ALTERNATIVE ESTIMATORS IN PPS SAMPLING

M.P. Singh[1]

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as the time passes the initial size measure used to determine the initial selection probabilities becomes more and more out of date resulting in loss of correlation and hence the loss in efficiency of the survey estimates.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 $Y = \sum_1^N y_i$ is

$$\hat{Y}_p = \frac{1}{n} \sum_1^n \frac{y_i}{p_i} \tag{2.1}$$

with variance

$$V_p = \frac{1}{n} \sum_1^N \frac{y_i^2}{p_i} - \frac{Y^2}{n} \tag{2.2}$$

where $p_i = x_i/X$ and N is the number of units in the population, $X=\sum_1^N x_i$.

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

$$\hat{Y}_S = \frac{N}{n} \sum_1^n y_i \tag{2.3}$$

with variance

$$V_S = \frac{1}{n} (N \sum_1^N y_i^2 - Y^2) \tag{2.4}$$

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 $\hat{Y}_p$ 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 $\hat{Y}_p$. Thus Rao's estimator is

$$\hat{Y}_o = \frac{N}{n} \sum_1^n y_i \tag{2.5}$$

and has variance

$$V_o = \frac{N^2}{n} [\sum_1^N y_i^2 \, p_i - (\sum_1^N y_i \, p_i)^2]. \tag{2.6}$$

Note that although $\hat{Y}_S$ 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 $\hat{Y}_p$ with PPS in contrast to $\hat{Y}_S$ with SRS, we consider an alternative estimator

$$\hat{Y}_{p'} = \frac{1}{n} \sum_1^n \frac{y_i}{p'_i} \tag{2.7}$$

where

$$p'_i = \frac{z_i}{Z} \, , \; Z = \sum_1^N z_i \, .$$

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

$$B_{p'} = \sum_1^N y_i \left(\frac{p_i}{p'_i} - 1\right) \tag{2.8}$$

and

$$B_o = \sum_1^N y_i \; (Np_i - 1). \tag{2.9}$$

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

$$V_{p'} = \frac{1}{n} \left[\sum_1^N \frac{y_i^2 \, p_i}{p_i'^2} - \left(\sum_1^N \frac{y_i \, p_i}{p_i'} \right)^2 \right]. \tag{2.10}$$

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 Δ_1 often used when y is highly positively correlated with z is

$$y_i = z_i + \eta_i \qquad i = 1, 2, \ldots, N \tag{3.1}$$

where

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

and

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

The smybol ε_1 denotes the average overall finite populations that can be drawn from the super-population.

Under this model Δ_1

$$\varepsilon_1 \; (B_{p'}) = \beta\Sigma \; z_i \; (\frac{p_i'}{p_i} - 1)$$

$$= 0 \quad \text{for any } p_i, \text{ by substituting } \mathbf{z}_i = p_i' \; Z.$$

Thus, $\hat{Y}_{p'}$ is unbiased under the model. However, in general, $\hat{Y}_{p'}$, 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 $\hat{Y}_{p'}$ under the model Δ_1 are

$$\varepsilon_1(V_o) = V_o^* = \beta^2 \; V(\hat{Z}_o) + \frac{N^2 \; aZ^g}{n} \sum_1^N p_i'^g \; p_i(1-p_i) \tag{3.2}$$

and $$\varepsilon_1(V_{p'}) = V_{p'}^* = \frac{aZ^g}{n} \sum_1^N p_i'^{g-2} \quad p_i(1-p_i) \tag{3.3}$$

where $$\hat{Z}_o = \frac{N}{n} \sum_1^N z_i \qquad \text{and} \tag{3.4}$$

$$V(\hat{Z}_o) = \frac{N^2}{n} [\sum_1^N z_i^2 p_i - (\sum_1^N z_i p_i)^2]. \tag{3.5}$$

Further, in developing the estimators $\hat{Y}_o$ and $\hat{Y}_{p'}$ the underlying assumption is that y and x are unrelated. The super-population model Δ often used for this situation (Rao, [7]) for comparison of estimators is $y_i = \mu + e_i$, where $\varepsilon(e_i|x_i) = 0$, $\varepsilon(e_i^2|x_i) = b$, $b > o$ and $\varepsilon(e_i e_j|x_i,x_j)=0$ and ε is defined for Δ like ε_1. Since Rao [7] has shown that $\varepsilon(V_o) < \varepsilon(V_p)$, it is enough to compare the average variances of $\hat{Y}_o$ and $\hat{Y}_p$, under the model Δ_1. In order to facilitate this comparison, we shall use the following model Δ_2 for the characteristic x, similar to the model Δ for y.

Δ_2: $\quad p_i = m + e_i^{'} \qquad i = 1, 2, \ldots, N$

$$\varepsilon_2(e_i^{'}|z_i) = 0, \; \varepsilon_2(e_i^{'2}|z_i) = a', \; a'>o \tag{3.6}$$

$$\varepsilon_2(e_i^{'}e_i^{'}|z_i, z_j) = 0$$

where ε_2 is defined as ε_1.

Thus, the expected variances are

$$\varepsilon_2(V_o^*) = V_o^{**} = \beta^2\varepsilon_2 \; V(\hat{Z}_o) + \frac{N^2 \; aZ^g}{n} (m-m^2-a') \sum_1^N p_i^{'g} \qquad \text{and} \tag{3.7}$$

$$\varepsilon_2(V_{p'}^*) = V_{p'}^{**} = \frac{aZ^g}{n} (m-m^2-a') \sum_1^N p_i^{'g-2} \quad . \tag{3.8}$$

Therefore from (3.7) and (3.8)

$$V_o^{**} - V_{p'}^{**} = \beta^2\varepsilon_2 \; V(\hat{Z}_o) + \frac{aZ^g}{n} (m-m^2-a') \sum_1^N p_i^{'g} \left(N^2 - \frac{1}{p_i^{'2}}\right). \tag{3.9}$$

Since $\quad \varepsilon_2(p_i) = m, \; \varepsilon_2(p_i^2) = m^2 + a'$ (3.10)

and $\quad \varepsilon_2(p_i) > \varepsilon_2(p_i^2)$ because $p_i^2 > p_i$ for all possible values

except 1 or 0 ,

we have that $(m-m^2-a') \geq 0.$ (3.11)

Also, $\sum_1^N p_i^{'2} \geq \frac{1}{N}$, with equality with all $p_i^{'} = 1/N$. (3.12)

For g = 2, the second term in (3.9) becomes $\frac{aZ^2}{n} (m-m^2-a') \sum_1^N (N^2 p_i^{'2} - 1) \geq 0$ because of the inequalities in the expressions (3.11) and (3.12). Therefore, for g = 2, in the model Δ in (3.1), the suggested estimator $\hat{Y}_{p'}$ 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

$$\hat{Y}_{RP} = \frac{\hat{Y}_p}{\hat{Z}_p} Z \qquad (4.1)$$

where $\hat{Y}_p$ is defined in (2.1) and $\hat{Z}_p = n^{-1} \sum_1^n \frac{z_i}{p_i}$, $p_i = \frac{x_i}{X}$.

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

$$B_{RP} = Z^{-1} \left[RV(\hat{Z}_p) - R \operatorname{Cov}(\hat{Y}_p, \hat{Z}_p)\right] \qquad (4.2)$$

and

$$V_{RP} = V(\hat{Y}_p) + R^2 V(\hat{Z}_p) - 2R \operatorname{Cov}(\hat{Y}_p, \hat{Z}_p) \qquad (4.3)$$

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

$$V(\hat{Z}_p) = \frac{1}{n} \sum_1^N \frac{z_i^2}{p_i} - \frac{Z^2}{n},$$

and

$$\operatorname{Cov}(\hat{Y}_p, \hat{Z}_p) = \frac{1}{n} \sum_1^N \frac{y_i z_i}{p_i} - \frac{YZ}{n}.$$

It is of interest to note that $B_{p'}$ in (2.8) for PPS with replacement sampling may be approximated by (see Appendix)

$$B_{p'} = Z^{-1} \; [R \; V(\frac{Z_i}{p_i}) - Cov(\frac{y_i}{p_i} , \frac{Z_i}{p_i})] = n \; B_{RP} \tag{4.4}$$

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

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

$$\hat{Y}_{RO} = \frac{\hat{Y}_0}{\hat{Z}_0} \; Z \tag{4.5}$$

where $\hat{Y}_0$ and $\hat{Z}_0$ 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 $\hat{Y}_0$ and $\hat{Z}_0$ themselves are biased estimates of Y and Z respectively. Approximate bias and variance of $\hat{Y}_{RO}$ may be written as (see Appendix)

$$B(\hat{Y}_{RO}) = B_0 + RB_0^* + Z^{-1}(RB_0^{*2} - B_0B_0^*) + Z^{-1}[R \; V(\hat{Z}_0) - Cov(\hat{Y}_0\hat{Z}_0)] \tag{4.6}$$

and

$$V(\hat{Y}_{RO}) = V(\hat{Y}_0) + R^2 \; V(\hat{Z}_0) - 2R \; Cov(\hat{Y}_0,\hat{Z}_0), \tag{4.7}$$

where B_0, $V(\hat{Y}_0)$, $V(\hat{Z}_0)$ are as defined in (2.9) and(2.6) and (3.5) respectively.

Further,

$$B_0^* = B(\hat{Z}_0) = \sum_1^N z_i(Np_i-1) \tag{4.8}$$

$$Cov(\hat{Y}_0,\hat{Z}_0) = \frac{N^2}{n} [\sum_1^N y_i z_i p_i - (\sum_1^N y_i p_i)(\sum_1^N z_i p_i)]. \tag{4.9}$$

For comparing $\hat{Y}_{RP}$ and $\hat{Y}_{RO}$, we obtain their expected variances under the model Δ_1 (ie., assuming that y and z are highly correlated). We find that

$$\varepsilon_1 \ \mathrm{Cov}(\hat{Y}_p, \hat{Z}_p) = \beta \left(\frac{1}{n} \sum_1^N \frac{z_i^2}{p_i} - \frac{Z^2}{n}\right)$$

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

and

$$\varepsilon_1 \ \mathrm{Cov}(\hat{Y}_0, \hat{Z}_0) = \beta \ \frac{N^2}{n} \left[\sum_1^N z_i^2 \ p_i - (\Sigma \ z_i p_i)^2\right]$$

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

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

Thus, from (4.3) and (4.7), we have under model Δ_1

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

and

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

Further, if $\beta = R$ and

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

then,

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

and

$$\varepsilon_1 \ V(\hat{Y}_{RO}) = V(\hat{Y}_0)(1 - R^2), \tag{4.16}$$

which shows that under the condition (4.14)

$$\varepsilon_1 \; V(\hat{Y}_{RO}) < \varepsilon_1 \; V(\hat{Y}_{RP})$$

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

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

$$\varepsilon_1 \; V(Y_{RO}) \lesseqgtr \varepsilon_1 \; V(\hat{Y}_{Rp})$$

depending on

$$\{V(\hat{Y}_p) - V(\hat{Y}_0)\} + (R^2 - 2R\,\beta) \;\; V(\hat{Z}_p) - V(\hat{Z}_0)\} \lesseqgtr 0 \, .$$

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

Further, from (4.16), it is observed that $\hat{Y}_{R0}$ is more efficient than $\hat{Y}_0$ under usual conditions of ratio estimation. As both $\hat{Y}_{R0}$ 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:

$$e_{p0} = \frac{V_p}{mse(\hat{Y}_0)} \ , \ e_{pp'} = \frac{V_p}{mse(\hat{Y}_{p'})}$$

and

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

The following Table gives the bias and the relative efficency of the estimators. The correlation coefficients (δyx, δyz and δxz) 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

	Stratum				
	1	2	3	4	5
δyx	0.092	0.007	0.012	0.069	0.070
δyz	0.998	0.995	0.997	0.998	0.998
δxz	0.099	-0.031	0.008	0.074	0.069
Cx	61	72	60	58	84
Cy	60	39	40	65	51
Cz	55	36	38	60	49
ΣY_i	1,034	1,160	1,178	983	1,063
B_0	35.4	2.2	3.5	25.5	32.3
$B_{p'}$	-44.8	6.3	-6.5	-103.7	-19.1
e_{p0}	765	3,446	1,184	342	2,530
$e_{0p'}$	1,194	4,581	8,732	455	5,824
$e_{pp'}$	9,137	157,895	103,480	1,572	147,393

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, δ_{yx} and δ_{xz} are around zero and δ_{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 corrélée 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

[1] Cochran, W.G. (1963): "Sampling Techniques", John Wiley & Sons, Inc., New York.

[2] Drew, D. Choudhry, H., and Gray, G.B., "Some Methods For Updating Sample Survey Frames and Their Effects on Estimation", presented at the Annual Meeting of the American Statistical Association, August 14-17, 1978, San Diego, California, U.S.A.

[3] Fellegi, I.P., "Changing the Probabilities of Selection When Two Units Are Selected With PPS Without Replacement", proceedings of Social Statistics Section, American Statistical Association, 1966, 434-442.

[4] Keyfitz, N., "Sampling With Probability Proportional to Size - Adjustment for Changes in Size", Journal of the American Statistical Association 58, 1961, 183-201.

[5] Kish, L. and Scott, "Retaining Units After Changing Strata and Probabilities", Journal of the American Statistical Association 66, 1971, 461-470.

[6] Platek, R. and Singh, M.P., "A Strategy for Updating Continuous Surveys", Metrika, 1978, 25, 1-7.

[7] Rao, J.N.K., "Alternative Estimators in PPS Sampling for Multiple Characteristics", Sankhya, Series A, 28, 47-60.

[8] Singh, M.P., "Alternative Estimators in PPS Sampling", Abstract, Bulletin of the Institute of Mathematical Statistics, January 1975, 4.1, 29-30.

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).

$$B(\hat{Y}_{p'}) = B_{p'} = E(\hat{Y}_p - Y) = E(\frac{Z}{n} \sum_1^n \frac{y_i}{z_i} - Y)$$

$$= E[\frac{Z}{n} \sum_1^n \frac{y_i/p_i}{z_i/p_i} - Y]$$

$$= YE[(1+e_{1i})(1+e_{2i}) - 1]$$

where $e_{1i} = (y_i/p_i - Y)/Y$ and $e_{2i} = (z_i/p_i - Z)Z$. Thus, assuming $|e_{2i}|<1$, under usual approximation

$$B_{p'} = Y[E(e_{1i}) - E(e_{2i}) + E(e_{2i}^2) - E(e_{1i}e_{2i})] \quad . \tag{A.1}$$

For PPS with replacement sampling, we have

$$E(e_{1i}) = E(e_{2i}) = 0$$

$$E(e_{2i}^2 = Z^{-2} E(z_i/p_i - Z)^2 = Z^{-2} V(z_i/p_i)$$

$$E(e_{1i}e_{2i}) = (YZ)^{-1} \operatorname{Cov}(y_i/p_i, z_i/p_i).$$

Thus, under usual approximation $B_{p'}$ for PPS with replacement sampling is

$$B_{p'} = Z^{-1} \, [R \, V(z_i/p_i) - \mathrm{Cov}(y_i/p_i, z_i/p_i)], \text{ where } R = Y/Z. \tag{A.2}$$

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

(2) Bias of $\hat{Y}_{R0}$ in (4.6):

$$\begin{aligned} B(\hat{Y}_{R0}) &= E[\frac{\hat{Y}_0}{\hat{Z}_0} Z - Y] \\ &= E[\frac{n^{-1} \sum_1^n Ny_i}{n^{-1} \sum_1^n Nz_i} Z - Y] \\ &= YE[(1+e_3)(1+e_4)^{-1} - 1] \end{aligned}$$

where $e_3 = (\hat{Y}_0 - Y)/Y$, $e_4 = (\hat{Z}_0 - Z)/Z$. Again assuming $|e_4|<1$, $B(\hat{Y}_{R0})$ is approximated by

$$B(\hat{Y}_{R0}) = Y[E(e_3) - E(e_4) + E(e_4^2) - E(e_3 e_4)] \ . \tag{A.3}$$

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

$$\begin{aligned} YE(e_3) &= E(n^{-1} \sum_1^n Ny_i - Y) \\ &= \sum_1^N y_i(Np_i - 1) = B(\hat{Y}_0) = B_0 \end{aligned} \tag{A.4}$$

$$YE(e_4) = \sum_1^n z_i(Np_i - 1) = B(\hat{Z}_0) = B_0^* \tag{A.5}$$

$$Z^2 E(e_4^2) = E[n^{-1} \sum_1^n Nz_i - ENz_i + ENz_i - Z]^2$$

$$= E[n^{-1}\sum_1^n Nz_i - ENz_i + B_0^*]^2$$

$$= E(n^{-1}\sum_1^n Nz_i)^2 - (ENz_i)^2 + B_0^{*2}$$

$$= \frac{N^2}{n}[\sum_1^N z_i^2 p_i - (\sum_1^N z_i p_i)^2] + B_0^{*2}$$

$$= V(\hat{Z}_0) + B_0^{*2} . \tag{A.6}$$

Similarly, $Y^2 E(e_3^2) = V(\hat{Y}_0) + B_0^2$. (A.7)

$$YZ\, E(e_3 e_4) = E(n^{-1}\sum_1^n Ny_i - Y)(n^{-1}\sum_1^n Nz_i - Z)$$

$$= E(n^{-1}\sum_1^n Ny_i - E(Ny_i) + B_0)[n^{-1}\sum_1^n Nz_i - E(Nz_i) + B_0^*]$$

$$= E[n^{-1}\sum_1^n Ny_i - E(Ny_i)][n^{-1}\sum_1^n Nz_i - E(Nz_i)] + B_o B_o^*$$

$$= n^{-2}[nE(N^2 y_i z_i) - nE(Ny_i)E(Nz_i)] + B_0 B_0^*$$

$$= \frac{N^2}{n}[\sum_1^N y_i z_i p_i - (\sum_1^N y_i p_i)(\sum_1^N z_i p_i)] + B_0 B_0^*$$

$$= Cov(\hat{Y}_0, \hat{Z}_0) + B_0 B_0^* . \tag{A.8}$$

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

$$B(\hat{Y}_{RO}) = B_0 + RB_0^* + Z^{-1}(RB_0^{*2} - B_0B_0^*) + Z^{-1}[RV(\hat{Z}_0) - \mathrm{Cov}(\hat{Y}_0\hat{Z}_0)].$$

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

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

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

$$V(\hat{Y}_{RO}) = V(\hat{Y}_0) + R^2V(\hat{Z}_0) \; 2R \; \mathrm{Cov}(\hat{Y}_0,\hat{Z}_0)$$

where $R = Y/Z$.