Performance of Generalized Regression Estimator and Raking Estimator in the Presence of Nonresponse

Daifeng Han, Westat; Richard Valliant, JPSM at University of Maryland;
Jill M. Montaquila, Westat; Keith Rust, Westat

FCSM Conference, Washington D.C.

Background

- Basic application of calibration is to reduce variance
  - No consideration of undercoverage or nonresponse.
  - Deville & Särndal (1992) prove that in the pure sampling context, alternative forms of calibration weighting are asymptotically equivalent. The generalized-regression estimator (GREG) is considered a good approximation of other calibration estimators.
- Calibration can be used to reduce nonresponse bias through one-step weighting
  - Särndal & Lundström (1999, 2005)
  - Kott & Chang (2006, 2008, 2010)
  - Slud & Thibaudeau Approach (2009)

Research Aims

Using calibration to reduce nonresponse bias through one-step weighting (Särndal & Lundström, 2005)

- Aim I: re-examine the results in Deville & Särndal (1992) through theoretical derivation

  Are the GREG estimator and the general calibration estimators still asymptotically identical in the presence of nonresponse? (Design-based)

- Aim II: evaluate how three widely used calibration estimators behave under different population and response models through simulation

  What are the impacts of the outcome variable model and the response mechanism? (Model-based and design-based)

Theoretical Re-examine of Deville & Särndal (1992) in the Presence of Nonresponse

- In the presence of nonresponse,

  $N^{-1}(\hat{t}_{cal} - \hat{t}_{GREG})$ does not necessarily converge to zero.

- Particular forms of calibration estimator and GREG estimator have (implicit) models under which they perform well.
- In what circumstances will GREG and a general calibration estimator perform differently?

Poststratification, Raking, and GREG_Main: The Need to Examine Models

- Poststratification
  - A special case of the GREG estimator
  - Accounting for both main and interaction effects of auxiliary variables
- Raking ratio estimator
  - A general calibration estimator
  - Calibration to margins using a non-GLS distance function
- GREG_Main
  - Based on a linear regression model accounting for only main effects
  - Calibration to margins using GLS distance function
- Need to examine the relationship between outcome variable model, response model, and covariates used in calibration weighting
  - Little and Vartivarian (2006) also looked at models for outcome and response.

Simulation Scope

- Estimator of population total for a single outcome variable
- Simple models for outcome variable and response propensity: both models contain the same two main effect covariates (both being categorical variables) + possibly an interaction effect term
- Missing at random (MAR) for response mechanism
- Simple random sampling (SRS)

Simulation Conceptual Framework

Outcome variable model (Y) and response propensity model (R) are varied by including or excluding the interaction effect term of the two main effect variables.

- General models are:

$$E_M(Y_{ijk}) = \mu_Y + \alpha_{Yi} + \beta_{Yj} + \gamma_{Yij}$$ Y_Main ($\gamma_{Yij} = 0$) vs Y_Int ($\gamma_{Yij} \neq 0$)

$$E_R(r_{ijk}) = \mu_R + \alpha_{Ri} + \beta_{Rj} + \gamma_{Rij}$$ R_Main ($\gamma_{Rij} = 0$) vs R_Int ($\gamma_{Rij} \neq 0$)

- Four scenarios under evaluation
  - Y_Main & R_Main: neither model includes interaction term
  - Y_Main & R_Int: only R model includes interaction term
  - Y_Int & R_Main: only Y model includes interaction term
  - Y_Int & R_Int: both models includes interaction term

Westat

50 years Improving lives through research

Simulation Parameters

- Number of simulation samples = 10,000
- Overall predictive power of outcome model and response model
- Substantive and statistical significance of the interaction effect in outcome variable model and response model
- Respondent sample size (SRS sample size: 8000, 2000, and 200)

Outcome variable model	$E_M(y_{11k})$	$E_M(y_{12k})$	$E_M(y_{21k})$	$E_M(y_{22k})$
Y_Main	700	950	1,200	1,450
Y_Int	800	1,250	1,900	2,650

Response propensity model	$E_R(r_{11k})$	$E_R(r_{12k})$	$E_R(r_{21k})$	$E_R(r_{22k})$
R_Main	0.30	0.50	0.35	0.55
R_Int	0.35	0.65	0.55	0.95

Evaluation Criteria

- Relative bias $RelBias(\hat{t}_{yw_s}) = (1/S)\sum_{s=1}^{S}(\hat{t}_{yw_s} - t_y) \big/ t_y$
- Relative standard error

$$RelSE(\hat{t}_{yw_s}) = \sqrt{var(\hat{t}_{yw_s})} \Big/ t_y = \sqrt{(1/S)\sum_{s=1}^{S} \; \hat{t}_{yw_s} - E_p(\hat{t}_{yw_s}) \;^{2}} \Big/ t_y$$

- Relative square root of Mean Squared Error (MSE)
- Coverage rate of 95% confidence interval (CI)

$$(1/S)\sum_{s=1}^{S} I \;\; \left|\hat{z}_j\right| \le z_{1-\alpha/2} \;\;, \text{ where } \alpha = 0.05 \text{ and } \hat{z}_j = (\hat{t}_{yw_s} - t_y) \Big/ \sqrt{var(\hat{t}_{yw_s})}$$

- Bias ratio

Properties over Repeated Sampling: Impacts of Outcome Variable and Response Models

SRS sample size = 8,000	**Relative bias** $RelBias(\hat{t}_{yw_s})$ **in 10^{-5}**		
	Poststratification	GREG_Main	Raking
Y_Main, 100% response	-0.1	-0.1	-0.1
Y_Main & R_Main	1.1	1.2	1.2
Y_Main & R_Int	1.8	1.9	1.8
Y_Int, 100% response	-0.1	-0.3	-0.3
Y_Int & R_Main	1.1	-135.1	-68.0
Y_Int & R_Int	1.0	-303.7	-65.0

SRS sample size = 8,000	**Relative standard error** $RelSE(\hat{t}_{yw_s})$ **in 10^{-4}**		
	Poststratification	GREG_Main	Raking
Y_Main, 100% response	2.8	2.8	2.8
Y_Main & R_Main	4.8	4.8	4.8
Y_Main & R_Int	4.0	4.0	4.0
Y_Int, 100% response	1.8	4.9	4.9
Y_Int & R_Main	3.1	9.0	8.3
Y_Int & R_Int	2.6	8.2	7.0

Properties over Repeated Sampling: Coverage Rate of 95% CI and Bias Ratio

	Poststratification	GREG_Main	Raking
SRS sample size = 200			
Y_Int & R_Main	94%	94%	96%
Y_Int & R_Int	94%	92%	98%
SRS sample size = 2,000			
Y_Int & R_Main	95%	88%	95%
Y_Int & R_Int	95%	55%	96%
SRS sample = 8,000			
Y_Int & R_Main	94%	63%	88%
Y_Int & R_Int	94%	3%	90%

$$t\text{-}statistic = \frac{\hat{t}_{yw_s} - t_y}{\sqrt{\mathrm{var}(\hat{t}_{yw_s})}} \approx N(0,1) + \frac{bias(\hat{t}_{yw_s})}{(\hat{t}_{yw_s})}$$

Westat®

50 years Improving lives through research

Properties over Repeated Sampling: Why does raking do better than GREG_Main?

	Odds ratio				
SRS sample size = 8,000	Population	Respondent sample	Post-stratification	GREG_ Main	Raking
R_Main, Regardless of Y model	0.99	0.93	0.99	0.88	0.93
R_Int, Regardless of Y model	0.99	0.93	0.99	0.76	0.93

Raking forces the weights to conform to the marginal totals without perturbing the associations in the unadjusted table (Haberman 1979).

Properties Conditioning on Sample: Defining a Distance Measure

Focusing on comparison between raking and poststratification

Assume that the outcome variable model contains an interaction term, then the correct calibration method should be poststratification.

If raking is used, then the model bias of the raking estimator is:

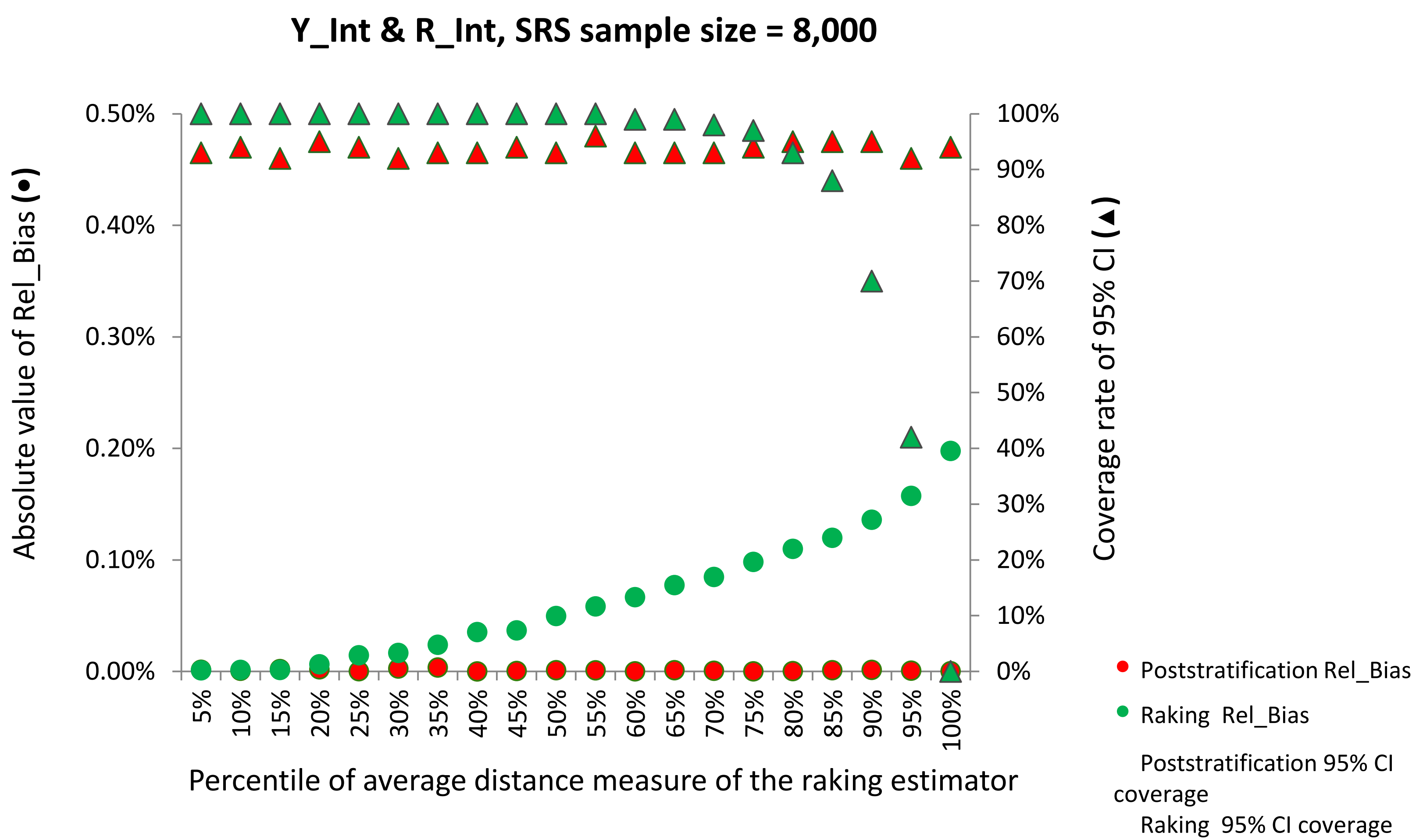
$$\mathrm{E}_{\mathrm{M}}(\hat{t}_{raking} - t) = \mu_Y(\hat{N} - N) + \sum_{i=1}^{2} \alpha_{Yi}(\hat{N}_{i\cdot} - N_{i\cdot}) + \sum_{j=1}^{2} \beta_{Yj}(\hat{N}_{\cdot j} - N_{\cdot j}) + \sum_{i=1}^{2}\sum_{j=1}^{2} \gamma_{Yij}(\hat{N}_{ij} - N_{ij})$$

The first three terms are approximately zero. The fourth term is often not zero.

Define distance measure as: $\mathrm{D}_{raking} = \sqrt{\sum_{i=1}^{2}\sum_{j=1}^{2}(\hat{N}_{ij} - N_{ij})^2}$

This measure is often computable and can help predict the model bias of raking estimator for a particular sample if poststratification is supposed to be the appropriate estimator.

Properties Conditioning on Sample: Rel_Bias (•) and Coverage of 95% CI ( )



Westat

Improving lives through research

Summary

- In general, GREG estimator and the general calibration estimators are not asymptotically equivalent in the presence of nonresponse.
- It is important to examine outcome variable model and response propensity model when choosing between poststratification and raking, and the outcome model seems to be the driving factor.
- Small relative bias associated with the inappropriate calibration estimator could result in very bad coverage rate of 95% CI. Increasing sample size only makes the situation worse.
- In survey practice where only a single sample is obtained, a distance measure could help gauge the potential consequence of choosing inappropriate estimator.