A Comparison of Model-Based and Model-Assisted Estimators under Ignorable and Non-Ignorable Nonresponse

Jill A. Dever[1] and Richard Valliant[1,2]
Joint Program in Survey Methodology, University of Maryland[1]
Institute for Social Research, University of Michigan[2]

Keywords: modified generalized regression estimator, best linear unbiased predictor, non-ignorable nonresponse

1. Introduction

The benefits of using model-based estimators, such as the best linear unbiased predictor (BLUP), for survey inference over model-assisted estimators, such as the generalized regression estimator (GREG), as well as the converse have been noted in the literature. The BLUP, used with prediction theory and well documented in classical statistics, has the desirable property of minimum model variance among the set of unbiased predictors. This model-based estimator was most notably proposed by Royall and Herson (1973) for use in survey statistics. As discussed by Godambe (1955), traditional survey estimators generally do not attain the BLU condition and therefore produce inefficient estimates. Additionally, theory states the likelihood generated by the sampling design is uninformative for survey inference since it is a constant for every sample and does not describe the underlying likelihood function of the population (Godambe 1966, Basu 1969).

General regression estimation, one of the model-assisted techniques from probability sampling theory, uses a set of auxiliary variables to produce efficient survey estimates (e.g., Chapter 6 in Särndal, et al 1992). Repeated sampling (or design) variance in the estimate, calculated as its average squared difference from the true population value, is affected by the strength of the relationship between the variable of interest and the auxiliary variables. These estimators incorporate analysis weights and components of the sampling design for inference instead of relying on the correct specification of the underlying superpopulation model as with the BLUP (e.g. Hansen, et al 1983). Additionally, the calibration property of the GREG ensures that certain survey estimates tabulate to their known population values.

Most of the research to date comparing model-based and model-assisted estimators has been conducted under the assumption of either 100% response or ignorable nonresponse after controlling for a specified list of known auxiliary variables (e.g., Gerow and McCulloch 2000, Brewer 1999, and Hansen, et al 1983). A recent exception is Laaksonen and Chambers (2006) who compare design-based and model-based estimation for two-phase designs where interviews are obtained from all first-phase nonrespondents in the second phase. Surveys of the general population rarely if ever achieve complete response and in many instances suffer from low response rates (de Leeuw and de Heer 2002).

The purpose of this article is to provide a preliminary empirical comparison of the bias and confidence interval coverage rates of four estimators in predicting a population total under various levels of ignorable and non-ignorable nonresponse. The estimators include one model-based estimator (a BLUP) and three model-assisted estimators – a GREG using only design weights irrespective of the level of nonresponse; a modified GREG using nonresponse-adjusted design weights (MGREG); and a regression estimator using only weights associated with the probability of response (RARE). We implement this research through a simulation study on three model-generated populations.

In section 2, we discuss the research produced to date on various aspects of our study. We provide details of the simulation and the analyses in section 2 followed by the analytical results in section 3. Limitations of the current study along with the proposed next steps in this investigation are included in the final section.

2. Background

2.1 Model-based vs. Model-Assisted Estimation

The prediction theory approach uses tools from classical statistics such as the model-based BLUP. Denote the population of units by U and the set of sample units by S. The form of the BLUP for an estimated population total under 100% response is

$$\begin{aligned}\hat{T}_{BLUP} &= T_S + \mathbf{X}_{U-S}\hat{\boldsymbol{\beta}}_{BLUP} \\ \hat{\boldsymbol{\beta}}_{BLUP} &= [\mathbf{X}_S'\mathbf{V}_S^{-1}\mathbf{X}_S]^{-1}\mathbf{X}_S'\mathbf{V}_S^{-1}\mathbf{Y}_S,\end{aligned} \tag{1}$$

where T_S is the total calculated from the sample S;

$\mathbf{X}_S$ is a matrix of sample values for the auixiliary x-variables; and $\mathbf{X}_{U-S}$ is a vector of values from population units *excluded* from the sample; $\hat{\boldsymbol{\beta}}_{BLUP}$ is the vector of generalized least squares (GLS) model coefficients estimated from the sample; $\mathbf{V}_s$ is the specified error covariance matrix; and $\mathbf{Y}_s$ is the vector of values for the analysis variable. For studies with some nonresponse, $\hat{T}_{BLUP}$ is defined only for the responding subset of the sample (S_R). We distinguish estimators under complete and less than complete response by using a subscript S and R, respectively. For example, $\mathbf{X}_{U-R}$ is the vector of population totals for the auxiliary variables excluding the responding sample members.

The characteristics of model-based estimators are derived with respect to a specified model. For example, the prediction bias for $\hat{T}_{BLUP}$ is evaluated based on the working model (M^*),

$$E_{M^*}(\hat{T}_{BLUP} - T) = E_{M^*}(\hat{T}_{BLUP}) - E_{M^*}(T). \quad (2)$$

Therefore, this type of inference is heavily dependent on M^* closely matching the true superpopulation model M. If the BLUP is derived using M^*, then (2) is zero and $\hat{T}_{BLUP}$ is prediction-unbiased under M^*. However, if some other model, say $\tilde{M}$, is correct (or at least a better description of y than M^*), then $\hat{T}_{BLUP}$ is biased.

Royall and Herson (1973) offered balanced sampling as one mechanism for protecting against certain types of misspecification of the superpopulation model. Balanced samples are selected so that a pre-specified number of sample moments for at least one relevant auxiliary variable equate to their corresponding population moments, thereby effectively reducing the model-based bias (2) to zero (see Chapters 3 and 4 in Valliant, et al 2000). These samples are achieved either exactly or approximately by purposive sampling (though criticized), controlled selection, and randomization. When a single x is available, M^* may be a low order polynomial. If a higher order polynomial however is more nearly correct, weighted balanced samples can protect an estimator against bias. Weighted balanced samples are selected such that the sample moments weighted by the inverse of the square root of the specified variance $(n^{-1}\sum_s v_k^{-1/2} x_k^j)$ are equivalent to the population values, $\sum_U x_i^j (\sum_U v_i^{-1/2})^{-1}$.

The BLUP by definition has the minimum model variance among the set of unbiased estimators, a desirable property for producing efficient estimates (e.g., Casella and Berger 2002). Therefore, in theory the BLUP will always be more efficient than a design-based estimator. Also, Royall and Herson (1973) argued that the focus of estimation should be on the characteristics of a particular sample instead of the method for selecting the sample, as with design-based estimation.

Conversely, probability sampling (or design-based) theory avoids explicit reliance on correctly specified models and instead uses the randomization principle which states that the random sampling distribution is the only means by which valid inferences are made. The design weights (i.e., the inverse of the unit inclusion probabilities) as well as the properties of the estimator over repeated samples are key to inference about quantities in the underlying population of interest. Neither concept is considered with the BLUP. A model-assisted technique such as generalized regression estimation (GREG) is one of many estimation methods used by design-based practitioners. The GREG relies on a set of auxiliary variables to produce efficient survey estimates (e.g., Chapter 6 in Särndal, et al 1992). The GREG for a population total, under 100% response, is defined as

$$\begin{aligned} \hat{T}_{GREG} &= \hat{T}_{y\pi} + \hat{\mathbf{B}}'_{GREG}(\mathbf{T_x} - \hat{\mathbf{T}}_{\mathbf{x}\boldsymbol{\pi}}) \\ \hat{\mathbf{B}}_{GREG} &= [\mathbf{X}'_s \mathbf{V}_s^{-1} \boldsymbol{\Pi}_s^{-1} \mathbf{X}_s]^{-1} \mathbf{X}'_s \mathbf{V}_s^{-1} \boldsymbol{\Pi}_s^{-1} \mathbf{Y}_s \end{aligned} \quad (3)$$

where $\hat{T}_{\alpha\pi}$ is the Horvitz-Thompson (HT) estimator for the total of the α variable ($\alpha = y, x$); $\mathbf{T_x}$ is the vector of known population totals for the auxiliary x-variables; and $\boldsymbol{\Pi}_s$ is the diagonal matrix of inclusion probabilities. The remaining components in (3) are similar to the variables defined in (1) above. In the presence of nonresponse, the subscripts in $\hat{\mathbf{B}}_{GREG}$ change from S to R.

As alluded to above, the estimation bias for $\hat{T}_{GREG}$ is taken with respect to the sampling distribution (δ) instead of a model: $E_\delta(\hat{T}_{GREG}) - T$. The "model" in model-assisted estimation is used only to define the structure of $\hat{\mathbf{B}}_{GREG}$, the vector of GLS model coefficients. Historically, the approximate design-unbiasedness of the GREG was declared to render it robust against model misspecification (Särndal, et al 1992). However, more recent research indicates a need for model diagnostics to improve the accuracy of the model because of an actual sensitivity to model misspecification (Hedlin, et al 2001).

The efficiency of $\hat{T}_{GREG}$ is dependent on the strength of the linear relationship between y and the auxiliary x-variables (Särndal, et al 1992). This form of the GREG is also known as a calibration estimator because $\sum_s g_k \mathbf{x}_k = \mathbf{T_x}$ with the g-weights defined below.

$$\hat{T}_{GREG} = \sum_s \left(1 + c_k\right) y_k \pi_k^{-1} = \sum_s g_k \breve{y}_k \quad (4)$$

where $c_k = [\mathbf{T_x} - \hat{\mathbf{T}}_{\mathbf{x}\boldsymbol{\pi}}]' (\sum_s \mathbf{x}_k \mathbf{x}'_k / \pi_k v_k)^{-1} \mathbf{x}_k v_k^{-1}$ and $\breve{y}_k = y_k \pi_k^{-1}$.

2.2 Ignorable and Non-ignorable Nonresponse

Nonresponse is classified as either ignorable or non-ignorable. Nonresponse is labeled ignorable when the response propensities are independent of the distribution of y. This type of nonresponse is further classified into either missing completely at random (MCAR) or missing at random (MAR). MCAR occurs when the respondents are still a representative (random) sample from the population of interest. MAR represents a MCAR situation but only after controlling for a set of auxiliary variables (i.e., the distribution given the x-variables ($y|\mathbf{x}$) is independent of the response mechanism). Conversely, non-ignorable nonresponse (NINR) occurs when the $y|\mathbf{x}$ distribution is dependent on the response propensities. Additional information on the nonresponse patterns is found in several sources (e.g., Little and Rubin 2002; Pfeffermann 1993).

2.3 Weight Adjustments

In the presence of nonresponse, most design-based practitioners choose to adjust the design weights in an attempt to reduce any existing nonresponse bias to a negligible level (e.g., Chapter 8 in Lessler and Kalsbeek 1992). Many types of nonresponse adjustments are cited in the literature (e.g., Kalton and Kasprzyk 1986). For our research, we focus on two specific methods – weighting classes and response propensities. Weighting classes are formed by cross-classifying the sample units by a set of categorical variables that are correlated with the probability to respond as well as y. These data must be known for both respondents and nonrespondents. The resulting weighting classes are usually required to contain a sufficient number of cases so that variability in the class-specific nonresponse adjustments is controlled. The adjustment for a particular weighting class (c) is the sum of all design weights divided by the sum of the design weights for the respondents.

Response propensities are calculated using the predicted values from a binary regression model with the dependent variable set to a 0/1 response indicator. As with the weighting class adjustment, the ideal model covariates are ones associated both with the probability of response and with a set of key analytic variables. This method is not restricted to categorical variables because the propensity model can allow for continuous variables. Including many variables into this model can reduce nonresponse bias. However, an over-specified model decreases the precision of the nonresponse adjustment. A cell-mean response propensity adjustment (a_k) is made by classifying the raw propensities into groups and using the inverse of the mean propensity within these groups. Little (1986) suggests the cell-mean propensity adjustment over the raw propensity adjustment to minimize the variance inflation associated with the highly variable weights.

The c_k adjustment (4) in the GREG also reduces nonresponse bias if the auxiliary variables are associated with response propensity and the y-variable (e.g., Chapter 6 in Särndal and Lundström 2005). Additionally, this adjustment can also account for any coverage bias from the sampling frame again based on the relationship between y and $\boldsymbol{x}$.

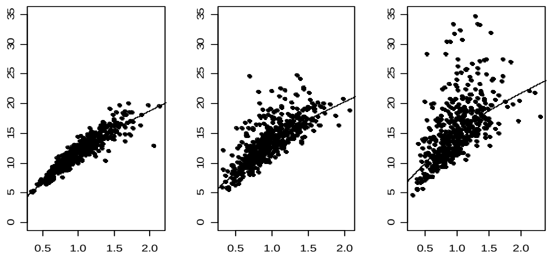
Given the discussion about nonresponse adjustments, there are two additional estimators to consider. First, we have a regression estimator similar in form to the GREG with the addition of a nonresponse adjustment instead of relying solely on the unadjusted design weight. The estimated total for this modified GREG (MGREG) using only the responding sample cases ($k \in S_R$) has the following form:

$$\hat{T}_{MGREG} = \hat{T}_{yw} + \hat{\mathbf{B}}'_{\boldsymbol{MGREG}} (\mathbf{T_x} - \hat{\mathbf{T}}_{\mathbf{x}\boldsymbol{w}}) \quad (5)$$

$$\hat{\mathbf{B}}_{\boldsymbol{MGREG}} = [\mathbf{X}'_R \mathbf{V}_R^{-1} \mathbf{\Pi}_R^{-1} \mathbf{A}_R \mathbf{X}_R]^{-1} \mathbf{X}'_R \mathbf{V}_R^{-1} \mathbf{\Pi}_R^{-1} \mathbf{A}_R \mathbf{Y}_R$$

where, $\mathbf{A}_R = diag(a_k)$. The components of this estimator are similar to those described in (3) with the use of nonresponse-adjusted weights $(\mathbf{\Pi}_R^{-1}\mathbf{A}_R)$ instead of design weights alone $(\mathbf{\Pi}_R^{-1})$. For example, $\hat{T}_{yw} = \sum_{S_R} \pi_k^{-1} a_k y_k = \sum_{S_R} w_k y_k$. If $1/a_k$ is a consistent estimator of the propensity of a unit to respond and each unit in the population has a non-zero response propensity, then $\hat{T}_{MGREG}$ is

Figure 1. Plot of 500 Records from Simulation Populations 1, 2, and 3.



approximately unbiased with respect to the sampling and response mechanisms. Many researchers today consider this estimator to be the standard regression estimator of choice.

The last estimator considered in our research is referred to as a response-adjusted regression estimator (RARE). RAREs have been used with studies where the probabilities of selection are unknown such as Web panel surveys. As with the MGREG (6), the RARE has the same basic structure as the GREG (3) for estimating totals except for the exclusion of the design weight:

$$\begin{aligned}\hat{T}_{RARE} &= \hat{T}_{ya} + \hat{\mathbf{B}}'_{\boldsymbol{RARE}}(\mathbf{T_x} - \hat{\mathbf{T}}_{\mathbf{x}\boldsymbol{a}}) \\ \hat{\mathbf{B}}_{\boldsymbol{RARE}} &= [\mathbf{X}'_R \mathbf{V}_R^{-1} \mathbf{A}_R \mathbf{X}_R]^{-1} \mathbf{X}'_R \mathbf{V}_R^{-1} \mathbf{A}_R \mathbf{Y}_R\end{aligned} \tag{6}$$

In (6), $\hat{T}_{\lambda a} = \sum_{S_R} a_k \lambda_k$ for $\lambda = y, x$. This estimator will be approximately unbiased with respect to the combination of the working model and the response mechanism.

3. Methods

The primary purpose of this paper is to compare the confidence interval coverage for a population total computed with the four estimators discussed in the previous section and with samples containing varying levels of nonresponse. We briefly describe the set-up of the simulation study, conducted in **R** (R Development Core Team 2005), from which our results are obtained. Summary information for the simulation study is provided in *Table 1.*

Three population list frames of size 100,000 were generated under the polynomial regression model $Y_k = x_k\beta_1 + x_k^2\beta_2 + z_k\beta_3 + e_k$ with $e_k \sim N(0, \sigma^2 x_k^2)$ and $x_k \sim Gamma(9, 9)$. Note that the variance structure is not constant because of the dependence on the auxiliary x-variable. We evaluated the model at $\sigma^2 = 1$. The model coefficients were (β_1, β_2) = (15,-3) for all the populations; β_3 was set to zero for Population 1 and to 1.0 for the other frames.

Only populations 2 and 3 incorporated the z-variable above with distributions $z_k \sim Gamma(1, 0.5)$ and $z_k \sim Gamma(1, 0.25)$, respectively. The existence of z, unknown for either sampling or weighting phases of the study, serves two purposes for our study. First, the presence of the z in Populations 2 and 3 dilutes the linear relationship between y and the auxiliary variables, a condition that affects the efficiency of the regression estimators. A graph of y by x for the three populations shows a linear relationship starting in Population 1 which degrades to a cloud of points by Population 3 (*Figure 1*). Second, because z is not included in the working (superpopulation) model of our estimators, these models are underspecified. Therefore, we can further examine the sensitivity of the BLUP and the GREG to the condition of the model. Note, however, that our two purposes are confounded so that we can not say which condition has the most negative effect on our estimates. The resulting x-y correlations for the three populations were 0.9, 0.7, and 0.5, respectively.

Two other auxiliary variables were added to the population list frames – a trichotomous variable created by dividing the x values into high, medium, and low categories by the 33rd and 66th percentiles (*x.cat3*), and a 6-level variable created by dividing the x values into the relevant percentiles (*x.bin*). Both variables were instrumental for the weighting class adjustments.

From each of the simulated populations, D=1,000 samples of size 1,200 were selected using two methods: simple random sampling without replacement (SRS), and probability proportional to size systematic sampling with size measure x, PPS(x). (Weighted balanced samples of size x_k^2 were selected in our study. However, the results from these samples did not differ greatly from those reported and are therefore excluded from the discussions.) The SRS selection method was implemented for comparisons with prior research results. The PPS method was selected to represent a likely single-stage sampling design (more complex than a simple random sample) that is used with design-based estimation. We constructed the corresponding design weights for the GREG and MGREG estimates.

As discussed in section 2.6 of Valliant, et al (2000), the designation of a design as ignorable or non-ignorable is one important aspect in choosing an estimation procedure. This designation can be affected by the level and pattern of nonresponse depending on the correlation between response propensity and the analysis variable. Ten response rates were incorporated into each simulation sample ranging from 15% to 90%. The three patterns of nonresponse were also incorporated into the simulation as described below.

MCAR was simulated by selecting a simple random subsample of respondents within the d^{th} sample at the specified level of response. A simple random subsample of respondents was chosen within *x.cat3* for MAR. To mimic a non-ignorable nonresponse situation, we chose the respondent subsamples through a PPS selection with size measure *y*, PPS(*y*), thereby creating in expectation a correlation between nonresponse and the analysis variable.

Five versions of the MGREG were created with differing nonresponse adjustments – three weighting class adjustments (overall, within *x.cat3*, and within *x.bin*) and two response propensity weight adjustments (raw and cell-mean) using logistic regression. Two types of RARE estimates were created using both the raw and cell-mean response propensity weights. By definition, nonresponse adjustments were not used for either the BLUP or the GREG though the estimates were calculated using only the respondents (S_R).

Table 1. Summary of the Simulation Components

Simulation Components	Component Values
Sampling Design (2)	SRS without replacement; PPS systematic sampling with size measure *x*
Response Rate (10)	15% to 90%
Pattern of Nonresponse (3)	Missing Completely at Random; Missing at Random; Non-Ignorable Nonresponse
Nonresponse Adjustment (3)	Weighting class; Propensity weighting (Raw, Cell-Mean)

The primary focus of the analyses was to determine the confidence interval (CI) coverage rates for the four types of predictors under various conditions summarized in *Table 1*. The bias ratio, $BR(\hat{T})$, is defined as the ratio of the absolute empirical bias of the estimated total, $\left|Bias(\hat{T})\right|$, to the root mean square error, $RMSE(\hat{T}) = D^{-1}\sqrt{\sum_{d=1}^{D}(\hat{T}_{(d)} - T_y)^2}$, where d is one of the D=1,000 samples. As suggested in section 5.2 of Särndal, et al (1992), we used an RMSE instead of a standard error due to bias exhibited in the simulation results. This bias ratio affects the desired CI coverage rates through the formula $P(|Z + BR(\hat{T})| \le z_{1-\alpha/2})$, where $Z = [\hat{T} - E(\hat{T})] / RMSE(\hat{T})$. A negligible bias ratio has minimal affect on coverage while a bias ratio larger than one can greatly reduce the coverage rates. In the simulations, confidence intervals were based on pivots defined as $[\hat{T} - T] / RMSE(\hat{T})$. Thus, their coverage properties were determined by the properties of $\hat{T}$ alone and not affected by any variance estimator. Additionally, we examined the characteristics of the relative empirical bias ($relbias(\hat{T}) = (D^{-1}\sum_{d=1}^{D}\hat{T}_{(d)} - T_y)/T_y$) and the relative MSE ($D^{-1}\sum_{d=1}^{D}(\hat{T}_{(d)} - T_y)^2 / T_y^2$).

4. Simulation Results

The simulation results are very similar for the SRS and PPS sampling designs. Many of the general results are similar across the three populations; differences are discussed where appropriate. Because the most notable differences occur across the nonresponse patterns, we discuss each condition separately.

4.1 Missing Completely At Random (MCAR)

The results for the MCAR pattern of nonresponse suggests that specified level of coverage is maintained for the four estimators irrespective of the response rate and simulation condition (*Table 1*). Negligible bias ratios range from 0.001 to 0.054 with larger values for Population 3.

4.2 Missing at Random (MAR)

Empirical results for Population 1 under a MAR pattern mimics those discussed under MCAR. Once the linear relationship between *y* and *x* (Population 1) weakens (Populations 2 and 3), there is a slight reduction from a 95% to a 93% coverage rate for the GREG, and the raw propensity-adjusted MGREG and RARE estimates within *x.cat3*. Coverage actually increases for MGREG and RARE within *x.bin*. The RMSEs for all estimators increase from Populations 1 to 3.

4.3 Non-Ignorable Nonresponse (NINR)

The most notable results in our simulation study exist for the NINR condition. Only the BLUP maintains the desired 95% CI coverage rates for all populations under consideration. The coverage rates for the other estimates fall to levels as low as 83.4% (*Figure 2*). The *other* regression estimators (GREG, MGREG, and RARE) consistently have the lowest coverage rates for Population 3 followed by Population 2. The minimum and maximum values for the bias ratios calculated for the BLUP are provided in *Table 2* by population. The bias ratios for the other regression estimators are not distinguishable and therefore are reported in a single column.

Table 2. Minimum and Maximum Bias Ratios for BLUP and Other Regression Estimators by Population Across All Response Rates

Populations	BLUP	Other Regression Estimators
1	(0.001, 0.065)	(0.729, 0.932)
2	(0.001, 0.050)	(0.882, 0.971)
3	(0.001, 0.063)	(0.941, 0.981)

Figure 3 shows the various levels of relbias for the three populations by response rate. Note that the values for the MGREG and the RARE are not differentiable for each population; the MGREG has a slightly smaller empirical relbias and is represented by the dashed line. The $relbias(\hat{T}_{BLUP})$ fluctuates around zero with a maximum value of 0.08% for all levels of response. This is consistent with the information provided in *Table 2*. The graph of the corresponding RMSEs is similar to *Figure 3*. Under the "best" simulation scenario, the estimated bias for the BLUP under all populations and the regression estimators under Population 1 appears to be negligible across the range of response rates.

Because the model variance is contained within the column space of the design matrix (or auxiliary variables), the estimated total can be expressed as a function of the estimator-specific predictions, $\hat{T} = N\bar{\mathbf{X}}_U\hat{\beta}$, where for example $\hat{\beta} = \hat{\beta}_{BLUP}$ (1) for $\hat{T}_{BLUP}$ (Valliant, et al 2000). *Figure 4* plots the relbias of the predictions for the GREG and BLUP, defined as $[avg(\hat{y}_k) - E_M(y_k)]/E_M(y_k)$, versus $E_M(y_k)$ for Population 1 where $avg(\hat{y}_k)$ $= avg(\hat{\beta})'\mathbf{x}_k$ with $avg(\hat{y}_k)$ being the average over the 1,000 PPS(x) samples under a 25% response rate. The x-axis in *Figure 4*, $E_M(y_k)$, gives the expected values under the true model, $E_M(y_k) = 15x_k - 3x_k^2$. Although the individual prediction relbiases for the GREG are small, we see that this apparent negligible bias for the regression estimators can accumulate to a positive bias that impacts CI coverage. Expected values for the MGREG and RARE are nearly identical to the GREG on this scale and are therefore excluded.

The *theoretical explanation* for the high performance of the BLUP in comparison to the other regression estimators under a NINR response mechanism τ follows. The model expectation conditioned on a response mechanism τ in general can be expressed as $E_M(y_k \mid \tau) = \mathbf{x}_k'\beta + b_k$, a function of the true value plus a bias term (b_k) for $k \in S_R$. If the response mechanism is correlated with the variable of interest, say PPS(y) as in our simulations, then $b_k > 0$ (strictly) and b_k will increase as x_k increases. The model expectation for the BLUP has the following form where $w_{BLUP,k}$ is the "BLUP weight" for the k^{th} unit:

$$\begin{aligned} E_M(\hat{T}_{BLUP} \mid s_R) &= N\bar{\mathbf{X}}_U' E_M(\hat{\beta}_{BLUP} \mid s_R) \\ &= N\bar{\mathbf{X}}_U'[\mathbf{X}_R'\mathbf{V}_R^{-1}\mathbf{X}_R]^{-1}\mathbf{X}_R'\mathbf{V}_R^{-1}E_M(\mathbf{Y}_R \mid s_R) \\ &= N\bar{\mathbf{X}}_U'[\mathbf{X}_R'\mathbf{V}_R^{-1}\mathbf{X}_R]^{-1}\mathbf{X}_R'\mathbf{V}_R^{-1}(\mathbf{X}_R\beta + \boldsymbol{b}_R) \\ &= E_M(\hat{T}) + \sum_{s_R} w_{BLUP,k}b_k \end{aligned}$$

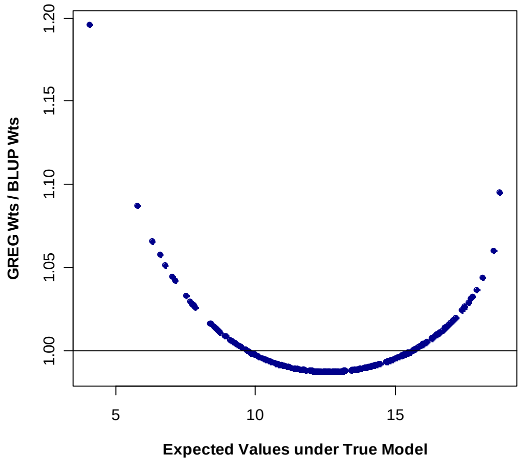
Similarly for the GREG, we have $E_M(\hat{T}_{GREG} \mid s_R) = E_M(\hat{T}) + \sum_{s_R} w_{GREG,k}b_k$.

A comparison of the estimator-specific weights (y-axis) with the expected values under the true model (x-axis) for one PPS(x) Population 1 sample is shown in *Figure 5*. The GREG weights for observations with either relatively low or high x values are higher than the corresponding BLUP weights due to the weight ratios exceeding the horizontal line at one. These observations also correspond with large predicted values under the true model. Therefore, in contrast to the BLUP, the GREG estimator assigns a relatively high-valued weight to observations with extreme predicted values thereby positively increasing the bias.

5. Conclusion and Future Work

An MCAR nonresponse pattern is the primary condition under which the confidence interval coverage is maintained for the estimators under investigation. The choice of the appropriate estimator is secondary. However, only the non-BLUP regression estimators (GREG, MGREG, and RARE) appear to be sensitive under some MAR and all NINR situations. This suggests that the nonresponse adjustments are not eliminating bias. Additionally, calibrating to known control totals as in the regression estimators does not eliminate bias in our research as purported in many standard survey

Figure 5. Comparison of True Predictions to Ratios of GREG and BLUP Weights for One Simulation Sample from Population 1.



textbooks. This problem exists even for estimates with small relbias and high correlation between the analysis and auxiliary variables (Population 1). However, our research suggests that the BLUP is robust to the conditions within our simulation.

The generalizability of our findings is limited to those conditions under study. Of note is the mechanism we used to incorporate NINR into our simulation. Some readers have suggested that PPS(y) is the best condition for the BLUP and that additional NINR mechanisms should be included. Also, samples were selected only using two single-stage designs; additional work is needed with designs that include stratification and multi-stage components with disproportionate sampling. An extension of the theory would naturally follow.

References

Basu, D. (1969), “Role of the Sufficiency and Likelihood Principles in Sample Survey Theory,” *Sankhyā A*, 31, pp. 441-454.

Brewer, K.R.W. (1999), “Design-based or Prediction Inference? Stratified Random vs. Stratified Balanced Sampling,” *International Statistical Review*, 67, pp. 35-47.

Casella, G. and Berger, R.L. (2002), *Statistical Inference*, Duxbury Press.

de Leeuw, E. and de Heer, W. (2002), “Trends in Household Survey Nonresponse: A Longitudinal and International Comparison,” Chapter 3 in *Survey Nonresponse*, Eds: Groves, R.M. Dillman, D.A., Eltinge, J.L., and Little, R.J.A, New York: John Wiley & Sons, Inc.

Gerow, K. and McCulloch, C.E. (2000), “Simultaneously Model-Unbiased, Design-Unbiased Estimation,” *Biometrics*, 56(3), pp. 873-878.

Godambe, V.P. (1955), “A Unified Theory of Sampling from Finite Populations,” *Journal of the Royal Statistical Society B*, 17, pp. 269-278.

_____ (1966), “A New Approach to Sampling from Finite Populations I: Sufficiency and Linear Estimation,” *Journal of the Royal Statistical Society B*, 28, pp. 310-319.

Hansen, M.H., Madow, W.G., and Tepping, B.J. (1983), “An Evaluation of Model-Dependent and Probability Sampling Inferences in Sample Surveys,” *Journal of the American Statistical Association*, 78, pp. 776-793.

Hedlin, D., Dalvey, H., Chambers, R., and Kokic, P. (2001), ”Does the Model Matter for GREG Estimation? A Business Survey Example, ” *Journal of Official Statistics*, 17(4), pp. 527-544.

Kalton, G. and Kasprzyk, D. (1986), ”The Treatment of Missing Survey Data,” *Survey Methodology*, 12(1), pp. 1-16.

Laaksonen, S. and Chambers, R. (2006), “Survey Estimation under Informative Nonresponse with Follow-up,” *Journal of Official Statistics*, 22(1), pp. 81-95.

Lessler, J.T. and Kalsbeek, W.D. (1992), *Nonsampling Error in Surveys*, New York: John Wiley & Sons, Inc.

Little, R.J.A. and Rubin, D.B. (2002), *Statistical Analysis with Missing Data*, New Jersey: John Wiley & Sons, Inc.

Little, R.J.A. (1986), "Survey Nonresponse Adjustments for Estimates of Means," International Statistical Review, 54(2), pp. 139-157.

Pfeffermann, D. (1993), “The Role of Sampling Weights when Modeling Survey Data,” *International Statistical Review*, 61, pp. 317-337.

R Development Core Team (2005). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, http://www.R-project.org.

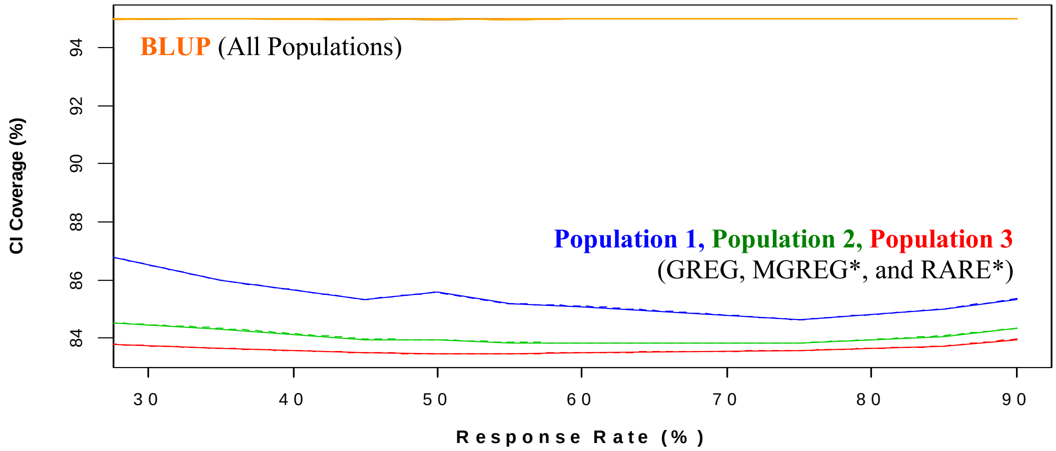
Royall, R.M. and Herson (1973), “Robust Estimation in Finite Populations I,” *Journal of the American Statistical Association*, 68(344), pp. 880-889.

Särndal, C-E. and Lundström, S. (2005), *Estimation in Surveys with Nonresponse*, London: John Wiley & Sons, Inc.

Särndal, C-E., Swensson, B., and Wretman, J. (1992), *Model Assisted Survey Sampling*, Springer-Verlag New York Inc.

Valliant, R., Dorfman, A.H., and Royall, R.M. (2000), *Finite Population Sampling and Inference: A Prediction Approach*, New York: John Wiley & Sons, Inc.

Figure 2. Confidence Interval Coverage Rates by Estimator Group, Population, and Response Rate for Non-Ignorable Nonresponse.



*MGREG and RARE estimates include those adjusted by weighting class and response propensity methods.

Figure 3. RelBias by (NINR) Response Rate by Population for GREG, MGREG, and RARE estimates.

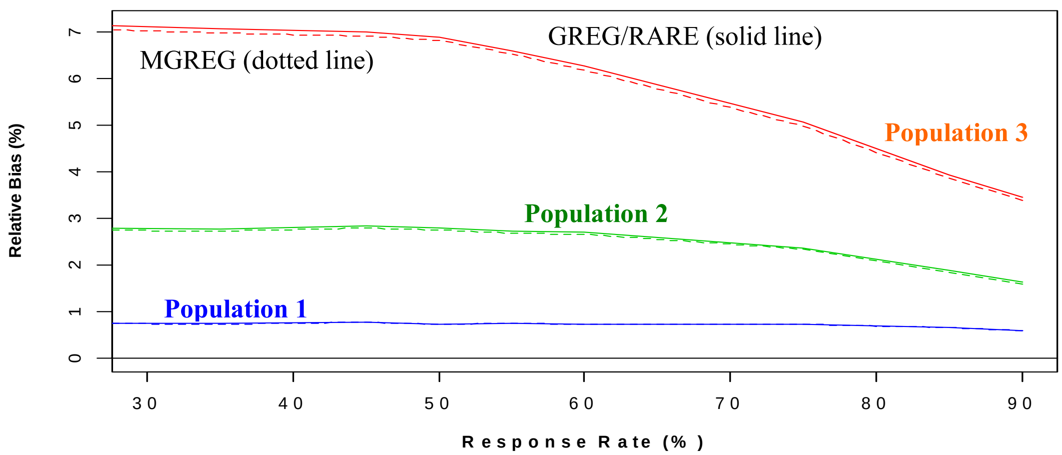


Figure 4. Comparison of True and Estimated Predictions for PPS(x) Samples from Population 1 with a 25% Response Rate.