Effect of extreme design weights on small area estimates

Michael Elliott[1,2]

[1]Department of Biostatistics, [2]Survey Methodology Program University of Michigan, USA

Outline

- Motivation for weight trimming.
- Baysian finite population inference.
- Weight smoothing models.
- SAE with design weights: applying “implied” trimmed weights under model to model-assisted SAE.
- Simulation study: comparing fully weighted, weight smoothing, “crude” weight trimming estimators with respect to bias, MSE, and estimated vs. true MSE.
- Discussion.

Introduction

- If data values are associated with probability of selection, non-SRS sample designs can yield biased estimates of population quantities such as the population mean.
- Sample weights equal to inverse of probability of selection ($w_i = \pi_i^{-1}$) often used to reduce or remove bias.
  - Ex: Estimate population total $T = \sum_{i=1}^{N} y_i$ by $\hat{T} = \sum_{i \in s} w_i y_i$ (Horvitz and Thompson 1952).
- Use of weights usually increases estimator's variance.

$$\left.\begin{array}{l} Var(w) \text{ large} \\ n \text{ small} \\ Corr(y,w) \text{ weak} \end{array}\right\} \begin{array}{l} \text{Increase in} \\ \text{variance over-} \\ \text{whelms bias} \end{array} \rightarrow \text{increased MSE}$$

  - Small n might be a particularly important issue in small area estimation.

Weight Trimming ("Winsorization")

- Common approach to dealing with overly variable estimators in weighted data is *weight trimming* or winsorization (Potter 1990, Kish 1992, Alexander et al. 1997).
    - Weights larger than some value w_0 are fixed as w_0. with the remaining weights are adjusted upward by a constant so that the trimmed and untrimmed weighted sample sizes are equal.
    - Introduce bias to reduce variance $\Rightarrow$ overall reduction in MSE.
    - Calibration literature (Deville and Sarndal 1992, Folsom and Singh 2000) has developed methods for bounding design weights in generalized poststratification and raking procedures, but choice of bound is still arbitrary.
    - Reflects a traditional design-based approach to survey inference.

Design-based vs. model-based approach(Pfefferman 1993)

Design-based	Model-Based
Treat y_i as fixed and sampling indicator I_i as random.	$y_i \sim f(y_i;\theta)$.
Estimate population parameters $T = \sum_1^N y_i$ or $\overline{Y} = T/N$.	Estimate "superpopulation" parameters θ.
Use weighted estimators for point estimates; develop $\sim$unbiased estimates of variance accounting for sample design	Incorporate sample design, including information in weights, into model.

Finite Population Bayesian Inference

- Focus on population quantities of interest $Q(Y)$: population means $\overline{Y}$; population least-squares regression slopes $\min_{B_0,B_1} \sum_{i=1}^{N} (Y_{i1} - B_0 - B_1 Y_{i2})^2$.
- Make inference about $Q(Y)$ from marginal posterior predictive distribution (Ericson 1969, Rubin 1983, Little 1993):

$$p(Q(Y) \mid y) \propto \int\int\int f(Q(Y) \mid \theta, \psi, I, y) f(I \mid \theta, \psi, y) f(y \mid \theta, \psi) p(\theta, \psi) d\theta d\psi dI$$

- If ψ, θ a priori independent, ψ governs I only, and $I \perp Y_{-y}$, then

$$p(Q(Y) \mid y) \propto \int f(Q(Y) \mid \theta, y) f(y \mid \theta) p(\theta) d\theta \int\int f(I \mid \psi, y) p(\psi) d\psi dI$$

$$\propto f(Q(Y) \mid \theta, y) p(\theta \mid y) d\theta$$

- These conditions require sufficiently detail in the likelihood and prior model structure to accommodate the sample design.

Example: estimating a population mean under one-stage unequal probability sampling

- Stratify by probability of selection: $P(I_{hi} = \pi_h)$, $w_h = 1/\pi_h$, $h = 1, ..., H$.
  - If all probabilities differ, $H = n$, or can collapse observations in strata with approximately equal weights.
- $\overline{Y} = N^{-1} \sum_h N_h \overline{Y}_h = N^{-1} \sum_h N_h \left[\frac{n_h}{N_h} \overline{y}_h + \frac{N_h - n_h}{N_h} \overline{Y}_{nobs,h} \right]$
- $E(\overline{Y} \mid y) = N^{-1} \sum_h N_h \left[\frac{n_h}{N_h} \overline{y}_h + \frac{N_h - n_h}{N_h} E(\overline{Y}_{nobs,h} \mid y) \right]$
- Assume $y_{hi} \sim N(\mu_h, \sigma_h^2)$, $p(\mu_h, \sigma_h^2) \propto 1$ Then

$$E(\overline{Y} \mid y) = N^{-1} \sum_h N_h \left[\frac{n_h}{N_h} \overline{y}_h + \frac{N_h - n_h}{N_h} E(\mu_h \mid y) \right] =$$

$$N^{-1} \sum_h N_h \left[\frac{n_h}{N_h} \overline{y}_h + \frac{N_h - n_h}{N_h} \overline{y}_h \right] = N^{-1} \sum_h N_n \overline{y}_h$$

Example: estimating a population mean under one-stage unequal probability sampling

- If disproportional stratified design,

$$E(\overline{Y} \mid y) = \sum_h \sum_i w_{hi} y_{hi} / \sum_h \sum_i w_{hi}, \; w_{hi} \equiv w_h = N_h / n_h$$

$$V(\overline{Y} \mid y) = \sum_h P_h^2 (1 - f_h)^2 V(\mu_h \mid y) = \sum_h P_h^2 (1 - f_h)^2 \frac{s_h^2}{n_h}, \; P_h = N_h / N, f_h = n_h / N_h$$

- For more general design, could obtain via simulation by drawing $P_h \sim DIR(n_1^*, ..., n_H^*)$, $n_h^* = n * (n_h w_h) / \sum_h (n_h w_h)$, $\mu_h \mid y \sim t_{n_h - 1}(\overline{y}, s^2 / n_h)$ and $\overline{Y} \mid y$ by computing $\sum_h P_h (f_h \overline{y}_h + (1 - f_h) \mu_h)$.

Model-based winsorization: weight smoothing

- Models equivalent to fully weighted estimators assume total separation of weight stratum means, or more generally interactions between the weight strata and the model parameters of interest (Little 1993; Elliott 2007).
- A way to reduce the impact of highly variable weights is to model the weight strata/interactions as random effects ("weight smoothing" models) (Lazzaroni and Little 1998, Elliott and Little 2000):

$$y_{ih} \mid \mu_h, \sigma^2 \sim N(\mu_h, \sigma^2)$$

$$\mu \mid \phi, D \sim N(\phi, D)$$

where ϕ, σ^2,D have non-informative prior distributions.

Model-based winsorization: weight smoothing

- Under these assumptions,

$$E(\overline{Y} \mid y) = \sum_h [n_h \overline{y}_h + (N_h - n_h)\hat{\mu}_h]/N_+$$

  where $\hat{\mu_h} = E(\overline{Y}_h \mid y) = E(\mu_h \mid y)$.
- This model allows of shrinkage of the means across weight strata, with the shrinkage a function of the data and model structure.
  - Focuses on shrinking estimates of poorly estimated interaction terms, rather than shrinking the weights
  - Contrast with Beaumont (2008), which focuses on estimating weights as a function of outcome.

Model-based winsorization: weight smoothing

Structure can be added through either the mean or the variance components (Elliott and Little 2000):

Model	ϕ	D
Exchangable random effects (XRE)	$\mu 1_H$	$\tau^2 I_H$
AR1	$\mu 1_H$	$D_{jk} = \tau^2 \{\rho^{\lvert j-k \rvert}\}$
Linear	$\begin{pmatrix} 1 & 1 \\ \vdots & \vdots \\ 1 & H \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$	$\tau^2 I_H$
Nonparameteric mean	$f(h)$	0

$f(h)$ is a twice differentiable smooth function that minimizes the residual sum of squares plus a roughness penalty:

$$\min_{f(h)} \sum_{h,i} (y_{hi} - f(h))^2 + \lambda \int (f^{(2)}(u))^2 du$$

Model-based winsorization: weight smoothing

Under this formulation:

$$\hat{\mu} = X\hat{\beta} + Z\hat{u},$$

$$\hat{\beta} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}\overline{y}$$

$$\hat{u} = \hat{G}Z'\hat{V}^{-1}(\overline{y} - X\hat{\beta})$$

- Obtain estimates of G and σ^2, and thus of β and u, by ML or REML methods.

Model-based winsorization: weight smoothing

- Can be extended to exponential family distributions, linear and generalized linear models (Elliott 2007). Indexing the inclusion stratum by h we have

$$g(E[y_{hi} \mid \beta_h]) = x_{hi}^T \beta_h$$

$$(\beta_1^T, \ldots, \beta_H^T)^T \mid \beta^*, G \sim N_{Hp}(\beta^*, G)$$

$$p(\phi, \beta^*, G) \propto p(\zeta)$$

- Target population parameter of interest B is slope that solves the population score equation $U_N(B) = 0$ where

$$U_N(\beta) = \sum_{i=1}^{N} \frac{\partial}{\partial \beta} \log f(y_i; \beta) = \sum_{h=1}^{H} \sum_{i=1}^{N_h} \frac{(y_{hi} - g^{-1}(\mu_i(\beta)))x_{hi}}{V(\mu_{hi}(\beta))g'(\mu_{hi}(\beta))}$$

Model-based winsorization: weight smoothing

- Results of approach (Elliott and Little 2000) showed that simple models (like XRE and LIN) could have substantial efficiency gains, but were at risk of being overly biased when "signal" of relationship between outcome and probability of selection approximately equal to residual variance.
- More complex models (AR1, NPAR) had smaller efficiency gains, but better balanced the bias-variance tradeoff.
- Extension to regression modeling (Elliott 2007) showed smaller efficiency gains, but allowed simpler models to be more “robust”.

Deriving implied weights from model-based winsorization

- Note that, in the linear model setting where $f_h \approx 0$ for all h, the posterior mean of the population mean is given by $\sum_h P_h \hat{\mu}_h$, where $\hat{\mu}_h$ is obtained from the vector

$$\hat{\mu} = X\hat{\beta} + Z\hat{u} =$$
$$X(X'V^{-1}X)^{-1}X'V^{-1}\overline{y} + ZGZ'V^{-1}(\overline{y} - X(X'V^{-1}X)^{-1}X'V^{-1}\overline{y}) =$$
$$\underbrace{[X(X'V^{-1}X)^{-1}X'V^{-1} + ZGZ'V^{-1}(I - X(X'V^{-1}X)^{-1}X'V^{-1}]}_{A}\overline{y}$$

  where $\overline{y}$ consists of the stacked elements of y_h.
- Thus we can compute the posterior mean of the population mean as a reweighted mean estimator, with

$$\overline{y}_{wt} = \sum_h P_h \hat{\mu}_h = N^{-1} \sum_h N_h \sum_{m=1}^{H} A_{hm} \overline{y}_m = N^{-1} \sum_h \sum_i w_h^* y_{hi}$$

  where $w_h^* = \frac{N_h}{n_h} \sum_{m=1}^{H} A_{hm} = w_h \sum_{m=1}^{H} A_{hm}$.
- For exponential family, use A matrix resulting from fit to $g(E[y]) = X\hat{\beta} + Z\hat{u}$.

Model-based winsorization in small area estimation

- Proposed two-step approach:
  1. Use model-based approach to develop “data driven” weight trimming estimator for outcome of interest.
  2. Use standard design-based approach for weighted small area estimates, replacing original weights with winsorized weights.
- Hybrid of design- and model-based estimation.

Small area estimation using design weights (Prasad and Rao 1999)

- Begin with standard components-of-variance model:

$$y_{ij} = \mu + v_i + e_{ij},\ j = 1, ..., n_i,\ i = 1, ..., m$$

$$E(v_i) = E(e_{ij}) = 0, V(v_i) = \sigma_v^2, V(e_{ij}) = \sigma^2$$

$$v_i \perp e_{ij}$$

- Fully weighted direct estimator given by $\overline{y}_{iw} = \sum_j \tilde{w}_{ij} y_{ij}$ where $\tilde{w}_{ij} = w_{ij} / \sum_j w_{ij}$ are the case weights normalized to sum to 1 within a small area.
- Prasad and Rao develop "pseudo" EBLUB estimators:
  - $\overline{y}_{iw} = \mu + v_i + \overline{e}_{iw},\ V(\overline{e}_{iw}) = \delta_i = \sigma^2 \sum_j \tilde{w}_{ij}^2$
  - BLUP estimator of $\theta_i = \mu + v_i$ given by $\hat{\mu}_w = \sum_i \gamma_{iw} \overline{y}_{iw} / \sum_i \gamma_{iw}$ where $\gamma_{iw} = \frac{\sigma_v^2}{\sigma_v^2 + \delta_i}$.
  - Estimate σ^2 by $\frac{\sum_i \sum_j (y_{ij} - \overline{y}_i)^2}{n-m}$
  - Estimate σ_v^2 by $\frac{\sum_i n_i (\overline{y}_i - \overline{y})^2 - (m-1)\hat{\sigma^2})}{n - (\sum_i n_i^2)/n}$

Small area estimation using design weights (Prasad and Rao 1999)

- Mean square error of pseudo BLUP estimator given by

$$MSE(\hat{\theta}_i; \sigma_v^2, \sigma^2) = (1-\gamma_{iw})\sigma_v^2 \left(1 + \frac{(1-\gamma_{iw})}{\sum_i \gamma_{iw}}\right)$$

- Estimating using predictors of σ_v^2 and σ^2 leads to underestmation of MSE; add second-order correction:

$$M\hat{S}E(\hat{\theta}_i; \hat{\sigma_v^2}, \hat{\sigma^2}) = MSE(\hat{\theta}_i; \hat{\sigma_v^2}, \hat{\sigma^2}) +$$

$$\hat{\gamma}_{iw}(1-\hat{\gamma}_{iw})^2/\hat{\sigma}_v^2 \left\{ \hat{V}(\hat{\sigma}_v^2) + 2(\hat{\sigma}_v^2/\hat{\sigma}^2)C\hat{o}v(\hat{\sigma}_v^2, \hat{\sigma}^2) + (\hat{\sigma}_v^2/\hat{\sigma}^2)^2 \hat{V}(\hat{\sigma}_v^2) \right\}$$

Simulation Study Design

- Generate population of size M_i, from multinomial size 1 with probability ω_i, $i = 1, ..., 20$, where $\omega_i \sim DIR(1, 2, ..., 20)$.
- Let $W_{ij} = i$ be the individual-level small area identifying variable, $j = 1, ..., M_i$. Then

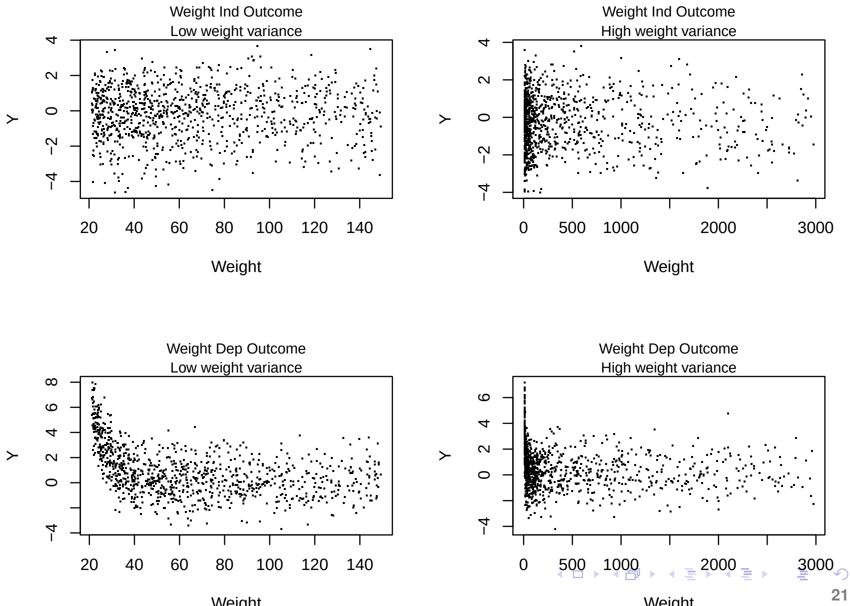
$$Y_{ij} = \beta_0 + \sum_i \beta_{1i} I(W_{ij} = i) + f(\pi_{ij}) + \varepsilon_{ij}$$

  where $\beta_0 = 0$, $\beta_i \sim N(0, \sigma_v^2)$, $f(\pi_{ij})$ is a function of the probability of selection, and $\varepsilon_{ij} \sim N(0, 1)$.
- Binomial sampling, with probability of selection $\pi_{ij} = \frac{\alpha_0 + \alpha_1 Z_{ij}}{1 + \alpha_0 + \alpha_1 Z_{ij}}$, where Z_{ij} is a fully-observed covariate.

Simulation Study Design

- We considered a $2 \times 2 \times 2 \times 2$ study design:
  - $\sigma_v^2 = 1$ (moderate within-small-area correlation), $\sigma_v^2 = 4$ (high within-small-area correlation)
  - $(\alpha_0 = -4, \alpha_1 = .5)$ (moderate variability in sample weights), $(\alpha_0 = -5, \alpha_1 = 1.5)$ (large variability in sample weights)
  - $Z_{ij} \sim N(-2, 2)$ (selection probability unrelated to small area), $Z_{ij} \sim N\left(\frac{-W_i}{10}, \frac{21-W_i}{10}\right)$ (selection probability higher in small areas with smaller population sizes)
  - $f(\pi_{ij}) = 0$ (selection probability unrelated to outcome), $f(\pi_{ij}) = 2\left(\frac{\pi_{ij}-\min(\pi_{ij})}{\max(\pi_{ij})-\min(\pi_{ij})}\right)^3$ (selection probability higher for larger values).

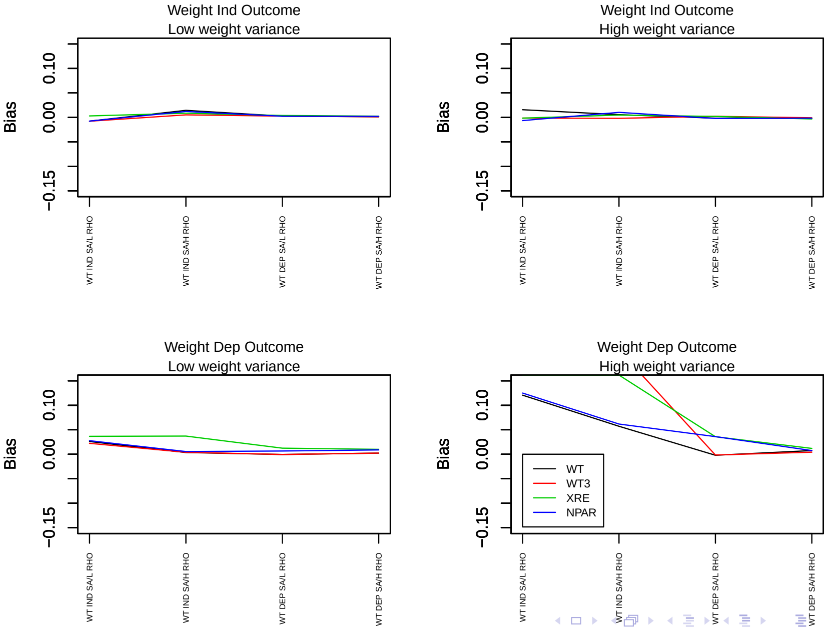
Simulation Study:Population Plot of Weight vs. Outcome



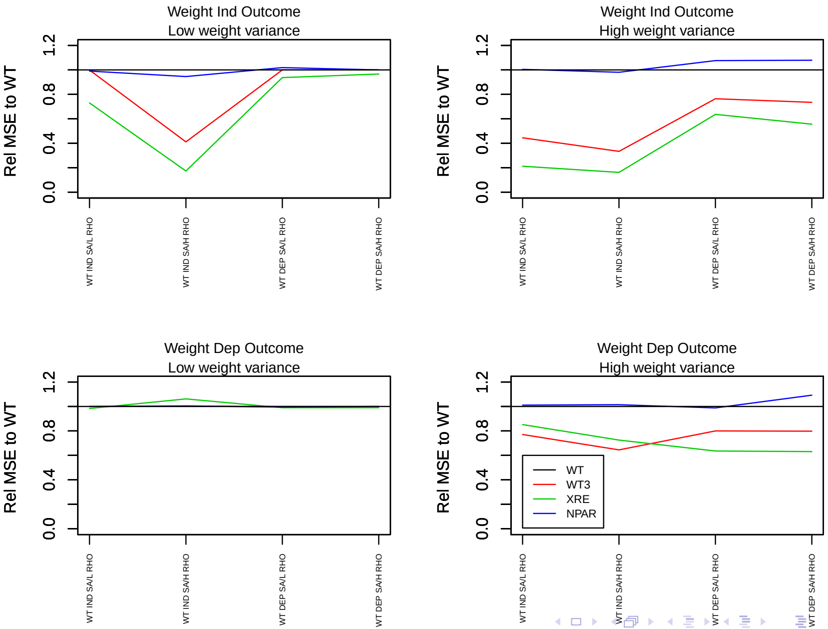
Simulation Study Design

- For each simulation, we compute four SAEs:
  - Fully-weighted
  - Crude-trimmed to a maximum value of 3 times the mean
  - Windsorized weights based on exchangable model (XRE)
  - Windsorized weights based on nonparametric regression model (NPAR)

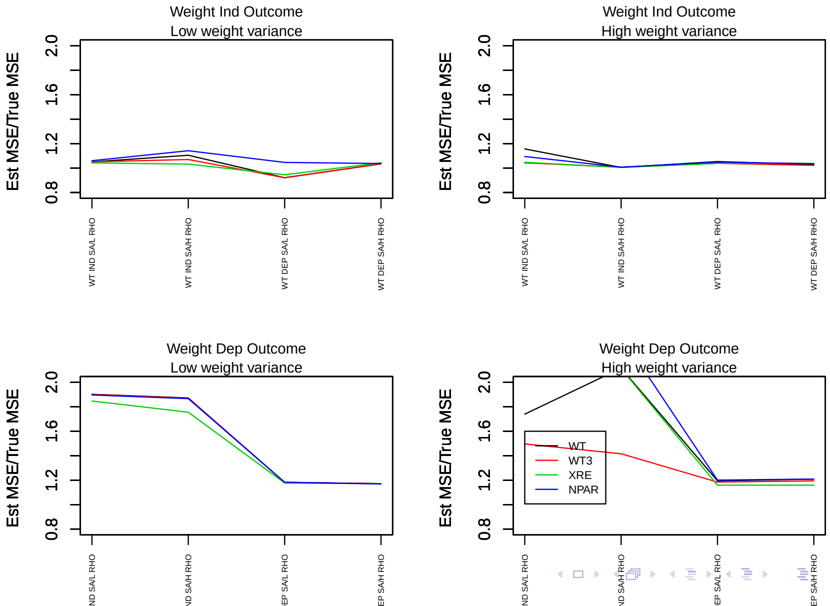
Simulation Study:Median SAE Bias



Simulation Study:Median SAE Relative MSE



Simulation Study:Median SAE Ratio of Est. MSE to True MSE



Simulation Study: Discussion

- Large reduction in MSE relative to fully weighted model for exchangable model when variance of weights in high or weights unrelated to outcome.
  - Extremely large reduction in MSE for exchangable model if outcome also independent of weights.
  - MSE reduction using crude trimming estimator, but usually less substantial
  - Less reduction in MSE for exchangable model when variance of weights in lower if weights vary by SAE
- XRE robust: only slight increase in MSE relative to fully weighted model if variance of weights is low and weights unrelated to outcome.
- NPAR has no MSE gain relative to fully-weighted model.

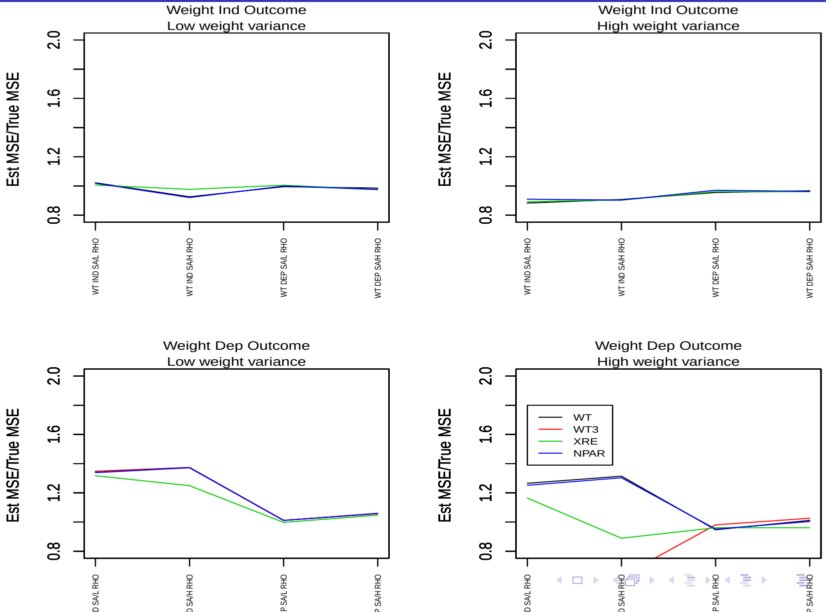
Simulation Study: Discussion

- When weight is not associated with the outcome, the MSE estimator slightly upwardly biased ($\sim$ 2%).
- When weight is associated with the outcome <u>and</u> the small areas, the MSE estimator is moderately upwardly biased ($\sim$ 20%).
- When weight is associated with the outcome <u>but not</u> the small areas, the MSE estimator is severely upwardly biased ($\sim$ 100%).

Simulation Study: Discussion

- Replace unweighted estimators of σ^2 and σ_v^2 with weighted equivalents $\sigma_w^2 = \frac{\sum_i \sum_j w_i j(y_{ij} - \overline{y}_{iw})^2}{n_w - m}$ and $\sigma_{v_w}^2 = \frac{\sum_i n_{iw}(\overline{y}_{iw} - \overline{y}_w)^2 - (m-1)\hat{\sigma_w^2})}{n_w - (\sum_i n_{iw}^2)/n_w}$ for $n_{iw} = \sum_j w_{ij}$ and $n_w = \sum_i n_{iw}$.

Simulation Study:Median SAE Ratio of Est. MSE to True MSE



Simulation Study: Discussion

- Removes slight and moderate upward bias in MSE estimator when weight is not associated with the outcome or when associated with the outcome <u>and</u> the small areas
- Substantially reduces bias in MSE estimator when weight is associated with the outcome <u>but not</u> the small areas (XRE model almost unbiased; crude trimming model substantially downward biased).
- Modest <u>downward</u> bias (5-10%) in MSE when is highly variables and not associated with the outcome.
- Consider Jiang and Lahiri (2006) variance estimator for better second-order properties.

Summary

- When outcome is associated with probability of inclusion, use of sampling weights can reduce bias.
- When weights are highly variable, bias reduction might be "outweighed" by increased variance, especially if outcomes are weakly associated with probability of inclusion.
  - Might be especially acute in SAE, with small samples and highly variable weights.
- Use of "data-driven" weight trimming via weight smoothing allows balancing of bias-variance tradeoff in a two-stage setting:
  - Model to determine trimmed weights "tuned" to the SAE outcome of interest, then application of trimmed weights in "standard" model-assisted setting.
  - Evidence that simpler weight smoothing models are sufficiently robust, while more complex weight smoothing models simply mimic fully-weighted approach.

Future Work

- Prasad and Rao not only method to incorporate design into SAE: quantile regression approach of Chambers and Tzavidis (2006) and extensions could also use windsorized weights.
- When (linear) model for SAE is used, iterative application of calibration weighting scheme such as generalized regression (GREG) could be used to obtain "implied" weights.
  - Begin with initial design weights $w_h^{(0)}$.
  - Update as follows:

$$w_h^{(t+1)} = w_h^{(t)}(1 + (\hat{B}_{w^\infty} - \hat{B}_{w^t})T_u^{(t)}u_h^{(t)})$$

where $u_h^{(t)} = \left[\sum_h w_h^{(t)} \sum_{i=1}^{n_h} x_{hi}x_{hi}^T\right]^{-1} \sum_{i=1}^{n_h} x_{hi}y_{hi}$, $T_u^{(t)} = \sum_h w_h^{(t)} u_h^{(t)}(u_h^{(t)})^T$, and $\hat{B}_{w^\infty}$ is the windsorized estimator of the population slope B relating covariates x_i to the mean of y_i: $B = \min_\beta \sum_{i=1}^{N}(y_i - x_i^T\beta)^2$.

Future Work

- An alternative unified, fully-model based approach incorporates the weights directly into the analysis (Gelman 2006).
  - Use flexible model relating probability of selection to outcome (Zheng and Little 2005, Chen et al. 2010) as standard SAE model with fixed effect predictors.

$$y_{ij} = \beta_0 + \sum_{k=1}^{3} \beta_k \pi_{ij}^3 \sum_{l=1}^{m} b_i(\pi_{ij} - \kappa_l)_+^3 + v_i + e_{ij}, j = 1, ..., n_i, i = 1, ..., m$$

$$b_l \sim N(0, \tau^2), \; v_i \sim N(0, \sigma_v^2), \; e_{ij} \sim N(0, \sigma^2)$$

$$b_l \perp v_i \perp e_{ij}$$

  where κ_l are prespecified knots for a cubic P-spline regression.
    - Assumes no interaction between weight and small area
    - Assumes no interaction with an additional covariates incorporated in SAE model
  - Interaction can be incorporated but typically data will not be sufficient in small area to estimate (Huang 2011).

Ackowledgements

The author would like to thank Qixuan Chen for her review and suggestions.