Modifying Weights to Improve Survey Estimates using Regression Models

Qixuan Chen

Department of Biostatistics
Columbia University

May 6, 2014

Weighting in Surveys Working Group
SAMSI CMSS Transition Workshop

Survey Samples

- The survey sample often differs in important ways from the target population, with differences arising from
  - sampling design
  - nonresponse
  - undercoverage
- For estimation of a population quantity, a standard approach is to use sample weights to reduce bias.

Notations

- Consider a finite population U of size N
  - $Y^T = (y_1, \ldots, y_N)$, y_i denotes the survey variable
  - $I^T = (I_1, \ldots, I_N)$, I_i denotes the sample inclusion indicator
  - $Z^T = (z_1, \ldots, z_N)$, z_i denotes the design variable
  - a sample s of size n is selected from the population
- It is assumed that $P(I|Z, Y) = P(I|Z)$
  - $\pi_i = E(I_i|Z_i, Y_i) = E(I_i|Z_i)$.

Horvitz-Thompson Estimator

- Horvitz-Thompson Estimator (HT, 1952) for the population total, $t_Y = \sum_{i=1}^{N} y_i$, is

$$\hat{t}_{HT} = \sum_{i=1}^{N} I_i w_i y_i = \sum_{i=1}^{n} w_i y_i$$

  - The sample weight $w_i = 1/\pi_i$, namely the inverse of the probability of inclusion.
  - The HT estimator is design unbiased.

The HT model

- Although the HT estimator is a design-based estimator, it can be regarded as a projective estimator for the following linear model:

$$y_i = \beta\pi_i + \pi_i\epsilon_i, \ \epsilon_i \overset{iid}{\sim} N(0, \sigma^2) \tag{1}$$

where $\hat{\beta} = n^{-1} \sum_{i=1}^{N} I_i y_i / \pi_i$ and $\hat{t}_{HT} = \sum_{i=1}^{N} \hat{y}_i = \sum_{i=1}^{N} \hat{\beta}\pi_i$.

Horvitz-Thompson Estimator

- The bias reduction in the HT estimator typically comes at the cost of increased variance.
- The increase in the variance can overwhelm the reduction in the bias, especially when
  - the sample weights are highly variable
  - the association between the survey variables and the sample weights are weak

Weight trimming

- To reduce the variability in the sample weight, large weights are often trimmed.
- Weight trimming typically proceeds by
  - establishing a priori cutpoint w_0
  - modifying the weights larger than w_0 to w_0
  - multiplying the remaining weights by a normalizing constant
- The choice of cutpoint w_0 is often ad-hoc
  - weight distribution (Potter 1988, 1990)
  - the "NAEP procedures" (Berund et al. 1978)
  - the MSE minimization (Kokic and Bell 1994)

Modifying weights to improve survey estimates

- The standard weight trimming methods
  - use a single tuning constant
  - in a multipurpose survey, the choice of the tuning constant can pose problem
- We review the approaches that modify the sample weights based on regression models
  - models for the sample weight
  - models for the survey variables

Models for the sample weight

- Beaumont (2008) proposed the weight modelling method
  - the objective is to improve the HT estimator without requiring specification of a tuning constant
  - instead of only modifying the largest weights, it modifies the weights of all the sample cases
  - it models the sample weight, and the inference is conditional on the survey variables

Beaumont's procedure

- Beaumont first considered the smoothed random variable

$$\tilde{T}_{SHT} = E(\hat{T}_{HT}|I, Y) = E(\sum_{i=1}^{N} I_i w_i y_i|I, Y) = \sum_{i=1}^{n} \tilde{w}_i y_i$$

where

  - $\tilde{w}_i = E(w_i|I, Y)$ is a smoothed weight for unit $i \in s$
  - the basic idea is to reduce the variability of w_i by taking their conditional expectation
  - w_i needs to be modelled to obtained an estimate $\hat{w}_i$ of $\tilde{w}_i$, so that $\hat{T}_{SHT} = \sum_{i=1}^{n} \hat{w}_i y_i$

Beaumont's procedure

- Beaumont proposed two simple models
  - Model 1: $w_i = h_i^T\beta + \nu_i^{1/2}\epsilon_i$, $\epsilon_i \overset{iid}{\sim} N(0,\sigma^2)$, $i \in s$
    - $\tilde{w}_i = h_i^T\beta$
    - $\hat{w}_i = h_i^T\hat{\beta}$
    - h_i and ν_i are known functions of y_i
    - $\hat{w}_i$ may be smaller than 1, when the true value of w_i is greater than 1
  - Model 2: $w_i = 1 + \exp(h_i^T\beta + \nu_i^{1/2}\epsilon_i)$, $i \in s$
    - $\tilde{w}_i = 1 + \exp(h_i^T\beta)E[\exp(\nu_i^{1/2}\epsilon_i)]$
    - $\hat{w}_i = 1 + \exp(h_i^T\beta)\left\{n^{-1}\sum_{l=1}^{n}\exp[\nu_i^{1/2}\epsilon_l(\hat{\beta})]\right\}$
    - $\epsilon_l(\hat{\beta}) = \{\log(w_l - 1) - h_l^T\hat{\beta}\}/\nu_l^{1/2}$
- Beaumont showed that $V(\tilde{t}_{SHT}) \le V(\hat{t}_{HT})$.

Kim and Skinner's procedure

- Kim and Skinner (2013) extended Beaumont's procedure to linear regression
  - Consider a model $y_i = x_i^T \beta + e_i$, with $E(e_i|x_i) = 0$, $i = 1, \dots, N$
  - Estimate the weighted estimators of β by solving the estimating equation $\sum_{i=1}^{n} d_i(y_i - x_i^T \beta)x_i = 0$
  - Instead of letting $d_i = w_i$, Kim and Skinner proposed using $d_i = \tilde{w}_i q_{i*}$, where $q_{i*} = [E(\tilde{w}_i e_i^2)]^{-1}$
  - $\tilde{w}_i = \tilde{w}(x_i, y_i, \phi)$ was estimated by modeling w_i on x_i and y_i

Models for the survey variables

- Modifications of weights arising from models for the survey variables
  - weight pooling models
  - weight smoothing models
  - penalized-spline predictive models

Weight pooling models

- The weight pooling method uses a variable selection that mimics weight trimming, but the trimming cutpoint was treated as an unknown parameter (Elliott and Little 2000)

Weight pooling models

- Assuming a disproportionately stratified or poststratified sample design, to estimate a population mean, they use

$$\begin{aligned} y_{hi}|\mu_h, \sigma^2 &\sim N(\mu_h, \sigma^2),\ h < l \\ y_{hi}|\mu_h, \sigma^2 &\sim N(\mu_l, \sigma^2),\ h \geq l \\ P(L = l) &= 1/H \\ p(\sigma^2|l) &= (1/\sigma^2)^{l/2+1} \\ p(\beta|\sigma^2, l) &= (2\pi)^{-l} \end{aligned} \tag{2}$$

where $\mu_1 = \beta_0, \ldots, \mu_l = \beta_0 + \beta_{l-1}$, $h = 1, \ldots, H$ orders the strata by their associated weight from smallest to largest.

Weight pooling models

- The posterior mean of the population mean is given by

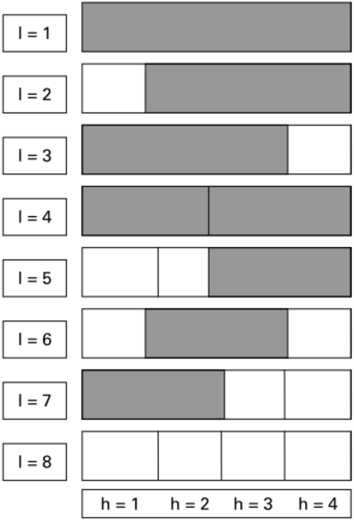
$$E(\bar{Y}|y) = N^{-1} \sum_{l=1}^{H} \left(\sum_{h=1}^{l-1} N_h \bar{y}_h + \bar{y}_l \sum_{h=l}^{H} N_h \right) P(L = l|y) \quad (3)$$

  where
  - $\bar{y}_l = \left(\sum_{h=l}^{H} n_h\right)^{-1} \sum_{h=l}^{H} n_h \bar{y}_h$ is the pooled mean for the trimmed strata,
  - $P(L = l|y)$ is determined via Bayesian variable selection method.

Weight pooling models

- Elliott (2008, 2009) extended the weight pooling model to linear and generalized regression models by
  - allowing for interactions between the regression parameters and the weight strata
  - allowing pooling of all coterminous weight strata - more robust than the previously developed
- Elliott found the weight pooling models approximate
  - fully weighted estimators when bias reduction is important
  - unweighted estimators when variance reduction is important

Weight pooling models



Weight smoothing models

- The weight smoothing method is to model the weight stratum means directly as random effects:

$$\begin{aligned} y_{hi}|\mu_h &\sim N(\mu_h, \sigma^2) \\ \boldsymbol{\mu} &\sim N_H(\phi, D) \\ p(\phi, D) &\propto 1 \end{aligned} \tag{4}$$

  where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_H)$, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_H)$, h indexes the weight strata with constant inclusion probabilities within each stratum.
- Different from the weight pooling method, the weight strata do not need to be ordered by probability of inclusion.

Weight smoothing models

- Elliott and Little (2000) considered four special cases of the model
  - Exchangeable random effects: $\phi_h = \mu$ for all h, $D = \tau^2 \boldsymbol{I}_H$
  - Autoregressive: $\phi_h = \mu$ for all h, $D = \tau^2 \{\rho^{|i-j|}\}$
  - Linear: $\phi_h = \alpha + \beta h$, $D = \tau^2 \boldsymbol{I}_H$
  - Nonparametric: $\phi_h = f(h)$, $D = 0$, and $f(h)$ minimizes $\sum_h \sum_i (y_{hi} - f(h))^2 + \lambda \int (f^{(2)}(\mu))^2 du$

Weight smoothing models

- The posterior mean of the population mean is given by

$$E(\bar{Y}|\boldsymbol{y}) = N^{-1}\sum_{h=1}^{H}(n_h\bar{y}_h + (N_h - n_h)\hat{\mu}_h) \qquad (5)$$

where

  - $\hat{u}_h = E(\mu_h|\boldsymbol{y})$,
  - The weight smoothing models allow compromises between weighted and unweighted estimates through the values of τ^2 and λ.

Weight smoothing models

- Elliott (2007) extended the weight smoothing models to linear and generalized linear regression models.
- Elliott and Little (2000) found that the weight smoothing model based on a nonparametric spline function for the underlying weight stratum means performs well in simulations as compared with alternative estimators including weight pooling and weight smoothing models.

Penalized-spline predictive models

- To estimate the population mean, Zheng and Little (2003) considered a penalized-spline predictive model

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j \pi_i^j + \sum_{l=1}^{m} \beta_{l+p} (\pi_i - \kappa_l)_+^p + \epsilon_i \qquad (6)$$

$$\beta_{l+p} \overset{iid}{\sim} N(0, \tau^2), l = 1, \ldots, m; \; \epsilon_i \overset{ind}{\sim} N(0, \pi_i^{2k} \sigma^2)$$

where

- π_i is the probability of inclusion for sample case i
- the constants $\kappa_1 < \cdots < \kappa_m$ are selected fixed knots
- $(u)_+^p = u^p \boldsymbol{I}(u \geq 0)$
- k (usually taking $0, .5, 1$) models error heteroscedasticity

Penalized-spline predictive models

- The penalized-spline predictive model can be viewed as an extension of the model implied by the HT estimator
- The model-based predictive estimator of the population total is

$$\hat{T} = \sum_{i=1}^{n} y_i + \sum_{i=n+1}^{N} E(Y_i|\pi_i) \tag{7}$$

where $E(Y_i|\pi_i) = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j \pi_i^j + \sum_{l=1}^{m} \hat{\beta}_{l+p}(\pi_i - \kappa_l)_+^p$

Penalized-spline predictive models

- The parameters in the model were estimated
  - using restricted maximum likelihood algorithm
  - implemented with standard software such as SAS Proc Mixed and the S-plus function lme()
- The variance of $\hat{T}$ was estimated using (Zheng and Little 2005)
  - model-based empirical Bayes estimator
  - jackknife estimator (recommended)
  - balanced repeated replicate (BRR) estimator

Penalized-spline predictive models

- Chen, Elliott, and Little (2012) extended the penalized spline model:

$$Y_i \overset{ind}{\sim} N\Big(f_1(\pi_i, m_1), \sigma_i^2\Big) \tag{8}$$
$$\log(\sigma_i^2) \overset{ind}{\sim} N(f_2(\pi_i, m_2), \sigma_A^2\Big)$$

where,

$f_1(\pi_i, m_1) = \beta_0 + \sum_{j=1}^{p_1} \beta_j \pi_i^j + \sum_{l=1}^{m_1} b_l(\pi_i - \kappa_l)_+^{p_1}, \boldsymbol{b} \sim N_L(0, \tau_b^2 \boldsymbol{I})$

$f_2(\pi_i, m_2) = \lambda_0 + \sum_{j=1}^{p_2} \lambda_j \pi_i^j + \sum_{l=1}^{m_2} \nu_l(\pi_i - \kappa_l)_+^{p_2}, \boldsymbol{\nu} \sim N_L(0, \tau_\nu^2 \boldsymbol{I})$

and

$\boldsymbol{\beta}, \boldsymbol{\lambda} \sim N(0, 10^6),\ \tau_b^2, \tau_\nu^2 \sim IG(0.01, 0.01)$

Penalized-spline predictive models

- The predictive estimator for the population mean and α-quantile:

$$\hat{\mu} = N^{-1}\left(\sum_{i=1}^{n} y_i + \sum_{j=n+1}^{N} \hat{y}_j\right) \tag{9}$$

$$\hat{\theta}(\alpha) = inf\left\{t; N^{-1}\left(\sum_{i=1}^{n} \Delta(t - y_i) + \sum_{j=n+1}^{N} \Delta(t - \hat{y}_j)\right) \geq \alpha\right\} \tag{10}$$

  - $\hat{y}_j$ was a draw from its posterior predictive distribution
  - the posterior mean and posterior variance were used

Penalized-spline predictive models

- Chen, Elliott, and Little (2010) also extended the penalized spline model to estimate the population proportion:

$$\Phi^{-1}(P(y_i = 1)) = \beta_0 + \sum_{j=1}^{p} \beta_j \pi_i^j + \sum_{l=1}^{m} b_l(\pi_i - \kappa_l)_+^p \qquad (11)$$

$$b_l \overset{iid}{\sim} N(0, \tau^2), l = 1, \ldots, m$$

$$\beta \sim N(0, 10^6),\ \tau^2 \sim IG(0.01, 0.01)$$

Penalized-spline predictive models

- The predictive estimator for the population proportion:

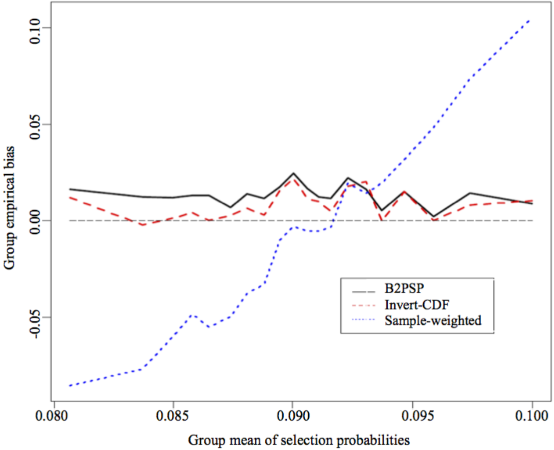
$$\hat{p} = N^{-1}\left(\sum_{i=1}^{n} y_i + \sum_{j=n+1}^{N} \hat{y}_j\right) \quad (12)$$

  - $\hat{y}_j$ was a draw from its posterior predictive distribution
  - the posterior mean and posterior variance were used

Penalized-spline predictive models

- Simulations showed that
  - The p-spline predictive estimators are generally more efficient than the HT estimator in terms of the mean squared error.
  - In simulations that most favor the HT estimator, the p-spline predictive estimators have comparable efficiency.
  - The jackknife variance estimator (Zheng and Little 2005) and the hierarchical Bayes models (Chen, Elliott, and Little 2010, 2012) both provided 95% CI with better confidence coverages and shorter interval widths.
  - The conditional bias of the p-spline predictive estimators is more stable than that of the HT estimator (Chen, Elliott, and Little, 2012).

Penalized-spline predictive models



Models for the sample weight

- Weight modeling approach can be used to remove the noise in the sample weights through an appropriate model for the sample weight
    - the validity of a model is required to obtain valid inferences
    - nonparametric models for the estimation of smoothed weights could be useful
- The objective of weight modeling is to improve the HT estimator. The smoothed HT estimator is a design-based estimator.

Models for the survey variables

- Weight modification approach arising from models for the survey variables yields model-based estimators that have smaller mean squared errors than the HT estimator.
- Their 95% CI generally have better confidence coverage and shorter interval.

Pooling, smoothing, or penalized-spline models?

- When to use weight pooling and smoothing models
  - disproportionately stratified sample design
  - poststratified sample
- When to use penalized-spline predictive models
  - probability-proportional-to-size (PPS) sample
  - sample weights for nonresponse adjustments
  - it requires that probability of inclusion is known for all the units in the population

References

1. Beaumont, J-F. (2008). A new approach to weighting and inference in sample surveys. *Biometrika*, **95**(3), 539-553.
2. Chen, Q, Elliott, MR, Little, RJA. (2010). Bayesian penalized spline model-based inference for finite population proportion in unequal probability sampling. *Survey Methodology*, **36**(1), 23-34.
3. Chen, Q, Elliott, MR, Little, RJA. (2012). Bayesian inference for finite population quantiles from unequal probabilty samples. *Survey Methodology*, **38**(2), 203-214.
4. Elliott, MR, Little, EJA. (2000). Model-based alternatives to trimming survey weights. *Journal of Official Statistics*, **16**(3), 191-209.
5. Elliott, MR. (2007). Bayesian weight trimming for generalized linear regression models. *Survey Methodology*, **33**(1), 23-34.
6. Elliott, MR. (2008). Model averaging methods for weight trimming. *Journal of Official Statistics*, **24**(4), 517-540.
7. Elliott, MR. (2009). Model averaging methods for weight trimming in generalized linear regression models. *Journal of Official Statistics*, **25**(1), 1-20.
8. Kim, JK, Skinner, CJ. (2013). Weighting in survey analysis under informative sampling. *Biometrika*, **100**(2), 385-398.
9. Zheng, H, Little, RJA. (2003). Penalized spline model-based estimation of the finite populations total from probability-proportional-to-size samples. *Journal of Official Statistics*, **19**(2), 99-117.
10. Zheng, H, Little, RJA. (2005). Inference for the population total from probability-proportional-to-size samples based on predictions from a penalized spline nonparametric model. *Journal of Official Statistics*, **21**(1), 1-20.

Thank you!