The Hájek Estimator Revisited

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

Keywords: balanced sample, best linear unbiased predictor, Horvitz-Thompson estimator, ratio estimator, robustness, superpopulation model.

1. The Hájek Estimator

Consider a population U of N units indexed by i. Suppose Y_i, $i \in U$ are values in the population of interest; in particular, suppose we wish to estimate their mean $\bar{Y} = N^{-1}\sum_{i=1}^{N} Y_i$ based on a sample s taken from the population U. Assume that the sample is taken according to a randomization scheme having inclusion probabilities $\pi_i = \Pr(i \in s)$. When the π_i are proportional to a positive quantity x_i available over U, and s has a predetermined sample size n, then $\pi_i = nx_i/N\bar{x}$, where $\bar{x} = \sum_{i=1}^{N} x_i \Big/ N$, and the sampling scheme is said to be *probability proportional to size* (*pps*).

Under this scheme, a well known and popular estimator attributed to Hájek (1971) is defined by

$$\hat{\bar{Y}}_{Haj} = \frac{\sum_s Y_i/\pi_i}{\sum_s 1/\pi_i}. \qquad (1)$$

He suggested this estimator in response to an observation by Basu (1971) on paradoxical behavior of the *pps*-unbiased Horvitz-Thompson (1952) estimator

$$\hat{\bar{Y}}_{HT} = N^{-1}\sum_s Y_i/\pi_i. \qquad (2)$$

Särndal, Swenson, and Wretman (1992, p. 182, referred to below as SSW) give several reasons for regarding the Hájek as "usually the better estimator", namely the relative behavior of $\hat{\bar{Y}}_{Haj}$ and $\hat{\bar{Y}}_{HT}$ when (a) the Y_i are relatively homogeneous, or (b) sample size is not fixed, or (c) the π_i are weakly or negatively correlated with the Y_i. The Hájek estimator can be derived from the theory of optimal estimating equations if we regard $\bar{Y}$ as the "induced finite population parameter" under the superpopulation model $Y_i \sim (\mu, \sigma^2)$ with the Y_i's independent (Godambe and Thompson (1986), *Example 1*).

Our present purpose is to examine the Hájek estimator, and the Horvitz-Thompson (HT) estimator as well, in the light of recent results using the "model-based" or *prediction* approach to survey sampling. In particular, we investigate consequences of a theorem connecting optimality and weighted balanced samples (Royall 1992, Theorem 2), stated below.

2. Background: Simple Balanced Samples

We continue to consider the problem of estimating the population mean $\bar{Y} = N^{-1}\sum_{i=1}^{N} Y_i$. Assume that a single auxiliary x_i is associated with and known for each unit i in the population, and that Y_i and x_i are related by the polynomial model

$$Y_i = \sum_{j=0}^{J} \delta_j \beta_j x_i^j + \varepsilon_i v_i^{1/2} \qquad (3)$$

where the errors are $\varepsilon_i \sim (0, \sigma^2)$ and uncorrelated, $\{\beta_j\}_{j=1}^{J}$ are a set of unknown parameters, and $\{\delta_j\}_{j=1}^{J}$ are 0-1 variables indicating whether the jth power term is in the model or not. Let $M(\delta_0, \delta_1, \ldots, \delta_J : v)$ denote model (3) and $\hat{\bar{Y}}(\delta_0, \delta_1, \ldots, \delta_J : v)$ denote the *BLU* predictor under that model, following the convention in Royall and Herson (1973). For example, $M(0,1:x)$ refers to the model $Y_i = \beta_1 x_i + \varepsilon_i x_i^{1/2}$ and $\hat{\bar{Y}}(0,1:x)$ is readily shown to be the well known ratio estimator $\hat{\bar{Y}}_R = \bar{Y}_s \bar{x}/\bar{x}_s$.

One of the principal concerns of the model-based opus is the question of *robustness*: how well does an estimator perform when the hypothesized model is incorrect, and what measures can be taken to guard against degradation of its performance (in terms, say, of root mean square error) under this almost inevitable circumstance? It is helpful to determine the behavior of an estimator defined in terms of one (usually fairly simple) model—the *working model*—and under another (usually more complicated) *true model*. Note: we will denote expectations with respect to a model by E_M and with respect to a sampling plan by E_π.

Thus, for the ratio estimator, we ask about its bias under (3), i.e., under $M(\delta_0, \delta_1, \ldots, \delta_J : v)$, and find that

$$E_M\left[\hat{\bar{Y}}(0,1:x) - \bar{Y}\right] = \bar{x}\sum_{j=1}^{J} \delta_j \beta_j \left[\frac{\bar{x}_s^{(j)}}{\bar{x}_s} - \frac{\bar{x}^{(j)}}{\bar{x}}\right] \qquad (4)$$

where $\bar{x}_s^{(j)} = \sum_s x_i^j/n$ and $\bar{x}^{(j)} = \sum_{i=1}^{N} x_i^j/N$. Note that there is no contribution from the term j=1, a fact

not surprising, since the model underlying the ratio estimator contains β_1. If

$$\frac{\bar{x}_s^{(j)}}{\bar{x}_s} = \frac{\bar{x}^{(j)}}{\bar{x}}, \quad j = 0,\ldots, J$$

then the ratio estimator $\hat{T}(0,1:x)$ is unbiased under the broad model (3). Let $s(J)$ denote any sample satisfying the condition above, that is, $s(J)$ is any sample for which

$$\bar{x}_s^{(j)} = \bar{x}^{(j)} \tag{5}$$

for $j = 1,\ldots, J$. Such samples are called *balanced samples (of order J)* (Royall and Herson 1973). We shall refer to them as *simple* or *unweighted balanced samples,* in the light of more general results to be discussed in the next section. The main point here is that by *deliberately* selecting one's sample to meet the criterion (5), one protects oneself against model failure, at least of a certain (perhaps not uncommon) form.

Now a somewhat subtle observation is in order. Suppose the sampling scheme was simple random sampling (*srs*). Then both the Hájek and Horvitz-Thompson estimators are the sample mean $\bar{Y}_s$. If the sample chosen is balanced (whether deliberately or through the chance result of our sampling scheme), then it is readily seen that the ratio estimator *also* reduces to the sample mean $\bar{Y}_s$. Furthermore, under *srs,* samples can be seen to be balanced *in expectation,* that is

$$E_\pi\left(\bar{x}_s^{(j)}\right) = n^{-1}\sum_{i=1}^{N} \pi_i x_i^j = \bar{x}^{(j)}, \tag{6}$$

where $E_\pi(\bullet)$ refers to expectation under the random sampling scheme. Thus the combination of *srs* and the Hájek/HT estimator would seem to receive added support from model-based theory.

But is this the case? To judge the relative merits of estimators, we might ask about their *sensitivity to imbalance.* How well does their unbiasedness hold up, if the balance aimed at by randomization is not quite achieved? It is readily shown that the ratio estimator is less sensitive to deviation from balance than the Hájek. For example, suppose the "true model" is $M(1,1,1:v)$, that is a quadratic model with intercept. Suppose $\bar{x}_s = (1+e_1)\bar{x}$ and $\bar{x}_s^{(2)} = (1+e_2)\bar{x}^{(2)}$, where, more often than not the deviations e_i will be of the same sign. Then one readily shows that the model bias of the Hájek/HT is $E_M(\bar{Y}_s - \bar{Y}) = \beta_1 \bar{x} e_1 + \beta_2 \bar{x}^{(2)} e_2$ and the bias of the ratio estimator is $E_M\left(\hat{\bar{Y}}_R - \bar{Y}\right) \approx -\beta_0 e_1 + \beta_2 \bar{x}^{(2)}(e_2 - e_1)$. Typically, the second term of the Hájek bias will be larger in absolute value than the corresponding term of the ratio, because of likely cancellation in the latter. Also, where Y tends to change with x—the typical circumstance in which we would be tempted to use the ratio estimator—it will be unusual for the first term of the ratio-estimator-bias to be as large in absolute value as the first term of the Hájek-bias. Clearly there are dangers in not having strict balance for either, but the Hájek will usually be *more* sensitive to imbalance than the ratio estimator. Thus, if one uses the Hájek, it is desirable to take a deliberately balanced sample, and not rely on *srs.* The frequently made claim that randomization is one's best protection against model failure does not seem well supported when we look at the matter under the lens of alternate models.

3. Weighted Balanced Samples

By a sampling *strategy* we shall mean a combination of choice of sample and of estimator. Under the model $M(0,1:x)$, it is *not* the case that selecting a balanced sample, and using the ratio estimator, is the most efficient procedure. The greater variance of Y_i at larger x_i dictates we sample units with larger x_i more heavily than simple balance will allow. The question arises whether there is a procedure that is bias-robust and most efficient under the working model.

We proceed with some degree of generality. Consider the general linear model with a *diagonal* covariance matrix:

$$E_M(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}, \quad \mathrm{var}_M(\mathbf{Y}) = \mathbf{V}\sigma^2 \tag{7}$$

which will be referred to as $M(\mathbf{X}:\mathbf{V})$, where $\boldsymbol{\beta}$ is $p\times 1$, $\mathbf{X}$ is $N\times p$, and $\mathbf{V}$ is $N\times N$. The matrix of auxiliaries can be partitioned between the sample and non-sample units as $\mathbf{X} = (\mathbf{X}_s', \mathbf{X}_r')'$. The *BLU* predictor under this model is $\hat{\bar{Y}}(\mathbf{X}:\mathbf{V}) = N^{-1}\left(\mathbf{1}_s'\mathbf{Y}_s + \mathbf{1}_r'\mathbf{X}_r\hat{\boldsymbol{\beta}}\right)$ where $\mathbf{1}_s$ and $\mathbf{1}_r$ are vectors of n 1's and N-n 1's, and $\hat{\boldsymbol{\beta}} = \mathbf{A}_s^{-1}\mathbf{X}_s'\mathbf{V}_{ss}^{-1}\mathbf{Y}_s$ with $\mathbf{V}_{ss}$ the $n\times n$ diagonal covariance matrix for the sample units, and $\mathbf{A}_s = \mathbf{X}_s'\mathbf{V}_{ss}^{-1}\mathbf{X}_s$. Let $\mathcal{M}(\mathbf{X})$ denote the linear manifold generated by the columns of $\mathbf{X}$, i.e. the vector space spanned by all linear combinations of the columns of $\mathbf{X}$. We will also need 1_N, an N-vector of 1's.

The collection of samples that satisfy

$$\frac{1}{n}\mathbf{1}_s'\mathbf{W}_s^{-1/2}\mathbf{X}_s = \frac{\mathbf{1}_N'\mathbf{X}}{\mathbf{1}_N'\mathbf{W}^{1/2}\mathbf{1}_N} \tag{8}$$

will be denoted $B(\mathbf{X}:\mathbf{W})$ and is said to be *balanced with respect to the weights root*(**W**) or to be *root*(**W**) *balanced.* Here **W** is an $N \times N$ 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 $\mathbf{V} = \mathbf{W}\sigma^2$.

When $\mathbf{W} = \mathbf{I}$, $B(\mathbf{X}:\mathbf{I})$ is the set of samples that are balanced on the columns of **X**, i.e. $\mathbf{1}_s'\mathbf{X}_s/n = \mathbf{1}_N'\mathbf{X}_N/N$. If the model for Y is a polynomial in x, as in model (3), then $B(\mathbf{X}:\mathbf{I})$ is the set of samples satisfying $\bar{x}_s^{(j)} = \bar{x}^{(j)}$, the simple balance conditions introduced in section 2. Thus, weighted balance contains our previous notion of balance as a special case.

***Theorem* 1** (Royall 1992). Under $M(\mathbf{X}:\mathbf{V})$ if both

$$\mathbf{V}\mathbf{1}_N \text{ and } \mathbf{V}^{1/2}\mathbf{1}_N \in \mathcal{M}(\mathbf{X}), \qquad (9)$$

then $\text{var}_M\left[\hat{\bar{Y}}(\mathbf{X}:\mathbf{V}) - \bar{Y}\right] \geq$

$$N^{-2}\left[n^{-1}\left(\mathbf{1}_N'\mathbf{V}^{1/2}\mathbf{1}_N\right)^2 - \mathbf{1}_N'\mathbf{V}\mathbf{1}_N\right]\sigma^2. \qquad (10)$$

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

$$\hat{\bar{Y}}(\mathbf{X}:\mathbf{V}) = N^{-1}n^{-1}\left(\mathbf{1}_N'\mathbf{V}^{1/2}\mathbf{1}_N\right)\left(\mathbf{1}_s'\mathbf{V}_{ss}^{-1/2}\mathbf{Y}_s\right)$$

$$= \frac{\bar{v}^{(1/2)}}{n}\sum_s \frac{Y_i}{v_i^{1/2}}, \qquad (11)$$

Note that, under the weighted balance condition, neither the estimator itself nor its variance depend explicitly on the **X** matrix. This has important implications. Suppose the columns of **X** are included in those of another matrix $\mathbf{X}^*$. Suppose $M(\mathbf{X}:\mathbf{V})$ is the working model, and $M(\mathbf{X}^*:\mathbf{V})$ the true model, that the conditions (9) on the standard deviations and variances are met, and that balance holds for the wider model, that is, $s \in B(\mathbf{X}^*:\mathbf{V})$; then the estimator based on the working model $M(\mathbf{X}:\mathbf{V})$ will be *BLU* under $M(\mathbf{X}^*:\mathbf{V})$. In other words, the estimator will still be unbiased, and nothing is lost in efficiency.

Example 1. Suppose the working model is the quadratic model $M(1,1,1:x^2)$ with variance proportional to x^2. The condition $\mathbf{V}\mathbf{1}_N$ and $\mathbf{V}^{1/2}\mathbf{1}_N \in \mathcal{M}(\mathbf{X})$ are met since both x and x^2 are in the model for $E_M(Y)$. The lower bound on the variance is

$$\frac{1}{N^2}\left[\frac{(N\bar{x})^2}{n} - \sum_{i=1}^{N} x_i^2\right]\sigma^2. \qquad (12)$$

This bound is achieved in any balanced sample with

$$\bar{x}_s^{(j-1)} = \bar{x}^{(j)}/\bar{x} \qquad (13)$$

for j = 0, 1, and 2; the j = 1 condition is fulfilled automatically, and (13) at j = 0 says that the harmonic mean of sample x's equals their population arithmetic mean. Bias protection against more general polynomial models is obtained at no cost in efficiency under the working model by balancing on additional powers j = 3, 4, ..., J. We refer to such balance as *root*(x^2) balance or just x-balance (of order J). It is also known in the literature as π-balance (Cumberland and Royall 1981). The *BLU* predictor reduces to $\hat{\bar{Y}} = \bar{x}\sum_s y_i/(nx_i)$, the "mean-of-ratios estimator" — a result first derived by Kott (1984). It is also the Horvitz-Thompson estimator for a fixed size sampling design with $\pi_i \propto x_i$. Furthermore, under the balance condition (13) corresponding to j = 0, the *BLU* predictor can also be written in the form of a Hájek estimator $\hat{Y} = \dfrac{\sum_s y_i/x_i}{\sum_s 1/x_i}$. Thus *at* balance, the *BLU* predictor under the model coincides with both the Hájek and Horvitz-Thompson estimators corresponding to *pps* sampling with size variable x, a sampling plan which gives (13) in design-expectation, that is, $E_\pi\left(\bar{x}_s^{(j-1)}\right) = \bar{x}^{(j)}/\bar{x}$. However, with only minor departures from balance, the three estimators begin to diverge, and can behave quite differently, as we show by analysis in Section 4 and by a simulation study in Section 5.

Example 2. Suppose the working model is the through-the-origin model $M(0,1:x)$ with variance proportional to x, which leads to the ratio estimator. The condition $\mathbf{V}^{1/2}\mathbf{1}_N \in \mathcal{M}(\mathbf{X})$ of the Theorem is *not* met. This suggests (Royall 1992) that as an alternative to the ratio estimator (which is bias robust under *un*weighted balance) we take root(x) balance and the estimator corresponding to the minimal model that meets the conditions of the Theorem.

Definition: The *minimal model* for given variance matrix **V** is $M_{\min}(V) = M(\mathbf{X}_V:\mathbf{V})$, where $\mathbf{X}_V = \left(\mathbf{V}^{1/2}\mathbf{1}_N, \mathbf{V}\mathbf{1}_N\right)$.

Given a particular variance structure, this is the smallest model to guarantee that the conditions (9) of the theorem are met.

With variance proportional to x, the minimal model, given by $E_M(Y_i) = \beta_{1/2}x_i^{1/2} + \beta_1 x_i$, can be denoted by $M_{\min}(x) = M(x^{1/2}, x:x)$. The lower bound on the variance is

$$\frac{1}{N^2}\left[\frac{\left(N\bar{x}^{(1/2)}\right)^2}{n}-\sum_{i=1}^{N}x_i\right]\sigma^2. \qquad (14)$$

This bound—which is readily shown to be lower than the variance of the ratio estimator under simple balance—is achieved under $M_{\min}(x)$ in any sample balanced in the sense that $\bar{x}_s^{(1/2)}=\bar{x}/\bar{x}^{(1/2)}$. Bias protection against more general polynomial models is obtained by balancing on additional powers:

$$\bar{x}_s^{(j-1/2)}=\bar{x}^{(j)}/\bar{x}^{(1/2)} \text{ for } j=0,1,2,\ldots,J. \qquad (15)$$

As before, the estimator reduces to the Horvitz-Thompson estimator for *pps* sampling with $x^{1/2}$ as size variable, and also to the Hájek if (15) is met, for $j = 0$.

In general, consider a sampling plan having fixed n, *pps* sampling with size variable $v_i^{1/2}$, that is, the inclusion probability of unit i is

$$\pi_i=nv_i^{1/2}/\mathbf{1}_N'\mathbf{V}^{1/2}\mathbf{1}_N. \qquad (16)$$

We shall also refer to this plan as $pp(v^{1/2})$ sampling. Then

(1) Weighted balance is met in design-expectation, that is $E_\pi\left(\frac{1}{n}\mathbf{1}_s'\mathbf{V}_s^{-1/2}\mathbf{X}_s\right)=\frac{\mathbf{1}_N'\mathbf{X}}{\mathbf{1}_N'\mathbf{V}^{1/2}\mathbf{1}_N}$ for any matrix $\mathbf{X}$ of regressors (even unknown ones.).

(2) The *BLU* predictor $\hat{Y}(\mathbf{X}:\mathbf{V})=N^{-1}n^{-1}\left(\mathbf{1}_N'\mathbf{V}^{1/2}\mathbf{1}_N\right)\left(\mathbf{1}_s'\mathbf{V}_{ss}^{-1/2}\mathbf{Y}_s\right)$ in a root(v) balanced sample is just the Horvitz-Thompson estimator $\hat{\bar{Y}}_{HT}=N^{-1}\sum_s Y_i/\pi_i$. The variance bound is the one established by Godambe and Joshi (1965, Theorem 6.1) for the model-based expectation of the design-based variance of the Horvitz-Thompson estimator.

(3) Under the condition

$$\frac{n}{\mathbf{1}_s'\mathbf{V}_s^{-1/2}\mathbf{1}_s}=\frac{1}{N}\mathbf{1}_N'\mathbf{V}^{1/2}\mathbf{1}_N, \qquad (17)$$

the *BLU* predictor (11) can be written as the Hájek estimator $\hat{\bar{Y}}_{Haj}=\frac{\sum_s Y_i/\pi_i}{\sum_s 1/\pi_i}$.

One method of selecting weighted balanced samples is to use the *pps* sampling plan above, but to reject any sample that is insufficiently balanced on particular moments. Figure 1 depicts a weighted balanced sample from the Hospitals population used in the simulation study described in section 5. The population was randomly ordered and systematic $pp(x^{1/2})$ samples of size $n = 30$ were selected using the Hartley-Rao (1962) method. The sample in the figure is approximately balanced in the sense of (15) for $j =$ 0, 1, and 2.

4. Comparison of Estimators in Unbalanced Samples

For simplicity we focus on the situation in *Example 1*; the basic ideas are very general. The minimal model when variance is proportional to x^2 is $M(0,1,1:x^2)$; the corresponding *BLU* estimator is, for any sample (not just a balanced sample)

$$\hat{\bar{Y}}_{BLU}=\frac{1}{n\left(\bar{x}_s^{(2)}-\left(\bar{x}_s\right)^2\right)}\left\{\left(\bar{x}\bar{x}_s^{(2)}-\bar{x}^{(2)}\bar{x}_s\right)\sum_s\frac{Y_i}{x_i}+\left(\bar{x}^{(2)}-\bar{x}\bar{x}_s\right)\sum_s Y_i\right\}$$

The Hájek and HT estimators under $pp(x)$ sampling are given in *Example 1* above.

Now suppose that sampling yields the following "near-balance" conditions relating sample to population x-moments:

$$\bar{x}_s^{(j-1)}=\frac{\bar{x}^{(j)}}{\bar{x}}\left(1+e_j\right),$$

for $j = 0,1,2,\ldots,J$. The "errors" e_j represent the distance the sample is from balance. Note that (in the present example) $e_1 = 0$. Typically the e_j with $j > 0$ will tend to have the same sign, and be opposite in sign from e_0.

We consider the bias $E_M\left(\hat{\bar{Y}}-\bar{Y}\right)$ under a polynomial model of order J, for the Hájek estimator, Horvitz-Thompson estimator, and minimal model *BLU* predictor. We find

$$\text{Bias}_{Haj}\sim\sum_{j=0}^{J}\beta_j\bar{x}^{(j)}\left(e_j-e_0\right) \qquad (18)$$

$$\text{Bias}_{HT}=\sum_{j=0}^{J}\beta_j\bar{x}^{(j)}e_j \qquad (19)$$

$$\text{Bias}_{BLU}\sim$$

$$\sum_{j=0}^{J}\beta_j\left\{\bar{x}^{(j)}e_j-\frac{\bar{x}^{(j+1)}-\left(\bar{x}^{(2)}\right)\bar{x}^{(j)}/\bar{x}}{\bar{x}^{(3)}-\left(\bar{x}^{(2)}\right)^2/\bar{x}}\bar{x}^{(2)}e_2\right\}. \qquad (20)$$

Notice that the multipliers of β_0 in (18), of β_1 in (19), and of β_1 and β_2 in (20) are all 0. Thus, if all coefficients except the intercept are 0, the Hájek can be expected to be least affected by being away from balance. This harmonizes with the first reason given by SSW for preferring it (see above). Where there exists a continuous non-constant dependency of Y on x, however, the HT will be less biased than the Hájek, because of the likely opposition of sign of e_0 and e_j with $j > 0$. Examination of the second expression in (20) suggests that in each term, both for $j = 0$, and $j >$ 2, some cancellation will take place, so that the minimal estimator will invariably be less biased than

the Horvitz-Thompson. These conjectures are borne out in a small simulation study, in the next section.

5. Simulation Study

We compared the Hájek, Horvitz-Thompson, and minimal model *BLU* estimators in a simulation study using three populations. Two of the populations, Hospitals and Cancer, are well-known in the survey literature (Royall and Cumberland 1981). Scatterplots reveal a strong relationship between y and x in both of these populations. The third was generated to be favorable for the Hájek estimator using a model with a common mean: $Y_i = \mu + \varepsilon_i x_i$ with $\varepsilon_i \sim (0,1)$, $\mu = 200$, and the x's coming from the Hospitals population. Two sets of 2,000 samples were selected from each population—one with probabilities proportional to $x^{1/2}$ and the other with probabilities proportional to x. In both cases, we used n=30 and the random-order/systematic-sampling method studied by Hartley and Rao (1962).

Table 1 shows the root mean square errors (*rmse*'s) for $\hat{\bar{Y}}_{Haj}$, $\hat{\bar{Y}}_{HT}$, and two model-based estimators. Each *rmse* is computed as $rmse\left(\hat{\bar{Y}}\right) = \sqrt{\sum_{s=1}^{S}\left(\hat{\bar{Y}}_s - \bar{Y}\right)^2 / N}$ where S=2,000 and $\hat{\bar{Y}}$ is one of the estimates from sample s. For a model with $\mathrm{var}_M(Y) \propto x^\gamma$, the minimal model is $M_{\min}(x^{\gamma/2}, x^\gamma : x^\gamma)$ with *BLU* predictor $\hat{\bar{Y}}_{\min}(x^{\gamma/2}, x^\gamma : x^\gamma)$. The second model-based estimator was constructed by adding an intercept to the minimal model: $\hat{\bar{Y}}_{1,\min}(1, x^{\gamma/2}, x^\gamma : x^\gamma)$ which is *BLU* under $M(1, x^{\gamma/2}, x^\gamma : x^\gamma)$. When the model has $\mathrm{var}_M(Y) \propto x^\gamma$, $pp(x^{\gamma/2})$ sampling produces, in design-expectation, the type of weighted balance required for optimality in Theorem 1, as noted in section 3.

As Table 1 shows, for the Hospitals and Cancer populations, the Hájek and HT estimators are far worse than either of the model-based estimators in $pp(x^{1/2})$ sampling. In contrast, the Hájek is unbiased under the model used to generate the Artificial population and has the smallest *rmse* there. This finding verifies the analysis of the preceding section regarding the SSW observation that the Hájek will perform well when the Y's are relatively homogeneous. The minimal estimator $\hat{\bar{Y}}_{\min}(x^{\gamma/2}, x^\gamma : x^\gamma)$ has the smallest *rmse* in both types of sampling for Hospitals and Cancer but is substantially worse in the Artificial population. Its poor showing stems from the fact that, except at balance, $\hat{\bar{Y}}_{\min}(x^{\gamma/2}, x^\gamma : x^\gamma)$ is biased under a model where Y has a common mean. Note, however, that $\hat{\bar{Y}}_{1,\min}(1, x^{\gamma/2}, x^\gamma : x^\gamma)$, which adds an intercept, has *rmse*'s much nearer those of the Hájek in the Artificial population, even though $\hat{\bar{Y}}_{1,\min}(1, x^{\gamma/2}, x^\gamma : x^\gamma)$ involves estimates of superfluous parameters for $x^{\gamma/2}$ and x^γ. The HT estimator does well in $pp(x)$ samples in the first two populations but poorly otherwise, and, generally, shows itself to be a risky procedure.

Table 1. Empirical root mean square errors of four estimators of the population mean in 2,000 $pp(x^{\gamma/2})$ samples selected from three populations.

Population	$\gamma = 1$	$\gamma = 2$
Hospitals		
$\hat{\bar{Y}}_{Haj}$	116.3	171.9
$\hat{\bar{Y}}_{HT}$	51.0	36.9
$\hat{\bar{Y}}_{\min}(x^{\gamma/2}, x^\gamma : x^\gamma)$	37.2	34.7
$\hat{\bar{Y}}_{1,\min}(1, x^{\gamma/2}, x^\gamma : x^\gamma)$	39.2	40.0
Cancer		
$\hat{\bar{Y}}_{Haj}$	7.2	9.6
$\hat{\bar{Y}}_{HT}$	3.7	1.7
$\hat{\bar{Y}}_{\min}(x^{\gamma/2}, x^\gamma : x^\gamma)$	1.7	1.7
$\hat{\bar{Y}}_{1,\min}(1, x^{\gamma/2}, x^\gamma : x^\gamma)$	1.8	1.9
Artificial		
$\hat{\bar{Y}}_{Haj}$	2.5	2.8
$\hat{\bar{Y}}_{HT}$	18.5	44.6
$\hat{\bar{Y}}_{\min}(x^{\gamma/2}, x^\gamma : x^\gamma)$	9.9	36.8
$\hat{\bar{Y}}_{1,\min}(1, x^{\gamma/2}, x^\gamma : x^\gamma)$	2.7	3.5

The source of the differences in the estimators is clarified by conditional analyses. We sorted each set of 2,000 samples by the sample mean $\bar{x}_s^{(\gamma/2)}$, which has design-expectation $\bar{x}^{(\gamma)} / \bar{x}^{(\gamma/2)}$ in $pp(x^{\gamma/2})$ sampling. The samples were broken into 10 groups of 200 samples each and the bias and *rmse* calculated in each group for each estimator of the mean. The bias and *rmse* were plotted against the group average of $\bar{x}_s^{(\gamma/2)}$ for the three populations, for both $pp(x^{1/2})$ samples and $pp(x)$ samples, producing figures of the sort laid out in Royall and Cumberland (1981). Because of space limitations of these *Proceedings,* the

figures are not reproduced here; a fuller version of this paper which includes the figures, may be found on the Web at *http://stats.bls.gov*. In Hospitals and Cancer, the Hájek has an egregious bias that runs from negative to positive over the range of $\bar{x}_s^{(\gamma/2)}$, which translates to the large, unconditional *rmse*'s in Table 1. Only at or near weighted balance is the Hájek estimator unbiased, but, as observed in section 4, minor departures from balance lead to major biases. The Horvitz-Thompson estimator also has a substantial, systematic bias in Hospitals for both methods of sampling and in Cancer for $pp(x^{\gamma/2})$ sampling, though the bias is smaller than that of the Hájek. The minimal estimator has uniformly small bias and *rmse* throughout the range of $\bar{x}_s^{(\gamma/2)}$ in the first two populations.

In the Artificial population, on the other hand, the Hájek does well in all groups of samples as expected. The HT estimator is conditionally quite biased since it makes no allowance for an intercept. The minimal estimator is also conditionally biased in extreme samples though much less so than the Horvitz-Thompson. Examination of individual samples shows that $\hat{\bar{Y}}_{\min}(x^{\gamma/2}, x^{\gamma}:x^{\gamma})$ fits a nonsensical, inverted ∪-shaped curve to data that follow a horizontal straight-line model, and, thus, is a poor choice. Addition of an intercept term in $\hat{\bar{Y}}_{1,\min}(1, x^{\gamma/2}, x^{\gamma}:x^{\gamma})$ largely eliminates this problem. (The curves for $\hat{\bar{Y}}_{1,\min}(1, x^{\gamma/2}, x^{\gamma}:x^{\gamma})$ are not shown in Figure 4 to simplify the plots.) It is well to keep in mind, however, that *all* of these estimators are the same at strict balance on the appropriate moments.

6. Conclusions

The Hájek estimator (and to a lesser extent, the Horvitz-Thompson estimator) which corresponds to a given sampling plan has been shown analytically and empirically to be bias sensitive to deviations of the sample selected from weighted balance, even though the sampling plan achieves weighted balance *in expectation.* The notable exception is when the variable of interest is unrelated to the auxiliary variable.

It is preferable to use the corresponding minimal estimator. Restricting the sample to be in the class of weighted balanced samples is best. In unbalanced samples, the BLU estimator corresponding to the minimal model augmented by an intercept can be efficient both for *Y* variables related to the auxiliary and for *Y*'s having a common mean.

Acknowledgments

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

References

Basu, D. (1971), "An Essay on the Logical Foundations of Survey Sampling, Part One" in *The Foundations of Survey Sampling,* Godambe, V.P. and Sprott, D.A. eds., 203-233, Holt, Rinehart, and Winston, Toronto.

Cumberland, W.G. and Royall, R.M. (1981), "Prediction Models and Unequal Probability sampling," *Journal of the Royal Statistical Society, Series B,* **43**, 353-367.

Godambe, V.P. and Joshi, V.M. (1965), "Admissibility and Bayes Estimation in Sampling Finite Populations," *Annals of Mathematical Statistics*, **36**, 1707-1722.

Godambe, V.P. and Thompson, M.E. (1986), "Parameters of superpopulation and survey population: Their Relationships and Estimation ," *International Statistical review*, 54, 127-138.

Hájek, J. (1971) Comment on "An Essay on the Logical Foundations of Survey Sampling, Part One," in *The Foundations of Survey Sampling,* Godambe, V.P. and Sprott, D.A. eds., 236, Holt, Rinehart, and Winston.

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

Horvitz, D. G. and Thompson, D.J. (1952) "A Generalization of Sampling without Replacement from a Finite Universe," *Journal of the American Statistical Association*, **47**, 663-685.

Kott, P.S. (1984), "A Fresh Look at Bias-robust Estimation in a Finite Population," *Proceedings of the Section on Survey Methods Research,* American Statistical Association, 176- 178.

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 I," *Journal of the American Statistical Association*, **68**, 880-889.

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