Stratification by Size Revisited

Alan H. Dorfman and Richard Valliant, U.S. Bureau of Labor Statistics
Richard Valliant, Room 4915, 2 Massachusetts Ave. NE, Washington DC 20212

Keywords: balanced sample, best linear unbiased predictor, robustness, superpopulation model.

Stratification is one of the most widely used techniques in finite population sampling. Strata are disjoint subdivisions of a population, the union of which exhaust the universe, each of which contains a portion of the sample. Two of 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. Using model-based analysis, we show that, in the situation where stratification by size is generally used, optimal allocation of a weighted balanced sample achieves exactly the same variance as unstratified, best linear unbiased (*BLU*) prediction coupled with weighted balanced sampling. In other words, stratification by size has no advantage over the optimal, unstratified procedure. This and other theoretical findings are illustrated with simulations using real populations.

1. A Stratified Linear Model and Weighted Balanced Samples

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

$$E_M(\mathbf{Y}_h)=\mathbf{X}_h\boldsymbol{\beta}_h,\ \mathrm{var}_M(\mathbf{Y}_h)=\mathbf{V}_h s_h^2 \qquad (1)$$

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 $\boldsymbol{\beta}_h$ is a $p_h\times 1$ parameter vector. The model in stratum h is denoted by $M(\mathbf{X}_h:\mathbf{V}_h)$ 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

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

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 sample satisfying (2) will be denoted by $B(\mathbf{X}_h:\mathbf{V}_h)$, and, when (2) is satisfied in each stratum, the entire sample is a *stratified weighted balanced sample*.

If the model has a certain structure given in Theorem 1 below, then a weighted balanced sample is the best that can be selected in the sense of making the error variance of the *BLU* predictor small. Let $M(\mathbf{X}_h)$ be the vector space generated by the columns of $\mathbf{X}_h$. A straightforward application of Theorem 2 in Royall (1992) yields the following stratified result.

***Theorem* 1.** Suppose that model (1) holds in stratum h for $h=1,\ldots,H$. If both $\mathbf{V}_h\mathbf{1}_{Nh}$ and $\mathbf{V}_h^{1/2}\mathbf{1}_{Nh}\in M(\mathbf{X}_h)$, 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

$$\hat{T}=\sum_{h=1}^{H}N_h\bar{v}_h^{(1/2)}\frac{1}{n_h}\sum_{i\in s_h}\frac{Y_{hi}}{v_{hi}^{1/2}}, \qquad (3)$$

and the error variance is

$$\mathrm{var}_M(\hat{T}-T)=\sum_h\left[\frac{1}{n_h}\left(N_h\bar{v}_h^{(1/2)}\right)^2-N_h\bar{v}_h\right]s_h^2 \quad (4)$$

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

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 $\hat{T}_{MRS}(v^{1/2})$.

2. Optimal Allocation for Stratified Balanced Sampling

The optimum allocation to the strata of a weighted balanced sample can be easily calculated. Assume that the cost of sampling is $C(\mathbf{n}) = C_0 + \sum_h c_h n_h$ where C_0 is a fixed cost and c_h is the cost per unit sampled in stratum *h*.

***Theorem* 2.** Assume that model (1) holds, that $\mathbf{V}_h \mathbf{1}_{Nh}$ and $\mathbf{V}_h^{1/2} \mathbf{1}_{Nh} \in M(\mathbf{X}_h)$, 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 $C(\mathbf{n}) = C_0 + \sum_h c_h n_h$, is

$$\frac{n_h}{n} = \frac{N_h \bar{v}_h^{(1/2)} s_h \big/ \sqrt{c_h}}{\sum_{h'} N_{h'} \bar{v}_{h'}^{(1/2)} s_{h'} \big/ \sqrt{c_{h'}}} \text{ for } h = 1, \ldots, H .$$

When optimal allocation is used and all costs are equal, the *BLU* predictor (3) becomes

$$\hat{T} = \frac{1}{n} \left(\sum_h N_h \bar{v}_h^{(1/2)} s_h \right) \sum_h \sum_{s_h} \frac{Y_{hi}}{v_{hi}^{1/2} s_h} \qquad (5)$$

and its error variance (4) can be rewritten as

$$\mathrm{var}_M \left(\hat{T} - T \right) = \frac{1}{n} \left(\sum_h N_h \bar{v}_h^{(1/2)} s_h \right)^2 - \sum_h N_h \bar{v}_h s_h^2 \qquad (6)$$

3. The Case of a Single Model for the Population

An important special case is having a single model that fits the whole population. Assume the model in each stratum is

$$E_M (\mathbf{Y}_h) = \mathbf{X}_h \mathbf{b} , \ \mathrm{var}_M (\mathbf{Y}_h) = \mathbf{V}_h s^2 , \qquad (7)$$

with $\mathbf{X}_h$ and $\mathbf{V}_h$ defined as in (1). Expression (7) is just another way of writing the unstratified model $M(\mathbf{X}{:}\mathbf{V})$. Thus, strata can be ignored in calculating the *BLU* predictor and its error variance. If $\mathbf{V1}_N$ and $\mathbf{V}^{1/2} \mathbf{1}_N \in M(\mathbf{X})$, then, by Theorem 1 with *H*=1, a weighted balanced sample $s \in B(\mathbf{X}{:}\mathbf{V})$ is optimal for the *BLU* predictor. In that case, the *BLU* reduces to

$$\hat{T} = \frac{1}{n} N \bar{v}^{(1/2)} \sum_s Y_i \big/ v_i^{1/2} \qquad (8)$$

with error variance

$$\mathrm{var}_M \left(\hat{T} - T \right) = s^2 \left[n^{-1} \left(N \bar{v}^{(1/2)} \right)^2 - N \bar{v} \right].$$

On the other hand, suppose we select a stratified weighted balanced sample and use the optimal allocation given in Theorem 2 above for the equal cost case. Using (5) with $s_h = s$, the *BLU* predictor with the optimal allocation is

$$\hat{T} = \frac{1}{n} \left(\sum_h N_h \bar{v}_h^{(1/2)} \right) \sum_h \sum_{s_h} Y_{hi} \big/ v_{hi}^{1/2}$$

which is exactly equal to (8). In other words, stratification with optimal allocation of a stratified weighted balanced sample gains nothing at all compared to the strategy of selecting an unstratified sample with overall weighted balance.

A situation where a common model may hold for the whole population is one where a single auxiliary variable *x* is available. The auxiliary can be used for stratification by size as well as for estimation. Strata are formed by ordering the units from low to high based on *x* so that the first stratum contains the N_1 units with the smallest *x* values, the second stratum contains the next N_2 smallest units, and so on. Take the special case of model (7) given by

$$E_M (\mathbf{Y}_h) = \mathbf{X}_h \mathbf{b} , \ \mathrm{var}_M (Y_{hi}) = s^2 x_{hi}^g \qquad (9)$$

where, in many populations, $0 \le g \le 2$. When (9) is true, we can use the idea of the *minimal model* introduced by Dorfman and Valliant (1997). The minimal model is the one with least variables satisfying $\mathbf{V1}_N$ and $\mathbf{V}^{1/2} \mathbf{1}_N \in M(\mathbf{X})$. When $\mathrm{var}_M (Y_{hi}) = s^2 x_{hi}^g$, this is $E_M (Y_{hi}) = b_{g/2} x_{hi}^{g/2} + b_g x_{hi}^g$. We denote this model by $M\left(x^{g/2}, x^g : x^g \right)$. With the variance specification in (9), the optimum allocation in Theorem 2 becomes

$$\frac{n_h}{n} = \frac{N_h \bar{x}_h^{(g/2)} \big/ \sqrt{c_h}}{\sum_{h'} N_{h'} \bar{x}_{h'}^{(g/2)} \big/ \sqrt{c_{h'}}} \qquad (10)$$

When the optimal allocation (10) is used and costs are all equal, the error variance (4) of the *BLU* predictor in a stratified weighted balanced sample reduces to the variance for an unstratified, weighted balanced sample:

$$\mathrm{var}_M \left(\hat{T} - T \right) = s^2 \left[\frac{1}{n} \left(N \bar{x}^{(g/2)} \right)^2 - N \bar{x}^{(g)} \right]. \qquad (11)$$

4. Comparisons with Other Strategies

In this section we denote the polynomial model having $E_M (Y_i) = d_0 + d_1 x_i + \cdots + d_J x_i^J$ and $\mathrm{var}_M (Y_i) = s^2 v_i$ by $M(d_0, \ldots, d_J : v)$.

When strata are formed on the basis of a size measure *x*, an oft-studied procedure is the separate ratio estimator, defined as

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

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

$$\text{var}_M\left(\hat{T}_{RS} - T\right) = s^2 \sum_h \frac{N_h^2}{n_h}\left(1 - f_h\right)\frac{\bar{x}_{hr}\bar{x}_h}{\bar{x}_{hs}},$$

where $f_h = n_h/N_h$ and $\bar{x}_{hr} = \sum_{i\notin s_h} x_{hi}/\left(N_h - n_h\right)$. If one is 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 of the simple ratio estimator 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 $M\left(d_0,\ldots,d_J : x\right)$, then the estimator has a bias:

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

where $\bar{x}_h^{(j)} = \sum_{i=1}^{N_h} x_{hi}^j/N_h$ and $\bar{x}_{hs}^{(j)} = \sum_{i\in s_h} x_{hi}^j/n_h$. 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 $s(J)$. When the working model $M(0,1:x)$ holds, the optimal allocation of the sample to strata for $\hat{T}_{RS}$ is $n_h \propto N_h\sqrt{\bar{x}_h}$.

Protection against bias under the polynomial model $M\left(d_0,\ldots,d_J : x\right)$ is afforded either by simple balanced sampling with the ratio estimator or stratified balanced sampling with the separate ratio estimator. Royall and Herson (1973) showed that if $n_h \propto N_h\sqrt{\bar{x}_h}$, then under $M\left(d_0,\ldots,d_J : x\right)$ the strategy $\left[s^*(J), \hat{T}_{RS}\right]$ is more efficient than $\left[s(J), \hat{T}(0,1:x)\right]$ in the sense that $E_M\left[\hat{T}(0,1:x) - T\right]^2 \geq E_M\left(\hat{T}_{RS} - T\right)^2$.

But, because the separate ratio estimator does not flow from a model satisfying the conditions of Theorem 1, the strategy $\left[s^*(J), \hat{T}_{RS}\right]$ is not the best that we can do. When $\text{var}_M\left(Y_i\right) = s^2 x_i$, as in $M(0,1:x)$, the minimal model is $M\left(x^{1/2}, x:x\right)$. Now, suppose that the correct model contains some higher order polynomial terms. Specifically, let $M\left(d_0, d_{1/2},\ldots,d_J : x\right)$ denote the model with $E_M\left(Y_i\right) = d_0 + d_{1/2}x_i^{1/2} + d_1 x_i + \cdots + d_J x_i^J$ and $\text{var}_M\left(Y_i\right) = s^2 x_i$. If the sample has weighted balance—equation (2) above—so that

$$\frac{1}{n}\sum_s \frac{x_i^j}{x_i^{1/2}} = \frac{\bar{x}^{(j)}}{\bar{x}^{(1/2)}} \text{ for } j = 0, ½, 1,\ldots,J, \qquad (12)$$

then the *BLU* predictor $\hat{T}\left(x^{1/2}, x:x\right)$ under $M\left(x^{1/2}, x:x\right)$ is protected against bias if the model is really $M\left(d_0, d_{1/2},\ldots,d_J : x\right)$. By Theorem 1, when (12) is satisfied, $\hat{T}\left(x^{1/2}, x:x\right)$ reduces to the mean-of-ratios estimator $\hat{T}_{MR}\left(x^{1/2}\right) = N\bar{x}^{(1/2)} n^{-1}\sum_s Y_i/x_i^{1/2}$ and has error variance

$$\text{var}_M\left[\hat{T}_{MR}\left(x^{1/2}\right) - T\right] = s^2\left[\frac{1}{n}\left(N\bar{x}^{(1/2)}\right)^2 - N\bar{x}\right]. \quad (13)$$

By Theorem 1, this error variance will be less than or equal to any that can be achieved under $M\left(d_0, d_{1/2},\ldots,d_J : x\right)$ using $\hat{T}_{RS}$.

5. Formation of Strata

A 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 2 and $\mathbf{V1}_N$, $\mathbf{V}^{1/2}\mathbf{1}_N \in M(\mathbf{X})$, we know that the *BLU* predictor with a weighted balanced sample is the best strategy. That is, stratification in this common circumstance is unnecessary. However, various methods of strata formation are used in practice, and it is interesting to investigate their properties.

One set of methods are known as *equal aggregate size* rules. Units are sorted from low to high based on *x*. Strata are then formed in such a way that each contains about the same total of the size variable or a monotone transformation of it. Equalizing $N_h \bar{x}_h^{(g/2)}$ leads to several stratification rules. When $g = 0$, equal values of $N_h \bar{x}_h^{(g/2)}$ correspond to equal numbers of units N_h in each stratum. When $g = 1$, we have equal aggregate square root of size, and $g = 2$ gives equal aggregate *x*.

The equal aggregate size rules can be derived using only model-based arguments. Due to limited space, we will only summarize the result. When the model is $E_M(\mathbf{Y}_h) = \mathbf{X}_h \boldsymbol{\beta}$, $\text{var}_M(Y_{hi}) = \sigma^2 x_{hi}^g$ with $\mathbf{V1}_N$, $\mathbf{V}^{1/2}\mathbf{1}_N \in M(\mathbf{X})$, the sample is $s_h \in B(\mathbf{X}_h : \mathbf{V}_h)$, and an equal number of units is allocated to each stratum, then the error variance of the *BLU* predictor is minimized if strata are constructed to have equal $N_h \bar{x}_h^{(g/2)}$ in each. Moreover, if strata are constructed in this manner, an equal allocation is the optimal allocation in the equal cost case. However, an optimally allocated stratified, weighted balanced sample yields exactly the same variance as an unstratified sample with weighted balance.

Another method of stratification is known as the cum $\sqrt{f}$ rule due to Dalenius and Hodges (1959) that we will include in the simulation reported in section 6 but will not describe in detail here.

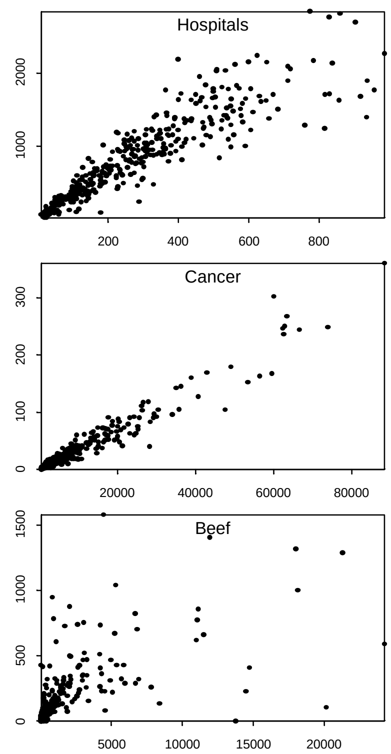
6. Some Empirical Results on Strata Formation

In this section we will illustrate the different methods of strata formation and their effects on estimation in a simulation study. The three populations used are known in the literature as Hospitals, Cancer (Royall and Cumberland 1981), and Beef (Chambers and Dunstan 1986). Figure 1 shows scatterplots of the three. Four methods of stratification were used with *H*=5:

(1) equal numbers of units N_h in each stratum,
(2) equal cum $\sqrt{f}$ in each stratum.
(3) equal aggregate total of $\sqrt{x}$ in each stratum (the cum $\sqrt{x}$ rule),
(4) equal aggregate total of *x* in each stratum (the cum *x* rule).

We used the four methods of stratification listed above and also did unstratified sampling. Five

Figure 1. Scatterplots of three populations.



combinations of estimators and sample selection methods were used:

(a) $\hat{T}(x^{g/2}, x^g : x^g)$, which is minimal when $\text{var}_M(Y_i) = \sigma^2 x_i^g$, and $pp(x^{g/2})$ sampling with $g = 1, 2$,
(b) $\hat{T}_{MRS}(x^{g/2})$, the stratified mean-of-ratios estimator and $pp(x^{g/2})$ sampling with $g = 1, 2$,
(c) $\hat{T}_{0S}$, the stratified expansion estimator, and stratified simple random sampling (*stsrs*) without replacement,
(d) $\hat{T}_{RS}$, the separate ratio estimator, and *stsrs* without replacement, and
(e) $\hat{T}_{LS}$, the separate regression estimator, defined below, and *stsrs* without replacement.

The separate regression estimator is defined as $\hat{T}_{LS} = \hat{T}_{0S} + \sum_h N_h b_{hs}(\bar{x}_h - \bar{x}_{hs})$ with $b_{hs} = \sum_{s_h}(x_{hi} - \bar{x}_{hs}) y_{hi} \Big/ \sum_{s_h}(x_{hi} - \bar{x}_{hs})^2$. Note that $\hat{T}_{MRS}(x^{g/2})$ is the Horvitz-Thompson estimator in $pp(x^{g/2})$ sampling and that $\hat{T}_{0S}$ is both the Horvitz-Thompson and the Hájek estimator in *stsrs*.

For each method of stratification a sample of *n*=30 was divided equally among the five strata giving $n_h = 6$ in each stratum. As noted in section 5, when strata are formed to equalize

$N_h \bar{x}_h^{(g/2)}$, costs are all equal, and $\text{var}_M(Y_{hi}) = s^2 x_{hi}^g$, then an equal allocation is optimal. In addition, equal allocation is one method traditionally used with the cum $\sqrt{f}$ method.

Both unrestricted and restricted sampling techniques were used in the simulation. Unrestricted $pp(x^{g/2})$ was implemented using the random order, systematic method described by Hartley and Rao (1962). Restricted $pp(x^{g/2})$ 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

$$e_j(s_h) = \left| \frac{\sqrt{n}\left(\bar{x}_{sh}^{(j-g/2)} - \bar{x}_h^{(j)} \big/ \bar{x}_h^{(g/2)}\right)}{s_{jxh}} \right|, \; j = 0, \tfrac{1}{2}, 1, 2$$

were calculated in each stratum where $s_{jxh} = \left[\sum_{i=1}^{N_h} p_{hi} \left(x_{hi}^{j-g/2} - \bar{x}_h^{(j)} \big/ \bar{x}_h^{(g/2)} \right)^2 \right]^{1/2}$ and $p_{hi} = x_{hi}^{g/2} \big/ \left(N_h \bar{x}_h^{(g/2)} \right)$. For the pairs $(j = \tfrac{1}{2}, g = 1)$ and $(j = 1, g = 2)$, $e_j(s_h) = 0$, and balance on those moments is trivially satisfied. For the non-trivial cases, if $e_j(s_h) \le 0.1256613$ for all measures in every stratum, then the sample was retained; otherwise, it was discarded and another drawn. This technique retains only about 10% of the best-balanced samples.

Balancing on the other moments above, in addition to $j = g$, protects the minimal estimator against different polynomial terms not in a minimal working model without losing any precision under the working model. With the weighted balance conditions above, the mean-of-ratios estimator $\hat{T}_{MRS}(x^{g/2})$ is equal to the minimal estimator $\hat{T}(x^{g/2}, x^g : x^g)$, but in unbalanced samples there may be important differences—a point that the simulation results will illustrate.

Unrestricted and restricted *stsrs* samples were used for estimators (c)-(e) 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 $\bar{x}_{sh}^{(j)}$, $j = 0, \tfrac{1}{2}, 1, 2$. As above, only about 10% of the best-balanced samples were retained.

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 square errors for each estimator were computed as $\text{rmse}(\hat{T}) = \left[\sum_{s=1}^{1000} (\hat{T} - T)^2 \big/ 1000 \right]^{1/2}$. 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 the stratified, minimal estimator with $pp(x^{1/2})$ sampling is best.
- 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, $\hat{T}_{MRS}(x^{g/2})$ can have much higher *rmse*'s than $\hat{T}(x^{g/2}, x^g : x^g)$ in unrestricted $pp(x^{g/2})$ 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 4.
- 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. Beef is the exception because restricted, stratified

sampling achieved better overall, weighted balance than unstratified sampling.

- In contrast, stratification with balanced sampling can substantially improve the expansion, ratio, and regression estimators.

7. Discussion

Rules for stratification by size have been in the literature for many years, e.g., Mahalanobis (1952). More recently (Wright 1983) the method has been justified as a means of approximating the optimum selection probabilities derived by Godambe and Joshi (1965). Thus, there has been some recognition that stratification by size may entail a loss of efficiency, but the method remains a common tool of practitioners. 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 or controlling for differential costs.

Acknowledgements

Any opinions expressed are those of the authors and do not constitute policy of the Bureau of Labor Statistics.

References

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

Chambers, R.L., and Dunstan, R. (1986), “Estimating Distribution Functions from Survey Data,” *Biometrika*, **73**, 597-604.

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

Dorfman, A.H. and Valliant, R. (1997), “The Hájek Estimator Revisited,” *Proceedings of the Section on Survey Methods Research*, American Statistical Association, to be published.

Godambe, V.P., and Joshi, V.M. (1965), "Admissibility and Bayes Estimation in Sampling Finite Populations - I," *Annals of Mathematical Statistics*, **36**, 1707-1723.

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

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.

Wright, R. (1983), “Finite Population Sampling with Multivariate Auxiliary Information,” *Journal of the American Statistical Association*, 78, 879-884.

Figure 2. Ratios of rmses for estimators in stratified and unstratified sampling to the minimum for the Hospitals, Cancer, and Beef populations; S=1000, n=30, nh=6 in all strata for stratified