ALTERNATIVE ESTIMATORS IN PPS SAMPLING

M.P. Singh

Some estimators alternative to the usual PPS estimator are suggested in this paper for situations where the size measure used for PPS sampling is not correlated with the study variable and where data are available on another supplementary variable (size measure). Properties of these estimators are studied under super-population models and also empirically.

1. INTRODUCTION

It is well known that selection with probability proportional to size (PPS) generally improves the efficiency of the estimate of the population total for the characteristic under study provided the auxiliary variable (x) used as size measure is highly positively correlated with the study variable. Usually, therefore, in large scale multipurpose surveys where data are collected on several characteristics on a continuous basis, PPS sampling is used. The size measure (x) chosen for PPS selection in such surveys is such that it is highly correlated with the most important variable(s) under study at the time of designing the survey. However, as the time passes the initial size measure used to determine the initial selection probabilities becomes more and more out of date resulting in loss of correlation and hence the loss in efficiency of the survey estimates. In order to prevent such decline in efficiency quite often more up to date data on new size measure (z) are collected. Such data may be used either for reselection (updating) of the sample or for improving the estimation procedure. Use of new size measures in updating the sample has been discussed earlier for different sampling schemes by Keyfitz [4], Fellegi [3], Kish and Scott [5], Platek and Singh [6] and

1 M.P. Singh, Household Surveys Development Division, Statistics Canada.
Drew, Choudhry and Gray [2]. In this paper, data on new size measures have been used at the estimation stage and the properties of the estimators which were introduced earlier by Singh [8] are studied.

Such estimators may also be used in the context of multi-purpose survey for those characteristics (y) that are not correlated with the size measure chosen for PPS sampling. Rao [7] has suggested an estimator alternative to the usual PPS estimator for such situations. The estimators suggested in this paper are compared with Rao's estimator and the usual PPS with replacement estimator under super-population models followed by an empirical study.

2. ALTERNATIVE ESTIMATOR

For a sample of size n selected with replacement with PPS of x, the usual unbiased estimator of the total \( \sum y_i \) is

\[
\hat{Y} = \frac{1}{n} \sum_{i=1}^{n} x_i y_i
\]

with variance

\[
V_p = \frac{1}{n} \sum_{i=1}^{N} \frac{N}{N} \left( \frac{y_i^2}{p_i} - \frac{Y^2}{n} \right)
\]

where \( p_i = x_i/X \) and \( N \) is the number of units in the population, \( X=\sum x_i \).

An unbiased estimator of \( Y \) in equal probability sampling (SRS) is

\[
\hat{Y}_s = \frac{N}{n} \sum_{i=1}^{n} y_i
\]

with variance

\[
V_s = \frac{1}{n} \left( N \sum_{i=1}^{N} \frac{y_i^2}{p_i} - \frac{Y^2}{n} \right)
\]
If y is uncorrelated with x then $V_s$ would be smaller than $V_p$ (Cochran, [1]). On this consideration, Rao [7] suggested an estimator alternative to $Y_p$ for situations where y and x are unrelated even if the sample is selected with PPS. Rao's estimator entails 'undoing' of the PPS weights and is obtained by replacing $x_i$ by 1 in the expression for $Y_p$. Thus Rao's estimator is

$$\hat{Y}_o = \frac{N}{n} \sum_{i=1}^{n} y_i$$  \hspace{1cm}(2.5)$$

and has variance

$$V_o = \frac{N^2}{n^2} \left[ \sum_{i=1}^{n} (y_i^2 p_i) - (\sum_{i=1}^{n} y_i p_i)^2 \right].$$  \hspace{1cm}(2.6)$$

Note that although $\hat{Y}_p$ and $\hat{Y}_o$ have the same form, their variances $V_s$ and $V_o$ are different due to difference in selection procedures.

Using the same reasoning, that is, whenever y and x are highly positively correlated substantial gains are achieved in using $Y_p$ with PPS in contrast to $Y_s$ with SRS, we consider an alternative estimator

$$\hat{Y}_p' = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{p_i}$$  \hspace{1cm}(2.7)$$

where

$$p_i = \frac{z_i}{Z}, \quad Z = \sum_{i=1}^{N} z_i.$$ 

Note that this estimator assumes the knowledge of an additional size measure z which is highly positively correlated with y.

The estimator $\hat{Y}_p$, like $\hat{Y}_o$, is biased and their biases respectively are

$$B_p = \frac{N}{n} \sum_{i=1}^{n} \frac{y_i}{p_i} \left( \frac{1}{p_i} - 1 \right)$$  \hspace{1cm}(2.8)$$
and

$$B_0 = \sum_1^N y_i (Np_i - 1).$$  \hspace{1cm} (2.9)

Variance of $$\hat{Y}_{p_i}$$ is obtained by simply replacing $$y_i$$ in (2.2) by $$y_i p_i / p_i$$. Thus

$$\text{Var}(\hat{Y}_{p_i}) = \frac{1}{n} \sum_1^N \left[ \frac{y_i^2 p_i}{p_i^2} - \left( \frac{N y_i p_i}{p_i} \right)^2 \right].$$  \hspace{1cm} (2.10)

In the following section we compare these estimators under super-population models and then two other estimators are suggested in section 4 for similar situations and compared among themselves.

3. COMPARISON UNDER SUPER-POPULATION MODEL

The super-population model $$\Lambda_1$$ often used when $$y$$ is highly positively correlated with $$z$$ is

$$y_i = z_i + \eta_i \quad \text{for } i = 1, 2, \ldots, N$$  \hspace{1cm} (3.1)

where

$$\varepsilon_1 (\eta_i | z_i) = 0, \quad \varepsilon_1 (\eta_i^2 | z_i) = a z_i^g,$$

and

$$\varepsilon_1 (\eta_i \eta_j | z_i, z_j) = 0, \quad a > 0, \quad g > 0.$$

The symbol $$\varepsilon_1$$ denotes the average overall finite populations that can be drawn from the super-population.
Under this model $\Delta_1$

$$\varepsilon_1(B_{p_i}) = \beta \sum z_i \frac{p_i}{r} (\frac{r}{p_i} - 1)$$

$$= 0 \text{ for any } p_i, \text{ by substituting } z_i = p_i Z.$$ 

Thus, $\hat{Y}_{p,1}$ is unbiased under the model. However, in general, $\hat{Y}_{p,1}$, like $\hat{Y}_o$, is biased and the bias does not depend on the sample size. Thus, neither estimator is consistent.

The expected variances of $\hat{Y}_o$ and $\hat{Y}_{p,1}$ under the model $\Delta_1$ are

$$\varepsilon_1(V_o) = V_o^* = \beta^2 V(Z_o) + \frac{N^2 a z^g}{n} \sum p_i g p_i (1-p_i)$$

and

$$\varepsilon_1(V_{p,1}) = V_{p,1}^* = \frac{a z^g}{n} \sum p_i g p_i (1-p_i)$$

where

$$\hat{Z}_o = \frac{N}{n} \sum \frac{N}{1} z_i$$

and

$$V(Z_o) = \frac{N^2}{n} \left[ \sum z_i^2 p_i - \left( \sum \frac{N}{1} z_i p_i \right)^2 \right].$$

Further, in developing the estimators $\hat{Y}_o$ and $\hat{Y}_{p,1}$, the underlying assumption is that $y$ and $x$ are unrelated. The super-population model $\Delta$ often used for this situation (Rao, [7]) for comparison of estimators is $y_i = u + e_i$, where $\varepsilon(e_i|x_i) = 0$, $\varepsilon(e_i^2|x_i) = b$, $b > 0$ and $\varepsilon(e_i e_j|x_i, x_j) = 0$ and $\varepsilon$ is defined for $\Delta$ like $\varepsilon_1$. Since Rao [7] has shown that $\varepsilon(V_o) < \varepsilon(V_{p,1})$, it is enough to compare the average variances of $\hat{Y}_o$ and $\hat{Y}_{p,1}$, under the model $\Delta_1$. In order to facilitate this comparison, we shall use the following model $\Delta_2$ for the characteristic $x$, similar to the model $\Delta$ for $y$. 
\[ \Delta_2: \quad p_i = m + e_i^i \quad i = 1, 2, \ldots, N \]

\[ \varepsilon_2(e_i^i|z_i) = 0, \quad \varepsilon_2(e_i^i e_j^j|z_i) = a', \quad a' > 0 \]  

\[ \varepsilon_2(e_i^i e_j^j|z_i, z_j) = 0 \]

where \( \varepsilon_2 \) is defined as \( \varepsilon_1 \).

Thus, the expected variances are

\[ \varepsilon_2(V_o^*) = V_o^{**} = \beta^2 \varepsilon_2 V(\hat{Z}_o) + \frac{N^2}{n} \frac{aZ^g}{n} (m - m^2 - a') \sum_{i} p_i^g \] and

\[ \varepsilon_2(V_{p_i}^*) = V_{p_i}^{**} = \frac{aZ^g}{n} (m - m^2 - a') \sum_{i} p_i^g \]  

Therefore from (3.7) and (3.8)

\[ V_o^{**} - V_{p_i}^{**} = \beta^2 \varepsilon_2 V(\hat{Z}_o) + \frac{N^2}{n} \frac{aZ^g}{n} (m - m^2 - a') \sum_{i} p_i^g (N^2 - \frac{1}{p_i^2}) \]  

Since \( \varepsilon_2(p_i) = m, \quad \varepsilon_2(p_i^2) = m^2 + a' \)

and \( \varepsilon_2(p_i) > \varepsilon_2(p_i^2) \) because \( p_i^2 > p_i \) for all possible values except 1 or 0,

we have that \( (m - m^2 - a') \geq 0 \).

Also, \( \sum_{i} p_i^2 \geq \frac{1}{N} \), with equality with all \( p_i^i = 1/N \).

For \( g = 2 \), the second term in (3.9) becomes \( \frac{aZ^2}{n} (m - m^2 - a') \sum_{i} (N^2 p_i^2 - 1) \geq 0 \) because of the inequalities in the expressions (3.11) and (3.12). Therefore, for \( g = 2 \), in the model \( \triangle \) in (3.1), the suggested estimator \( \hat{Y}_{p_i} \) performs better than Rao's estimator \( \hat{Y}_o \).
The conditions for the choice of $\hat{Y}_p$, over $\hat{Y}_o$ for other values of $g$ are quite complex to interpret in practice. However, as seen from the empirical study, considerable gains would be achieved in using the suggested estimator for situations where $y$ and $z$ are highly correlated and the coefficient of variation for $x$ is relatively higher than that of $z$.

4. RATIO ESTIMATION

Two estimators of $z$, namely $\hat{Z}_p$ and $\hat{Z}_o$, similar to $\hat{Y}_p$ in (2.1) and $\hat{Y}_o$ in (2.5) may be obtained using data on the new size measure $z$. These estimators are used to construct ratio estimators $\hat{Y}_{RP}$ and $\hat{Y}_{RO}$ for PPS with replacement sampling. Thus $\hat{Y}_{RP}$ is

$$\hat{Y}_{RP} = \frac{\hat{Y}_p}{\hat{Z}_p}$$

where $\hat{Y}_p$ is defined in (2.1) and $\hat{Z}_p = n^{-1} \sum_{i=1}^{n} \frac{z_i}{p_i}, \ p_i = \frac{x_i}{x}$.  

$\hat{Y}_{RP}$ has usual ratio estimation bias and variance which are approximated by

$$B_{RP} = Z^{-1} \left[ RV(\hat{Z}_p) - R \text{Cov}(\hat{Y}_p, \hat{Z}_p) \right]$$

and

$$V_{RP} = V(\hat{Y}_p) + R^2 V(\hat{Z}_p) - 2R \text{Cov}(\hat{Y}_p, \hat{Z}_p)$$

where $R = Y/Z$, $V(\hat{Y}_p) = V_p$ in (2.2),

$$V(\hat{Z}_p) = \frac{1}{n} \sum_{i=1}^{N} \frac{z_i^2}{p_i} - \frac{Z^2}{n},$$

and

$$\text{Cov}(\hat{Y}_p, \hat{Z}_p) = \frac{1}{n} \sum_{i=1}^{N} \frac{y_i z_i}{p_i} - \frac{YZ}{n}.$$
It is of interest to note that $B_{p'}$ in (2.8) for PPS with replacement sampling may be approximated by (see Appendix)

$$B_{p'} = Z^{-1} \left[ R \left( V \left( \frac{Y_i}{P_i} \right) - \text{Cov} \left( \frac{Y_i}{P_i}, \frac{Z_i}{P_i} \right) \right) \right] = n B_{RP}$$

(4.4)

and $V_{p'}$ in (2.10) may be approximated by $V_{RP}$ in (4.3). Therefore, $\hat{Y}_{RP}$ may be preferred over $\hat{Y}_{p'}$ on account of having less bias.

An alternative ratio estimator for situations when $y$ and $x$ are unrelated is

$$\hat{Y}_{RO} = \frac{\hat{Y}_O}{\hat{Z}_O} Z$$

(4.5)

where $\hat{Y}_O$ and $\hat{Z}_O$ are as defined in (2.5) and (3.4) respectively.

$\hat{Y}_{RO}$ like $\hat{Y}_{RP}$ is biased but it will contain additional terms in the bias due to the fact that $\hat{Y}_O$ and $\hat{Z}_O$ themselves are biased estimates of $Y$ and $Z$ respectively. Approximate bias and variance of $\hat{Y}_{RO}$ may be written as (see Appendix)

$$B(\hat{Y}_{RO}) = B_0 + R B_0^* + Z^{-1}(R B_0^* B_0^* - B_0 B_0^*) + Z^{-1} \left[ R \left( V \left( \hat{Z}_O \right) - \text{Cov}(\hat{Y}_O, \hat{Z}_O) \right) \right]$$

(4.6)

and

$$V(\hat{Y}_{RO}) = V(\hat{Y}_O) + R^2 V(\hat{Z}_O) - 2R \text{Cov}(\hat{Y}_O, \hat{Z}_O),$$

(4.7)

where $B_0$, $V(\hat{Y}_O)$, $V(\hat{Z}_O)$ are as defined in (2.9) and (2.6) and (3.5) respectively.

Further,

$$B_0^* = B(\hat{Z}_O) = \sum_{i=1}^{N} z_i (Np_i - 1)$$

(4.8)

$$\text{Cov}(\hat{Y}_O, \hat{Z}_O) = \frac{n}{N} \left[ \sum_{i=1}^{N} y_i z_i p_i - \left( \sum_{i=1}^{N} y_i p_i \right) \left( \sum_{i=1}^{N} z_i p_i \right) \right].$$

(4.9)
For comparing $\hat{Y}_{RP}$ and $\hat{Y}_{RO}$, we obtain their expected variances under the model $\Delta_1$ (i.e., assuming that $y$ and $z$ are highly correlated).

We find that

$$\varepsilon_1 \text{Cov}(\hat{Y}_p, \hat{Z}_p) = \beta \left( \frac{1}{n} \sum_{i=1}^{N} \frac{z_i^2}{p_i} - \frac{Z^2}{n} \right)$$

$$= \beta V(\hat{Z}_p)$$ (4.10)

and

$$\varepsilon_1 \text{Cov}(\hat{Y}_0, \hat{Z}_0) = \beta \frac{N^2}{n} \left[ \sum_{i=1}^{N} z_i^2 p_i - (\sum z_i p_i)^2 \right]$$

$$= \beta V(\hat{Z}_0).$$ (4.11)

Both (4.10) and (4.11) are obtained by substituting (3.1) and noting that $\text{E} n_i | i = 0$.

Thus, from (4.3) and (4.7), we have under model $\Delta_1$

$$\varepsilon_1 V(\hat{Y}_{RP}) = V(\hat{Y}_p) + V(\hat{Z}_p)(R^2 - 2R\beta)$$ (4.12)

and

$$\varepsilon_1 V(\hat{Y}_{RO}) = V(\hat{Y}_0) + V(\hat{Z}_0)(R^2 - 2R\beta)$$ (4.13)

Further, if $\beta = R$ and

$$V(\hat{Y}_p) = V(\hat{Z}_p), \quad V(\hat{Y}_0) = V(\hat{Z}_0)$$ (4.14)

then,

$$\varepsilon_1 V(\hat{Y}_{RP}) = V(\hat{Y}_p)(1 - R^2)$$ (4.15)

and

$$\varepsilon_1 V(\hat{Y}_{RO}) = V(\hat{Y}_0)(1 - R^2),$$ (4.16)
which shows that under the condition (4.14)

\[ \epsilon_1 V(\hat{Y}_{RO}) < \epsilon_1 V(\hat{Y}_{RP}) \]

since \( V(\hat{Y}_0) < V(\hat{Y}_p) \) under the model \( \Delta \) (Rao [7]).

However, in general, that is if (4.14) is not satisfied, then,

\[ \epsilon_1 V(\hat{Y}_{RO}) \geq \epsilon_1 V(\hat{Y}_{RP}) \]

depending on

\[ \{V(\hat{Y}_p) - V(\hat{Y}_0)\} + (R^2 - 2R \beta) V(Z_p) - V(Z_0) \geq 0. \]

Note that this comparison does not depend on the value of \( g \).

Further, from (4.16), it is observed that \( \hat{Y}_{RO} \) is more efficient than \( \hat{Y}_0 \) under usual conditions of ratio estimation. As both \( \hat{Y}_{RO} \) and \( \hat{Y}_{RP} \) are biased, the choice between them may be made on the basis of their biases as well. These estimators may be made unbiased or almost unbiased following usual techniques of bias reduction. In the following section, examples are given in which efficiency of \( \hat{Y}_0 \) and \( \hat{Y}_p \) are compared.

5. NUMERICAL EXAMPLES

We have constructed 5 sets of data using two digit random numbers and each set is treated as a stratum. In each stratum, \( N = 20 \) random numbers are first drawn (designated as \( x \)) and then independently another 20 numbers are drawn (designated as \( y \)) so that \( y \) and \( x \) are unrelated. Further, the corresponding values of \( z \) are obtained by selecting 20 single digit random numbers and adding them to the numbers designated as \( y \) in order that \( y \) and \( z \) are highly correlated. Relative efficiencies of the estimates \( \hat{Y}_0 \) and \( \hat{Y}_p \) are defined as:
\[ e_{p0} = \frac{V_p}{\text{mse}(\hat{Y}_0)}, \quad e_{pp'} = \frac{V_p}{\text{mse}(\hat{Y}_{p'})} \]

and

\[ e_{0p'} = \frac{\text{mse}(\hat{Y}_0)}{\text{mse}(\hat{Y}_{p'})}. \]

The following Table gives the bias and the relative efficiency of the estimators. The correlation coefficients (\(\delta yx, \delta yz\) and \(\delta xz\)) and the coefficient of variations \(C_x, C_y\) and \(C_z\) are also given. The sample size in each stratum is assumed to be 2.

**Table: Relative Bias and Efficiency of Alternative Estimators**

<table>
<thead>
<tr>
<th>Stratum</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\delta yx)</td>
<td>0.092</td>
<td>0.007</td>
<td>0.012</td>
<td>0.069</td>
<td>0.070</td>
</tr>
<tr>
<td>(\delta yz)</td>
<td>0.998</td>
<td>0.995</td>
<td>0.997</td>
<td>0.998</td>
<td>0.998</td>
</tr>
<tr>
<td>(\delta xz)</td>
<td>0.099</td>
<td>-0.031</td>
<td>0.008</td>
<td>0.074</td>
<td>0.069</td>
</tr>
<tr>
<td>(C_x)</td>
<td>61</td>
<td>72</td>
<td>60</td>
<td>58</td>
<td>84</td>
</tr>
<tr>
<td>(C_y)</td>
<td>60</td>
<td>39</td>
<td>40</td>
<td>65</td>
<td>51</td>
</tr>
<tr>
<td>(C_z)</td>
<td>55</td>
<td>36</td>
<td>38</td>
<td>60</td>
<td>49</td>
</tr>
<tr>
<td>(\Sigma y_i)</td>
<td>1,034</td>
<td>1,160</td>
<td>1,178</td>
<td>983</td>
<td>1,063</td>
</tr>
<tr>
<td>(B_0)</td>
<td>35.4</td>
<td>2.2</td>
<td>3.5</td>
<td>25.5</td>
<td>32.3</td>
</tr>
<tr>
<td>(B_{p'})</td>
<td>-44.8</td>
<td>6.3</td>
<td>-6.5</td>
<td>-103.7</td>
<td>-19.1</td>
</tr>
<tr>
<td>(e_{p0})</td>
<td>765</td>
<td>3,446</td>
<td>1,184</td>
<td>342</td>
<td>2,530</td>
</tr>
<tr>
<td>(e_{0p'})</td>
<td>1,194</td>
<td>4,581</td>
<td>8,732</td>
<td>455</td>
<td>5,824</td>
</tr>
<tr>
<td>(e_{pp'})</td>
<td>9,137</td>
<td>157,895</td>
<td>103,480</td>
<td>1,572</td>
<td>147,393</td>
</tr>
</tbody>
</table>

Although Rao's estimate is highly efficient compared to the usual PPS estimator \((e_{p0})\), substantial gains are further achieved by utilizing information on \(z\) in the suggested estimator \((e_{0p'})\) for all the 5 strata.
The correlation patterns in the 5 strata are the same, that is, $\delta_{yx}$ and $\delta_{xz}$ are around zero and $\delta_{yz}$ is around 0.99 but stratum 2, 3 and 5 show considerably higher gains than those in stratum 1 and 4. This may be explained by the relative magnitude of coefficient of variation in these strata. In strata 1 and 4, $C_x$, $C_y$ and $C_z$ are approximately equal, but for strata 2, 3 and 5, we have $C_y = C_z = C_{x/2}$, which implies that the alternative estimators will perform much better if the model is satisfied and in addition if $C_x$ is relatively higher than $C_y$ and $C_z$. Bias in both the estimators seems to be usually small relative to the population total being estimated.

ACKNOWLEDGEMENT

The author wishes to thank the referee and Dr. M. Hidiroglou for their comments.

RESUME

On suggère dans cet article que certains estimateurs pourraient remplacer l'estimateur habituel basé sur l'échantillonnage avec probabilité proportionnelle à la taille dans le cas où la mesure de taille utilisée dans l'échantillonnage avec probabilité proportionnelle à la taille n'est pas corrlée avec la variable étudiée et où l'on dispose de données sur une autre variable supplémentaire (mesure de taille). On étudie les propriétés de ces estimateurs dans le contexte des modèles basés sur une population infinie, ainsi qu'empiriquement.

REFERENCES


APPENDIX

Approximate expressions for bias and variance are derived here using Taylor's series expansion and considering terms of second order only, as is usually the case with ratio estimation.

(1) Bias of $\hat{Y}_p$, in (4.4).

\[
B_p(\hat{Y}_p) = B_p = E(\hat{Y}_p - Y) = E\left(\frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{z_i} - Y\right)
\]

\[= E\left[\frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{p_i} - \frac{z_i}{p_i} - Y\right]
\]

\[= Y\left[1 + e_{11} + e_{21} - 1\right]
\]

where $e_{11} = (y_i / p_i - Y) / Y$ and $e_{21} = (z_i / p_i - Z)Z$. Thus, assuming $|e_{21}| < 1$, under usual approximation

\[
B_p = Y[E(e_{11}) - E(e_{21}) + E(e_{21}^2) - E(e_{11}e_{21})]
\]

(A.1)

For PPS with replacement sampling, we have

\[
E(e_{11}) = E(e_{21}) = 0
\]

\[
E(e_{21}^2) = Z^{-2} E((z_i / p_i - Z)^2) = Z^{-2} V(z_i / p_i)
\]

\[
E(e_{11}e_{21}) = (YZ)^{-1} \text{Cov}(y_i / p_i, z_i / p_i).
\]
Thus, under usual approximation $B_{pi}$ for PPS with replacement sampling is

$$B_{pi} = Z^{-1} \left[ R V(z_i/p_i) - \text{Cov}(y_i/p_i,z_i/p_i) \right], \text{ where } R = Y/Z. \quad (A.2)$$

Similarly, it is easy to show that $V_{pi} = \frac{1}{n} \sum_{i=1}^{n} (e_{1i} - e_{2i})^2 = V_{RP}$ in (4.3).

(2) Bias of $\hat{Y}_{RO}$ in (4.6):

$$B(\hat{Y}_{RO}) = E[\frac{\hat{Y}_0}{Z_0} Z - Y]$$

$$= E[\frac{1}{n} \sum_{i=1}^{n} N_{y_i} - Y]$$

$$= \frac{1}{n} \sum_{i=1}^{n} N_{z_i} - Y$$

$$= YE[(1+e_3)(1+e_4)^{-1} - 1]$$

where $e_3 = (\hat{Y}_0 - Y) / Y$, $e_4 = (\hat{Z}_0 - Z) / Z$. Again assuming $|e_4| < 1$, $B(\hat{Y}_{RO})$ is approximated by

$$B(\hat{Y}_{RO}) = YE[\hat{e}_3 - E(e_4) + E(e_4^2) - E(e_3 e_4)]. \quad (A.3)$$

Expressions for the expectations involved in (A.3) are computed below for cases of PPS with replacement scheme.

$$YE(e_3) = E(n^{-1} \sum_{i=1}^{n} N_{y_i} - Y)$$

$$= \sum_{i=1}^{N} Y_i (N_{p_i} - 1) = B(\hat{Y}_0) = B_0 \quad (A.4)$$

$$YE(e_4) = \sum_{i=1}^{n} z_i (N_{p_i} - 1) = B(\hat{Z}_0) = B_0^* \quad (A.5)$$

$$Z^2 E(e_4^2) = E(n^{-1} \sum_{i=1}^{n} N_{z_i} - EN_{z_i} + EN_{z_i} - Z)^2$$
\[ = E \left[ n^{-1} \sum_{1}^{n} N_{i} - E(Nz_{i}) + B_{0}^{*} \right]^{2} \]

\[ = E(n^{-1} \sum_{1}^{n} N_{i})^{2} - (E(Nz_{i}))^{2} + B_{0}^{*2} \]

\[ = \frac{N^{2}}{n} \left[ \sum_{1}^{N} z_{i}^{2} p_{i} - \left( \sum_{1}^{N} z_{i} p_{i} \right)^{2} \right] + B_{0}^{*2} \]

\[ = V(\hat{Z}_{0}) + B_{0}^{*2} . \]  

(A.6)

Similarly,

\[ Y^{2} E(e_{3}^{2}) = V(Y_{0}) + B_{0}^{2} . \]  

(A.7)

\[ YZ E(e_{3}^{2}e_{4}) = E(n^{-1} \sum_{1}^{n} N_{i} - Y)(n^{-1} \sum_{1}^{n} N_{i} - Z) \]

\[ = E(n^{-1} \sum_{1}^{n} N_{i} - E(Ny_{i}) + B_{0})[n^{-1} \sum_{1}^{n} N_{i} - E(Nz_{i}) + B_{0}^{*}] \]

\[ = E(n^{-1} \sum_{1}^{n} N_{i} - E(Ny_{i}))[n^{-1} \sum_{1}^{n} N_{i} - E(Nz_{i})] + B_{0}B_{0}^{*} \]

\[ = n^{-2}[nE(N^{2}y_{i}z_{i}) - nE(Ny_{i}E(Nz_{i})) + B_{0}B_{0}^{*}] \]

\[ = \frac{N^{2}}{n} \left[ \sum_{1}^{N} y_{i}z_{i} p_{i} - \left( \sum_{1}^{N} y_{i} p_{i} \right) \left( \sum_{1}^{N} z_{i} p_{i} \right) \right] + B_{0}B_{0}^{*} \]

\[ = \text{Cov}(Y_{0}, \hat{Z}_{0}) + B_{0}B_{0}^{*} . \]  

(A.8)
Using A.4, A.5, A.6 and A.8 bias for $\hat{Y}_{R0}$ in A.3 is

$$B(\hat{Y}_{R0}) = B_0 + RB_0^* + Z^{-1}(RB_0^* - B_0B_0^*) + Z^{-1}[RV(Z_0) - \text{Cov}(\hat{Y}_0^*, Z_0)].$$

Similarly, approximate expression for the mean square error of $\hat{Y}_{R0}$ is

$$M(\hat{Y}_{R0}) = Y^2[E(e_3^2) + E(e_4^2) - 2E(e_3e_4)].$$

Again using A.6, A.7 and A.8, and ignoring the bias terms, approximate expression for the variance of $\hat{Y}_{R0}$ is

$$V(\hat{Y}_{R0}) = V(\hat{Y}_0) + R^2V(Z_0) + 2R\text{ Cov}(\hat{Y}_0, Z_0)$$

where $R = Y/Z$. 