Survey Research Methods (2009)
Vol.3, No.1, pp. 27-44
ISSN 1864-3361
http://www.surveymethods.org

Comparing Sampling and Estimation Strategies in Establishment Populations

Kimberly Henry
Internal Revenue Service

Richard Valliant
University of Michigan/University of Maryland

Population structure is a key determinant of the efficiency of sampling plans and estimators. Variables in many establishment populations have structures that can be described by simple linear models with a single auxiliary variable and a variance related to some power of that auxiliary. If a working model can be devised that is a good approximation to the population structure, then very efficient sample designs and estimators are possible. This study compares alternative strategies of (i) selecting a pilot study to estimate the variance power and using that estimate to select a main sample and (ii) selecting only a main sample based on an educated guess about the variance power. We also examine a number of sampling plans, including probability proportional to size, deep stratification based on a measure of size, and weighted balanced sampling. Population totals are estimated by best linear unbiased predictors, general regression estimators, and some other choices often used in practice.
Keywords: best linear unbiased predictor, deep stratification, general regression estimator, measure of heteroscedasticity, optimal sample, robust variance estimation, weighted balance

1 Introduction

Estimating totals is often an objective in samples of businesses and other types of establishments. Regardless of whether one uses a design-based or model-based approach to sampling and estimation, one factor that can affect the variance and bias of estimated totals is the superpopulation structure. We consider cases where an analysis variable's variance is proportional to some power of an auxiliary variable. This type of structure is often present in establishment or accounting populations where quantitative variables like revenues or capital assets are collected. The purpose of this paper is to compare alternative ways of estimating that structure and the impact of those alternatives when estimating population totals using a model-based approach.

Various strategies conceivable in this situation include: (1) selection of a relatively small, pilot sample to make preliminary structural parameter estimates, (2) selection of a larger, main sample based on either pilot results or educated guesses about population parameters, and (3) use of either a model-based or design-based estimator of the total. For various single-stage sample designs, sample sizes, and estimators, we compare alternative strategies for estimating values of the variance power for simulated population data. The strategies' effects on estimates of totals and their variances are then evaluated.

This paper is organized into five sections. After the introduction, Section 2 contains descriptions of our superpopulation model and generated populations and related practical applications. Section 3 includes our simulation setup details, while results are discussed in Section 4. Conclusions and limitations are in Section 5.

Contact information: Kimberly Henry, Internal Revenue Service, P.O. Box 2608, Washington DC 20013-2608, e-mail: Kimberly.A.Henry@irs.gov

2 Superpopulation Model and Generated Populations

The populations we study can be described by a fairly simple model structure introduced in section 2.1. This structure fits a variety of real populations, examples of which are presented. For the simulations, we generated some artificial populations that mimic real ones as discussed in section 2.2. Using generated populations allows us to compute nearly optimal estimators based on parameters that are under our control. These estimators serve as a basis of comparison to more realistic ones for which estimators of unknown quantities must be used.

2.1 Model Theory

Given a study variable Y and an auxiliary variable X , we consider a superpopulation with the following structure:

$$E_M(y_i|x_i) = \beta_0 + \beta_1 x_i$$

$$var_M(y_i|x_i) = \sigma^2 x_i^{\gamma} \tag{1}$$

The x_i's are assumed to be known for each unit i in the finite population. In related literature, the exponent γ in model (1)'s conditional variance has been referred to as a *measure of heteroscedasticity* (Foreman 1991), or *coefficient of heteroscedasticity* (Brewer 2002). Applications using models like (1) include companies using "cost segregation" to report depreciable assets on their U.S. Internal Revenue Service Corporate Income Tax Form (e.g., Allen and Foster 2005 and Strobel 2002) and comparing inventory data values versus actual values (e.g., Roshwalb 1987 and Godfrey, Roshwalb and Wright 1984). The current U.S. tax law allows businesses to depreciate capital assets over the life of the asset. Depreciation is recorded as an operating expense and will reduce a business's tax bill. Cost segregation is the process of separating different depreciable assets into different categories by years. For example, buildings can be depreciated over a 39-year period; smaller items, like cash registers and shelving, over a 5-year period. Other asset classes are 7-year and 15-year. The result of cost segregation can be a gain in the company's short-term deduction reported on their tax return, which creates a larger amount of cash flow.

Figure 1 gives *x-y* plots for five real populations that have a structure like that of model (1). The upper left panel shows the U.S. dollar value of 39-year depreciable assets for a set of stores plotted versus total cost of assets for the stores.[1] In the upper right is a plot of patient discharge data for the hospital population in Valliant, Dorfman and Royall (2000). U.S. county employment data are plotted in the second row left panel (see Census Bureau 2007). The fourth plot shows the number of students in New York state public schools versus the number of teachers in each school (National Center for Education Statistics 2007). The last is a plot of expenditures versus number of inpatient beds for a subset of the sample from the 1998 Survey of Mental Health Organizations (SMHSA 2007). The red line in each panel is a nonparametric smoother (lowess) showing that *x-y* relationships can be reasonably described by linear models.

Each panel of Figure 1 shows the estimate of γ in model (1), calculated using the algorithm described later in section 3.4. The range of γ estimates is about 0.6 to 2, which is fairly common in these types of populations.

The variance parameter γ is of interest since, in some populations, having a reasonable γ estimate can be used to produce nearly optimal sample designs and estimators of totals along with their variances. First, the selection probabilities that minimize the anticipated variance of the general regression (GREG) estimator under (1) are proportional to $x_i^{\gamma/2}$ (Särndal, Swensson and Wretman 1992, Sec. 12.2). Second, if the model is

$$E_M\left(y_i|x_i\right) = \beta_{\gamma/2}x_i^{\gamma/2} + \beta_\gamma x_i^\gamma$$

$$var_M\left(y_i|x_i\right) = \sigma^2 x_i^\gamma, \qquad (2)$$

then a *weighted balanced sample*, defined below in (3), will minimize the error variance of the best linear unbiased predictor (BLUP) of the finite population total of Y, i.e. $E_M\left(\hat{T} - T\right)^2$ will be minimized where $\hat{T}$ is the BLUP and T is the total (Theorem 4.2.1 Valliant, et al. 2000). The key requirement of that theorem is satisfied when $E_M\left(y_i|x_i\right)$ is proportional to a linear combination of both $\sqrt{var_M\left(y_i|x_i\right)}$ and $var_M\left(y_i|x_i\right)$, as is the case in (2). The optimal, weighted balanced sample under model (2) has

$$\bar{x}_s^{(\gamma/2)} = \bar{x}^{(\gamma)} \Big/ \bar{x}^{(\gamma/2)} \qquad (3)$$

where $\bar{x}_s^{(\gamma/2)}$ is the sample mean of $x_i^{\gamma/2}$, and $\bar{x}^{(\gamma)}$ and $\bar{x}^{(\gamma/2)}$ are the population means of x_i^γ and $x_i^{\gamma/2}$. However, this type of sample will not be optimal under model (1), since $E_M\left(y_i|x_i\right)$ does not contain $x_i^{\gamma/2}$ or x_i^γ.

Weighted balanced samples can be selected in various ways. Chauvet and Tillé (2006) and Tillé (2006) give some efficient algorithms. In this study, we used a less sophisticated rejective method described in section 3.2.

Model (2), called the *minimal model* (Valliant et al. 2000:100), is associated with the conditional variance, $var_M(y_i|x_i) = \sigma^2 x_i^\gamma$. If (1) were unknown, but the intercept is small, then working model (2) may be a reasonable starting place to determine a sample size.

2.2 Generated Populations

We created four unstratified versions of the population described in Hansen, Madow and Tepping (1983, denoted HMT hereafter) that follow model (1). In an HMT population, x has a gamma distribution with density, $f(x) = (x/25)\exp(-x/5)$; y also has a gamma density conditional on x:

$$g(y; x) = \left[b^c \Gamma(c)\right]^{-1} y^{c-1} \exp(-y/b)$$

where

$$b = (5/4d)x^\gamma(8 + 5x)^{-1},$$
$$c = (1/25)\, x^{-\gamma}\, (8 + 5x)^2, \text{ and}$$
$$d = 1/4 \text{ when } \gamma = 3/4 \text{ and } 1/2 \text{ when } \gamma = 2.$$

With this structure we have $\left(\beta_0, \beta_1, \sigma^2\right) = (1/10,\ 1/16,\ 1/256)$ when $\gamma = 3/4$ and $\left(\beta_0, \beta_1, \sigma^2\right) = (2/10,\ 1/8,\ 1/64)$ when $\gamma = 2$.

Figure 2 shows the X, Y values for each generated population of size 10,000, while smaller populations of size 200 are shown in Figure 3. These populations are similar to some of the real ones in Figure 1 and give us test cases where all parameters are known.

The first population in Figure 2 has a relatively strong dependence between y_i and x_i, while the second one has a much weaker relationship. The smaller populations in Figure 3 were designed to resemble those encountered in accounting applications. In a cost segregation problem, x_i is typically the total monetary value of total capital assets in a store while y_i is the value of 5-year, 7-year, 15-year, or 39-year assets. A depreciation class that accounts for a large part of total assets will have a strong relationship between y_i and x_i, shown in Figure 3(a), while a smaller asset class tends to have a weaker relationship like the one in Figure 3(b).

There is often a huge incentive to use optimal samples and estimators in the applications we consider due to high data collection costs. In a cost segregation study, for example, experts such as lawyers, engineers, and accountants may be

[1] To protect confidentiality, the source of the store data cannot be revealed. The asset values have also been perturbed while maintaining the original structure.

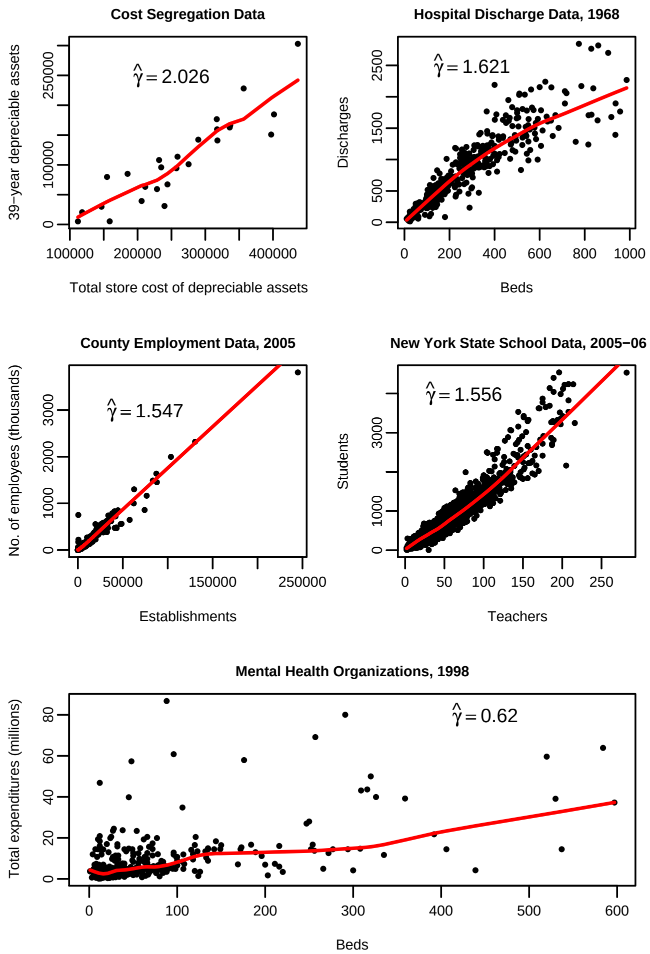


Figure 1. Five real populations of Stores, Hospitals, Counties, Schools and Mental Health Organizations.

needed to assign capital goods to depreciation classes. Time-consuming assessments may have to be done on-site at establishments plus travel and personnel costs can be high; so, the smaller the sample size that yields desired precision, the better (e.g., Rotz, Joshee, and Yang 2006).

3 Simulation Setup

This section describes the details of our simulation study, including working models, sample designs, simulation strategies, and the method of estimating γ. The *working model* is the one used for sample design planning and estimation, but may be misspecified in the mean function, the variance function, or both.

An important practical consideration is how sensitive different estimators and sampling plans are to model misspecification. Note that all the populations in Figures 2 and 3 have a small non-zero intercept, which resulted in some model-based estimators being biased in the earlier HMT study.

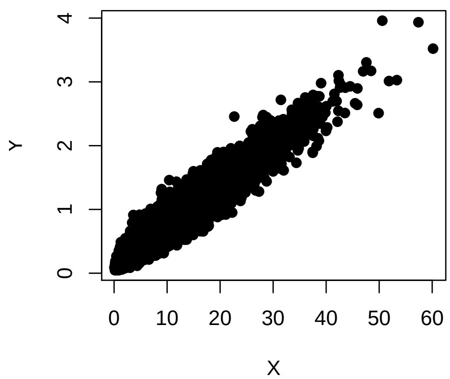
3.1 Models

Using the Valliant et. al (2000) notation, we use $M(\delta_0, \delta_1, \ldots, \delta_J : v)$ to denote the polynomial model $Y_i = \sum_{j=1}^{J} \delta_j \beta_j x_i^j + \varepsilon_i v_i^{1/2}$ with δ_j equal to 1 or 0 depending on whether the term x_i^j is in the model or not. The errors ε_i are independent and identically distributed with mean 0 and variance σ^2 and v_i is an unknown variance parameter. We are specifically interested in $v_i \propto x_i^{\gamma}$. By extension, $M(x^{\gamma/2}, x^{\gamma} : x^{\gamma})$ denotes the model in (2).

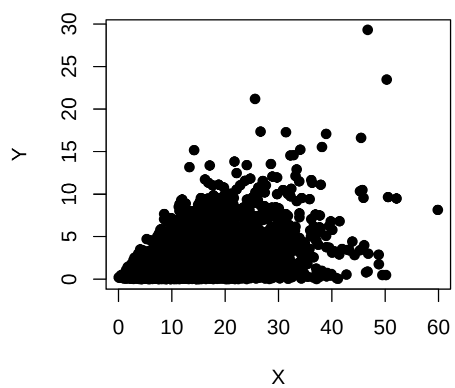
We based estimators of totals on the four working models listed below:

$$\begin{aligned} &M(1, 1 : x^{\gamma}), \text{ or} \\ &\quad E_M(y_i|x_i) = \beta_0 + \beta_1 x_i \\ &\quad Var_M(y_i|x_i) = \sigma^2 x_i^{\gamma} \end{aligned} \tag{4}$$

$$\begin{aligned} &M(x^{\gamma/2}, x^{\gamma} : x^{\gamma}), \text{ or} \\ &\quad E_M(y_i|x_i) = \beta_1 x_i^{\gamma/2} + \beta_2 x_i^{\gamma} \\ &\quad Var_M(y_i|x_i) = \sigma^2 x_i^{\gamma} \end{aligned} \tag{5}$$

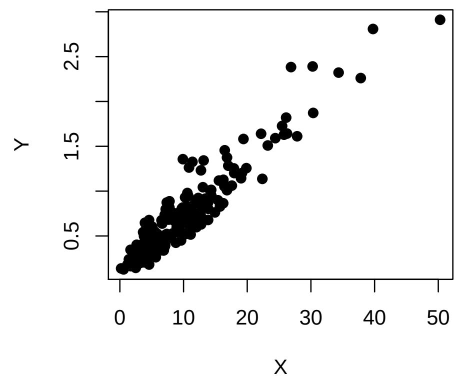


(a) γ=3/4

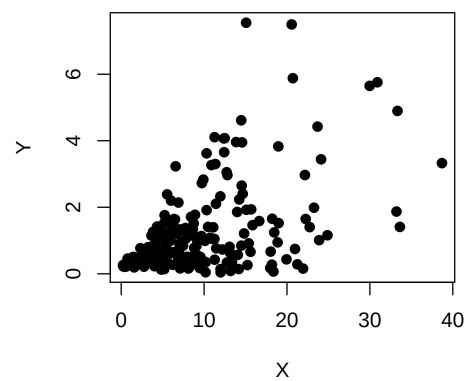


(b) γ=2

Figure 2. Generated Large Populations (N=10,000)



(a) γ=3/4



(b) γ=2

Figure 3. Generated Small Populations (N=200)

$$
\begin{aligned}
&M(0,1:x^{\gamma}), \text{ or} \\
&\quad E_M(y_i|x_i) = \beta_1 x_i \\
&\quad Var_M(y_i|x_i) = \sigma^2 x_i^{\gamma}
\end{aligned} \tag{6}
$$

$$
\begin{aligned}
&M(0,1:x), \text{ or} \\
&\quad E_M(y_i|x_i) = \beta_1 x_i \\
&\quad Var_M(y_i|x_i) = \sigma^2 x_i
\end{aligned} \tag{7}
$$

Model (4) is the correct working model, i.e., the one equivalent to model (1). Model (5) is also the minimal model noted earlier in (2), which may be a reasonable starting place to determine a sample size and an efficient sampling plan.

Model (6) is a special case of model (1) with a zero intercept. Model (7), the *ratio model*, corresponds to the special case of (6) with $\gamma = 1$ in the variance structure. Estimators generated from models (5), (6), and (7) do have model biases under (4), which will increase their mean square errors when averaged over all of the samples that are possible under each design plan.

3.2 Sample Designs

We consider four without replacement (*wor*) sample designs:

(1) *srswor*: simple random sampling without replacement.

(2) *ppswor*: the Hartley-Rao (1962) method with probabilities of selection proportional to a measure of size (MOS).

(3) *ppstrat*: strata are formed in the population by cumulating an MOS and forming strata with equal total size. A simple random sample of one unit is selected from each stratum.

(4) *wtd.bal*: weighted balanced sampling. *ppswor* samples using an MOS are selected that satisfy particular conditions on the population and sample moments of x_i.

Next, we denote a sampling plan that uses probability proportional to a particular MOS by $pp\,(MOS)$. Weighted balanced samples were selected by a rejective algorithm based on probability proportional to size sampling. A $pp\ (\sqrt{x^{\gamma}})$ sample satisfies the balance condition (3) in expectation since

$$
E_\pi\left[\bar{x}_s^{(\gamma/2)}\right] = \bar{x}^{(\gamma)} \Big/ \bar{x}^{(\gamma/2)}
$$

where E_π denotes expectation with respect to repeated sampling. The rejective algorithm consists of three steps:

(i) Select a $pp\ (\sqrt{x^{\gamma}})$ sample s by the method of Hartley-Rao (1962);

(ii) Calculate the balance measures

$$e_k(s) = \left| \sqrt{n}\left[\bar{x}_s^{(k)} - E_\pi\left(\bar{x}_s^{(k)}\right)\right] \Big/ s_{kx} \right| (k = \gamma/2, \gamma),$$

where $s_{kx}\big/\sqrt{n}$ is the standard error of $\bar{x}_s^{(k)}$ in repeated *pp* ($\sqrt{x^\gamma}$) with-replacement sampling; and

(iii) Determine whether $e_k(s) \leq 0.125$ for $k = \gamma/2, \gamma$. If so, the sample was considered to have weighted balance and was retained; if not, steps (i) and (ii) were repeated until $e_k(s) \leq 0.125$. Using the 0.125 cut point leads to rejecting about 90% of samples.

The rejective sampling algorithm defined by (i)-(iii) above is denoted *wtd.bal*($\sqrt{x^\gamma}$). Three types of MOS were used for weighted balanced sampling: $\gamma = 1/2$, γ equal to the true value used to generate a population, and γ equal to the estimated value from a pilot study.

As noted in section 2.1, the selection probabilities that minimize the anticipated variance of the GREG under model (1) are proportional to $\sqrt{x^\gamma}$. When the MOS is $\sqrt{x^\gamma}$, the *ppstrat* design approximates optimal *pp* ($\sqrt{x^\gamma}$) selection and *wtd.bal* ($\sqrt{x^\gamma}$) sampling. The design *ppstrat* is also similar to "deep stratification" (e.g, Bryant et al. 1960; Cochran 1977:124-126; Sitter and Skinner 1994), which is used in accounting applications (Batcher and Liu 2002). More specific details on these designs are given in pages 66-67 of Valliant et al. (2000).

3.3 Strategies

The strategies we examined consisted of (i) selecting a pilot study to get a preliminary estimate of γ followed by a main sample or (ii) only selecting a main sample. Both options were crossed with the possibility of rounding $\hat{\gamma}$ to the nearest one-half or using the original estimate. Rounding $\hat{\gamma}$ is one plausible method for limiting the effect of unstable point estimates of the variance parameter. Thus, our main comparisons concern four strategies:

A: draw a small *pp* ($\sqrt{x}$) pilot sample, estimate γ, select a main sample using *pp* ($\sqrt{x^{\hat{\gamma}}}$), *ppstrat* ($\sqrt{x^{\hat{\gamma}}}$), and *wtd.bal* ($\sqrt{x^{\hat{\gamma}}}$) samples, and construct estimates using the pilot estimate, $\hat{\gamma}$.

B: draw *srswor*, *ppswor* ($\sqrt{x}$), *ppstrat* ($\sqrt{x}$), and *wtd.bal* ($\sqrt{x}$) main samples only, estimate γ in each, and construct estimates using that $\hat{\gamma}$.

C: strategy A, rounding $\hat{\gamma}$ to the nearest one-half.

D: strategy B, rounding $\hat{\gamma}$ to the nearest one-half.

Note that, in strategy A, using the pilot $\hat{\gamma}$ is necessary to construct the GREG, defined in section 3.5, with its appropriate selection probabilities.

There is no *srswor* used for strategies A and C. Also, B and D correspond to assuming $\gamma = 1$ for selecting the *ppswor*, *ppstrat*, and *wtd.bal* samples. The $\gamma = 1$ specification does not match our population γ's, but will be a reasonable advance choice for sampling in many populations. We consider the rounding in C and D to see if reducing variability in the $\hat{\gamma}$'s leads to improved estimates of totals and their variances. The estimates, $\hat{\gamma}$'s, described above, are used in some of our estimates of totals and variances, as described in sections 3.5 and 3.6.

The combinations of pilot and main sample sizes are:

Population	Pilot sample size	Main sample size
Small	10	25, 50
Large	10	50, 100
Small	No pilot	25, 50
Large	No pilot	50, 100

3.4 Estimation of γ

When the expected value, $E_M(y_i|x_i)$, is specified correctly and the model is fit by ordinary least squares (OLS), the squared residual can be used to estimate the unknown model variance. In particular, we define $r_i = y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}$, where $\boldsymbol{x}_i$ is the vector of x's in the working model and $\hat{\boldsymbol{\beta}}$ is the OLS estimate of the associated model parameters. Then $E_M(r_i^2)$ is an approximately unbiased estimate of $var_M(y_i|x_i)$. When $var_M(y_i|x_i) = \sigma^2 x_i^\gamma$, this leads to the following algorithm (e.g., see Roshwalb 1987):

(a) Fit the model $E_M(y_i|x_i) = \boldsymbol{x}_i'\boldsymbol{\beta}$ by OLS and obtain the residuals.
(b) Regress the log of the squared residuals on $\log(x)$, i.e. $\log(r_i^2) = \alpha + \gamma \log(x_i)$ using OLS.
(c) Fit the model $E_M(y_i|x_i) = \boldsymbol{x}_i'\boldsymbol{\beta}$ by weighted least squares with weights, $1\big/x_i^{\hat{\gamma}}$ using $\hat{\gamma}$ from step (b). Obtain the residuals.
(d) Repeat steps (b) and (c) until the relative change in $\hat{\gamma}$ is small.

For the probability samples, an option would be to fit the models in the algorithm by sample-weighted least squares; we did not pursue that here although in practice it may be a reasonable step to account for an informative design. We also experimented with fitting the nonlinear model $r_i^2 = \sigma^2 x_i^\gamma + \varepsilon_i$ but this led to severe convergence problems in many samples, and this method was discarded.

3.5 Estimation of Totals

We consider three kinds of estimators for totals: the *Horvitz-Thompson* (HT) estimator, *best linear unbiased predictors* (BLUP), and *general regression estimators* (GREG). These were selected to span the design- and model-based choices that are of theoretical and practical interest. The HT estimator is given by

$$\hat{T}_\pi = \sum_s y_i/\pi_i,$$

where π_i is the probability of selection for unit i and s is the set of sample units. In weighted balanced samples, we used the same value of π_i as appropriate for *ppswor* samples. This is, of course, not necessarily the actual selection probability because of the restrictive nature of balanced samples. This approach was also used for the GREG estimators computed from weighted balanced samples.

The general form of the BLUP estimator is

$$\hat{T} = \sum_s y_i + \sum_r \boldsymbol{x}_i' \hat{\boldsymbol{\beta}},$$

where r is the set of nonsample units, $\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}$ is the prediction for y_i $(i \in r)$ using the working model, and $\hat{\boldsymbol{\beta}}$ is estimated using the sample units $(i \in s)$. For example, the BLUP using the correct model (1) has

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}_s'\boldsymbol{V}_{ss}^{-1}\boldsymbol{X}_s\right)^{-1} \boldsymbol{X}_s'\boldsymbol{V}_{ss}^{-1}\boldsymbol{y}_s, \tag{8}$$

where $\boldsymbol{X}_s$ is an $n\times 2$ matrix with rows $(1, x_i)$, $\boldsymbol{V}_{ss} = diag(x_i^{\gamma})$, and $\boldsymbol{y}_s$ is the n-vector of sample data values. The BLUP for a well-fitting model has the advantage of being very efficient compared to other options.

The general form of the GREG estimator is

$$\hat{T}_{GR} = \sum_s g_i y_i / \pi_i,$$

where $g_i = 1 + \left(\boldsymbol{X} - \hat{\boldsymbol{X}}_\pi\right)' \boldsymbol{A}^{-1}\boldsymbol{x}_i/v_i$ is the "g-weight" for unit i (Särndal et al. 1992), where $\boldsymbol{A} = \boldsymbol{X}_s'\boldsymbol{\Pi}^{-1}\boldsymbol{V}_{ss}^{-1}\boldsymbol{X}_s$, $\hat{\boldsymbol{X}}_\pi$ is the HT estimate for the population total of $\boldsymbol{X}$, and $\boldsymbol{\Pi} = diag\,(\pi_i)$. The main advantages of the GREG are its approximate design-unbiasedness, even if the working model is wrong, and its high efficiency if the working model is correct.

We combined these estimators with the four working models, true value of γ, and estimates of γ to form eight estimators of totals. We use notation for estimated totals that parallels the earlier notation for models. In particular, $\hat{T}(\delta_0, \delta_1, \ldots, \delta_J : v)$ denotes the BLUP under $M(\delta_0, \delta_1, \ldots, \delta_J : v)$ while $\hat{T}_{GR}(\delta_0, \delta_1, \ldots, \delta_J : v)$ is the GREG under that model and whatever sample design is used. For model (4), we have $\hat{T}(1, 1 : x^{\gamma})$, $\hat{T}(1, 1 : x^{\hat{\gamma}})$, and $\hat{T}_{GR}(1, 1 : x^{\hat{\gamma}})$. All three of these are model-unbiased under (4), and, therefore, design/model-unbiased in the sense that $E_\pi E_M(\hat{T} - T) = 0$, where E_π denotes the expectation with respect to a given sample design. $\hat{T}_{GR}(1, 1 : x^{\hat{\gamma}})$ and the other GREGs mentioned below are approximately design-unbiased (regardless of underlying model) as long as correct selection probabilities are used. The estimators $\hat{T}(x^{\hat{\gamma}/2}, x^{\hat{\gamma}} : x^{\hat{\gamma}})$ and $\hat{T}_{GR}(x^{\hat{\gamma}/2}, x^{\hat{\gamma}} : x^{\hat{\gamma}})$ correspond to model (5), while $\hat{T}_{GR}(0, 1 : x^{\hat{\gamma}})$ is for model (6); all three estimators are model-biased under (4). $\hat{T}_\pi$ and $\hat{T}_{GR}(0, 1 : x)$, which do not involve γ, are the seventh and eighth estimators considered. Each of these is also model-biased under (4). The ratio estimator, $\hat{T}(0, 1 : x)$, is approximately design-unbiased only under *srswor*. $\hat{T}_\pi$ and $\hat{T}_{GR}(0, 1 : x)$ are (approximately) design-unbiased in any probability sampling plan if the correct selection probabilities are used to construct the estimators. Note that *wtd.bal*($\sqrt{x^{\gamma}}$) sampling is theoretically optimal only for $\hat{T}(x^{\gamma/2}, x^{\gamma} : x^{\gamma})$ under model $M(x^{\gamma/2}, x^{\gamma} : x^{\gamma})$. For other estimators, like $\hat{T}(0, 1 : x)$, the ratio estimator, *wtd.bal* ($\sqrt{x^{\gamma}}$) or *wtd.bal*($\sqrt{x}$) sampling can give biased, inefficient estimates. Note that the true γ is not available in any real situation; the estimator $\hat{T}(1, 1 : x^{\gamma})$ has the smallest theoretical error variance of any estimator under model (1) and is used as a standard of comparison for the other choices.

3.6 Variance Estimation

There are a number of alternative design- and model-based choices for variance estimation. The ones used in our simulation are described in this section. For the HT estimator, the variance estimator is:

$$v_0\left(\hat{T}_\pi\right) = \left(1 - \frac{n}{N}\right)\frac{n}{n-1}\sum_s \left(\frac{y_i}{\pi_i} - \frac{1}{n}\sum_s \frac{y_i}{\pi_i}\right)^2.$$

This variance expression assumes with replacement sampling, but uses the finite population correction adjustment $1 - n/N$ to approximately account for *wor* sampling.

For the BLUP estimators, we used a robust *leverage-adjusted variance* estimate:

$$v_D\left(\hat{T}\right) = \sum_s \frac{a_i^2 r_i^2}{1 - h_{ii}} + \frac{\sum_r x_i^{\hat{\gamma}}}{\sum_s x_i^{\hat{\gamma}}} \sum_s r_i^2,$$

where $h_{ii} = \boldsymbol{x}_i'\left(\boldsymbol{X}_s'\boldsymbol{V}_{ss}^{-1}\boldsymbol{X}_s\right)^{-1}\boldsymbol{x}_i/v_i$ is the leverage for unit i, $a_i = \boldsymbol{1}_r'\boldsymbol{X}_r\left(\boldsymbol{X}_s'\boldsymbol{V}_{ss}^{-1}\boldsymbol{X}_s\right)^{-1}\boldsymbol{x}_i/v_i$, and $r_i = y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}$ with $\hat{\boldsymbol{\beta}}$ defined in the associated working model. The second term in v_D accounts for variability in the population units that are not in the sample.

For the GREG's, we include variance estimators with the following form (e.g., see Valliant 2002, expression 2.4)

$$v_{D,GR}(\hat{T}_{GR}) = \left(1 - \frac{n}{N}\right)\sum_s \frac{g_i^2 r_i^2}{\pi_i^2\,(1 - h_{ii})}$$

with $r_i = y_i - \boldsymbol{x}_i'\hat{\boldsymbol{B}}$, $\hat{\boldsymbol{B}} = \left(\boldsymbol{X}_s'\boldsymbol{\Pi}^{-1}\boldsymbol{V}_{ss}^{-1}\boldsymbol{X}_s\right)^{-1}\boldsymbol{X}_s'\boldsymbol{\Pi}^{-1}\boldsymbol{V}_{ss}^{-1}\boldsymbol{y}_s$, and $h_{ii} = \boldsymbol{x}_i'\left(\boldsymbol{X}_s'\boldsymbol{V}_{ss}^{-1}\boldsymbol{\Pi}^{-1}\boldsymbol{X}_s\right)^{-1}\boldsymbol{x}_i/(v_i\pi_i)$. Since the ratio estimator is a special case of a GREG, the particular form of $v_{D,GR}$ was used for its variance estimator.

Note that v_D and $v_{D,GR}$ can both be affected by the use of an estimate of γ, since $\hat{\gamma}$ enters in both the leverages and the residuals. The exception to this is the ratio estimator, where γ is set to 1.

The same variance estimators were used for all sample designs, except for the *ppstrat* design-based variances for the HT and GREG estimators. In *ppstrat*, one sample unit is selected per stratum. Thus, adjacent strata had to be collapsed to estimate a variance. That is, successive pairs of sample units were formed, variances were calculated within each stratum, and strata variances were cumulated to give a stratified version of v_0. Since all of our working models were specified over all strata, v_D was used for samples selected using *ppstrat* sampling in estimating the variance of the BLUP.

We summarize our simulation factors, the number of levels for each factor, and the associated levels in Table 1.

4 Simulation Results

Results from the simulations are summarized in this section. Since the results are extensive, we mainly present summary statistics graphically, including estimates of γ, relative biases of estimated totals, root mean square errors, and confidence interval properties. For each combination of simulation parameters described in sections 3.2 and 3.3, we selected 1,000 samples.

4.1 γ Estimates

Since we are interested in the practicality of the various procedures, it is worth noting some of the numerical difficulties that can arise when estimating γ. Generally, the iterative procedure described in section 3.4 converged and produced reasonable answers. However, four different problems occurred during the simulations:

(1) The procedure produced negative γ estimates;
(2) $\hat{\gamma}$ diverged to positive or negative infinity;
(3) The procedure did not diverge but did not converge to an estimate;
(4) The procedure converged to an unreasonably large $\hat{\gamma}$.

These are well-known phenomena in numerical analysis (Gentle 2002), but it is worthwhile to understand them in more depth for this simple problem. Problems (1), (3), and (4) occurred for all four populations, while (2) only occurred for the small ones. We illustrate (1)-(3) with examples.

Example 1: Negative $\hat{\gamma}$ The first example is a *pp* ($\sqrt{x}$) pilot of size 50 selected from the large population with $\gamma = 3/4$. The *x-y* sample plot, shown in Figure 4, does not appear to be problematic, but this is an apparently innocuous configuration of sample data that can still lead to convergence problems. The "problems" associated with this type of sample are that the population units with larger x_i's are missing and the variability in the y_i's is greater for smaller values of x_i than larger ones. This resulted in the estimate of γ becoming negative by the fifth iteration of the program (see Figure A.1 in Appendix A).

Example 1 also resulted in problem (3): the iterative program did not converge to a solution within 100 iterations. This problem, which also occurred for positive estimates of γ in some samples, is shown in Figure 5. Estimates oscillate within a fairly narrow range of negative values but do not converge.

For all strategies, if $\hat{\gamma} \leq 0$, $\hat{\gamma}$ was forced to 1, which corresponds to *pp* ($\sqrt{x}$) sampling. Rejected alternatives included forcing $\hat{\gamma} = 0$, implying homoscedasticity, or dropping these samples, both of which are unrealistic in practice. As an illustration of the extent to which this was necessary, Table B.1 in Appendix B shows the number of these occurrences

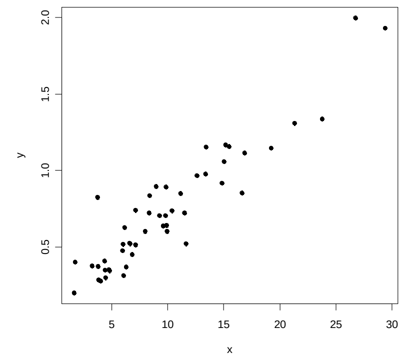


Figure 4. Sample Plot of Problem Example 1

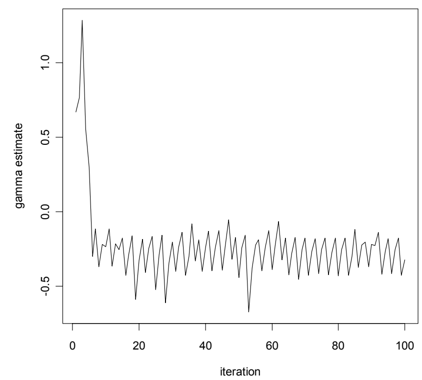


Figure 5. Estimates of γ for the first 100 iterations, Problem Example 1

for the $\gamma = 3/4$ population (there were less than 5 cases for each strategy for the $\gamma = 2$ population). In this table, strategy A and B's numbers are the number of negative $\hat{\gamma}$'s. For C and D, the numbers include cases where small positive $\hat{\gamma}$'s were rounded down to zero. The numbers in parentheses are the number of negative $\hat{\gamma}$'s. The rounding used for C and D leads to fewer negative estimates than in A and B, but rounding does not offer overall improvement, as discussed in a later section. Strategies B and D produced fewer negative $\hat{\gamma}$'s than A and C since B and D use 50 and 100 units, as opposed to pilot samples of size 50 in A and C. Also, depending on the strategy, there were considerably more negative $\hat{\gamma}$'s using model (5) or (6) compared to using (4).

The percentage of samples in Table B.1 with $\hat{\gamma}$ reset to 1 range from 0.07% for ($M(1, 1 : x^{\gamma})$, B, *srswor*, main n=100) to 42.6% for ($M(x^{\gamma/2}, x^{\gamma} : x^{\gamma})$, C, *ppswor*, pilot n=10). None of the sampling methods – *ppswor*, *ppstrat*, *wtd.bal* – is less prone to the problem than the others.

Example 2: Diverging $\hat{\gamma}$ Our second example, a *pp*($\sqrt{x}$) pilot of size 10 from the small population with $\gamma = 3/4$, shows problem (2). The *x-y* plot, in Figure 6, looks well-behaved, but this, again, is deceptive.

The combination of points with small x's together with the three largest x's results in divergence. After the first five iter-

Table 1: Simulation Factors, Number of Levels (#), and Levels, by Simulation Factor

Factor	# Levels	Levels
Population sizes	2	200 10,000
γ	2	3/4 2
Strategies	4	A pilot/no rounding $\hat{\gamma}$ B main sample $\hat{\gamma}$ C - pilot/rounding $\hat{\gamma}$ D - main sample/rounding $\hat{\gamma}$
Sample designs	4	srswor (B, D) ppswor ppstrat wtd.bal
Sample sizes	2	
Pilot sample size/ Main sample size combinations (A, C)		Small populations: 10/25, 50 Large populations: 10/50, 100
Main sample size (B, D)		Small populations: 25, 50 Large populations: 50, 100
Estimators of Totals	9	$\hat{T}_\pi$, $\hat{T}_{BLU}(0, 1 : x)$, $\hat{T}_{GREG}(0, 1 : x)$, $\hat{T}_{BLU}(0, 1 : x^{\hat{\gamma}})$, $\hat{T}_{BLU}(1, 1 : x^{\gamma})$, $\hat{T}_{BLU}(1, 1 : x^{\hat{\gamma}})$, $\hat{T}_{BLU}(x^{\hat{\gamma}/2}, x^{\hat{\gamma}} : x^{\hat{\gamma}})$, $\hat{T}_{GREG}(1, 1 : x^{\hat{\gamma}})$, $\hat{T}_{GREG}(x^{\hat{\gamma}/2}, x^{\hat{\gamma}} : x^{\hat{\gamma}})$, $\hat{T}_{GREG}(0, 1 : x^{\hat{\gamma}})$
Variance estimators	1	$v_0\left(\hat{T}_\pi\right)$, $v_D\left(\hat{T}\right)$, or $v_{D,GR}(\hat{T}_{GR})$

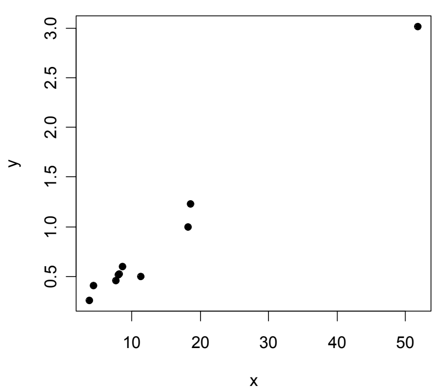


Figure 6. Sample Plot of Problem Example 2

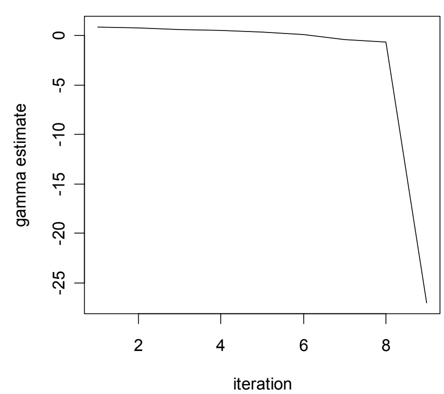


Figure 7. Estimates of γ for the first 9 Iterations, Problem Example 2

ations (see Figure A.2 in Appendix A), $\hat{\gamma}$ becomes progressively more negative until iteration ten, when the estimate is negative infinity. Figure 6 shows the estimates for the first 9 iterations. In the simulations, the number of iterations was limited and negative $\hat{\gamma}$'s were treated as previously described.

To resolve problem (4) for all strategies, if $\hat{\gamma} > 3$ (including infinity), then it was forced to equal 3 to avoid unreasonably large $\hat{\gamma}$'s. Table B.2 contains the number of these occurrences for the (Large, $\gamma = 2$) population; there were none of these cases for the $\gamma = 3/4$ population. Strategies B and D produced many fewer large $\hat{\gamma}$'s than A and C. Rounding in C and D also produced fewer large $\hat{\gamma}$'s because only a main sample is used for them. Model (6) also produced fewer large $\hat{\gamma}$'s than model (4), but noticeably more than model (5) for strategies A and C, which use pilot studies.

The percentages of pilot samples in Table B.2 where $\hat{\gamma}$ was reset to 3 range from 0% when n=100 for several sample designs for working models $M(x^{\gamma/2}, x^{\gamma} : x^{\gamma})$ and $M(0, 1 : x^{\gamma})$ to 39.0% for strategies A and C with pilots of n=10 and *wtd.bal* samples). As in Table B.1, none of the sampling methods – *ppswor*, *ppstrat*, *wtd.bal* – seems noticeably better for avoiding this problem.

Clearly, the small sized pilots (A and C) are more prone to unrealistic estimates of γ than are the larger main samples (B and D). This is no surprise, but does call attention to a risk associated with pilots or with drawing extremely small samples from populations similar to the ones we used. Figure 8 is a set of boxplots of $\hat{\gamma}$'s for main samples of size n= 50 from the (Small, $\gamma = 3/4$) population for Strategies B and D. The characteristics of the distribution of the $\hat{\gamma}$'s across the

1,000 simulated samples are different for the (Small, $\gamma = 2$) population and the large population, but Figure 8 gives some of the flavor of what may be experienced in practice. The horizontal reference line is drawn at 3/4, which is the target. If estimation of γ were the main goal, then a good strategy would have the box centered at 3/4, would have a short interquartile range, and few, if any, outliers.

Strategy D has a median $\hat{\gamma}$ of 1 for all sample designs and estimators due to rounding and resetting of negative estimates to 1. Thus, rounding introduces a bias into $\hat{\gamma}$, as might be expected. Simple random sampling produces more extreme estimates for each working model than the other plans. Even when the working model correctly contains an intercept and x, the median $\hat{\gamma}$ is not equal to 3/4, although the interquartile range is usually among the smallest for the $M(1, 1 : x^{\gamma})$ model. On the whole, attempts to estimate the variance parameter in (Small, $\gamma = 3/4$) were not particularly effective. This is generally consistent with the suggestion of Brewer (2002) that extremely large sample sizes may be needed to get good estimates of γ.

4.2 Relative Biases and RMSEs of Estimated Totals

Our primary focus is how estimating $\hat{\gamma}$ effects estimates of totals and their variances. First, consider the biases of estimated totals. We summarize the results here without showing detailed tables or graphics. The empirical relative bias (relbias) of an estimator is defined as

$$relbias\left(\hat{T}\right) = 100 \sum_{k=1}^{1000} \left(\hat{T}_k - T\right) \Big/ (1000T)$$

where $\hat{T}_k$ is one of the estimated totals from simulation sample k. Relbiases in the two large populations and in (Small, $\gamma = 3/4$) were all less than 1.4% in absolute value. The range in relbiases in the (Small pop, $\gamma = 2$) population was -2.7% for ($\hat{T}\left(x^{\hat{\gamma}/2}, x^{\hat{\gamma}} : x^{\hat{\gamma}}\right)$, D, *srswor, n*=25) to 2.9% for ($\hat{T}(1, 1 : x^{\gamma})$, D, *srswor, n*=25). Although these relbiases are relatively large, both are within simulation error of zero. They are also indicative of the fact that *srswor* is an inefficient sampling plan in these types of populations.

A basic summary statistic for an estimated total is the root mean square error defined as

$$RMSE\left(\hat{T}\right) = \sqrt{\sum_{k=1}^{1000} \left(\hat{T}_k - T\right)^2 \Big/ 1000.}$$

To explore which factors were important in determining mean square errors (i.e., $RMSE^2$), we regressed $RMSE^2$ on factors with levels defined by:

- Population (4 levels),
- Strategy (A, B, C, D),
- Sample Design (*srswor*, *ppswor*, *ppstrat*, *wtd.bal*),
- Estimator (the eight choices in section 3.5),
- Two-way interactions of Population, Strategy, Sample Design, and Estimator.

Since $RMSE^2$ is an average of 1,000 observations, it expected to be roughly normally distributed. Main sample size acts as a blocking factor and is, naturally, a significant predictor of $RMSE^2$. Separate regressions were run for main samples sizes of 25, 50, and 100.

Based on the F-test for comparing a reduced model to a full model (e.g., see Searle 1971:116-120), the (Sample Design)*(Estimator) and (Strategy)*(Estimator) interactions could all be eliminated, yielding a model of the form

RMSE^2 = *Population Strategy*
(Sample Design) Estimator
*Population*Strategy*
Population(Sample Design)*
*Population*Estimator*
Strategy(Sample Design)*
*Strategy*Estimator*

The same reduction occurred for each of the samples sizes of 25, 50, and 100. The message from this analysis is that describing differences in root mean square errors is a fairly complex process. Graphical summaries will be the easiest to understand.

The theoretically best estimator of the total in each of the populations is $\hat{T}(1, 1 : x^{\gamma})$. Although not necessarily optimal, weighted balanced sampling, as defined in (3), should be efficient for this estimator because of its close relationship to $pp\left(x^{\gamma/2}\right)$ sampling. The optimal estimator uses the true value of γ and working model (1). One of our simulation choices was $\hat{T}(1, 1 : x^{\gamma})$ with *wtd.bal* ($\sqrt{x}$) sampling. This method of sampling is not necessarily the most efficient but should be one of the smaller variance choices in the study. Thus, we computed the precision of an estimator of the total relative to that of [$\hat{T}(1, 1 : x^{\gamma})$, *wtd.bal* ($\sqrt{x}$)] by taking the ratio of the root mean square errors (RMSEs), that is

$$RMSE(\hat{T}) \Big/ RMSE\left[\hat{T}(1, 1 : x^{\gamma}), wtd.bal(\sqrt{x})\right].$$

Separate ratios were computed for the different main sample sizes of 50 and 100 for the large populations and 25 and 50 for the small populations. This allows for a relative comparison between strategies and estimators, for a given sample size.

Figures 9-12 are dotplots of root mean square error ratios for the various (Estimator, Strategy, Sample Design) combinations for each of the four populations. Vertical reference lines are drawn at 1. All RMSEs in the left column of each Figure are divided by the same RMSE for $\hat{T}(1, 1 : x^{\gamma})$ in *wtd.bal* ($\sqrt{x}$) samples for the main size in that column (i.e., $n = 50$ in Figures 9-10 and $n = 25$ in Figures 11-12. Likewise, the right-hand column RMSEs are divided by the RMSE for $\hat{T}(1, 1 : x^{\gamma})$ in *wtd.bal* ($\sqrt{x}$) samples of size $n = 100$ in Figures 9-10 and $n = 50$ in Figures 11-12. In the row labels of these and subsequent figures, g.hat denotes $\hat{\gamma}$. Some general observations are:

(a) *ppswor* ($\sqrt{x}$) and *srswor* can be poor compared to the optimal estimator, $\hat{T}(1, 1 : x^{\gamma})$. This is true for both the large and small populations.[2]

[2] The combination (HT, srswor) is omitted in the figures because

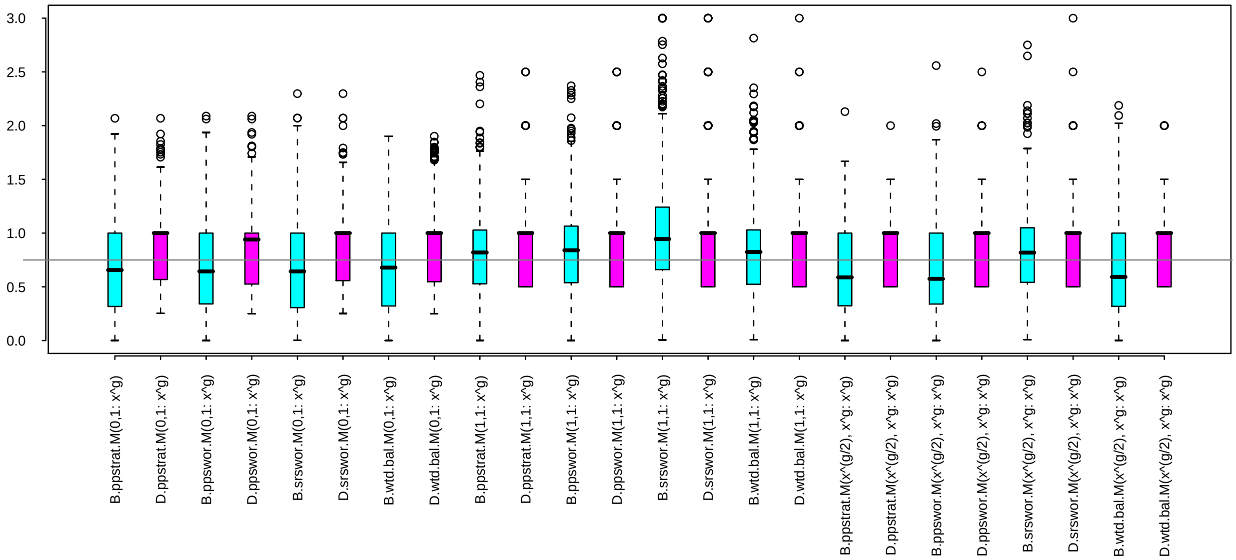


Figure 8. Boxplots of $\hat{\gamma}$'s for main samples of size n= 50 from the (Small, $\gamma = 3/4$) population for Strategies B and D.

(b) When no pilot sample is selected (B or D), *wtd.bal* sampling is often the best method. This is true even for the GREG estimators. The exception is the (Small, $\gamma = 2$) population where *ppstrat* ($\sqrt{x}$) is most efficient, but the difference from *wtd.bal* is only 5% or less.

(c) When a pilot is selected (A or C) from the less variable $\gamma = 3/4$ populations, γ estimated, and the main sample selected based on $\hat{\gamma}$, this is often worse in terms of higher RMSEs than directly selecting the main sample using $\sqrt{x}$ as the MOS (B or D) in *ppswor*, *ppstrat*, or *wtd.bal*.

(d) *ppstrat* is competitive with *wtd.bal,* particularly in B and D which use $\sqrt{x}$ as the MOS. In A and C for (Large pop, $\gamma = 2$), *ppstrat* $\left(x^{\hat{\gamma}/2}\right)$ is usually more efficient than *wtd.bal* $\left(x^{\hat{\gamma}/2}\right)$ when the main sample is 100. The same is true in A and C for (Small pop, $\gamma = 2$) at either main sample of 25 or 50.

(e) Rounding $\hat{\gamma}$ to the nearest (C and D) typically is ineffective in reducing RMSEs. In some cases, rounding makes things somewhat worse (see Small pop, $\gamma = 3/4$ for $\hat{T}(x^{\hat{\gamma}/2}, x^{\hat{\gamma}} : x^{\hat{\gamma}})$ at n= 25 with *ppstrat* or *wtd.bal*).

(f) For the smaller sample size in either the Large or Small populations with no pilot selected, weighted balanced sampling with $\sqrt{x}$ as the MOS is quite efficient.

(g) When the BLUP and GREG are based on the correct working model, $M(1, 1 : x^{\gamma})$ but with estimated γ, the BLUP usually has smaller RMSE. In contrast, when the wrong working model is used, $\hat{T}_{GR}(x^{\hat{\gamma}/2}, x^{\hat{\gamma}} : x^{\hat{\gamma}})$ can be more efficient than $\hat{T}(x^{\hat{\gamma}/2}, x^{\hat{\gamma}} : x^{\hat{\gamma}})$.

4.3 Confidence Interval Properties

The coverage of confidence intervals (CIs) depends on several factors, including the bias of an estimated total and consistency of the variance estimator. In this application, CI coverage is effected by using the estimated $\hat{\gamma}$ in some of the sampling plans, in the $\hat{T}$'s, and in their variance estimators. Using $\hat{\gamma}$ is a type of adaptive procedure whose variability is not reflected in variance estimation. In adaptive procedures, confidence intervals typically do not cover at the proper rates. For example, in stepwise regression, the usual standard error estimates are well-known to be too small (Hurvich and Tsai 1990, Zhang 1992), leading to confidence intervals that cover at less than nominal rates and significance tests with inflated Type I error levels.

Figures 13-14 are dotplots of the empirical proportions of 95% confidence intervals that covered the population totals in the simulations for the (Small, $\gamma = 3/4$) and (Small, $\gamma = 2$) populations. Findings were similar for the Large population. CIs using a particular $\hat{T}$ were computed using the corresponding variance estimate described in section 3.6. The general form of the CIs was $\hat{T} \pm 1.96\sqrt{v}$, where v is a variance estimate described in Section 3.6. Marginal increases in coverage rates could have been obtained by using t-intervals, particularly when $n = 25$. Summary observations are:

(a) In *wtd.bal* samples, there is a tendency for CIs to cover more than 95% of the time for the less variable, (Small, $\gamma = 3/4$) population for the ratio, HT, and $\hat{T}(x^{\hat{\gamma}/2}, x^{\hat{\gamma}} : x^{\hat{\gamma}})$ estimates. This occurs since weighted balance is not a theoretically appropriate design for these estimators.

(b) In the high variance (Small, $\gamma = 2$) population, coverage with (*srswor*, B or D) is generally among the lowest, although over 90% in all cases. This is another reflection that *srswor*is a poor design for these populations.

(c) Coverage rates generally increase for $n = 50$ compared to n= 25 for all strategies and estimators in (Small, $\gamma = 2$)

its RMSE ratios were far larger than the others; inclusion would have distorted the scales.

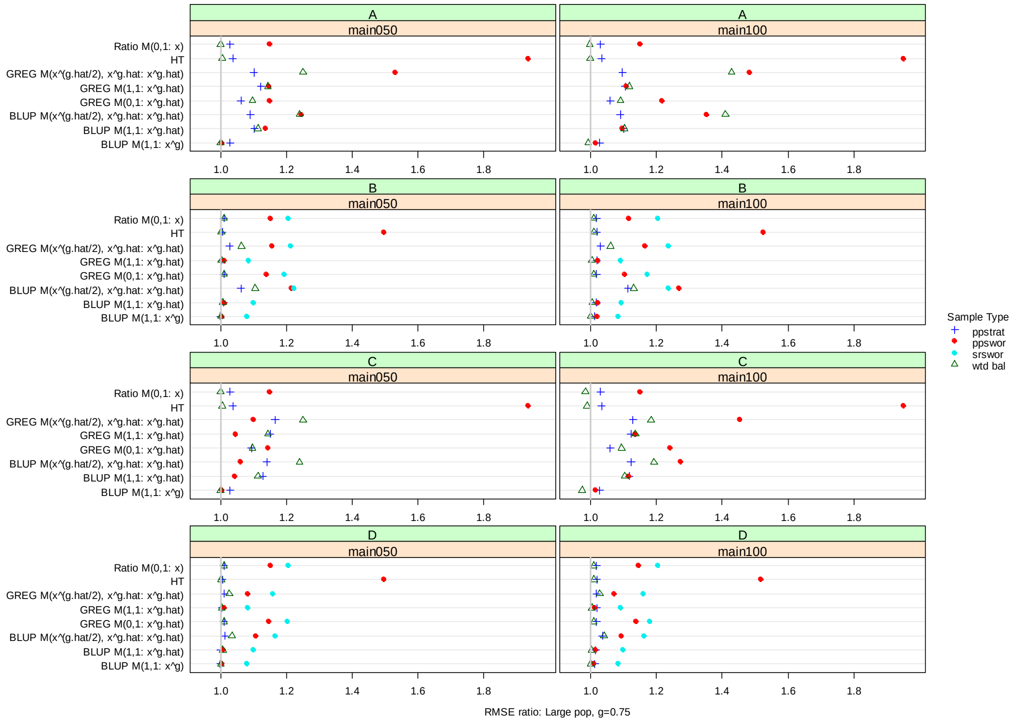


Figure 9. Root mean square error ratios for the various (Estimator, Strategy, Sample Design) combinations for the (Large, $\gamma = 3/4$) population.

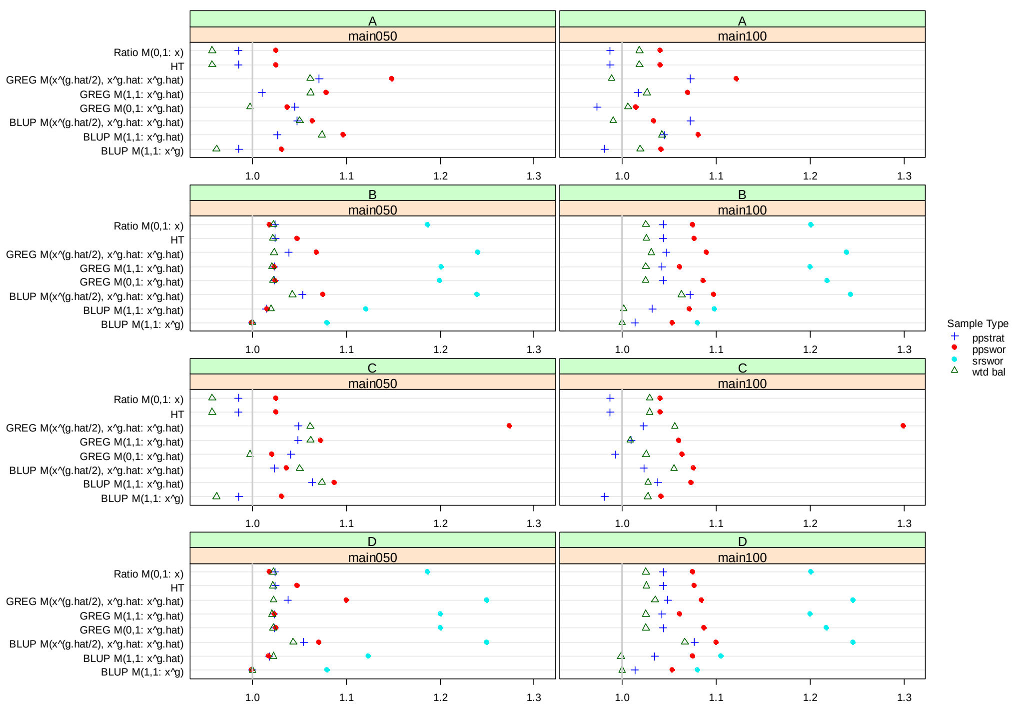


Figure 10. Root mean square error ratios for the various (Estimator, Strategy, Sample Design) combinations for the (Large, $\gamma = 2$) population.

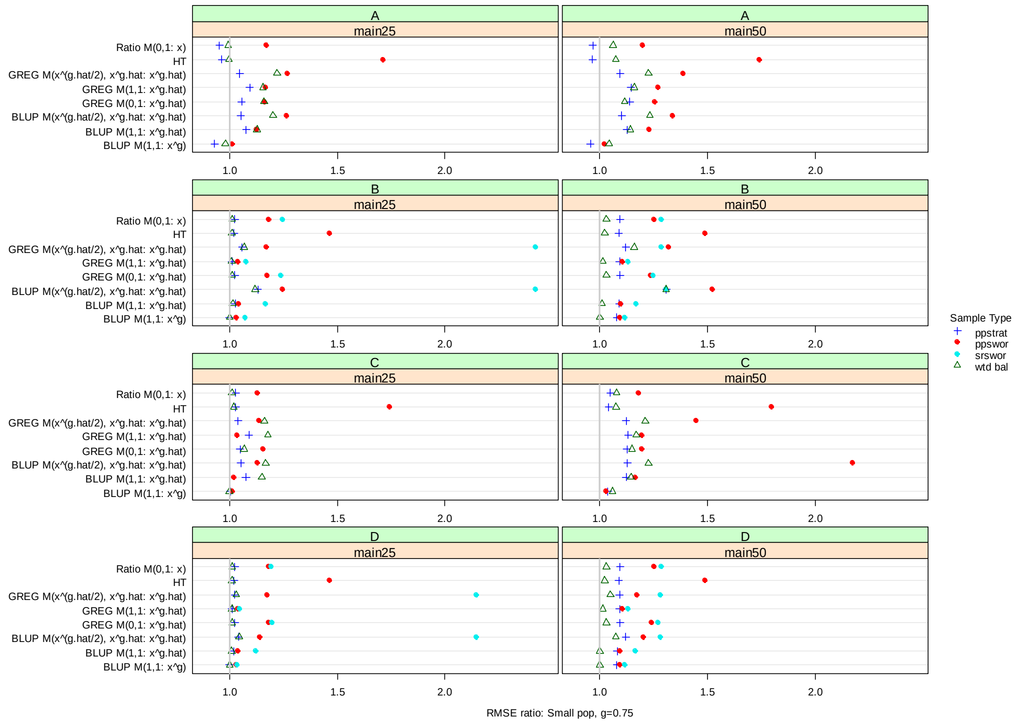


Figure 11. Root mean square error ratios for the various (Estimator, Strategy, Sample Design) combinations for the (Small, $\gamma = 3/4$) population.

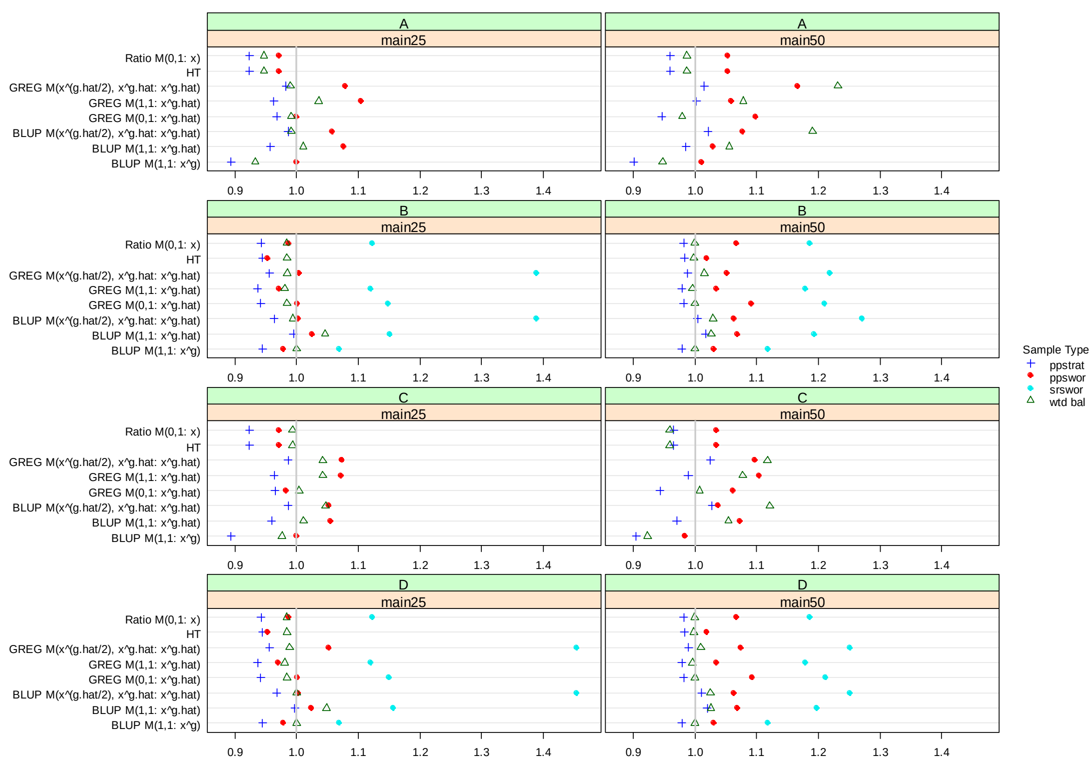


Figure 12. Root mean square error ratios for the various (Estimator, Strategy, Sample Design) combinations for the (Small, $\gamma = 2$) population.

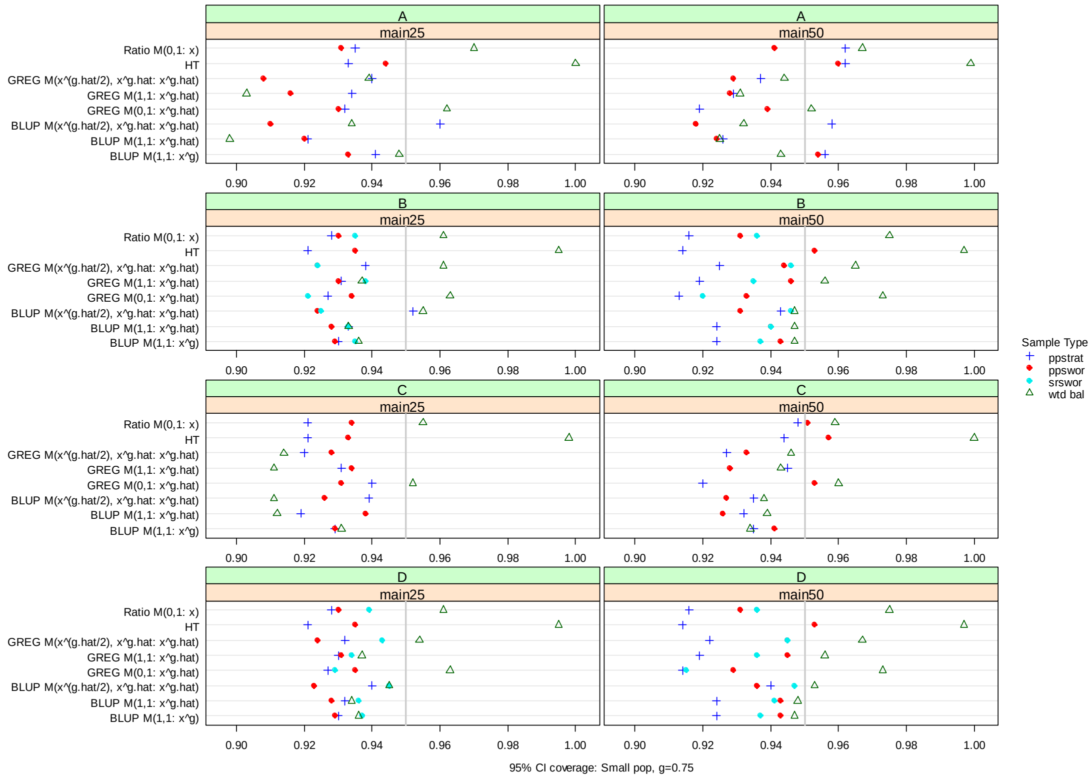


Figure 13. Empirical proportions of 95% confidence intervals that covered the population totals for the (Small, $\gamma = 3/4$) population.

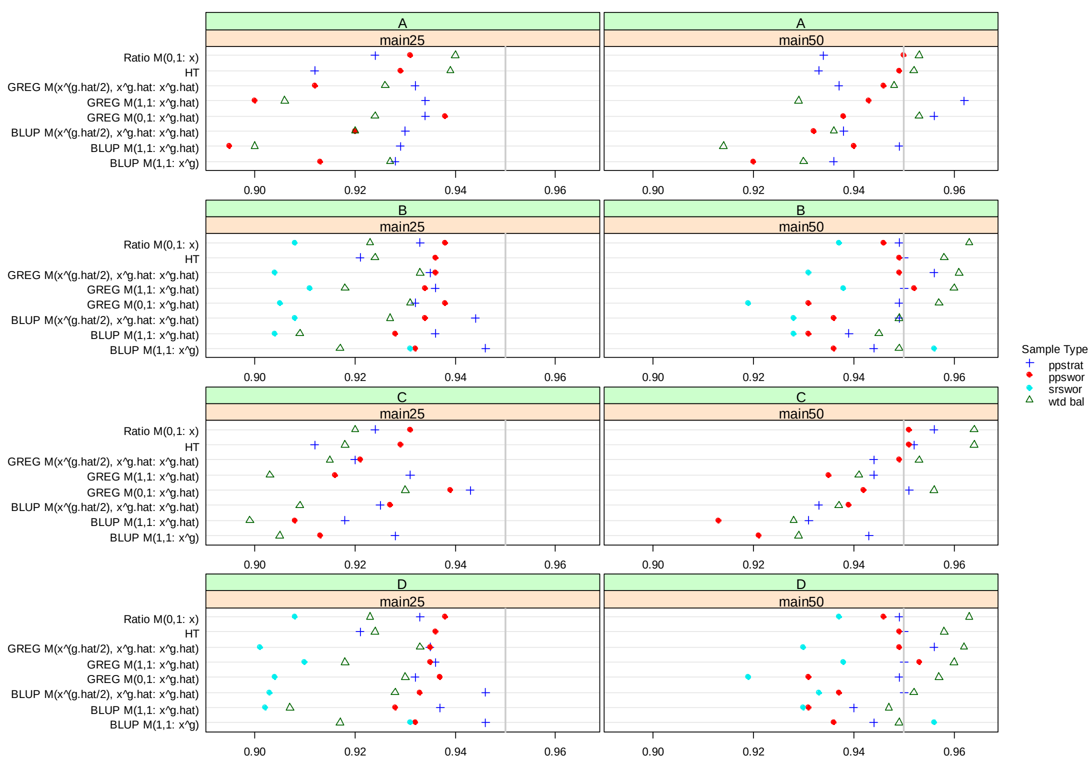


Figure 14. Empirical proportions of 95% confidence intervals that covered the population totals for the (Small, $\gamma = 2$) population.

Table 2: Coefficients of variation (CV) and ratios of CV's to smallest CV of estimated totals from Strategy A with pilot samples of 10 and main samples of 50 and Strategy B with samples of 50 from the Hospitals population.

Strategy	Sample Design	Estimator	CV (%)	Ratio to smallest
A	*wtd bal*	Ratio $\hat{T}(0,1:x)$	3.0	1.00
A	*wtd bal*	HT $\hat{T}_{\pi}$	3.1	1.01
A	*ppswor*	HT $\hat{T}_{\pi}$	3.2	1.04
B	*ppstrat*	GREG $\hat{T}_{GR}(0,1:x^{\hat{\gamma}})$; Ratio $\hat{T}(0,1:x)$; GREG $\hat{T}_{GR}(1,1:x^{\hat{\gamma}})$; GREG $\hat{T}_{GR}(x^{\hat{\gamma}/2},x^{\hat{\gamma}}:x^{\hat{\gamma}})$; BLUP $\hat{T}(1,1:x^{\gamma})$; HT $\hat{T}_{\pi}$	3.2	1.05
B	*wtd bal*	GREG $\hat{T}_{GR}(1,1:x^{\hat{\gamma}})$; Ratio $\hat{T}(0,1:x)$; GREG $\hat{T}_{GR}(0,1:x^{\hat{\gamma}})$; BLUP $\hat{T}(1,1:x^{\gamma})$;	3.2	1.05
A	*ppstrat*	Ratio $\hat{T}(0,1:x)$; HT $\hat{T}_{\pi}$;	3.2	1.06
B	*wtd bal*	GREG $\hat{T}_{GR}(x^{\hat{\gamma}/2},x^{\hat{\gamma}}:x^{\hat{\gamma}})$; HT $\hat{T}_{\pi}$;	3.2	1.06
A	*ppstrat*	GREG $\hat{T}_{GR}(x^{\hat{\gamma}/2},x^{\hat{\gamma}}:x^{\hat{\gamma}})$	3.2	1.07
A	*wtd bal*	BLUP $\hat{T}(1,1:x^{\gamma})$	3.3	1.07
A	*ppswor*	Ratio $\hat{T}(0,1:x)$	3.3	1.08
A	*ppstrat*	GREG $\hat{T}_{GR}(0,1:x^{\hat{\gamma}})$; BLUP $\hat{T}(1,1:x^{\gamma})$	3.3	1.09
B	*ppswor*	BLUP $\hat{T}(1,1:x^{\gamma})$	3.4	1.10
A	*wtd bal*	GREG $\hat{T}_{GR}(x^{\hat{\gamma}/2},x^{\hat{\gamma}}:x^{\hat{\gamma}})$; GREG $\hat{T}_{GR}(0,1:x^{\hat{\gamma}})$	3.4	1.10
B	*ppstrat*	BLUP $\hat{T}(x^{\hat{\gamma}/2},x^{\hat{\gamma}}:x^{\hat{\gamma}})$	3.4	1.11
B	*ppswor*	GREG $\hat{T}_{GR}(1,1:x^{\hat{\gamma}})$	3.4	1.12
A	*ppswor*	GREG $\hat{T}_{GR}(0,1:x^{\hat{\gamma}})$; Ratio $\hat{T}(0,1:x)$	3.4	1.12
A	*ppstrat*	GREG $\hat{T}_{GR}(1,1:x^{\hat{\gamma}})$	3.4	1.13
B	*ppswor*	GREG $\hat{T}_{GR}(x^{\hat{\gamma}/2},x^{\hat{\gamma}}:x^{\hat{\gamma}})$; GREG $\hat{T}_{GR}(0,1:x^{\hat{\gamma}})$	3.5	1.15
B	*wtd bal*	BLUP $\hat{T}(x^{\hat{\gamma}/2},x^{\hat{\gamma}}:x^{\hat{\gamma}})$	3.5	1.15
A	*ppstrat*	BLUP $\hat{T}(1,1:x^{\hat{\gamma}})$	3.5	1.15
A	*ppswor*	BLUP $\hat{T}(1,1:x^{\gamma})$	3.5	1.16
A	*wtd bal*	GREG $\hat{T}_{GR}(1,1:x^{\hat{\gamma}})$; BLUP $\hat{T}(1,1:x^{\hat{\gamma}})$	3.7	1.20
B	*ppswor*	BLUP $\hat{T}(x^{\hat{\gamma}/2},x^{\hat{\gamma}}:x^{\hat{\gamma}})$	3.7	1.22
B	*ppswor*	BLUP $\hat{T}(1,1:x^{\hat{\gamma}})$	3.8	1.25
B	*ppstrat*	BLUP $\hat{T}(1,1:x^{\hat{\gamma}})$	3.9	1.27
B	*wtd bal*	BLUP $\hat{T}(1,1:x^{\hat{\gamma}})$	3.9	1.27
A	*ppswor*	GREG $\hat{T}_{GR}(1,1:x^{\hat{\gamma}})$; BLUP $\hat{T}(1,1:x^{\hat{\gamma}})$	4.0	1.32
A	*ppswor*	GREG $\hat{T}_{GR}(x^{\hat{\gamma}/2},x^{\hat{\gamma}}:x^{\hat{\gamma}})$	4.2	1.37
B	*ppswor*	HT $\hat{T}_{\pi}$	4.7	1.53

and many combinations in (Small, $\gamma = 3/4$). This is expected due to large sample theory.

(d) Rounding of $\hat{\gamma}$ improved coverage somewhat for *ppswor* when a pilot was used (A and C). For example, see $\hat{T}_{GR}(1,1:x^{\hat{\gamma}})$ in both populations and main sample sizes. However, the improvement is only about 1 percentage point and is not significant.

(e) Rounding has little effect when no pilot is used (B and D).

(f) Coverage using the BLUP $\hat{T}(1,1:x^{\hat{\gamma}})$ with an estimated γ is less than with $\hat{T}(1,1:x^{\gamma})$. The difference in coverage is larger when a pilot is selected. The loss when only a main sample is used is minimal. For example, coverage with $\hat{T}(1,1:x^{\hat{\gamma}})$ is 90% in (Small, $\gamma = 2$, A, n=25, *wtd.bal*) but with $\hat{T}(1,1:x^{\gamma})$ is 92.7%. In (Small, $\gamma = 2$, B, n=25, *wtd.bal*) the same comparison is 90.9% versus 91.7%.

As an illustration with a real population, we compared estimated totals and sampling plans in the Hospitals population in Figure 1. This population has $N = 393$ units. The two best-fitting models for Hospitals, among several alternatives, are $M(x, x^2 : x^2)$ and $M(x^{1.6/2}, x^{1.6} : x^{1.6})$. We used Strategies A and B where $\sqrt{\text{Beds}}$ was the MOS in *ppstrat*, *ppswor*, and *wtd.bal*. For Strategy A we used pilot samples of size 10 and main samples of size 50 while for B we used main samples of size $n = 50$. For each strategy we selected 1,000 samples.

Table 2 shows the coefficients of variation (CV's) for estimated totals using the estimators in section 3.5, sorted in ascending order. The CV's are a convenient way of gauging relative precision. The "true" value of γ was taken to be 1.62, the estimate for the full population using model (1). The three best combinations do use a pilot study but their results are within simulation error of several of the Strategy B choices that use only a main sample. Thus, selecting a pilot sample is not worthwhile in this population.

Among the strategy B combinations, the *ppstrat* and *wtd.bal* plans are the most efficient. When strategy B samples are selected by *ppstrat* or *wtd.bal*, a number of estimators have CV's of about 3.2%, despite there being several different models underlying these estimators. The GREG's

$\hat{T}_{GR}(1,1 : x^{\hat{\gamma}})$, $\hat{T}_{GR}(0,1 : x^{\hat{\gamma}})$, and $\hat{T}_{GR}(x^{\hat{\gamma}/2}, x^{\hat{\gamma}} : x^{\hat{\gamma}})$ in *ppstrat* are among the most efficient. The BLUP's where γ is estimated, $\hat{T}(x^{\hat{\gamma}/2}, x^{\hat{\gamma}} : x^{\hat{\gamma}})$ and $\hat{T}(1,1 : x^{\hat{\gamma}})$, are somewhat less efficient in both *ppstrat* and *wtd.bal* samples. As expected, *srswor* is extremely inefficient and is not shown in Table 2. The worst combinations were strategy A coupled with (*ppswor,* $\hat{T}(x^{\hat{\gamma}/2}, x^{\hat{\gamma}} : x^{\hat{\gamma}})$), (*wtd.bal*, $\hat{T}(x^{\hat{\gamma}/2}, x^{\hat{\gamma}} : x^{\hat{\gamma}})$), and (*ppstrat*, $\hat{T}(x^{\hat{\gamma}/2}, x^{\hat{\gamma}} : x^{\hat{\gamma}})$). Each of these combinations had CV's larger than 15% and were omitted from the table.

5 General Conclusions and Limitations

We investigated some alternative strategies for sampling and estimation in populations where there is one target variable y, whose total is to be estimated, and one auxiliary x, which is known for every unit in the population. The variance of y is known to increase as x increases, but the exact form of the variance is unknown to the sampler. Modeling the variance as $Var_M(y_i|x_i) = \sigma^2 x_i^{\gamma}$ is assumed to be a good approximation to reality. We studied four options that might be considered for this type of problem: design of a pilot sample, design of a main sample, the method of sample selection, and choice of an estimator.

We found little evidence that a pilot study designed to get a preliminary estimate of γ would be worthwhile. For our versions of the HMT population, the smaller pilot studies often resulted in samples where there were practical problems in estimating γ. These included negative $\hat{\gamma}$'s and ones that diverged to either positive or negative infinity. In such cases, a pilot study would simply be a waste of time and resources.

In the less variable ($\gamma = 3/4$) populations we studied, conducting a pilot also did not result in lower root mean square errors for the totals than the alternative of using only a main sample with an educated guess about the size of γ. In the more variable populations ($\gamma = 2$), there were some minor gains in RMSE in some cases, e.g., when a small pilot (n = 10) was followed by a small main sample (n= 25). However, this does not consider the extra cost of conducting a pilot.

Rounding $\hat{\gamma}$ to the nearest half was not particularly helpful or harmful in estimating totals. The more serious computational issue is that legitimate estimates of γ often could not be obtained because of convergence problems. These problems occurred regardless of the working-model/estimator used and all methods of sample selection were affected. One method of combating this might be to use a robust regression technique like least median of squares (Rousseeuw 1984) to estimate γ. We did not pursue that here.

Among the sampling plans we considered, stratification based on cumulative $\sqrt{x^{\hat{\gamma}}}$ or $\sqrt{x}$ rules, denoted *ppstrat* here, were both reasonably efficient. The use of *wtd.bal* samples based on $\hat{\gamma}$'s was somewhat effective in reducing the root mean square errors of totals but not substantially more efficient than *ppstrat*, which can approximate an optimal weighted balanced sample. In addition, poor estimates of γ from a pilot will often rob *wtd.bal* of any theoretical efficiency it might have. Weighted balanced sampling is most useful when the sample size is small, tight control is needed over the sample configuration in order to be most efficient, and a reasonable advance guess about γ can be made to use in the measure of size.

A good overall strategy for this type of problem appears to be the following. Select a highly restricted probability proportional to $\sqrt{x}$ sample. This can be accomplished using the $ppstrat\left(\sqrt{x}\right)$ rule with one or two units selected per stratum. Then, estimate the total with either a BLUP or a GREG estimator based on a reasonable model for the population at hand. This general approach is similar to ones used by some accounting firms (e.g., Rotz et al. 2006) that conduct cost segregation studies.

Any simulation study is, of course, limited. We have attempted to mimic real-life applications, but populations that are less well-behaved than our HMT ones and the Hospital population may yield different results.

Appendix A: Iterative Gamma Estimate Plots

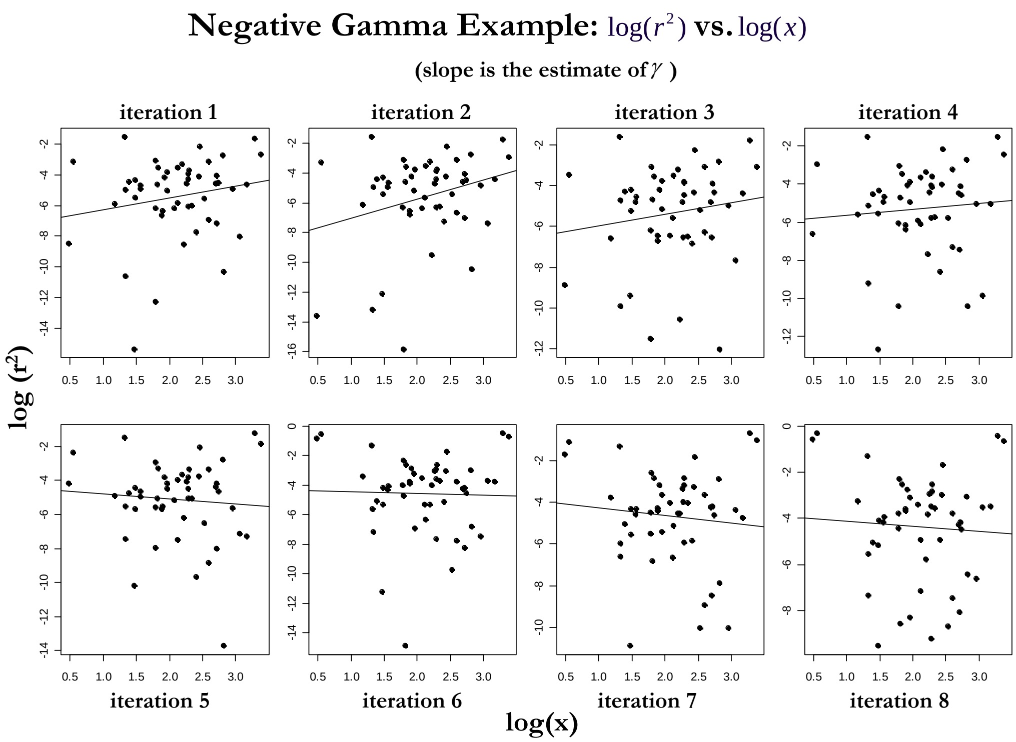


Figure A.1. First 8 Iterations of Gamma Estimation Program (Example 1)

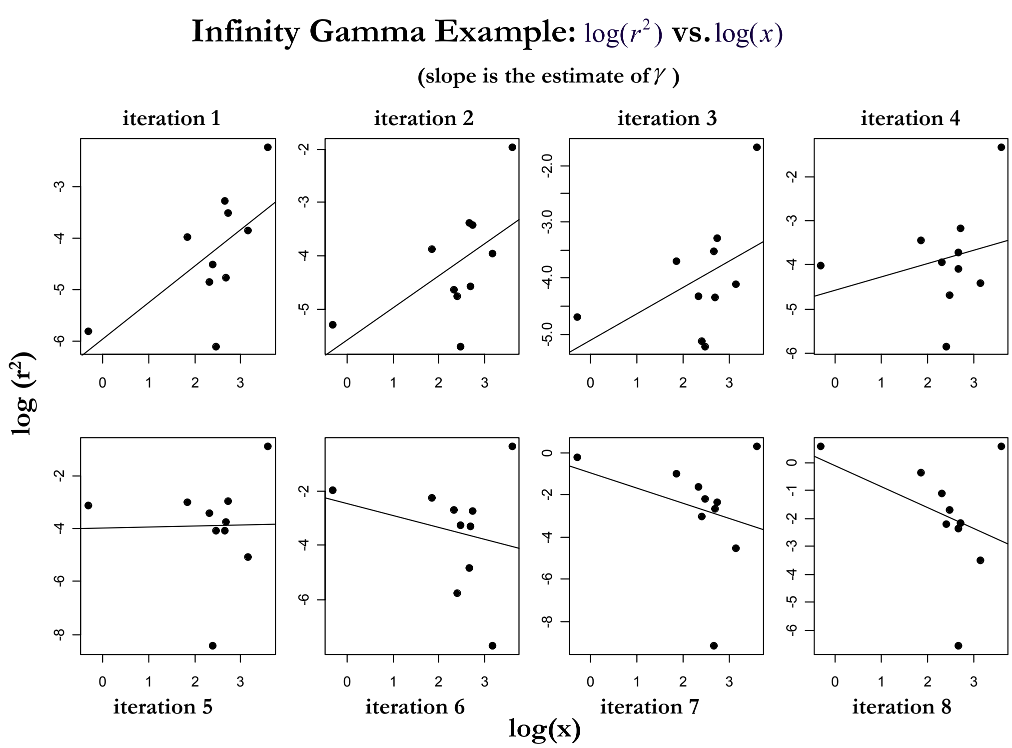


Figure A.2. First 8 Iterations of Gamma Estimation Program (Example 2)

Appendix B: Frequency of Problematic Gamma Estimates

Table A.1: Number of Times $\hat{\gamma}$ was reset to 1*, Large* $\gamma = 3/4$ *Population.*

		Working Model/Main Sample Size					
Strategy	Sample Design	$M(1,1:x^{\gamma})$		$M(x^{\gamma/2},x^{\gamma}:x^{\gamma})$		$M(0,1:x^{\gamma})$	
		n=50	n=100	n=50	n=100	n=50	n=100
A, Pilot=10	*ppswor*	285	285	303	303	345	345
	ppstrat	263	263	252	252	326	326
	wtd.bal	304	304	313	313	358	358
A, Pilot=50	*ppswor*	N/A	52	N/A	156	N/A	117
	ppstrat		68		146		123
	wtd.bal		66		179		125
B	*srswor*	30	7	134	67	129	42
	ppswor	53	19	164	86	86	40
	ppstrat	64	11	147	68	104	45
	wtd.bal	66	12	172	97	116	54
C, Pilot=10	*ppswor*	330	329	354	368	420	426
	ppstrat	327	327	331	331	402	402
	wtd.bal	342	342	381	381	412	412
C, Pilot=50	*ppswor*	N/A	136	N/A	345	N/A	253
	ppstrat		164		235		263
	wtd.bal		149		269		277
D	*srswor*	110	40	208	137	312	207
	ppswor	142	64	233	162	226	173
	ppstrat	143	53	239	160	242	164
	wtd.bal	171	57	261	182	256	144

Table A.2: Number of Times $\hat{\gamma}$ was reset to 3*, Large* $\gamma = 2$ *Population.*

		Working Model/Main Sample Size					
Strategy	Sample Design	$M(1,1:x^{\gamma})$		$M(x^{\gamma/2},x^{\gamma}:x^{\gamma})$		$M(0,1:x^{\gamma})$	
		n=50	n=100	n=50	n=100	n=50	n=100
A, Pilot=10	*ppswor*	359	359	158	158	211	211
	ppstrat	379	366	183	177	234	206
	wtd.bal	390	376	170	160	224	201
A, Pilot=50	*ppswor*	N/A	55	N/A	18	N/A	22
	ppstrat		54		16		23
	wtd.bal		66		24		27
B	*srswor*	43	5	5	1	15	4
	ppswor	53	7	24	3	43	4
	ppstrat	77	10	16	1	25	3
	wtd.bal	51	5	22	3	17	5
C, Pilot=10	*ppswor*	324	324	126	126	165	165
	ppstrat	339	326	143	145	196	156
	wtd.bal	390	308	170	127	224	158
C, Pilot=50	*ppswor*	N/A	30	N/A	10	N/A	6
	ppstrat		26		6		11
	wtd.bal		33		10		13
D	*srswor*	23	1	3	0	3	0
	ppswor	24	2	10	1	13	0
	ppstrat	39	1	6	0	6	0
	wtd.bal	27	2	6	0	6	2

References

Allen, W. E., & Foster, M. B. (2005). *Cost segregation applied.* Journal of Accountancy. http://www.aicpa.org/pubs/jofa/aug2005/allen.htm.

Batcher, M., & Liu, Y. (2002). *Ratio estimation of small samples using deep stratification.* American Statistical Association. Proceedings of the Business and Economic Statistics Section.

Brewer, K. (2002). *Combined Survey Sampling Inference: Weighing Basu's Elephants*. London: Arnold Publishing.

Bryant, E. C., Hartley, H. O., & Jessen, R. J. (1960). Design and Estimation in Two-Way Stratification. *Journal of the American Statistical Association*, *55*, 105-124.

Census Bureau. (2007). *County Business Patterns.* www.census.gov/epcd/cbp/view/cbpview.html.

Chauvet, G., & Tillé, Y. (2006). A fast algorithm of balanced sampling. *Computational Statistics*, *21*, 53-62.

Cochran, W. (1977). *Sampling Techniques* (3rd ed.). New York: John Wiley & Sons.

Foreman, E. K. (1991). *Survey Sampling Principles*. New York: Marcel Dekker, Inc.

Gentle, J. E. (2002). *Elements of Computational Statistics*. New York: Springer Science + Business Media.

Godfrey, J., Roshwalb, A., & Wright, R. (1984). Model-based stratification in inventory cost estimation. *Journal of Business & Economic Statistics*, *2*, 1-9.

Hansen, M. H., Madow, W. G., & Tepping, B. J. (1983). An evaluation of model-dependent and probability-sampling inferences in sample surveys. *Journal of the American Statistical Association*, *78*, 776-793.

Hartley, H. O., & Rao, J. N. K. (1962). Sampling with Unequal Probabilities and without Replacement. *Annals of Mathematical Statistics*, *33*, 350-374.

Hurvich, C. M., & Tsai, C.-L. (1990). The impact of model selection on inference in linear regression. *The American Statistician*, *44*, 214-217.

National Center for Education Statistics. (2007). *Common core of data.* http://nces.ed.gov/ccd/.

Roshwalb, A. (1987). *The estimation of a heteroscedastic linear model using inventory data.* American Statistical Association. Proceedings of the Business and Economic Statistics Section.

Rotz, W., Joshee, A., & Yang, J. (2006). *Confidence Interval Coverage in Model based Estimation.* American Statistical Association. Proceedings of the Survey Research Methods Section.

Rousseeuw, P. (1984). Least median of squares regression. *Journal of the American Statistical Association*, *79*, 871-880.

Särndal, Swensson, & Wretman. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Searle, S. R. (1971). *Linear Models*. New York: John Wiley.

Sitter, R. R., & Skinner, C. J. (1994). Multi-Way Stratification by Linear Programming. *Survey Methodology*, *20*, 65-73.

Strobel, C. (2002). New rapid write-off provisions for tangible and intangible assets. *Journal of Corporate Accounting and Finance*, *13*, 99-101.

Substance Abuse and Mental Health Services Administration. (2007). *National Expenditures for Mental Health Services and Substance Abuse Treatment 1991-2001.* http://www.samhsa.gov/spendingestimates/chapter1.aspx.

Tillé, Y. (2006). *Sampling Algorithms*. New York: Springer.

Valliant, R. (2002). Variance estimation for the general regression estimator. *Survey Methodology*, *28*, 103-114.

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

Zhang, P. (1992). Inference after variable selection in linear regression models. *Biometrika*, *79*, 741-746.