
1.075m_0	0.7	0.79	0.93	0.58	0.74	0.83
1.1m_0	0.85	0.95	0.98	0.75	0.84	0.96
1.15m_0	0.97	1	1	0.92	0.99	1
1.2m_0	1	1	1	0.98	1	1

Table 7. Power of the bootstrap test. Source: own calculations.

Results of the simulation analysis H_0'

The significant level: $\alpha = 0.1$

H1	Alg_5 r=0.85			Alg_5 r=0.75		
	n=30	n=60	n=90	n=30	n=60	n=90
$1.025m_0$	0.43	0.59	0.46	0.36	0.47	0.37
$1.05m_0$	0.61	0.68	0.72	0.51	0.60	0.69
$1.075m_0$	0.76	0.84	0.95	0.66	0.8	0.88
$1.1m_0$	0.89	0.97	0.99	0.81	0.89	0.97
$1.15m_0$	0.98	1	1	0.94	0.99	1
$1.2m_0$	1	1	1	0.99	1	1

Table 8. Power of the bootstrap test. Source: own calculations.

Results of the simulation analysis H_0'

The significant level: $\alpha = 0.2$

H1	Alg_5 r=0.85			Alg_5 r=0.75		
	n=30	n=60	n=90	n=30	n=60	n=90
$1.025m_0$	0.53	0.68	0.56	0.47	0.57	0.47
$1.05m_0$	0.7	0.76	0.8	0.61	0.7	0.78
$1.075m_0$	0.82	0.89	0.97	0.74	0.87	0.92
$1.1m_0$	0.94	0.98	1	0.87	0.93	0.98
$1.15m_0$	0.99	1	1	0.97	1	1
$1.2m_0$	1	1	1	1	1	1

Table 9. Power of the bootstrap test. Source: own calculations.

Results of the simulation analysis H_0

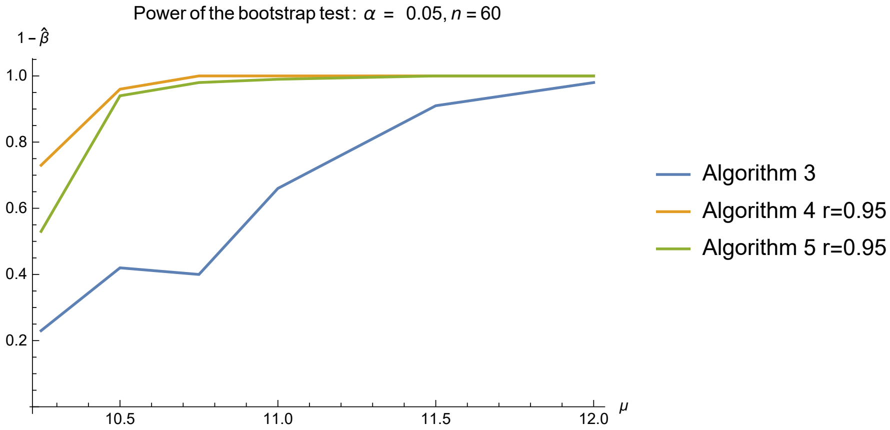

Figure 1: Power of the bootstrap test. $n = 60$, $\alpha = 0.05$. Source: Based on Tables 1, 4 and 7.

Results of the simulation analysis H_0'

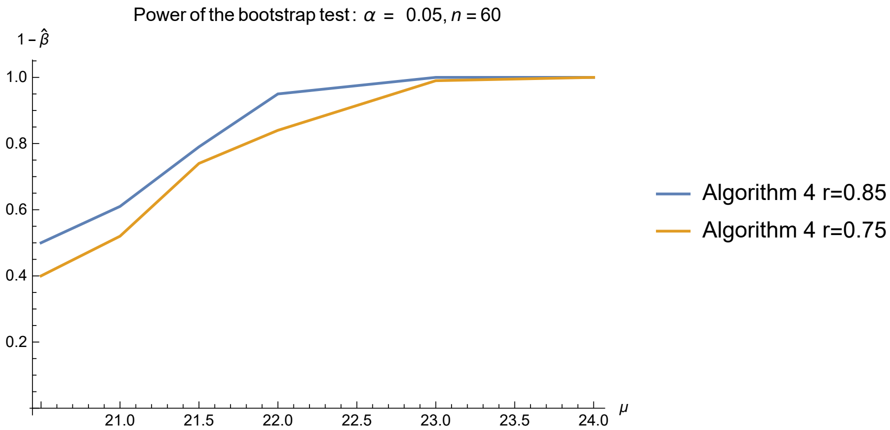

Figure 2: Power of the bootstrap test. $n = 60$, $\alpha = 0.05$. Source: Based on Tables 4.

Results of the simulation analysis H_0

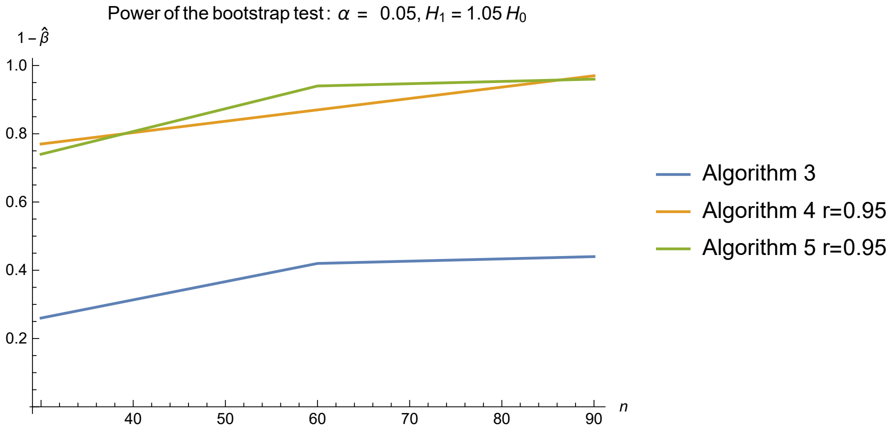

Figure 3: Power of the bootstrap test. $\alpha = 0.05$. $\mu_1 = 1.05\mu_0$. Source: Based on Tables 1, 4 and 7.

Conclusions

- The power of the mean from the simple random sample $(T_S, h_2(y))$ is shorter than the powers of $(\hat{T}_S, \hat{f}(x))$ and $(\tilde{T}_S, \tilde{f}(x))$;
- The power of $(\hat{T}_S, \hat{f}(x))$ is usually slightly better than $(\tilde{T}_S, \tilde{f}(x))$;
- The power of $(\hat{T}_S, \hat{f}(x))$ andn $(\tilde{T}_S, \tilde{f}(x))$ are close to one in the cases when the sample sizes are not very large;
- The strategy $(\hat{T}_S, \hat{f}(x))$ could be useful only in the case when the data are distributed according to McKay's bivariate gamma distribution;
- The strategy $(\tilde{T}_S, \tilde{f}(x))$ could be applied in the case when the both variable has continuous distribution. Therefore this strategy useful in practice;
- The results could be useful especially in statistical auditing where we are able to observe large number of auxiliary variable data.

Reference

- Cordy, C. B. (1993). An extension of the Horvitz-Thompson theorem to point sampling from a continuous universe. *Statistics and Probability Letters*, vol. 18, 353-362.
- Cox D. R., Snell E. J. (1979). On sampling and the estimation of rare errors. *Biometrika*, vol. 66, 1, pp. 125-32.
- Hall P. (1992). Bootstrap and Edgeworth Expansion. Springer-Verlag, New York.
- Horvitz D. G., Thompson D.J. (1952). A generalization of the sampling without replacement from finite universe. *Journal of the American Stat. Assoc.*, vol. 47, 663-685.
- Wywiał, J. L. (2020). Estimating the population mean using a continuous sampling design dependent on an auxiliary variable. *Statistics in Transition new series*, vol. 21, 5, pages: 1-16. DOI: 10.21307/stattrans-2020-052, https://sit.stat.gov.pl/Article/157
- Wywiał J. L. (2021). Sampling designs dependent on sample parameters of auxiliary variables, Second Edition. SpringerBriefs in Statistics.

Thank you for your attention