Journal of Official Statistics, Vol. 16, No. 2, 2000, pp. 139–154

# Stratification by Size Revisited

*Alan H. Dorfman*[1] *and Richard Valliant*[2]

Stratification by size is used in finite population sampling as a means of producing efficient estimators. The technique has also been recognized as a method of approximating optimum selection probabilities for a variety of estimators. Using prediction theory, we show that an unstratified, weighted balanced sample yields the same variance as stratification by size with optimum allocation of a stratified, weighted balanced sample when using the best linear unbiased predictor of the population total. Stratification by size can, thus, be viewed as nothing more than a way of selecting a sample with overall weighted balance. A practical method of selecting weighted balanced samples is to use restricted randomization in which poorly balanced samples are rejected. We illustrate by simulation the superiority of weighted balanced sampling in three real populations.

*Key words:* Balanced sample; best linear unbiased predictor; minimal model; restricted random sampling; robustness; superpopulation model.

## 1. Introduction

Stratification is one of the most widely used techniques in finite population sampling. Strata are disjoint subdivisions of a population, the union of which exhausts the universe, each of which contains a portion of the sample. Its essential statistical purposes are to:

(1) allow for efficient estimation, especially in the case of stratification by size, and
(2) deal statistically with subpopulations or domains by controlling their sample allocations.

Stratification by size is typically considered as serving purpose (1) by creating strata in an efficient way and optimally allocating the sample to the strata. Dalenius and Hodges (1959) suggested a way of constructing strata based on an hypothesized density function for the variable of interest. Wright (1983) and Godfrey, Roshwalb, and Wright (1984) studied stratification by size as a way of approximating optimum selection probabilities, thereby achieving the greatest lower bound on the variance of certain estimators, including the Horvitz-Thompson, regression, difference, and ratio estimators. Bethel (1989) studied the relative efficiency of the Horvitz-Thompson estimator under a variety of methods of stratification. A synopsis of these efforts can be found in Särndal, Swensson, and Wretman (1992, Sections 12.2–12.6).

[1] U.S. Bureau of Labor Statistics, Room 4915, 2 Massachusetts Ave., NE, Washington, DC 20212, U.S.A. E-mail: dorfman_a@bls.gov
[2] Westat Inc., 1650 Research Boulevard, Rockville, MD 20850-3129, U.S.A. E-mail: valliar1@westat.com
**Acknowledgment:** Any opinions expressed by the authors are their own and do not constitute policy of the U.S. Bureau of Labor Statistics. The authors thank the Associate Editor for his useful comments.

Using model-based analysis, we show that, in the situation where stratification by size is generally used, best linear unbiased (*BLU*) prediction coupled with a certain kind of weighted balanced sampling – which requires *no* stratification – achieves minimal variance. In other words, stratification by size has no theoretical advantage over the optimal unstratified procedure. This in effect makes moot earlier concerns about how best to stratify. Our results also supersede those on model-based stratification due to Royall and Herson (1973). The theoretical findings are illustrated with simulations using real populations. There are, of course, often administrative advantages to using strata for control of survey costs and workloads, but we do not address those concerns here.

A key tool in our investigation is a result of Royall (1992, *Theorem* 2) that we restate here as *Theorem* 1 below. Consider a finite population with $N$ elements, each element $i$ having associated with it a vector $\mathbf{x}_i$ of $p$ auxiliaries. The $N \times p$ matrix of those auxiliaries is $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)'$. A sample $s$ of size $n$ is selected and the realization of a target variable $Y_i$ is observed for each sample unit along with its $\mathbf{x}_i$. The matrix of $x$'s for the sample units is $\mathbf{X}_s$. Suppose that $\mathbf{Y} = (Y_1, \ldots, Y_N)'$ follows the linear model

$$E_M(\mathbf{Y}) = \mathbf{X}\beta, \quad \text{var}_M(\mathbf{Y}) = \mathbf{V}\sigma^2 \tag{1}$$

where $\beta$ is an unknown $p \times 1$ parameter, and $\mathbf{V} = diag(v_i)$ is a known $N \times N$ diagonal covariance matrix. Denote this model as $M(\mathbf{X} : \mathbf{V})$. The vector of $Y$'s for sample units is $Y_s$ and their $n \times n$ diagonal covariance matrix is $\mathbf{V}_{ss}$.

Let $\mathbf{1}_s$ and $\mathbf{1}_N$ be vectors of $n$ and $N$ 1's, respectively. The collection of samples that satisfy

$$\frac{1}{n}\mathbf{1}_s'\mathbf{W}_s^{-1/2}\mathbf{X}_s = \frac{\mathbf{1}_N'\mathbf{X}}{\mathbf{1}_N'\mathbf{W}^{1/2}\mathbf{1}_N} \tag{2}$$

is denoted $B(\mathbf{X} : \mathbf{W})$ and such samples are said to be *balanced with respect to the weights root*$(\mathbf{W})$ or to be *root*$(\mathbf{W})$ *balanced*. Here $\mathbf{W}$ is an $N \times N$ diagonal matrix and $\mathbf{W}_s$ is the $n \times n$ submatrix for the sample units. This form of balance turns out to be appropriate when the variance matrix of the model is given by $\mathbf{V} = \mathbf{W}\sigma^2$. In *Theorem* 1, the notation $\mathcal{M}(\mathbf{X})$ refers to the vector space generated by the columns of $\mathbf{X}$, and $\hat{T}(\mathbf{X} : \mathbf{V})$ is the best linear unbiased (*BLU*) predictor under model (1).

***Theorem* 1** (Royall 1992). Under $M(\mathbf{X} : \mathbf{V})$ if both $\mathbf{V}\mathbf{1}_N$ and $\mathbf{V}^{1/2}\mathbf{1}_N \in \mathcal{M}(\mathbf{X})$, then

$$\begin{aligned}\text{var}_M[\hat{T}(\mathbf{X} : \mathbf{V}) - T] &\geq [n^{-1}(\mathbf{1}_N'\mathbf{V}^{1/2}\mathbf{1}_N)^2 - \mathbf{1}_N'\mathbf{V}\mathbf{1}_N]\sigma^2 \\ &= \sigma^2[n^{-1}(N\bar{v}^{(1/2)})^2 - N\bar{v}] \end{aligned} \tag{3a}$$

where $\bar{v}^{(1/2)} = \Sigma_{i=1}^N v_i^{1/2}/N$ and $\bar{v} = \Sigma_{i=1}^N v_i/N$. The bound is achieved if and only if $s \in B(\mathbf{X} : \mathbf{V})$, in which case

$$\begin{aligned}\hat{T}(\mathbf{X} : \mathbf{V}) &= n^{-1}(\mathbf{1}_N'\mathbf{V}^{1/2}\mathbf{1}_N)(\mathbf{1}_s'\mathbf{V}_{ss}^{-1/2}\mathbf{Y}_s) \\ &= \frac{1}{n}N\bar{v}^{(1/2)}\Sigma_s Y_i/v_i^{1/2} \end{aligned} \tag{3b}$$

The Estimator (3b) to which the *BLU* predictor reduces in a $B(\mathbf{X} : \mathbf{V})$ sample is a mean of ratios estimator conveniently denoted $\hat{T}_{MR}(v^{1/2})$. It equals the Horvitz-Thompson

estimator $\hat{T}_\pi = \Sigma_s y_i/\pi_i$ when the selection probability of unit $i$ is proportional to the model standard deviation $v_i^{1/2}$. The lower bound on the error variance of the *BLU* predictor does not depend on the particular sample that is selected. This means that *any* sampling strategy that achieves (3a) is an optimal strategy under a model that satisfies the conditions of Theorem 1.

There are two main points to note. The first regards *bias-robustness*: the estimator $\hat{T}(\mathbf{X} : \mathbf{V})$ will be unbiased for $T$ under a different model $M(\mathbf{X}^* : \mathbf{V}^*)$ if $s \in B(\mathbf{X}, \mathbf{X}^* : \mathbf{V})$, that is, if the sample is balanced on all columns of both $\mathbf{X}$ and $\mathbf{X}^*$. To see this, note that if the model is $M(\mathbf{X} : \mathbf{V})$, then $\hat{T}(\mathbf{X} : \mathbf{V})$, as defined by (3b), has a bias equal to

$$E(\hat{T} - T) = [n^{-1}(\mathbf{1}_N'\mathbf{V}^{1/2}\mathbf{1}_N)(\mathbf{1}_s'\mathbf{V}_{ss}^{-1/2}\mathbf{X}_s) - \mathbf{1}_N'\mathbf{X}]\beta$$

The bias under the model $M(\mathbf{X}^* : \mathbf{V}^*)$ is equal to the expression above with $\mathbf{X}_s$ replaced by $\mathbf{X}_s^*$ and $\mathbf{X}$ replaced by $\mathbf{X}^*$ (and different $\beta$). From (2) the balance condition, $B(\mathbf{X}, \mathbf{X}^* : \mathbf{V})$, is

$$n^{-1}\mathbf{1}_s'\mathbf{V}_{ss}^{-1/2}[\mathbf{X}_s \quad \mathbf{X}_s^*] = \frac{\mathbf{1}_N'[\mathbf{X} \quad \mathbf{X}^*]}{\mathbf{1}_N'\mathbf{V}^{1/2}\mathbf{1}_N}$$

and this is exactly what is needed to eliminate the biases under both $M(\mathbf{X} : \mathbf{V})$ and $M(\mathbf{X}^* : \mathbf{V}^*)$. Given this sort of balance, misspecification of the model does not lead to inefficiency under the working model nor bias under a broader model.

The second point, noted by Royall (1992), is that the lower bound on the error variance in Theorem 1 is the same as that derived by Godambe and Joshi (1965) for the model-based expectation of the random sampling variance of $\hat{T}_\pi$. Wright (1983) extended their result to the large sample, anticipated variance of the general regression estimator.

It is convenient to define the *Minimal Model* $M(v^{1/2}, v : v)$ having $E_M(Y_i) = \beta_{1/2}v_i^{1/2} + \beta_1 v_i$, which for a given variance matrix $\mathbf{V} = diag(v_i)$, has the simplest specification of $E_M(\mathbf{Y})$ that satisfies the conditions of Theorem 1. The estimator constructed in accord with this (typically oversimplified) model will be the *BLU* estimator under a broader model so long as the appropriate weighted balance is maintained, as indicated above.

Using the variance specification in a model to dictate the form of $E_M(\mathbf{Y})$ may seem to be ''putting the cart before the horse'' to many readers. Normally, the working model is our best assessment of the relationship of the survey variable to the auxiliary variables. In certain cases where the variance is related to the mean, Theorem 1 allows us to reverse the usual order of thinking. For example, if the variance of $Y$ is proportional to a single auxiliary $x$, the minimal model for $E_M(\mathbf{Y})$ would contain terms for $x^{1/2}$ and $x$. Suppose that a reasonable working model contains, in addition, an intercept, a quadratic term in $x$, and another variable $z$, i.e., $E_M(\mathbf{Y}) = \beta_0 + \beta_{1/2}x^{1/2} + \beta_1 x + \beta_2 x^2 + \gamma z$. By Theorem 1, the smallest error variance of the *BLU* predictor under the minimal model, $E_M(Y) = \beta_{1/2}x^{1/2} + \beta_1 x$, is achieved by selecting a sample with weighted balance on $x^{1/2}$ and $x$. If we further balance on the intercept, $x^2$, and $z$, we protect ourselves against model bias but lose no efficiency at all under the working model. In the rest of this article, we study whether this is a useful approach and how it can be applied in populations that are usually stratified by size.

## 2. A Stratified Linear Model and Weighted Balanced Samples

Consider now a population divided into $H$ strata. Let $h$ denote a stratum and $i$ a unit within the stratum. The target variable for unit $hi$ is $Y_{hi}$. Assume that the population contains $H$ strata with the number of units in each stratum $N_h (h = 1, \ldots, H)$ and the population size $N = \Sigma_{h=1}^{H} N_h$. A sample of $n_h$ units is selected from stratum $h$ with the total sample size being $n = \Sigma_h n_h$. Denote the set of sample units in stratum $h$ as $s_h$ and the set of non-sample units as $r_h$. Assume that a separate linear regression model holds within each stratum:

$$E_M(\mathbf{Y}_h) = \mathbf{X}_h \beta_h, \text{var}_M(\mathbf{Y}_h) = \mathbf{V}_h \sigma_h^2 \tag{4}$$

where $\mathbf{Y}_h$ is $N_h \times 1$, $\mathbf{X}_h$ is $N_h \times p_h$, $\mathbf{V}_h = diag(v_{hi})$ is $N_h \times N_h$, and $\beta_h$ is a $p_h \times 1$ parameter vector. The model in stratum $h$ is $M(\mathbf{X}_h : \mathbf{V}_h)$ and the *BLU* predictor is then the sum of the *BLU* predictors in each stratum:

$$\hat{T} = \sum_{h=1}^{H} \hat{T}(\mathbf{X}_h : \mathbf{V}_h)$$

In stratum $h$ define a *root*($v$)-balanced sample to be one that satisfies

$$\frac{1}{n_h} \mathbf{1}'_{sh} \mathbf{V}_{sh}^{-1/2} \mathbf{X}_{sh} = \frac{\mathbf{1}'_{Nh} \mathbf{X}_h}{\mathbf{1}'_{Nh} \mathbf{V}_h^{1/2} \mathbf{1}_{Nh}} \tag{5}$$

where $\mathbf{1}_{sh}$ is a vector of $n_h$ 1's, $\mathbf{1}_{Nh}$ is a vector of $N_h$ 1's, $\mathbf{V}_{sh}$ is the $n_h \times n_h$ diagonal covariance matrix for the sample units, and $\mathbf{X}_{sh}$ is the $n_h \times p_h$ matrix of auxiliaries for the sample units. Any stratum sample satisfying (5) will be denoted by $B(\mathbf{X}_h : \mathbf{V}_h)$, and, when (5) is satisfied in each stratum, the entire sample is a *stratified weighted balanced sample*. A straightforward application of *Theorem* 1 yields the following result.

***Theorem* 2**. Suppose that Model (4) holds in stratum $h$ for $h = 1, \ldots, H$. If both $\mathbf{V}_h \mathbf{1}_{Nh}$ and $\mathbf{V}_h^{1/2} \mathbf{1}_{Nh} \in \mathcal{M}(\mathbf{X}_h)$, then the *BLU* predictor achieves its minimum variance when each stratum sample is $B(\mathbf{X}_h : \mathbf{V}_h)$. In that case, the *BLU* predictor reduces to

$$\hat{T} = \sum_{h=1}^{H} N_h \bar{v}_h^{(1/2)} \frac{1}{n_h} \sum_{i \in s_h} \frac{Y_{hi}}{v_{hi}^{1/2}} \tag{6a}$$

and the error variance is

$$\text{var}_M(\hat{T} - T) = \Sigma_h \left[ \frac{1}{n_h} (N_h \bar{v}_h^{(1/2)})^2 - N_h \bar{v}_h \right] \sigma_h^2 \tag{6b}$$

where $\bar{v}_h^{(1/2)} = \Sigma_{i=1}^{N_h} v_{hi}^{1/2} / N_h$ and $\bar{v}_h = \Sigma_{i=1}^{N_h} v_{hi} / N_h$

In a stratified weighted balanced sample, the optimal estimator thus reduces to a sum of mean-of-ratios estimators, which, for later reference, we will write as $\hat{T}_{MRS}(v^{1/2})$. Thus, as in the unstratified case, a weighted balanced sample is the best that can be selected in the sense of making the error variance of the *BLU* predictor small. And again, to guard against bias under more elaborate models, we can balance on additional $x$ factors that are not in the working model, without introducing any bias or losing any precision under the working model.

## 3. Optimal Allocation for Stratified Balanced Sampling

The optimum allocation to the strata of a weighted balanced sample can be easily calculated using standard methods. Assume that the cost of sampling is $C = C_0 + \Sigma_h c_h n_h$ where $C_0$ is a fixed cost and $c_h$ is the cost per unit sampled in stratum $h$.

***Theorem* 3**. Assume that Model (4) holds, that $\mathbf{V}_h\mathbf{1}_{Nh}$ and $\mathbf{V}_h^{1/2}\mathbf{1}_{Nh} \in \mathcal{M}(\mathbf{X}_h)$, and that a weighted balanced sample $B(\mathbf{X}_h : \mathbf{V}_h)$ is selected in each stratum. The allocation of the sample to the strata that minimizes the error variance of the *BLU* predictor, subject to the cost constraint $C = C_0 + \Sigma_h c_h n_h$, is

$$\frac{n_h}{n} = \frac{N_h \bar{v}_h^{(1/2)} \sigma_h / \sqrt{c_h}}{\Sigma_{h'} N_{h'} \bar{v}_{h'}^{(1/2)} \sigma_{h'} / \sqrt{c_{h'}}} \quad \text{for} \quad h = 1, \ldots, H \tag{7}$$

When optimal allocation is used and all costs are equal, the *BLU* predictor (6a) becomes

$$\hat{T} = \frac{1}{n}(\Sigma_h N_h \bar{v}_h^{(1/2)} \sigma_h)\Sigma_h \Sigma_{s_h} \frac{Y_{hi}}{v_{hi}^{1/2}\sigma_h} \tag{8a}$$

and its error variance (6b) can be rewritten as

$$\mathrm{var}_M(\hat{T} - T) = \frac{1}{n}(\Sigma_h N_h \bar{v}_h^{(1/2)} \sigma_h)^2 - \Sigma_h N_h \bar{v}_h \sigma_h^2 \tag{8b}$$

## 4. The Case of a Single Model for the Population

For the purpose of selecting strata, typically a single model is assumed to fit the whole population. Suppose the model in each stratum is

$$E_M(\mathbf{Y}_h) = \mathbf{X}_h\beta, \quad \mathrm{var}_M(\mathbf{Y}_h) = \mathbf{V}_h\sigma^2 \tag{9}$$

Expression (9) is just another way of writing the model $M(\mathbf{X} : \mathbf{V})$. Thus, strata can be ignored in calculating the *BLU* predictor and its error variance. If $\mathbf{V}\mathbf{1}_N$ and $\mathbf{V}^{1/2}\mathbf{1}_N \in \mathcal{M}(\mathbf{X})$, then by Theorem 1, a weighted balanced sample $s \in B(\mathbf{X} : \mathbf{V})$ is optimal for the *BLU* predictor, and the *BLU* predictor and its error variance reduce to (3a) and (3b) above.

In the eventuality of differential costs, an overall balanced sample for a given value of $n$ may violate the cost constraint even though it is optimal in the sense of Theorem 1. A stratified allocation is then needed to account for costs even when a single model fits the entire population.

On the other hand, suppose we select a stratified weighted balanced sample and use the optimal allocation given in Theorem 3 for the equal cost case. Using (8a) with $\sigma_h = \sigma$, the *BLU* predictor with the optimal allocation is

$$\hat{T} = \frac{1}{n}(\Sigma_h N_h \bar{v}_h^{(1/2)})\Sigma_h \Sigma_{s_h} Y_{hi}/v_{hi}^{1/2} = \frac{1}{n}(N\bar{v}^{(1/2)})\Sigma_h \Sigma_{s_h} Y_{hi}/v_{hi}^{1/2}$$

which equals (3b). In other words, stratification with optimal allocation of a stratified weighted balanced sample here gains nothing at all compared to the strategy of selecting an unstratified sample with overall weighted balance.

In fact, for the equal cost, equal $\sigma_h$ case, an optimally allocated, stratified weighted balanced sample has overall weighted balance. Let $\mathbf{x}_{hi}$ be the $i^{\text{th}}$ row of $\mathbf{X}_h$, and define the vectors of population means $\bar{\mathbf{x}}_h = \Sigma_{i=1}^{N_h} \mathbf{x}_{hi}/N_h$ and $\bar{\mathbf{x}} = \Sigma_{h=1}^{H} \sum_{i=1}^{N_h} \mathbf{x}_{hi}/N$. Multiply both sides of (5) by $n_h n^{-1}$ and on the right hand side substitute (7) with $\sigma_h = \sigma$ and $c_h = c$; sum over $h$ to get

$$\frac{1}{n}\Sigma_h \Sigma_{s_h} \frac{\mathbf{x}_{hi}}{v_{hi}^{1/2}} = \frac{\Sigma_h N_h \bar{x}_h}{\Sigma_h N_h \bar{v}_h^{(1/2)}} = \frac{\bar{\mathbf{x}}}{\bar{v}^{(1/2)}}$$

which is equivalent to (2) with $\mathbf{W} = \mathbf{V}$.

In particular, a common model is often assumed to hold for the whole population when a single auxiliary variable $x$ is available, and the auxiliary is then used for stratification by size as well as for estimation. Strata are formed by ordering the units from low to high based on $x$ and partitioning the population into $H$ groups. A variety of methods of partitioning have been proposed, which are recapitulated in Section 6 below. An important special case is given by the polynomial model

$$E_M(Y_i) = \delta_0 + \delta_1 x_i + \ldots + \delta_J x_i^J, \quad \text{var}_M(Y_{hi}) = \sigma^2 x_{hi}^{\gamma} \tag{10}$$

where $\delta_i = 1$ if the $j^{\text{th}}$ order term is in the model and 0 if not. This model is denoted by $M(\delta_0, \ldots, \delta_J : v)$ and the corresponding *BLU* predictor by $\hat{T}(\delta_0, \ldots, \delta_J : v)$. In many populations, $0 \leq \gamma \leq 2$ (see, e.g., Brewer 1963 and Scott, Brewer, and Ho 1978).

When the variance is as specified in (10), the *minimal model* (see Section 1 above) has $E_M(Y_{hi}) = \beta_{\gamma/2} x_{hi}^{\gamma/2} + \beta_{\gamma} x_{hi}^{\gamma}$. With the variance specification in (10), the optimum allocation in Theorem 3 becomes

$$\frac{n_h}{n} = \frac{N_h \bar{x}_h^{(\gamma/2)} / \sqrt{c_h}}{\Sigma_{h'} N_{h'} \bar{x}_{h'}^{(\gamma/2)} / \sqrt{c_{h'}}} \tag{11}$$

The error variance of the *BLU* predictor in a stratified weighted balanced sample is

$$\text{var}_M(\hat{T} - T) = \sigma^2 \Sigma_h \left[ \frac{1}{n_h} (N_h \bar{x}_h^{(\gamma/2)})^2 - N_h \bar{x}_h^{(\gamma)} \right] \tag{12}$$

which, when the optimal allocation (11) is used and costs are all equal, reduces to the variance for an unstratified, weighted balanced sample:

$$\text{var}_M(\hat{T} - T) = \sigma^2 \left[ \frac{1}{n} (N \bar{x}^{(\gamma/2)})^2 - N \bar{x}^{(\gamma)} \right] \tag{13}$$

as it must, by the general result described at the beginning of this section.

## 5. Comparison with Other Model-based Strategies

Two other strategies for protecting estimators against bias under polynomial models (10) are due to Royall and Herson (1973) and Scott, Brewer, and Ho (1978). We compare those approaches to weighted balanced sampling in this section.

### *5.1. Royall-Herson strategy*

We consider further the polynomial model $M(\delta_0, \ldots, \delta_J : v)$, given by (10). When strata are formed on the basis of a size measure $x$, the separate ratio estimator is frequently recommended, namely

$$\hat{T}_{RS} = \sum_{h=1}^{H} N_h \bar{Y}_{hs} \frac{\bar{x}_h}{\bar{x}_{hs}}$$

where $\bar{x}_h = \Sigma_{i=1}^{N_h} x_{hi}/N_h$, $\bar{Y}_{hs} = \Sigma_{i \in s_h} Y_{hi}/n_h$, and $\bar{x}_{hs} = \Sigma_{i \in s_h} x_{hi}/n_h$. Under the working model $M(0, 1 : x)$, $\hat{T}_{RS}$ is unbiased with variance equal to

$$\mathrm{var}_M(\hat{T}_{RS} - T) = \sigma^2 \sum_h \frac{N_h^2}{n_h}(1 - f_h)\frac{\bar{x}_{hr}\bar{x}_h}{\bar{x}_{hs}}$$

where $f_h = n_h/N_h$ and $\bar{x}_{hr} = \Sigma_{i \notin s_h} x_{hi}/(N_h - n_h)$. If one were completely confident that $M(0, 1 : x)$ is correct, then the optimal sample for $\hat{T}_{RS}$ would be to pick the $n_h$ units with the largest $x$'s in each stratum. Even more extreme is the globally optimal strategy that uses the simple ratio estimator, $\hat{T}(0, 1 : x)$, and the $n$ largest units in the population.

Confidence in any single model is seldom this high and having protection against model failure is usually prudent. If the true model is $M(\delta_0, \ldots, \delta_J; x)$, then the estimator has a bias:

$$E_M(\hat{T}_{RS} - T) = \sum_h N_h \bar{x}_h \sum_{j=0}^{J} \delta_j \beta_j \left[\frac{\bar{x}_{hs}^{(j)}}{\bar{x}_{hs}} - \frac{\bar{x}_h^{(j)}}{\bar{x}_h}\right]$$

where $\bar{x}_h^{(j)} = \Sigma_{i=1}^{N_h} x_{hi}^j/N_h$ and $\bar{x}_{hs}^{(j)} = \Sigma_{i \in s_h} x_{hi}^j/n_h$. If a *stratified* (*unweighted*) *balanced sample*, i.e., one that is balanced in each stratum ($\bar{x}_{hs}^{(j)} = \bar{x}_h^{(j)}$ for $j = 1, \ldots, J$), is selected, then $\hat{T}_{RS}$ reduces to the stratified expansion estimator

$$\hat{T}_{0S} = \sum_{h=1}^{H} N_h \bar{Y}_{hs}$$

Denote a stratified (unweighted) balanced sample by $s^*(J)$ and a simple (unstratified) balanced sample of order $J$ by $s(J)$. Royall and Herson (1973) showed that if $n_h \propto N_h\sqrt{\bar{x}_h}$, then under $M(\delta_0, \ldots, \delta_J : x)$ the strategy $[s^*(J), \hat{T}_{RS}]$ is more efficient than $[s(J), \hat{T}(0, 1 : x)]$ in the sense that $E_M[\hat{T}(0, 1 : x) - T]^2 \geq E_M(\hat{T}_{RS} - T)^2$.

However, because the separate ratio estimator does not correspond to a model satisfying the conditions of Theorem 2, the strategy $[s^*(J), \hat{T}_{RS}]$ is not the best that we can use. When $\mathrm{var}_M(Y_i) = \sigma^2 x_i$, as in $M(0, 1 : x)$, the minimal model is $M(x^{1/2}, x : x)$. Now suppose that the correct model contains some higher order polynomial terms. Specifically, let $M(\delta_0, \delta_{1/2}, \ldots, \delta_J : x)$ denote the model with $E_M(Y_i) = \delta_0 + \delta_{1/2}x_i^{1/2} + \delta_1 x_i + \ldots + \delta_J x_i^J$ and $\mathrm{var}_M(Y_i) = \sigma^2 x_i$. If the sample has weighted balance so that

$$\frac{1}{n}\sum_s \frac{x_i^j}{x_i^{1/2}} = \frac{\bar{x}^{(j)}}{\bar{x}^{1/2}} \quad \text{for} \quad j = 0, 1/2, 1, \ldots, J \tag{14}$$

then the *BLU* predictor $\hat{T}(x^{1/2}, x : x)$ under $M(x^{1/2}, x : x)$ is protected against bias if the model is really $M(\delta_0, \delta_{1/2}, \ldots, \delta_J : x)$, and therefore *a fortiori* if $M(\delta_0, \ldots, \delta_J : x)$ holds.

By Theorem 1, when (14) is satisfied, $\hat{T}(x^{1/2}, x : x)$ reduces to the mean-of-ratios estimator $\hat{T}_{MR}(x^{1/2}) = N\bar{x}^{(1/2)}n^{-1}\Sigma_s Y_i/x_i^{1/2}$ and has error variance

$$\text{var}_M[\hat{T}_{MR}(x^{1/2}) - T] = \sigma^2\left[\frac{1}{n}(N\bar{x}^{(1/2)})^2 - N\bar{x}\right] \tag{15}$$

By Theorem 1, this error variance will be less than or equal to any that can be achieved under $M(\delta_0, \delta_{1/2}, \ldots, \delta_J : x)$ using $\hat{T}_{RS}$. In fact, it can be shown that the error variance of $\hat{T}_{RS}$ exceeds (15) by at least $\sigma^2/n[(\Sigma_h N_h\sqrt{\bar{x}_h})^2 - (N\bar{x}^{(1/2)})^2]$.

### 5.2. *Overbalance*

Scott, Brewer, and Ho (1978) noted that under the model $M(\delta_0, \delta_1 \ldots, \delta_J : x^2)$ the *BLU* predictor $\hat{T}(0, 1 : x^2)$ is protected by an ''overbalanced'' sample in which

$$\bar{x}_s^{(j-1)} = \Sigma_r x_i^j/\Sigma_r x_i, \quad j = 0, 1, \ldots, J \tag{16}$$

where $r$ denotes the set of nonsample units. If the model is actually $M(\delta_0, \delta_1, \ldots, \delta_J : V(x))$ with $V(x) = \sigma_1^2 x + \sigma_2^2 x^2$, then the error variance of $\hat{T}(0, 1 : x^2)$ in an overbalanced sample is less than the error variance of the ratio estimator in an unweighted balanced sample (i.e., the Royall-Herson strategy for $H = 1$).

If the variance specification in the working model $M(0, 1 : x^2)$ is correct, then the error variance of $\hat{T}(0, 1 : x^2)$ in an overbalanced sample is $\sigma^2[(N\bar{x})^2/n - \Sigma_1^N x_i^2 + \Sigma_s(x_i - \bar{x})^2]$. As observed by Royall (1992), a better strategy when the variance is proportional to $x^2$ is to use the minimal model $M(0, 1, 1 : x^2)$, the estimator $\hat{T}(0, 1, 1 : x^2)$, and a weighted balanced sample. Balance condition (2) is $\bar{x}_s = \bar{x}^{(2)}/\bar{x}$ and Theorem 1 gives the minimum variance of $\hat{T}(0, 1, 1 : x^2)$ as $\sigma^2[(N\bar{x})^2/n - \Sigma_1^N x_i^2]$. Thus, $\hat{T}(0, 1, 1 : x^2)$ with weighted balance has a smaller error variance than $\hat{T}(0, 1 : x^2)$ with overbalance, and both estimators are protected against bias under the general polynomial model $M(\delta_0, \delta_1, \ldots, \delta_J : v)$ by their respective balance conditions.

Under the variance specification $V(x) = \sigma_1^2 x + \sigma_2^2 x^2$, the comparison of the overbalanced strategy with the weighted balanced strategy is less clear. However, if the sample and population are large and the sampling fraction is negligible $(N, n, (N - n) \to \infty$ and $n/N \to 0)$, then $\Sigma_r x_i^j/\Sigma_r x_i \cong \bar{x}^{(j)}/\bar{x}$ and the overbalance condition in (16) is the same as a weighted balance condition. Thus, the two strategies will be essentially the same.

## 6. Formation of Strata

The primary question traditionally posed when stratifying by size is how to form the strata. When a common model holds for the entire population as in Section 4 and $\mathbf{V1}_N$, $\mathbf{V}^{1/2}\mathbf{1}_N \in \mathcal{M}(\mathbf{X})$, we know that the *BLU* predictor with a weighted balanced sample is the best strategy. Stratification by size is then, at best, a mechanism for selecting a weighted balanced sample. However, various methods of strata formation are used in practice and it is interesting to investigate their properties.

Assume that Model (10) holds, and let $\mathbf{V} = diag(\mathbf{V}_h)$ with $\mathbf{V}_h = diag(x_{hi}^\gamma)$. When $\mathbf{V1}_N$, $\mathbf{V}^{1/2}\mathbf{1}_N \in \mathcal{M}(\mathbf{X})$ and the sample is $s_h \in B(\mathbf{X}_h : \mathbf{V}_h)$, the error variance of the *BLU* predictor is given by (12). The problem of how to create strata is most conveniently studied when an equal number of sample units, $n_h = n_0$, is allocated to each stratum. In that case, (12)

becomes

$$\text{var}_M(\hat{T} - T) = \frac{\sigma^2}{n_0}\Sigma_h(N_h\bar{x}_h^{(\gamma/2)})^2 - \sigma^2 N\bar{x}^{(\gamma)} \tag{17}$$

and optimal stratification occurs when $\Sigma_h(N_h\bar{x}_h^{(\gamma/2)})^2$ is minimized. Let $Z_h = N_h\bar{x}_h^{(\gamma/2)}$. Adding and subtracting $\sigma^2 H\bar{Z}^2/n_0$, where $\bar{Z} = \Sigma_{h=1}^{H} Z_h/H$, gives

$$\text{var}_M(\hat{T} - T) = \frac{\sigma^2}{n_0}S_Z^2 + \frac{\sigma^2}{n_0}\frac{(N\bar{x}^{(\gamma/2)})^2}{H} - \sigma^2 N\bar{x}^{(\gamma)} \tag{18}$$

where $S_Z^2 = \Sigma_h(Z_h - \bar{Z})^2$. The one term in (18) that depends on the formation of the strata is the first, which is eliminated by making the $Z_h$ all equal. Expression (18) then becomes $\text{var}_M(\hat{T} - T) = \sigma^2/n(N\bar{x}^{(\gamma/2)})^2 - \sigma^2 N\bar{x}^{(\gamma)}$. This is just the variance (13) with optimal allocation and equal costs, which, of course, it must be since allocating the same number of sample units to each stratum is optimal when the $Z_h$ are all equal.

From the results in Section 4 we know that a stratified weighted balanced sample, optimally allocated to strata, is also a sample with overall weighted balance. Thus, we again confirm that stratified, weighted balanced sampling gains nothing over weighted balanced sampling in the common model case.

Equalizing $Z_h = N_h\bar{x}_h^{(\gamma/2)}$ leads to several ''equal aggregate size'' rules for forming strata found in the literature, for example, Cochran (1977, p.172), Godfrey, Roshwalb, and Wright (1984). When $\gamma = 0$, equal values of $N_h\bar{x}_h^{(\gamma/2)}$ correspond to equal numbers of units $N_h$ in each stratum. When $\gamma = 1$, we have equal aggregate square root of size, and $\gamma = 2$ gives equal aggregate $x$. We will include several of these methods in the empirical study in Section 7.

Another method of stratification, the cum $\sqrt{f}$ rule due to Dalenius and Hodges (1959), derives from consideration of the theoretical density function of the $Y_i$. This method is described in Särndal, Swensson, and Wretman (1992, Section 12.6) and is also one of the methods used in Section 7.

## 7. Some Empirical Results on Strata Formation

In this section we illustrate the different methods of strata formation and their effects on estimation in a simulation study. The effects of different combinations of stratification and estimator are examined in three populations which are well known in the literature as Hospitals, Cancer (Royall and Cumberland 1981), and Beef (Chambers, Dorfman, and Wehrly 1993). Figure 1 shows scatterplots of the three.

### *7.1. Combinations of stratification, sample selection, and estimators*

The methods of stratification, sample selection, and estimators were combined in several ways. The complete list is given below. Note that these combinations are also shown as the rows in Figure 2, which summarizes our findings.

(1) equal numbers of units $N_h$ in each stratum
   (a) $\hat{T}_{0S}$, the stratified expansion estimator, combined with stratified simple random sampling (*stsrs*) without replacement

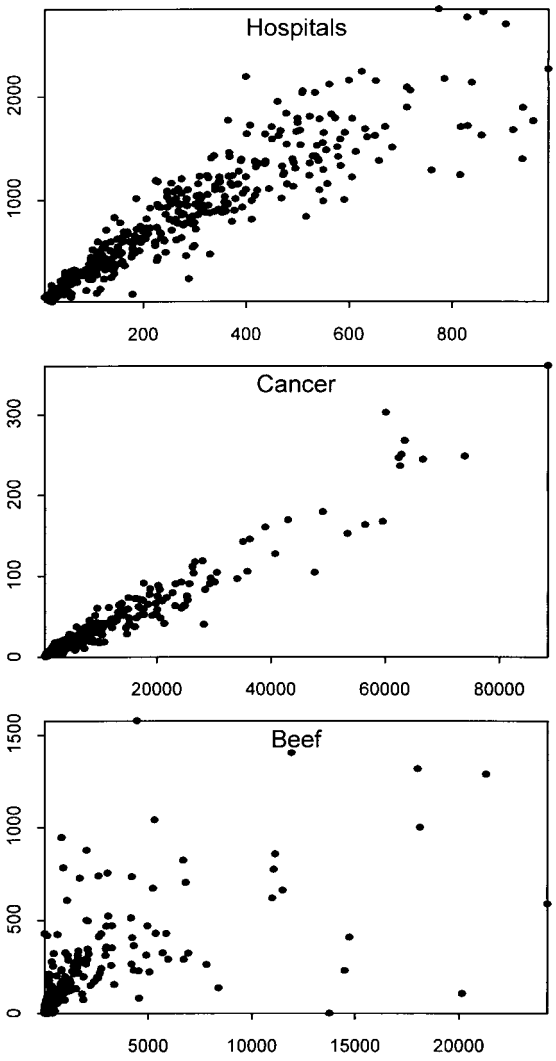


*Fig. 1. Scatterplots of three populations*

(b) $\hat{T}_{RS}$, the separate ratio estimator, combined with *stsrs* without replacement
(c) $\hat{T}_{LS}$, the separate regression estimator, defined below, combined with *stsrs* without replacement

(2) Equal cum $\sqrt{f}$ in each stratum
1(a), 1(b), and 1(c) as above

(3) equal aggregate total of $\sqrt{x}$ (cum $\sqrt{x}$) in each stratum
1(a), 1(b), and 1(c) as above and
(d) separate stratified version of $\hat{T}(x^{1/2}, x : x)$ combined with stratified $pp(x^{1/2})$ sampling
(e) $\hat{T}_{MRS}(x^{1/2})$, the stratified mean-of-ratios estimator, combined with $pp(x^{1/2})$ sampling

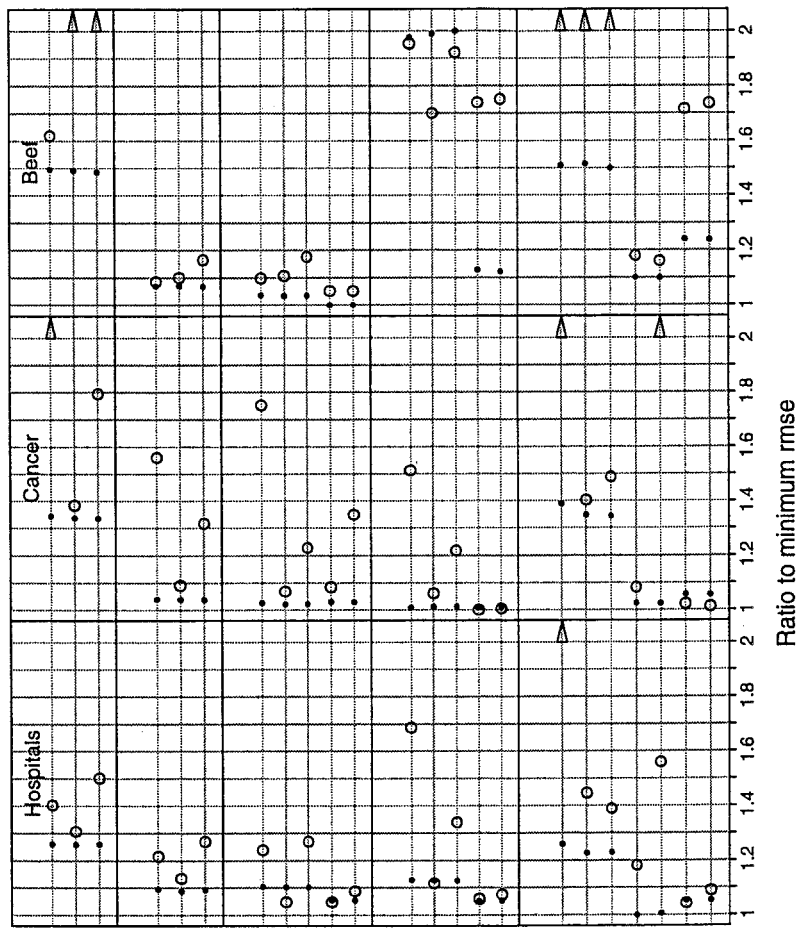


*Fig. 2. Ratios of rmses for estimators in stratified and unstratified sampling to the minimum rmse for the Hospitals, Cancer, and Beef populations; $S = 1{,}000$, $n = 30$, $nh = 6$ in all strata for stratified designs*

(4) equal aggregate total of $x$ (cum $x$) in each stratum
  1(a), 1(b), and 1(c) as above and
  (d) separate stratified version of $\hat{T}(x, x^2; x^2)$ combined with stratified $pp(x)$ sampling
  (e) $\hat{T}_{MRS}(x)$, the stratified mean-of-ratios estimator, combined with $pp(x)$ sampling
(5) unstratified sampling
  (a) $\hat{T}_0$, the expansion estimator combined with *srs* without replacement
  (b) $\hat{T}_R$, the ratio estimator, combined with *srs* without replacement
  (c) $\hat{T}_L$, the regression estimator, defined below, combined with *srs* without replacement
  (d) $\hat{T}(x^{1/2}, x : x)$, which is minimal when $\text{var}_M(Y_i) = \sigma^2 x_i$, combined with $pp(x^{1/2})$ sampling
  (e) $\hat{T}_{MR}(x^{1/2})$, a mean-of-ratios estimator, combined with $pp(x^{1/2})$ sampling
  (f) $\hat{T}(x, x^2 : x^2)$, which is minimal when $\text{var}_M(Y_i) = \sigma^2 x_i^2$, combined with $pp(x)$ sampling
  (g) $\hat{T}_{MR}(x)$, a mean-of-ratios estimator, combined with $pp(x)$ sampling.

Aggregation for all methods of stratification was done after sorting each population in ascending order on the auxiliary $x$. For the cum $\sqrt{f}$ rule, 100 equal length $x$-intervals were formed, and the running totals of $f_j^{1/2}$ were computed, where $f_j$ is the number of units in the $j^{\text{th}}$ interval. The intervals were then grouped into five strata. The separate regression estimator is defined as $\hat{T}_{LS} = \hat{T}_{0S} + \Sigma_h N_h b_{hs}(\bar{x}_h - \bar{x}_{hs})$ with $b_{hs} = \Sigma_{s_h}(x_{hi} - \bar{x}_{hs})(y_{hi} - \bar{y}_{hs})/\Sigma_{s_h}(x_{hi} - \bar{x}_{hs})^2$. The unstratified regression estimator $\hat{T}_L$ is equal to $\hat{T}_{LS}$ with $H = 1$.

The number of strata was taken as $H = 5$ in all cases. For each method of stratification, a sample of $n = 30$ was divided equally among the five strata giving $n_h = 6$ in each stratum. Remember from Section 6 that, when strata are formed to equalize $N_h \bar{x}_h^{(\gamma/2)}$, costs are all equal, and $\text{var}_M(Y_{hi}) = \sigma^2 x_{hi}^{\gamma}$, then an equal allocation is optimal. Thus, there is a logical consistency to using an equal allocation in each of methods (1), (3), and (4) of strata formation noted above. In addition, equal allocation is one method traditionally used with the cum $\sqrt{f}$ method.

### 7.2. *Methods of sample selection*

Both unrestricted and restricted sampling techniques were used in the simulation. Unrestricted $pp(x^{\gamma/2})$ was implemented using the random order, systematic method described by Hartley and Rao (1962). Restricted $pp(x^{\gamma/2})$ sampling was done by selecting a sample with the random-order method and then checking its closeness to weighted balance on four moments within each stratum. The balance measures

$$e_j(s_h) = \left| \frac{\sqrt{n}(\bar{x}_{sh}^{(j-\gamma/2)} - \bar{x}_h^{(j)}/\bar{x}_h^{(\gamma/2)})}{s_{jxh}} \right|, \quad j = 0, 1/2, 1, 2 \tag{19}$$

were calculated in each stratum, where $s_{jxh} = [\Sigma_{i=1}^{N_h} \pi_{hi}(x_{hi}^{j-\gamma/2} - \bar{x}_h^{(j)}/\bar{x}_h^{(\gamma/2)})^2]^{1/2}$ and $\pi_{hi} = x_{hi}^{\gamma/2}/(N_h \bar{x}_h^{(\gamma/2)})$. For the pairs $(j = 1/2, \gamma = 1)$ and $(j = 1, \gamma = 2)$, $e_j(s_h) = 0$, and balance on those moments is trivially satisfied. For the nontrivial cases, if $e_j(s_h) \leq 0.1256613$ for all measures in every stratum, then the sample was retained; otherwise, it was discarded and another drawn. With unstratified sampling, $H = 1$, so that only overall

sample balance was checked. The 55$^{th}$ quantile of the standard normal distribution is $q_{.55} = 0.1256613$. Thus, this technique retains only about 10% of the best-balanced samples.

Balancing on the other moments above, in addition to $j = \gamma$, protects the minimal estimator against polynomial terms not in a minimal working model without losing any precision under the working model. For instance, when $\gamma = 1$, balancing on the $j = 0$ and 2 terms protects the minimal estimator against the possibility that the correct model contains an intercept and a quadratic term. With the weighted balance conditions above, the mean-of-ratios estimator $\hat{T}_{MRS}(x^{\gamma/2})$ is equal to the minimal estimator $\hat{T}(x^{\gamma/2}, x^{\gamma} : x^{\gamma})$, but in unbalanced samples there may be important differences – a point that the simulation results will illustrate.

Unrestricted and restricted *stsrs* samples were used in stratification plans (1)–(4) for Estimators (a), (b), and (c) above. In the unrestricted samples, a simple random sample was selected without replacement in each stratum and retained regardless of its configuration. For restricted samples, a without-replacement *srs* was selected in each stratum and checked for simple balance on the moments $\bar{x}_{sh}^{(j)}, j = 0, 1/2, 1, 2$. For unstratified sampling – plan (5) – only overall balance was checked. As above, 10% of the best-balanced samples were retained. This type of sampling was also investigated by Herson (1976) and Royall and Cumberland (1981).

### *7.3. Simulation results*

For each combination of stratification, sampling method, and estimator, 1,000 samples were selected. For restricted samples this means that samples were selected until 1,000 were retained. The root mean squared errors for each estimator were computed as $rmse(\hat{T}) = [\Sigma_{s=1}^{1000}(\hat{T} - T)^2/1000]^{1/2}$. Figure 2 presents results, using a rowplot of the type devised by Carr (1994). In each column, the ratio of each *rmse* to the minimum *rmse* among the estimators for the population is plotted. Black dots represent restricted samples while open circles are for unrestricted samples. The narrow triangles are cases where the ratio was truncated at 2 to avoid scaling problems. Some observations are:

- In Hospitals and Cancer, the minimal estimator with unstratified, restricted $pp(x^{1/2})$ sampling has the smallest *rmse* or very near it. In Beef, where the relationship between $Y$ and $x$ is the most diffuse, the stratified, minimal estimator with $pp(x^{1/2})$ sampling is best. (See the further comment on Beef below.)
- Unrestricted sampling is generally inferior to restricted, balanced sampling.
- The minimal and mean-of-ratios estimators have about the same *rmse*'s in weighted balanced samples, as expected. In contrast, $\hat{T}_{MRS}(x^{1/2})$ can have much higher *rmse*'s than $\hat{T}(x^{1/2}, x : x)$ in unrestricted $pp(x^{1/2})$ sampling.
- The estimators used when sampling is *stsrs* – expansion, ratio, and regression – are improved by balanced sampling, but are generally inferior to the minimal estimator with weighted balance, as anticipated in Section 5.
- For a given selection method ($pp(x^{1/2})$ or $pp(x)$), stratification with weighted balance within strata yields *rmse*'s very near those of unstratified sampling and weighted balance for the minimal or mean-of-ratios estimator in Hospitals and Cancer. This is expected since the minimal and mean-of-ratios estimators are equal

in weighted balanced samples, and an optimally allocated, stratified, weighted balance sample also has overall balance.

- In Beef, stratified weighted balanced sampling is somewhat better than unstratified weighted balanced sampling for the minimal and mean-of-ratios estimators in both $pp(x^{1/2})$ or $pp(x)$ samples. This may be due to better overall balance being obtained using stratification and optimal allocation than for unstratified samples, since the restrictions (19) are tighter for $H = 5$ than for $H = 1$. Another potential explanation is that a single overall model is inadequate, or the relationship between $Y$ and $x$ is nonlinear beyond what a polynomial of low order can capture, in which case a stratum by stratum fit would be advantageous. In fact, Chambers, Dorfman, and Wehrly (1993) note that $\log(Y + 1)$ is approximately linearly related to $\log(\log(x))$ in Beef, implying that the relationship between $Y$ and $x$ is extremely nonlinear.
- In contrast to the weighted balanced results, stratification with simple balanced sampling does substantially improve the expansion, ratio, and regression estimators. In the Cancer population, these three estimators are competitive with the minimal estimator when the cum $\sqrt{f}$ or cum $\sqrt{x}$ rule is used for strata formation.

## 8. Discussion

Rules for stratification by size have been in the literature for many years; see e.g., Mahalanobis (1952). The relative merits of particular approaches to selecting stratum boundaries have been commonly studied under a single population model with uniform costs per unit sampled (e.g., Bethel 1989). A major point of the present article is that under this circumstance one *need not stratify* to get best results. A strategy of selecting a weighted balanced sample, with weights determined by the model's variance structure, combined with the best linear unbiased estimator under the model, yields a bias robust estimator having minimal variance among sample or estimator strategies. Exact model-based optimality *can* be obtained through stratified, weighted balanced sampling and optimum allocation, but the stratification by size is superfluous, unless the strata are needed for other reasons, such as estimating domain characteristics, controlling for differential costs, or coping with extreme nonlinearity.

One means of selecting weighted balanced samples is restricted probability proportional to size sampling in which probability samples that are not well-balanced are rejected. This type of plan has the advantage of maintaining the characteristic of impartiality associated with randomly selected samples. Such a restriction is, in fact, what practitioners have been roughly accomplishing for many years using systematic sampling from a list sorted by size.

Probability proportional to size sampling itself is usually approximated in practice either by some version of systematic sampling as in the Hartley-Rao (1962) approach, or by some version of stratified sampling as in Wright (1983). Thus, as an incidental tool stratified sampling retains its utility.

Our results verify that restricted sampling (either stratified or unstratified) yields smaller mean squared errors than a variety of stratified, unrestricted random sampling plans even when optimal allocation is used.

Choosing a sample and an estimator to control bias are key features of the approach studied here. Thus, in closing, it is useful to compare the model-based and design-based concepts of robustness. The design-based counterpart of robustness to model-failure is asymptotic design unbiasedness (ADU): as sample and population grow large, the given estimator is ADU if it is design-unbiased despite total failure of the model. For example, by requiring ADU, Wright (1983) determines criteria which lead to the choice of the generalized regression estimator (GREG), which has achieved so much prominence in recent years. Apart from the model-based versus design-based dichotomy, the difference between the two standards is that ADU is asymptotic, and all-purpose, in the sense that the ADU estimator is indifferent to the specifics of the alternative models that might hold, and to the specifics of the particular sample chosen, whereas the bias-robust estimator is specific as to what group of models it is robust against, under specific conditions of the particular (finite) sample chosen.

## 9. References

Bethel, J. (1989). Minimum Variance Estimation in Stratified Sampling. Journal of the American Statistical Association, 84, 260–265.

Brewer, K.R.W. (1963). Ratio Estimation and Finite Populations: Some Results Deducible from the Assumption of an Underlying Stochastic Process. Australian Journal of Statistics, 5, 93–105.

Carr, D. (1994). Topics in Scientific Visualization: Using Gray in Plots. Statistical Computing and Graphics Newsletter, 5, 11–14.

Cochran, W.G. (1977). Sampling Techniques, 3rd ed. New York: Wiley.

Chambers, R.L., Dorfman, A.H., and Wehrly, T.E. (1993). Bias Robust Estimation in Finite Populations Using Nonparametric Calibration. Journal of the American Statistical Association, 88, 268–277.

Dalenius, T. and Hodges, J.L., Jr. (1959). Minimum Variance Stratification. Journal of the American Statistical Association, 54, 88–101.

Godambe, V.P. and Joshi, V.M. (1965). Admissibility and Bayes Estimation in Sampling Finite Populations. I. Annals of Mathematical Statistics, 2, 1707–1722.

Godfrey, J., Roshwalb, A., and Wright, R. (1984). Model-based Stratification in Inventory Cost Estimation. Journal of Business and Economic Statistics, 2, 1–9.

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

Herson, J. (1976). An Investigation of Relative Efficiency of Least-Squares Prediction to Conventional Probability Sampling Plans. Journal of the American Statistical Association, 71, 700–703.

Mahalanobis, P.C. (1952). Some Aspects of the Design of Sample Surveys. Sankhya, 12, 1–7.

Royall, R.M. (1992). Robustness and Optimal Design Under Prediction Models for Finite Populations. Survey Methodology, 18, 179–185.

Royall, R.M. and Cumberland, W.G. (1981). An Empirical Study of the Ratio Estimator and Estimators of its Variance. Journal of the American Statistical Association, 76, 66–77.

Royall, R.M. and Herson, J. (1973). Robust Estimation in Finite Populations II: Stratification on a Size Variable. Journal of the American Statistical Association, 68, 890–893.

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

Scott, A.J., Brewer, K.R.W., and Ho, E.W.H. (1978). Finite Population Sampling and Robust Estimation. Journal of the American Statistical Association, 73, 359–361.

Wright, R.L. (1983). Finite Population Sampling with Multivariate Auxiliary Information. Journal of the American Statistical Association, 78, 879–884.

Received August 1998
Revised July 1999