Weight trimming and weight smoothing methods in surveys

David Haziza

Département de mathématiques et de statistique
Université de Montréal

SAMSI Transition Workshop

May 6, 2014

Basic concepts

- Let U be a finite population of size N.
- Goal: estimate the population total, $t_y = \sum_{i \in U} y_i$.

Basic concepts

- Let U be a finite population of size N.
- Goal: estimate the population total, $t_y = \sum_{i \in U} y_i$.
- A sample s is selected according to a sampling design $p(s)$ with inclusion probabilities $\pi_1, \ldots, \pi_N$.

Basic concepts

- Let U be a finite population of size N.
- Goal: estimate the population total, $t_y = \sum_{i \in U} y_i$.
- A sample s is selected according to a sampling design $p(s)$ with inclusion probabilities $\pi_1, \ldots, \pi_N$.
- We consider linear estimators of t_y of the form

$$\widehat{t}_y = \sum_{i \in s} w_i y_i,$$

where w_i is the survey weight attached to unit i.

Weighting in surveys

- Each unit i is assigned a basic (or design) weight. Basic weighting system: $\{d_i = \pi_i^{-1}; i \in s\}$.

Weighting in surveys

- Each unit i is assigned a basic (or design) weight. Basic weighting system: $\{d_i = \pi_i^{-1}; i \in s\}$.

- Nonresponse adjustment: The basic weights of responding units are adjusted to compensate for the non-responding units.

  Weighting system adjusted for nonresponse: $\{w_i^* = d_i/\hat{p}_i; i \in s_r\}$, where s_r is the set of sample respondents.

Weighting in surveys

- Each unit i is assigned a basic (or design) weight. Basic weighting system: $\{d_i = \pi_i^{-1}; i \in s\}$.

- Nonresponse adjustment: The basic weights of responding units are adjusted to compensate for the non-responding units.

  Weighting system adjusted for nonresponse: $\{w_i^* = d_i/\hat{p}_i; i \in s_r\}$, where s_r is the set of sample respondents.

- Calibration adjustment: The weights w_i^* are further adjusted to match known population counts.
  Final (or calibrated) weighting system: $\{w_i = w_i^* g_i; i \in s_r\}$.

$$\text{Steps: } d_i \xrightarrow{\text{nra}} w_i^* \xrightarrow{\text{calibration}} w_i$$

Classical estimators

- In the absence of nonsampling errors, the basic weighting system $\{d_i = \pi_i^{-1}; i \in s\}$ leads to the Horvitz-Thompson estimator:

$$\widehat{t}_{HT} = \sum_{i \in s} d_i y_i.$$

Classical estimators

- In the absence of nonsampling errors, the basic weighting system $\{d_i = \pi_i^{-1}; i \in s\}$ leads to the Horvitz-Thompson estimator:

$$\widehat{t}_{HT} = \sum_{i \in s} d_i y_i.$$

- Properties of the Horvitz-Thompson estimator:
  - design-unbiased for t_y: $E_p(\widehat{t}_{HT}) = t_y$.
  - Fixed size sampling designs: if $y_i \propto \pi_i$ then $\widehat{t}_{HT} = t_y$ for any sample s; i.e.,

$$V_p(\widehat{t}_{HT}) = 0.$$

Classical estimators

- In the absence of nonsampling errors, the basic weighting system $\{d_i = \pi_i^{-1}; i \in s\}$ leads to the Horvitz-Thompson estimator:

$$\widehat{t}_{HT} = \sum_{i \in s} d_i y_i .$$

- Properties of the Horvitz-Thompson estimator:
  - design-unbiased for t_y: $E_p(\widehat{t}_{HT}) = t_y$.
  - Fixed size sampling designs: if $y_i \propto \pi_i$ then $\widehat{t}_{HT} = t_y$ for any sample s; i.e.,

$$V_p(\widehat{t}_{HT}) = 0.$$

- We expect $\widehat{t}_{HT}$ to be very efficient if y is approximately proportional to π.

Classical estimators

- Suppose that a Q-vector of auxiliary variables $\mathbf{x} = (x_1, \ldots, x_Q)^\top$ is available for $i \in s$ and that the vector of population totals (benchmarks)
$$\mathbf{t_x} = (t_{x_1}, \ldots, t_{x_Q})^\top$$
is known, where $t_{x_q} = \sum_{i \in U} x_{qi}, \quad q = 1, \ldots, Q.$

Classical estimators

- Suppose that a Q-vector of auxiliary variables $\mathbf{x} = (x_1, \ldots, x_Q)^\top$ is available for $i \in s$ and that the vector of population totals (benchmarks)

$$\mathbf{t_x} = (t_{x_1}, \ldots, t_{x_Q})^\top$$

  is known, where $t_{x_q} = \sum_{i \in U} x_{qi}, \quad q = 1, \ldots, Q$.
- In the absence of nonsampling errors, the calibrated weighting system $\{w_i = d_i g_i; i \in s\}$ leads to the calibration estimator:

$$\widehat{t}_C = \sum_{i \in s} w_i y_i,$$

  where the calibrated weights are such that

$$\sum_{i \in s} w_i \mathbf{x}_i = \mathbf{t_x}.$$

Classical estimators

- Suppose that a Q-vector of auxiliary variables $\mathbf{x} = (x_1, \ldots, x_Q)^\top$ is available for $i \in s$ and that the vector of population totals (benchmarks)
$$\mathbf{t}_\mathbf{x} = (t_{x_1}, \ldots, t_{x_Q})^\top$$
is known, where $t_{x_q} = \sum_{i \in U} x_{qi}, \quad q = 1, \ldots, Q.$
- In the absence of nonsampling errors, the calibrated weighting system $\{w_i = d_i g_i; i \in s\}$ leads to the calibration estimator:
$$\widehat{t}_C = \sum_{i \in s} w_i y_i,$$
where the calibrated weights are such that
$$\sum_{i \in s} w_i \mathbf{x}_i = \mathbf{t}_\mathbf{x}.$$
- That is, the calibrated weighting system $\{w_i; i \in s\}$ ensures consistency with known population totals.

Classical estimators

- Properties of calibration estimators:
  - asymptotically design-unbiased for t_y: $E_p(\widehat{t}_C) \approx t_y$ provided that the sample size n is large enough.
  - If $y_i = \mathbf{x}_i^\top \boldsymbol{\beta}$ for a vector of constants $\boldsymbol{\beta}$ (i.e., there exists a perfect relationship between y and $\mathbf{x}$), then

$$\widehat{t}_C = t_y$$

for any sample s; i.e.,

$$MSE_p(\widehat{t}_C) = 0.$$

Classical estimators

- Properties of calibration estimators:
  - asymptotically design-unbiased for t_y: $E_p(\widehat{t}_C) \approx t_y$ provided that the sample size n is large enough.
  - If $y_i = \mathbf{x}_i^\top \boldsymbol{\beta}$ for a vector of constants $\boldsymbol{\beta}$ (i.e., there exists a perfect relationship between y and $\mathbf{x}$), then

$$\widehat{t}_C = t_y$$

    for any sample s; i.e.,

$$MSE_p(\widehat{t}_C) = 0.$$

- We expect $\widehat{t}_C$ to be very efficient if there is a linear relationship between y and $\mathbf{x}$ and the vector $\mathbf{x}$ explains the y-variable well.

Classical estimators

- Properties of calibration estimators:
  - asymptotically design-unbiased for t_y: $E_p(\widehat{t}_C) \approx t_y$ provided that the sample size n is large enough.
  - If $y_i = \mathbf{x}_i^\top \boldsymbol{\beta}$ for a vector of constants $\boldsymbol{\beta}$ (i.e., there exists a perfect relationship between y and $\mathbf{x}$), then

$$\widehat{t}_C = t_y$$

    for any sample s; i.e.,

$$MSE_p(\widehat{t}_C) = 0.$$

- We expect $\widehat{t}_C$ to be very efficient if there is a linear relationship between y and $\mathbf{x}$ and the vector $\mathbf{x}$ explains the y-variable well.
- The class of calibration estimators includes the Generalized REGression (GREG) estimator as a special case.

Classical estimators

- Simplest special case: $x_i = 1$ for all i and $t_x = N$. In this case, $\widehat{t}_C$ reduces to the Hàjek estimator

$$\widehat{t}_{HA} = \frac{N}{\widehat{N}_{HT}} \widehat{t}_{HT}$$

with $\widehat{N}_{HT} = \sum_{i \in s} d_i$.

Classical estimators

- Simplest special case: $x_i = 1$ for all i and $t_x = N$. In this case, $\widehat{t}_C$ reduces to the Hàjek estimator

$$\widehat{t}_{HA} = \frac{N}{\widehat{N}_{HT}} \widehat{t}_{HT}$$

  with $\widehat{N}_{HT} = \sum_{i \in s} d_i$.
- Properties of the Hàjek estimator:
  - asymptotically design-unbiased for t_y: $E_p(\widehat{t}_{HA}) \approx t_y$ provided that the sample size n is large enough.
  - If $y_i = \beta$ for some constant β then $\widehat{t}_{HA} = t_y$ for any sample s; i.e.,

$$MSE_p(\widehat{t}_{HA}) = 0.$$

Classical estimators

- Simplest special case: $x_i = 1$ for all i and $t_x = N$. In this case, $\widehat{t}_C$ reduces to the Hàjek estimator

$$\widehat{t}_{HA} = \frac{N}{\widehat{N}_{HT}} \widehat{t}_{HT}$$

  with $\widehat{N}_{HT} = \sum_{i \in s} d_i$.
- Properties of the Hàjek estimator:
  - asymptotically design-unbiased for t_y: $E_p(\widehat{t}_{HA}) \approx t_y$ provided that the sample size n is large enough.
  - If $y_i = \beta$ for some constant β then $\widehat{t}_{HA} = t_y$ for any sample s; i.e.,

$$MSE_p(\widehat{t}_{HA}) = 0.$$

- We expect the Hàjek estimator to perform well if y_i and π_i are unrelated.

Large variability of the weights

- Large variability in the survey weights is often associated with unstable estimators

Large variability of the weights

- Large variability in the survey weights is often associated with unstable estimators
- Often true but not necessarily true

Large variability of the weights

- Large variability in the survey weights is often associated with unstable estimators
- Often true but not necessarily true
- Counterexamples:
    - Use of $\widehat{t}_{HT}$ if $y_i \propto \pi_i$.
    - Use of $\widehat{t}_{HA}$ if $y_i = c$.

Large variability of the weights

- Large variability in the survey weights is often associated with unstable estimators
- Often true but not necessarily true
- Counterexamples:
  - Use of $\widehat{t}_{HT}$ if $y_i \propto \pi_i$.
  - Use of $\widehat{t}_{HA}$ if $y_i = c$.
- Combination of two factors can lead to inefficient estimators:
  - High variation in the weights
  - Poor correlation between the weights and the study variables.
- Early references: Rao (1966) and Basu (1971).

Large variability of the weights

- Large variability in the survey weights is often associated with unstable estimators
- Often true but not necessarily true
- Counterexamples:
  - Use of $\widehat{t}_{HT}$ if $y_i \propto \pi_i$.
  - Use of $\widehat{t}_{HA}$ if $y_i = c$.
- Combination of two factors can lead to inefficient estimators:
  - High variation in the weights
  - Poor correlation between the weights and the study variables.
- Early references: Rao (1966) and Basu (1971).
- In surveys with multiple characteristics, it is not rare to end up with highly dispersed weights that are poorly related (or even unrelated) to some study variables.

Large variability of the weights

- Large variation in the weights may result from
  - Unequal probability sampling
  - Nonresponse adjustments
  - Calibration adjustments

Large variability of the weights

- Large variation in the weights may result from
  - Unequal probability sampling
  - Nonresponse adjustments
  - Calibration adjustments
- At the nonresponse treatment stage, forming weighting classes may help preventing an extreme variation in the weights.

Large variability of the weights

- Large variation in the weights may result from
  - Unequal probability sampling
  - Nonresponse adjustments
  - Calibration adjustments
- At the nonresponse treatment stage, forming weighting classes may help preventing an extreme variation in the weights.
- At the calibration stage, it's possible to ensure that the calibration adjustment factors lie between pre-specified lower and upper bounds through an appropriate distance function $\rightarrow$ may also help preventing an extreme variation in the weights.

What can we do to limit the impact of dispersed weights?

- A number of techniques has been proposed in the literature:
  - weight trimming methods;
  - weight smoothing methods;

What can we do to limit the impact of dispersed weights?

- A number of techniques has been proposed in the literature:
  - weight trimming methods;
  - weight smoothing methods;
- Different in nature but share a same objective:
  - modify the survey weights so that the resulting estimators have a lower mean square error than that of the usual estimators (e.g., the Horvitz-Thompson estimator)
  - Reduction of the mean square error is generally achieved at the expense of introducing a bias.
  - Treatment of survey weights: can be viewed as a compromise between bias and variance.

What is weight trimming?

- Weight trimming consists of reducing the weight of units that are identified as influential.

What is weight trimming?

- Weight trimming consists of reducing the weight of units that are identified as influential.
- Influential units are those responsible for the high variance of classical estimators; i.e., the units that contribute significantly to the instability of the estimator.

What is weight trimming?

- Weight trimming consists of reducing the weight of units that are identified as influential.
- Influential units are those responsible for the high variance of classical estimators; i.e., the units that contribute significantly to the instability of the estimator.
- The concept of influential units is not always clearly defined in practice:
  - Units with a large weight w_i?
  - Units with large weighted value $w_i y_i$?
  - Later, we define an appropriate influence measure called the conditional bias of a unit.

What is weight trimming?

- Weight trimming consists of reducing the weight of units that are identified as influential.
- Influential units are those responsible for the high variance of classical estimators; i.e., the units that contribute significantly to the instability of the estimator.
- The concept of influential units is not always clearly defined in practice:
  - Units with a large weight w_i?
  - Units with large weighted value $w_i y_i$?
  - Later, we define an appropriate influence measure called the conditional bias of a unit.
- We distinguish between two types of weight trimming methods:
  - the methods based only on the distribution of the survey weights
  - the methods based on the distribution of the weights but that also account for the study variables.

Methods based only on the distribution of the weights

- Idea: units with an extreme weight w_i are influential and are "harmful".

Methods based only on the distribution of the weights

- Idea: units with an extreme weight w_i are influential and are "harmful".
- Let w_0 be a cutpoint.

Methods based only on the distribution of the weights

- Idea: units with an extreme weight w_i are influential and are "harmful".
- Let w_0 be a cutpoint.
- The weights after trimming are defined as

$$\widetilde{w}_i = \begin{cases} w_0 & \text{if} \quad w_i \geq w_0 \\ \gamma w_i & \text{if} \quad w_i < w_0 \end{cases}$$

where γ is a rescaling factor that ensures that

$$\sum_{i \in s} \widetilde{w}_i = \sum_{i \in s} w_i.$$

Methods based only on the distribution of the weights

- Idea: units with an extreme weight w_i are influential and are "harmful".
- Let w_0 be a cutpoint.
- The weights after trimming are defined as

$$\widetilde{w}_i = \begin{cases} w_0 & \text{if} \quad w_i \geq w_0 \\ \gamma w_i & \text{if} \quad w_i < w_0 \end{cases}$$

  where γ is a rescaling factor that ensures that

$$\sum_{i \in s} \widetilde{w}_i = \sum_{i \in s} w_i.$$

- The estimator of t_y after trimming is given by

$$\widehat{t}_{trim}(w_0) = \sum_{i \in s} \widetilde{w}_i y_i.$$

Methods based only on the distribution of the weights

- How determine the cutpoint value w_0?

Methods based only on the distribution of the weights

- How determine the cutpoint value w_0?
- Adhoc method: $w_0 = c\overline{w}$ for some constant c.

Methods based only on the distribution of the weights

- How determine the cutpoint value w_0?
- Adhoc method: $w_0 = c\overline{w}$ for some constant c.
- Potter (1988, 1990) provides one of the first formal treatment of weight trimming.

Methods based only on the distribution of the weights

- How determine the cutpoint value w_0?
- Adhoc method: $w_0 = c\overline{w}$ for some constant c.
- Potter (1988, 1990) provides one of the first formal treatment of weight trimming.
- Modeling the distribution of the weights: it assumes that the reciprocal of the the scaled sampling weights follows a beta distribution:

$$f(w_i) = \frac{n\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}(1/(nw_i))^{\alpha-1}(1-1/(nw_i))^{\beta-1}$$

- The weights from the upper tail of the distribution, say where $1 - F(w_i) < .01$, are trimmed to w_0 such that $1 - F(w_0) = .01$.

Methods based only on the distribution of the weights

- The NAEP procedure (Potter, 1990): termed NAEP because of its use in the National Assessment of Educational Progress.

Methods based only on the distribution of the weights

- The NAEP procedure (Potter, 1990): termed NAEP because of its use in the National Assessment of Educational Progress.
- It consists of trimming the weights w_i which are larger than

$$w_0 = \sqrt{c \frac{\sum_{i \in s} w_i^2}{n}}$$

for a fixed c.

Methods based only on the distribution of the weights

- The NAEP procedure (Potter, 1990): termed NAEP because of its use in the National Assessment of Educational Progress.
- It consists of trimming the weights w_i which are larger than

$$w_0 = \sqrt{c \frac{\sum_{i \in s} w_i^2}{n}}$$

  for a fixed c.
- The procedure is iterated until all the weights are less or equal to the threshold.

Methods based only on the distribution of the weights

- The NAEP procedure (Potter, 1990): termed NAEP because of its use in the National Assessment of Educational Progress.
- It consists of trimming the weights w_i which are larger than

$$w_0 = \sqrt{c \frac{\sum_{i \in s} w_i^2}{n}}$$

for a fixed c.

- The procedure is iterated until all the weights are less or equal to the threshold.
- The excess weight is redistributed among the non-trimmed units.

Methods based only on the distribution of the weights

- The NAEP procedure (Potter, 1990): termed NAEP because of its use in the National Assessment of Educational Progress.
- It consists of trimming the weights w_i which are larger than

$$w_0 = \sqrt{c \frac{\sum_{i \in s} w_i^2}{n}}$$

for a fixed c.

- The procedure is iterated until all the weights are less or equal to the threshold.
- The excess weight is redistributed among the non-trimmed units.
- Value of c: to be chosen empirically by examining the distribution of $nw_i^2 / \sum_{i \in S} w_i^2$.

Methods that account for the study variable

- Potter (1990) suggested to find the cutpoint w_0 that minimizes the estimated mean square error of the trimmed estimator $\widehat{t}_{trim}(w_0)$:

$$\widehat{MSE}(\widehat{t}_{trim}) = \widehat{V}(\widehat{t}_{trim}) + \left(\widehat{t}_{trim} - \widehat{t}_{HT}\right)^2 - \widehat{V}(\widehat{t}_{trim} - \widehat{t}_{HT}).$$

Methods that account for the study variable

- Potter (1990) suggested to find the cutpoint w_0 that minimizes the estimated mean square error of the trimmed estimator $\widehat{t}_{trim}(w_0)$:

$$\widehat{MSE}(\widehat{t}_{trim}) = \widehat{V}(\widehat{t}_{trim}) + \left(\widehat{t}_{trim} - \widehat{t}_{HT}\right)^2 - \widehat{V}(\widehat{t}_{trim} - \widehat{t}_{HT}).$$

- Optimal trimmed estimator: $\widehat{t}_{trim}(w_0^{opt})$.

Methods that account for the study variable

- Potter (1990) suggested to find the cutpoint w_0 that minimizes the estimated mean square error of the trimmed estimator $\widehat{t}_{trim}(w_0)$:

$$\widehat{MSE}(\widehat{t}_{trim}) = \widehat{V}(\widehat{t}_{trim}) + \left(\widehat{t}_{trim} - \widehat{t}_{HT}\right)^2 - \widehat{V}(\widehat{t}_{trim} - \widehat{t}_{HT}).$$

- Optimal trimmed estimator: $\widehat{t}_{trim}(w_0^{opt})$.
- Unlike the previous ones, this method accounts for the y-variable being estimated.

Methods that account for the study variable

- Potter (1990) suggested to find the cutpoint w_0 that minimizes the estimated mean square error of the trimmed estimator $\widehat{t}_{trim}(w_0)$:

$$\widehat{MSE}(\widehat{t}_{trim}) = \widehat{V}(\widehat{t}_{trim}) + \left(\widehat{t}_{trim} - \widehat{t}_{HT}\right)^2 - \widehat{V}(\widehat{t}_{trim} - \widehat{t}_{HT}).$$

- Optimal trimmed estimator: $\widehat{t}_{trim}(w_0^{opt})$.
- Unlike the previous ones, this method accounts for the y-variable being estimated.
- An optimal cutpoint w_0^{opt} for a given study variable may be far from optimal for another study variable.

Methods that account for the study variable

- Potter (1990) suggested to find the cutpoint w_0 that minimizes the estimated mean square error of the trimmed estimator $\widehat{t}_{trim}(w_0)$:

$$\widehat{MSE}(\widehat{t}_{trim}) = \widehat{V}(\widehat{t}_{trim}) + \left(\widehat{t}_{trim} - \widehat{t}_{HT}\right)^2 - \widehat{V}(\widehat{t}_{trim} - \widehat{t}_{HT}).$$

- Optimal trimmed estimator: $\widehat{t}_{trim}(w_0^{opt})$.
- Unlike the previous ones, this method accounts for the y-variable being estimated.
- An optimal cutpoint w_0^{opt} for a given study variable may be far from optimal for another study variable.
- Minimizing the estimated mean square error is not an easy task, in general.

Methods that account for the study variable

- Dalén-Tambay winsorization: Dalén (1987) and Tambay (1988) considered the winsorization

$$\widetilde{y}_i = \begin{cases} y_i \text{ if } w_i y_i \leq K \\ \frac{K}{w_i} + \frac{1}{w_i}(y_i - \frac{K}{w_i}) \text{ if } w_i y_i > K \end{cases}$$

Methods that account for the study variable

- Dalén-Tambay winsorization: Dalén (1987) and Tambay (1988) considered the winsorization

$$\widetilde{y}_i = \begin{cases} y_i \text{ if } w_i y_i \leq K \\ \frac{K}{w_i} + \frac{1}{w_i}(y_i - \frac{K}{w_i}) \text{ if } w_i y_i > K \end{cases}$$

- The Dalén-Tambay winsorized estimator is given by

$$\widehat{t}_{DT}(K) = \sum_{i \in s} w_i \widetilde{y}_i = \sum_{i \in s} \widetilde{w}_i y_i,$$

where

$$\widetilde{w}_i = 1 + (w_i - 1)\frac{min\left(y_i,\ \frac{K}{w_i}\right)}{y_i} \geq 1$$

.

Methods that account for the study variable

- Dalén-Tambay winsorization: Dalén (1987) and Tambay (1988) considered the winsorization

$$\widetilde{y}_i = \begin{cases} y_i \text{ if } w_i y_i \leq K \\ \frac{K}{w_i} + \frac{1}{w_i}(y_i - \frac{K}{w_i}) \text{ if } w_i y_i > K \end{cases}$$

- The Dalén-Tambay winsorized estimator is given by

$$\widehat{t}_{DT}(K) = \sum_{i \in s} w_i \widetilde{y}_i = \sum_{i \in s} \widetilde{w}_i y_i,$$

where

$$\widetilde{w}_i = 1 + (w_i - 1)\frac{min\left(y_i,\ \frac{K}{w_i}\right)}{y_i} \geq 1$$

.

- If $min\left(y_i,\ \frac{K}{w_i}\right) = y_i$, then $\widetilde{w}_i = w_i$.

Methods that account for the study variable

- Dalén-Tambay winsorization: Dalén (1987) and Tambay (1988) considered the winsorization

$$\widetilde{y}_i = \begin{cases} y_i \text{ if } w_i y_i \leq K \\ \frac{K}{w_i} + \frac{1}{w_i}(y_i - \frac{K}{w_i}) \text{ if } w_i y_i > K \end{cases}$$

- The Dalén-Tambay winsorized estimator is given by

$$\widehat{t}_{DT}(K) = \sum_{i \in s} w_i \widetilde{y}_i = \sum_{i \in s} \widetilde{w}_i y_i,$$

where

$$\widetilde{w}_i = 1 + (w_i - 1)\frac{min\left(y_i,\ \frac{K}{w_i}\right)}{y_i} \geq 1$$

.

- If $min\left(y_i,\ \frac{K}{w_i}\right) = y_i$, then $\widetilde{w}_i = w_i$.
- For an influential unit: $\widetilde{w}_i < w_i$.

Methods that account for the study variable

- Dalén-Tambay winsorization: Dalén (1987) and Tambay (1988) considered the winsorization

$$\widetilde{y}_i = \begin{cases} y_i \text{ if } w_i y_i \leq K \\ \frac{K}{w_i} + \frac{1}{w_i}(y_i - \frac{K}{w_i}) \text{ if } w_i y_i > K \end{cases}$$

- The Dalén-Tambay winsorized estimator is given by

$$\widehat{t}_{DT}(K) = \sum_{i \in s} w_i \widetilde{y}_i = \sum_{i \in s} \widetilde{w}_i y_i,$$

where

$$\widetilde{w}_i = 1 + (w_i - 1)\frac{min\left(y_i, \frac{K}{w_i}\right)}{y_i} \geq 1$$

.

- If $min\left(y_i, \frac{K}{w_i}\right) = y_i$, then $\widetilde{w}_i = w_i$.
- For an influential unit: $\widetilde{w}_i < w_i$.
- Nice property: $\widetilde{w}_i$ cannot be smaller than 1.

Methods that account for the study variable

- How to determine the cutpoint K for either the winsorized estimator?

Methods that account for the study variable

- How to determine the cutpoint K for either the winsorized estimator?
- A bad choice of K may affect the properties of the winsorized estimators considerably $\Rightarrow$ may even exhibit a mean square error larger than that of classical estimator (e.g., Horvitz-Thompson)!

Methods that account for the study variable

- How to determine the cutpoint K for either the winsorized estimator?
- A bad choice of K may affect the properties of the winsorized estimators considerably $\Rightarrow$ may even exhibit a mean square error larger than that of classical estimator (e.g., Horvitz-Thompson)!

- First criterion: the value of K is set a priori
  - Generally, not a good idea;
  - Should be determined using a more formal criterion.
- Second criterion: Determine K that minimizes the estimated mean square error of the winsorized estimator:

Methods that account for the study variable

- How to determine the cutpoint K for either the winsorized estimator?
- A bad choice of K may affect the properties of the winsorized estimators considerably $\Rightarrow$ may even exhibit a mean square error larger than that of classical estimator (e.g., Horvitz-Thompson)!
- First criterion: the value of K is set a priori
  - Generally, not a good idea;
  - Should be determined using a more formal criterion.
- Second criterion: Determine K that minimizes the estimated mean square error of the winsorized estimator:

Methods that account for the study variable

- How to determine the cutpoint K for either the winsorized estimator?
- A bad choice of K may affect the properties of the winsorized estimators considerably $\Rightarrow$ may even exhibit a mean square error larger than that of classical estimator (e.g., Horvitz-Thompson)!

- First criterion: the value of K is set a priori
  - Generally, not a good idea;
  - Should be determined using a more formal criterion.
- Second criterion: Determine K that minimizes the estimated mean square error of the winsorized estimator:
  - Requires historical information and/or a model for the study variable;
  - Often requires simplifying assumptions;
  - Complex to implement, in general;
  - Kokic and Bell (1994), Hulliger (1995) and Rivest and Hurtubise (1995).

Methods that account for the study variable

- Third criterion: find the value of K that minimizes the maximum estimated influence:
  - Does not require historical information or a model;
  - Does not require simplifying assumptions;
  - Easy to implement in practice;
  - see Beaumont, Haziza and Ruiz-Gazen (2013) and Favre-Martinoz, Haziza and Beaumont (2014).

What is an influential unit?

- So far, units exhibiting a large weight w_i (methods of Potter) or a large weighted value $w_i y_i$ (winsorized estimator) were considered as influential.

What is an influential unit?

- So far, units exhibiting a large weight w_i (methods of Potter) or a large weighted value $w_i y_i$ (winsorized estimator) were considered as influential.
- However, these influence measures do not seem to account for
  - the sampling design;
  - the estimator used to estimate t_y;
  - the parameter to be estimated (total, ratio, etc);
- Alternative influence measure?

What is an influential unit?

- Before defining the concept of influential unit, we define that of configuration.

What is an influential unit?

- Before defining the concept of influential unit, we define that of configuration.
- A configuration is a quadruplet, which consists of
  (1) a variable of interest;
  (2) a parameter of interest (to be estimated);
  (3) a sampling design;
  (4) a point estimator.

What is an influential unit?

- Before defining the concept of influential unit, we define that of configuration.
- A configuration is a quadruplet, which consists of
  1. a variable of interest;
  2. a parameter of interest (to be estimated);
  3. a sampling design;
  4. a point estimator.
- Examples of configurations:
  - (Revenue, total revenue, stratified simple random sampling, HT estimator);

What is an influential unit?

- Before defining the concept of influential unit, we define that of configuration.
- A configuration is a quadruplet, which consists of
  (1) a variable of interest;
  (2) a parameter of interest (to be estimated);
  (3) a sampling design;
  (4) a point estimator.
- Examples of configurations:
  - (Revenue, total revenue, stratified simple random sampling, HT estimator);
  - (Revenue, total revenue, stratified simple random sampling, calibration estimator);

What is an influential unit?

- Before defining the concept of influential unit, we define that of configuration.
- A configuration is a quadruplet, which consists of
  (1) a variable of interest;
  (2) a parameter of interest (to be estimated);
  (3) a sampling design;
  (4) a point estimator.
- Examples of configurations:
  - (Revenue, total revenue, stratified simple random sampling, HT estimator);
  - (Revenue, total revenue, stratified simple random sampling, calibration estimator);
  - (Revenue, total revenue, stratified Bernoulli, HT estimator);

What is an influential unit?

- Before defining the concept of influential unit, we define that of configuration.
- A configuration is a quadruplet, which consists of
  (1) a variable of interest;
  (2) a parameter of interest (to be estimated);
  (3) a sampling design;
  (4) a point estimator.
- Examples of configurations:
  - (Revenue, total revenue, stratified simple random sampling, HT estimator);
  - (Revenue, total revenue, stratified simple random sampling, calibration estimator);
  - (Revenue, total revenue, stratified Bernoulli, HT estimator);
  - (Revenue, Portion of the annual revenue for the first quarter, stratified simple random sampling, HT estimator).

Influential unit

- Let θ be a parameter to be estimated.

Influential unit

- Let θ be a parameter to be estimated.
- Let $\widehat{\theta}$ be an estimator of θ.

Definition

A unit is influential if, given a configuration, it has a significant impact on the sampling error, $\widehat{\theta} - \theta$.

Influential unit

- Let θ be a parameter to be estimated.
- Let $\widehat{\theta}$ be an estimator of θ.

Definition

A unit is influential if, given a configuration, it has a significant impact on the sampling error, $\widehat{\theta} - \theta$.

- An unit may be influential with respect to a configuration and have no influence with respect to another configuration.
- Modifying a single element of the configuration may have a drastic impact on the influence of an unit
- How measure the influence of an unit?

Measuring the influence: the conditional bias

- Let I_i be a sample selection indicator such that $I_i = 1$ if $i \in s$ and $I_i = 0$, otherwise.

Measuring the influence: the conditional bias

- Let I_i be a sample selection indicator such that $I_i = 1$ if $i \in s$ and $I_i = 0$, otherwise.
- Let i be a sample unit (i.e., $I_i = 1$).

Measuring the influence: the conditional bias

- Let I_i be a sample selection indicator such that $I_i = 1$ if $i \in s$ and $I_i = 0$, otherwise.
- Let i be a sample unit (i.e., $I_i = 1$).
- The conditional bias attached to unit i with respect to $\widehat{\theta}$ is defined as

$$B_{1i} = E_p(\widehat{\theta} - \theta | I_i = 1).$$

Measuring the influence: the conditional bias

- Let I_i be a sample selection indicator such that $I_i = 1$ if $i \in s$ and $I_i = 0$, otherwise.
- Let i be a sample unit (i.e., $I_i = 1$).
- The conditional bias attached to unit i with respect to $\widehat{\theta}$ is defined as

$$B_{1i} = E_p(\widehat{\theta} - \theta | I_i = 1).$$

- Interpretation of B_{1i}: average of the sampling error taken among all samples that contain unit i.

Measuring the influence: the conditional bias

- Let I_i be a sample selection indicator such that $I_i = 1$ if $i \in s$ and $I_i = 0$, otherwise.
- Let i be a sample unit (i.e., $I_i = 1$).
- The conditional bias attached to unit i with respect to $\widehat{\theta}$ is defined as

$$B_{1i} = E_p(\widehat{\theta} - \theta | I_i = 1).$$

- Interpretation of B_{1i}: average of the sampling error taken among all samples that contain unit i.
- Conditional bias B_{1i}: measure of influence of a sample unit.
- see Moreno-Rebollo, Munoz-Reyes and Munoz-Pichardo (1999) and Beaumont, Haziza and Ruiz Gazen (2014).

Measuring the influence: the conditional bias

- Let i be a nonsampled unit (i.e., $I_i = 0$). The conditional bias of unit i with respect to $\widehat{\theta}$ is defined as

$$B_{0i} \quad = \quad E_p(\widehat{\theta} - \theta | I_i = 0).$$

Measuring the influence: the conditional bias

- Let i be a nonsampled unit (i.e., $I_i = 0$). The conditional bias of unit i with respect to $\widehat{\theta}$ is defined as

$$B_{0i} \quad = \quad E_p(\widehat{\theta} - \theta | I_i = 0).$$

- Interpretation of B_{0i}: average of the sampling error taken among all samples that do not contain unit i.

Measuring the influence: the conditional bias

- Let i be a nonsampled unit (i.e., $I_i = 0$). The conditional bias of unit i with respect to $\widehat{\theta}$ is defined as

$$B_{0i} \quad = \quad E_p(\widehat{\theta} - \theta | I_i = 0).$$

- Interpretation of B_{0i}: average of the sampling error taken among all samples that do not contain unit i.
- Conditional bias B_{0i}: measure of influence of a nonsampled unit.

Measuring the influence: the conditional bias

- Let i be a nonsampled unit (i.e., $I_i = 0$). The conditional bias of unit i with respect to $\widehat{\theta}$ is defined as

$$B_{0i} \quad = \quad E_p(\widehat{\theta} - \theta | I_i = 0).$$

- Interpretation of B_{0i}: average of the sampling error taken among all samples that do not contain unit i.
- Conditional bias B_{0i}: measure of influence of a nonsampled unit.
- Although, a nonsample unit can have a large influence, nothing can be done at the estimation stage because its y-value is not observed. At the estimation stage, only the influence of sample units can be curbed down.

Conditional bias: Horvitz-Thompson estimator

- Goal: estimate $t_y = \sum_{i \in U} y_i$
- Horvitz-Thompson estimator of t_y:

$$\widehat{t}_{HT} = \sum_{i \in s} d_i y_i.$$

Conditional bias: Horvitz-Thompson estimator

- Goal: estimate $t_y = \sum_{i \in U} y_i$
- Horvitz-Thompson estimator of t_y:

$$\widehat{t}_{HT} = \sum_{i \in s} d_i y_i.$$

- Conditional bias of sample unit i with respect to $\widehat{t}_{HT}$:

$$\begin{aligned} B_{1i}^{HT} &= E_p(\widehat{t}_{HT} - t_y | I_i = 1) \\ &= \sum_{j \in U} \left(\frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} \right) y_j, \end{aligned}$$

where $\pi_{ij} = P(i \in s \;\&\; j \in s)$ denotes the joint inclusion probability of units i and j in the sample.

Remarks

- The conditional bias generally depends on the second-order inclusion probabilities, $\pi_{ij} \mapsto$ take the sampling design into account.

Remarks

- The conditional bias generally depends on the second-order inclusion probabilities, $\pi_{ij} \mapsto$ take the sampling design into account.
- The conditional bias is, in general, unknown $\rightarrow$ it must be estimated

Remarks

- The conditional bias generally depends on the second-order inclusion probabilities, $\pi_{ij} \mapsto$ take the sampling design into account.
- The conditional bias is, in general, unknown $\rightarrow$ it must be estimated
- The influence of unit depends on the y-variable being estimated as well as the sampling design.
- If $\pi_i = 1$ then
$$B_{1i}^{HT} = 0.$$

Remarks

- The conditional bias generally depends on the second-order inclusion probabilities, $\pi_{ij} \mapsto$ take the sampling design into account.
- The conditional bias is, in general, unknown $\rightarrow$ it must be estimated
- The influence of unit depends on the y-variable being estimated as well as the sampling design.
- If $\pi_i = 1$ then
$$B_{1i}^{HT} = 0.$$

Conditional bias: special cases

- Simple random sampling without replacement: $\pi_i = n/N$ for all i and $\pi_{ij} = n(n-1)/N(N-1)$ for all $i \neq j$.

$$B_{1i}^{HT} = \frac{N}{(N-1)} \left(\frac{N}{n} - 1 \right) (y_i - \overline{Y}),$$

where $\overline{Y} = Y/N$ is the population mean.

Conditional bias: special cases

- Simple random sampling without replacement: $\pi_i = n/N$ for all i and $\pi_{ij} = n(n-1)/N(N-1)$ for all $i \neq j$.

$$B_{1i}^{HT} = \frac{N}{(N-1)}\left(\frac{N}{n} - 1\right)(y_i - \overline{Y}),$$

where $\overline{Y} = Y/N$ is the population mean.

- Poisson sampling: $\pi_{ij} = \pi_i \pi_j$, $i \neq j$

$$B_{1i}^{HT} = (d_i - 1)y_i = (d_i - 1)(y_i - 0)$$

Conditional bias: special cases

- Simple random sampling without replacement: $\pi_i = n/N$ for all i and $\pi_{ij} = n(n-1)/N(N-1)$ for all $i \neq j$.

$$B_{1i}^{HT} = \frac{N}{(N-1)}\left(\frac{N}{n} - 1\right)(y_i - \overline{Y}),$$

where $\overline{Y} = Y/N$ is the population mean.

- Poisson sampling: $\pi_{ij} = \pi_i \pi_j$, $i \neq j$

$$B_{1i}^{HT} = (d_i - 1)y_i = (d_i - 1)(y_i - 0)$$

Conditional bias: special cases

Example

Consider a population of size $N = 100$ such that

$$y_1 = 0, \; y_2 = \cdots = y_{99} = 500 \text{ and } y_{100} = 1000 \Rightarrow \overline{Y} = 500.$$

Conditional bias: special cases

Example

Consider a population of size $N = 100$ such that

$$y_1 = 0,\ y_2 = \cdots = y_{99} = 500 \text{ and } y_{100} = 1000 \Rightarrow \overline{Y} = 500.$$

Simple random sampling without replacement: both units 1 and 100 have a large influence because they are both far from the population mean $\overline{Y} = 500$.

$$B_{1i}^{HT} = \frac{N}{(N-1)} \left(\frac{N}{n} - 1 \right) (y_i - \overline{Y})$$

Conditional bias: special cases

Example

Consider a population of size $N = 100$ such that

$$y_1 = 0,\ y_2 = \cdots = y_{99} = 500 \text{ and } y_{100} = 1000 \Rightarrow \overline{Y} = 500.$$

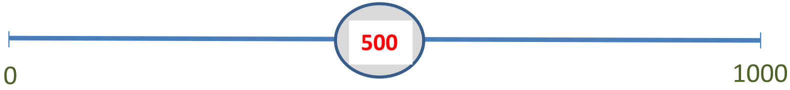
Simple random sampling without replacement: both units 1 and 100 have a large influence because they are both far from the population mean $\overline{Y} = 500$.

$$B_{1i}^{HT} = \frac{N}{(N-1)}\left(\frac{N}{n} - 1\right)(y_i - \overline{Y})$$

Poisson sampling: unit 1 has no influence, whereas unit 100 has a large influence.

$$B_{1i}^{HT} = (d_i - 1)(y_i - 0)$$

Poisson sampling vs. simple random sampling without replacement



Link between the conditional bias and the sampling error

- For Poisson sampling, the sampling error can be written as:

$$\widehat{t}_{HT} - t_y = \sum_{i \in s} B_{1i}^{HT} + \sum_{i \in U-s} B_{0i}^{HT}.$$

Link between the conditional bias and the sampling error

- For Poisson sampling, the sampling error can be written as:

$$\widehat{t}_{HT} - t_y = \sum_{i \in s} B_{1i}^{HT} + \sum_{i \in U-s} B_{0i}^{HT}.$$

- Interpretation: the conditional bias attached to unit i (sampled or not) is a measure of its contribution to the sampling error.

Link between the conditional bias and the sampling error

- For Poisson sampling, the sampling error can be written as:

$$\widehat{t}_{HT} - t_y = \sum_{i \in s} B_{1i}^{HT} + \sum_{i \in U-s} B_{0i}^{HT}.$$

- Interpretation: the conditional bias attached to unit i (sampled or not) is a measure of its contribution to the sampling error.
- For simple random sampling without replacement and high entropy sampling designs:

$$\widehat{t}_{HT} - t_y \approx \sum_{i \in s} B_{1i}^{HT} + \sum_{i \in U-s} B_{0i}^{HT}.$$

provided that the population size N is large.

Link between the conditional bias and the variance of $\widehat{t}_{HT}$

- For an arbitrary sampling design, the sampling variance of $\widehat{t}_{HT}$ is given by:

$$
\begin{aligned}
V_p\left(\widehat{t}_{HT}\right) &= \sum_{i\in U}\sum_{j\in U}\left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\right)y_iy_j \\
&= \sum_{i\in U} B_{1i}^{HT} y_i
\end{aligned}
$$

Calibration estimators

- Conditional bias of $\widehat{t}_C$ attached to sample unit i :

$$
\begin{aligned}
B_{1i}^{C} &= E_p(\widehat{t}_C - t_y | I_i = 1) \\
&\approx \sum_{j \in U} \left(\frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} \right) E_j,
\end{aligned}
$$

where $E_i = y_i - \mathbf{x}_i' \mathbf{B}$ denotes the census residual attached to unit i.

Calibration estimators

- Conditional bias of $\widehat{t}_C$ attached to sample unit i :

$$
\begin{aligned}
B_{1i}^{C} &= E_p(\widehat{t}_C - t_y | I_i = 1) \\
&\approx \sum_{j \in U} \left(\frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} \right) E_j,
\end{aligned}
$$

where $E_i = y_i - \mathbf{x}_i' \mathbf{B}$ denotes the census residual attached to unit i.

- Recall: For $\widehat{t}_C$, the conditional bias of unit i was given by the previous expression with y_i instead of E_i.

Calibration estimators

- Poisson sampling:

$$B_{1i}^{C} \approx (d_i - 1)E_i.$$

Calibration estimators

- Poisson sampling:

$$B_{1i}^{C} \approx (d_i - 1)E_i.$$

- Simple random sampling without replacement:

$$B_i^{C}\left(I_i = 1\right) \approx \frac{N}{(N-1)}\left(\frac{N}{n} - 1\right)\left(E_i - \bar{E}\right)$$

Calibration estimators

- Poisson sampling:

$$B_{1i}^{C} \approx (d_i - 1)E_i.$$

- Simple random sampling without replacement:

$$B_i^{C}\left(I_i = 1\right) \approx \frac{N}{(N-1)}\left(\frac{N}{n} - 1\right)\left(E_i - \bar{E}\right)$$

- Often, we have $\bar{E} = 0$.
- Conclusion: The conditional bias takes the type of estimator into account.

Method based on the conditional bias

- Proposed by Beaumont, Haziza and Ruiz-Gazen (2013).
- Consider a robust estimator of $\widehat{t}_{HT}$:

$$\widehat{t}_{BHR}(K) = \widehat{t}_{HT} - \sum_{i \in s} \widehat{B}_{1i}^{HT} + \sum_{i \in s} \psi_K\left(\widehat{B}_{1i}^{HT}\right)$$

where $\psi_K(\cdot)$ is the Huber function given by

$$\psi_K(x) = \begin{cases} K & \text{if } x > K \\ x & \text{if} |x| \leq K \\ -K & \text{if } x < -K \end{cases}$$

- Role of the ψ-function: reduce the influence of units that have a large influence.

Method based on the conditional bias

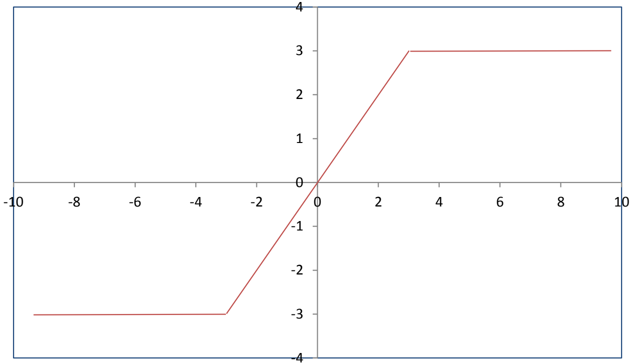
- The tuning constant K may be determined so that it minimizes the estimated mean square error of $\widehat{t}_{BHR}(K) \mapsto$ usually complex without simplifying assumptions.
- Alternative choice of K: determine K which minimizes $\max_{i \in s} \left\{ |\hat{B}^R_{1i}| \right\}$:

$$\widehat{t}_{BHR}(K_{opt}) = \widehat{t}_{HT} - \frac{1}{2}(\widehat{B}^{HT}_{min} + \widehat{B}^{HT}_{max}).$$

- Easy to implement.
- $\widehat{t}_{BHR}(K_{opt})$ is design-consistent for t_y.

Huber function $\psi_3(x)$ (with $K = 3$)

- Remark: $0 \leq \frac{\psi_K(x)}{x} \leq 1$.



Method based on the conditional bias

- The BHR estimator can be written as

$$\widehat{t}_{BHR}(K_{opt}) = \sum_{i \in s} \widetilde{d}_i y_i,$$

where

$$\widetilde{d}_i = d_i - \frac{\alpha_i}{y_i} \widehat{B}_{1i}^{HT},$$

where

$$\alpha_i = 1 - \psi_K(\widehat{B}_{1i}^{HT}) / \widehat{B}_{1i}^{HT}.$$

Method based on the conditional bias

- Poisson sampling: $B_{1i}^{HT} = (d_i - 1)y_i$.

Method based on the conditional bias

- Poisson sampling: $B_{1i}^{HT} = (d_i - 1)y_i$.

$$\widetilde{d}_i = \alpha_i + (1 - \alpha_i)d_i \Rightarrow 1 \leq \widetilde{d}_i \leq d_i$$

Method based on the conditional bias

- Poisson sampling: $B_{1i}^{HT} = (d_i - 1)y_i$.

$$\widetilde{d}_i = \alpha_i + (1 - \alpha_i)d_i \Rightarrow 1 \leq \widetilde{d}_i \leq d_i$$

- Simple random sampling without replacement (assuming $N/(N-1) = 1$):

Method based on the conditional bias

- Poisson sampling: $B_{1i}^{HT} = (d_i - 1)y_i$.

$$\widetilde{d}_i = \alpha_i + (1 - \alpha_i)d_i \Rightarrow 1 \leq \widetilde{d}_i \leq d_i$$

- Simple random sampling without replacement (assuming $N/(N-1) = 1$):

$$\widetilde{d}_i = \phi_i + \{1 - \phi_i\} \frac{N}{n},$$

where $\phi_i = \alpha_i \left(\frac{y_i - \bar{Y}}{y_i} \right)$
  - If $\alpha_i = 0 \Rightarrow \phi_i = 0 \Rightarrow \widetilde{d}_i = N/n$
  - If $\alpha_i = 1$

$$\Rightarrow \widetilde{d}_i = 1 + \left(\frac{N}{n} - 1 \right) \frac{\overline{Y}}{y_i}.$$

Weight smoothing

- Weight smoothing: introduced by Beaumont (2008).

Weight smoothing

- Weight smoothing: introduced by Beaumont (2008).
- Idea: model the weights as a function of the study variables $y_1, \ldots, y_p$. That is,

$$w_i = f(y_1, \ldots, y_p; \boldsymbol{\beta}) + \epsilon_i.$$

Weight smoothing

- Weight smoothing: introduced by Beaumont (2008).
- Idea: model the weights as a function of the study variables $y_1, \ldots, y_p$. That is,

$$w_i = f(y_1, \ldots, y_p; \boldsymbol{\beta}) + \epsilon_i.$$

- Obtain the fitted weights:

$$\widehat{w}_i = f(y_1, \ldots, y_p; \widehat{\boldsymbol{\beta}}).$$

Weight smoothing

- Weight smoothing: introduced by Beaumont (2008).
- Idea: model the weights as a function of the study variables $y_1, \ldots, y_p$. That is,

$$w_i = f(y_1, \ldots, y_p; \boldsymbol{\beta}) + \epsilon_i.$$

- Obtain the fitted weights:

$$\widehat{w}_i = f(y_1, \ldots, y_p; \widehat{\boldsymbol{\beta}}).$$

- Smoothed estimator:

$$\widehat{t}_{smooth} = \sum_{i \in s} \widehat{w}_i y_i.$$

Weight smoothing

- Weight smoothing: introduced by Beaumont (2008).
- Idea: model the weights as a function of the study variables $y_1, \ldots, y_p$. That is,

$$w_i = f(y_1, \ldots, y_p; \beta) + \epsilon_i.$$

- Obtain the fitted weights:

$$\widehat{w}_i = f(y_1, \ldots, y_p; \widehat{\beta}).$$

- Smoothed estimator:

$$\widehat{t}_{smooth} = \sum_{i \in s} \widehat{w}_i y_i.$$

Weight trimming vs.weight smoothing

- Weight trimming:trims only the weights of units that are identified as influential, whereas weight smoothing potentially modifies all the weights.

Weight trimming vs.weight smoothing

- Weight trimming:trims only the weights of units that are identified as influential, whereas weight smoothing potentially modifies all the weights.
- Weight trimming does not have to rely on the validity of a model to find the optimal cutpoint, whereas weight smoothing requires a model expressing the relationship between the weights and the study variables. If the model is wrong, the smoothed estimator may be biased.

Weight trimming vs.weight smoothing

- Weight trimming:trims only the weights of units that are identified as influential, whereas weight smoothing potentially modifies all the weights.
- Weight trimming does not have to rely on the validity of a model to find the optimal cutpoint, whereas weight smoothing requires a model expressing the relationship between the weights and the study variables. If the model is wrong, the smoothed estimator may be biased.
- Weight smoothing does require an optimal cutpoint, whereas weight trimming requires an optimal cutpoint for study variable.