*Sankhyā: The Indian Journal of Statistics*
1996, Volume 58, Series B, Pt. 1, pp. 90-104

# RAO-HARTLEY-COCHRAN STRATEGY IN SURVEY SAMPLING OF CONTINUOUS POPULATIONS

*By* V. R. PADMAWAR
*Indian Statistical Institute*

*SUMMARY.* Rao-Hartley-Cochran (RHC) strategy is defined for the continuous set-up and its efficiency is compared with some of the known strategies under the regression model.

## 1. Introduction

Early proponents of the survey sampling of continuous populations, C. M. Cassel and C. E. Särndal, were strongly motivated to use the continuous set-up as it facilitates the assessment especially of the mathematically cumbersome strategies (Cassel *et al.* (1977)). Särndal (1980) reiterates, 'the continuous variable formulation is an attempt to adapt Godambe's survey sampling set-up in continuous terms. This makes it easier to interpret and grasp some of the complexities of modern survey sampling theory of finite populations'.

Several sampling strategies for estimating the population mean have been considered in the literature of survey sampling of continuous populations (Cassel and Särndal (1972, 1974), Cassel *et al.* (1977), Särndal (1980), Padmawar (1982, 1984, 1994), Cordy (1993)). Results regarding nonexistence (Padmawar (1982)) and existence (Padmawar (1984)) of optimal strategies in certain classes of $p$-unbiased strategies are known. Padmawar (1994) compares several sampling strategies under the regression model.

Cordy (1993), motivated by environment related real life problems, develops a theory of estimation for sampling from continuous populations. He provides an interesting extension from the finite set-up, of the well known Horvitz-Thompson strategy, to the continuous frame-work.

Rao-Hartley-Cochran (RHC) strategy (vide Rao, Hartley and Cochran (1962)) is one of the important strategies in the finite set-up. This paper is an attempt to facilitate its assessment and comparison with other known strategies vis-a-vis the continuous set-up. In section 2, we first define Rao-Hartley-Cochran (RHC) strategy in the continuous set-up. We consider design based results

---

Paper received. June 1995; revised November 1995.

*AMS subject classification.* 62D05.

*Key words and phrases.* Rao-Hartley-Cochran strategy; continuous population; regression model; measure of uncertainty.

in this section and obtain the limiting variance of the RHC strategy under the assumption of a.e. continuity of $y(x)$. In section 3, we derive results under the regression model where $x$ is assumed to have gamma distribution. We compare the efficiency of the RHC strategy with that of other sampling strategies, including those considered by Cassel *et al.* (1977), Särndal (1980) and Padmawar (1994), under the regression model.

Consider a population of infinitely many pairs $(y(x), x); x \geq 0$, such that the joint distribution of $y(x), x \geq 0$, is known only partially. For convenience let us assume that $y(x), x \geq 0$, are defined on some probability space $(\Omega, \mathcal{A}, \xi)$. The distribution of $X$, whose observed values are $x$, assumed to be continuous and known is given by

$$F(x) = \int_0^x f(u)du; \quad x \geq 0.$$

$Y$ is called study variable and $X$ is called auxiliary variable.

In the continuous survey sampling set-up the label of a population unit is a continuous index $\lambda$ where for convenience $\lambda \in [0, 1)$. A more specific ordering is imposed on $\lambda$ by identifying it with the $\lambda$th quantile of the $X$-distribution. Having drawn and observed $n$ units the data is recorded as $(y(x_i), x_i)$, $i = 1, 2, \ldots, n$; or equivalently $(y(\mathbf{x}), \mathbf{x})$ where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. The problem under consideration is to estimate efficiently the population mean for the variate $Y$, namely

$$m_y = E_f(Y) = \int_0^\infty y(x)f(x)dx.$$

This, incidentally, defines the operator $E_f$.

Let $\mathcal{B}$ be the Borel $\sigma$-field of $\Re_n^+ = \{\mathbf{x} : x_i \geq 0, i = 1, 2, \ldots, n\}$. Any continuous probability measure $Q$ on $\mathcal{B}$ is called a sampling design. $Q(\mathbf{x})$ is the probability of drawing a sample such that the auxiliary variate value does not exceed $x_i$ in the ith draw, $1 \leq i \leq n$. If we write $q(\mathbf{x}) = \frac{\partial^n Q(\mathbf{x})}{\partial x_1 \partial x_2 \ldots \partial x_n}$, then $q(\mathbf{x})$ can be expressed as $q(\mathbf{x}) = p(\mathbf{x})f(\mathbf{x})$, where $f(\mathbf{x}) = \prod_{i=1}^{n} f(x_i)$. We would say that $p(\mathbf{x})$ is a design function giving rise to the sampling design $Q(\mathbf{x})$. One may also think of a two-stage sampling design.

Here we consider a specific superpopulation model, namely the regression model, induced by the probability space $(\Omega, \mathcal{A}, \xi)$, given by

$$Y(x) = \beta x + Z(x), x \geq 0$$

where for every fixed $x \geq 0$

$$E_\xi(Z(x)) = 0, E_\xi(Z^2(x)) = \sigma^2 x^g \qquad \ldots (1.1)$$

and for every $x \neq x', x, x' \geq 0$

$$E_\xi(Z(x)Z(x')) = 0$$

where $\sigma^2 > 0$ and $\beta$ are unknown and $g \in [0, 2]$ may be known or unknown.

A function $t$ of the observed data $(y(\mathbf{x}), \mathbf{x})$ is called an estimator of the population mean $m_y$,whereas $(p, t)$, an estimator together with a design function $p$ is called a strategy.

A strategy $(p, t)$ is said to be $p$-unbiased (design-unbiased) for $m_y$ if

$$E_p(t) = \int_{\Re_n^+} t(y(\mathbf{x}), \mathbf{x})p(\mathbf{x})f(\mathbf{x})d\mathbf{x} = \int_0^\infty y(x)f(x)dx = m_y$$

for every real valued $F$-integrable function $y(x)$. This defines the operator $E_P$.

A strategy $(p, t)$ is said to be $\xi$-unbiased (model-unbiased) for $m_y$ if

$$E_\xi[(t(Y(\mathbf{x}), \mathbf{x}) - m_y] = 0 \text{ a.e. } [Q].$$

A strategy $(p, t)$ is said to be $p\xi$-unbiased (model-design-unbiased) for $m_y$ if

$$E_p E_\xi[(t(Y(\mathbf{x}), \mathbf{x})] - E_\xi[m_y] = 0.$$

We assume that $Y(x)$ is square integrable w.r.t. the product probability $(F \times \xi)$. To judge the performance of a strategy $(p, t)$ we use the following measures of uncertainty :

$$M_1(p, t) = E_\xi E_P(t - m_y)^2 \qquad \ldots (1.2)$$

$$M_2(p, t) = E_\xi E_P(t - \mu_y)^2 \qquad \ldots (1.3)$$

where $\mu_y = E_\xi m_y = E_\xi \int_0^\infty y(x)f(x)dx = \beta E_f X = \beta\mu$ (say).

In this note we assume that the auxiliary variable $X$ has Gamma distribution with parameter $\alpha$.

In section 3, we compare strategies ( srs, $t_R$), $(P_M, t_R)$, ( ppx, $t_{HT}$), $(P_g, t_g)$ with the RHC strategy, here we are using the following :

(a) *sampling designs* :

(1) *srs* : simple random sampling for which $p(\mathbf{x}) \equiv 1$

(2) $ppx^a$ : design for which $p(\mathbf{x}) \propto \prod_{i=1}^n x_i^a$

(3) $P_M$ : the continuous analogue of the Midzuno-Sen sampling design with

$$p(\mathbf{x}) = \frac{1}{n\mu}\sum_{i=1}^n x_i, \text{ where } \mu = E_f(X) = \int_0^\infty xf(x)dx.$$

(4) $P_g$ : the sampling design with $p(\mathbf{x}) = \Lambda \prod_{i=1}^n x_i^{g-1} \sum_{i=1}^n x_i^{2-g}$

where $\Lambda = \frac{1}{n\mu}\left(\frac{\Gamma(\alpha)}{\Gamma(\alpha+g-1)}\right)^{n-1}$ ; $(\mu = \alpha)$ and $g \in [0,2], \alpha + g - 1 > 0$.

(b) *estimators* :

(1) $t_R$ : the ratio estimator $\mu \dfrac{\sum_{i=1}^{n} y(x_i)}{\sum_{i=1}^{n} x_i}$

(2) $t_{HT}$ : the Horvitz-Thompson estimator, (Cordy (1993)), based on $q(\mathbf{x})$ given by $\sum_{i=1}^{n} \frac{\omega(x_i)f(x_i)}{\pi(x_i)}$, where $\pi(x_i) = \sum_{j=1}^{n} q_j(x_i), \pi(x)$ assumed to be positive for $x > 0$, and $q_i(x_i) = \int_{\Re^+_{n-1}} q(\mathbf{x}) \prod_{j \neq i}^{n} dx_j, 1 \leq i \leq n$.

(3) $t_g$ : the estimator given by $\dfrac{\mu}{\sum_{i=1}^{n} x_i^{2-g}} \sum_{i=1}^{n} x_i^{1-g} y(x_i), g \in [0,2]$.

## 2. Rao-Hartley-Cochran strategy

The Rao-Hartley-Cochran (RHC) strategy that consists of design $P_{RHC}$ and estimator $t_{RHC}$ is defined as follows :

The design $P_{RHC}$ is a two stage design. At stage one fix an integer $k \geq 2$. Let $z_1, z_2, z_3, \ldots, z_{nk-1}, z_{nk}$ be such that $z_i$ is the $\frac{i}{nk}$-th percentile of the distribution of $X, 0 \leq i \leq nk$. Clearly $z_0 = 0$ and $z_{nk} = \infty$. Consider the intervals $[z_i, z_{i+1}); 0 \leq i \leq nk - 1$. Divide these intervals into $n$ groups of size $k$ each at random. Let $G_h$ be the $h$-th group containing the intervals $B_{h1}, B_{h2}, \ldots, B_{hk}$ say, $1 \leq h \leq n$. At the second stage one point each is chosen independently from the $n$ groups formed at the first stage using ppx sampling within each group. This describes the design $P_{RHC}$ completely. To construct the estimator $t_{RHC}$ define for the $h$-th group $G_h$ the density $f_h$ as follows

$$\begin{aligned} f_h(x) &= \frac{f(x)}{W_h} \quad \text{if } x \in G_h \\ &= 0 \quad \text{otherwise} \end{aligned}$$

where $W_h = \int_{G_h} f(x)dx$.

Let $\mu_h = \int_{G_h} x f_h(x)dx$ and $(y(x_h), x_h)$ be the observation from the $h$-th group, $1 \leq h \leq n$. The estimator $t_{RHC}$ is now given as $t_{RHC} = \sum_{h=1}^{n} W_h \frac{y(x_h)}{x_h/\mu_h}$.

Since we are dealing with a two-stage sampling let $E_1, V_1$ and $E_2, V_2$ denote the expectation and variance under the design $P_{RHC}$ at stage 1 and stage 2 respectively.

It is easy to prove the following

Theorem 2.1. *The RHC strategy is p-unbiased.*

*Proof.*

$$\begin{aligned}
E_{P_{RHC}}(t_{RHC}) &= E_1 E_2(t_{RHC}) \\
&= E_1 E_2\left(\sum_{h=1}^{n} W_h \frac{y(x_h)}{x_h/\mu_h}\right) \\
&= E_1 \sum_{h=1}^{n} E_2 W_h \frac{y(x_h)}{x_h/\mu_h} \\
&= E_1 \sum_{h=1}^{n} W_h \int_{G_h} \frac{y(x_h)}{x_h/\mu_h} \frac{x_h}{\mu_h} f_h(x_h) dx_n \\
&= E_1 \int_0^{\infty} y(x) f(x) dx \\
&= E_1 m_y \\
&= m_y.
\end{aligned}$$

Hence the RHC strategy is $p$-unbiased.

We now compute the sampling variance of the strategy RHC.

Theorem 2.2. *The sampling variance of the RHC strategy is given by*

$$V_{P_{RHC}}(t_{RHC}) = \left\{ \frac{(k-1)}{(nk-1)} \int_0^{\infty} \left(\frac{y(x)}{x/\mu} - m_y\right)^2 \frac{x}{\mu} \; f(x) dx \right.$$

$$\left. + \frac{(n-1)}{2n} \frac{nk}{(nk-1)} \sum_{i=1}^{nk} \int_{B_i \times B_i} \left(\frac{y(x)}{x} - \frac{y(x')}{x'}\right)^2 x x' f(x) f(x') dx dx' \right\} \quad \ldots (2.1)$$

*Proof.* The variance of the strategy RHC may be expressed as

$$
\begin{aligned}
V_{P_{RHC}}(t_{RHC}) &= E_1V_2(t_{RHC}) + V_1E_2(t_{RHC}) \\
&= E_1V_2(t_{RHC}) \text{ as } V_1E_2(t_{RHC}) = Vm_y = 0 \\
&= E_1V_2\left(\sum_{h=1}^{n} W_h \frac{y(x_h)}{x_h/\mu_h}\right) \\
&= E_1\sum_{h=1}^{n} V_2\left(W_h \frac{y(x_h)}{x_h/\mu_h}\right) \\
&= E_1\sum_{h=1}^{n} W_h^2 \left\{\int_{G_h}\left(\frac{y(x)}{x/\mu_h}\right)^2 \frac{x}{\mu_h} f_h(x)dx - \left(\int_{G_h}\frac{y(x)}{x/\mu_h}\frac{x}{\mu_h} f_h(x)dx\right)^2\right\} \\
&= E_1\left\{\sum_{h=1}^{n} W_h^2 \int_{G_h}\frac{y^2(x)}{x/\mu_h} f_h(x)dx - \sum_{h=1}^{n}\left(\int_{G_h} y(x)f(x)dx\right)^2\right\} \\
&= \frac{1}{2}E_1\sum_{h=1}^{n}\int_{G_h\times G_h}\left(\frac{y(x)}{x} - \frac{y(x')}{x'}\right)^2 xx'f(x)f(x')dxdx'.
\end{aligned}
$$

Since $G_h$ is the $h$-th group containing the intervals $B_{h1}, B_{h2}, \ldots, B_{hk}$.

$$
G_h \times G_h = \bigcup_{i=1}^{k} B_{hi}\times B_{hi} + \bigcup_{i\neq j=1}^{k} B_{hi}\times B_{hj}.
$$

$$
E_1\int_{G_h\times G_h}\left(\frac{y(x)}{x} - \frac{y(x')}{x'}\right)^2 \quad xx'f(x)f(x')dxdx'
$$

$$
\begin{aligned}
&= E_1\sum_{i=1}^{k}\int_{B_{hi}\times B_{hi}}\left(\frac{y(x)}{x} - \frac{y(x')}{x'}\right)^2 \quad xx'f(x)f(x')dxdx' \\
&+E_1\sum_{i\neq j=1}^{k}\int_{B_{hi}\times B_{hi}}\left(\frac{y(x)}{x} - \frac{y(x')}{x'}\right)^2 \quad xx'f(x)f(x')dxdx' \\
&= c_1\sum_{i=1}^{nk}\int_{B_i\times B_i}\left(\frac{y(x)}{x} - \frac{y(x')}{x'}\right)^2 \quad xx'f(x)f(x')dxdx' \\
&+c_2\sum_{i\neq j=1}^{nk}\int_{B_i\times B_j}\left(\frac{y(x)}{x} - \frac{y(x')}{x'}\right)^2 \quad xx'f(x)f(x')dxdx' \\
&= c_2\int_{\Re^+\times\Re^+}\left(\frac{y(x)}{x} - \frac{y(x')}{x'}\right)^2 \quad xx'f(x)f(x')dxdx' \\
&+(c_1-c_2)\sum_{i=1}^{nk}\int_{B_i\times B_i}\left(\frac{y(x)}{x} - \frac{y(x')}{x'}\right)^2 \quad xx'f(x)f(x')dxdx'
\end{aligned}
$$

where $c_1$ and $c_2$ are respectively the probabilities of inclusion of an interval $B_i$ and a distinct pair $B_i, B_j$ in the $h$-th group. Clearly these probabilities are independent of $i, j$ and $h$ and are given by

$$c_1 = \frac{(nk-1)c_{(k-1)}}{nkc_K} = \frac{1}{n}$$
$$c_2 = \frac{(nk-2)c_{(k-2)}}{nkc_K} = \frac{1}{n}\frac{k-1}{nk-1}.$$

Thus

$$V_{P_{RHC}}(t_{RHC}) = \frac{1}{2}\Big\{ \frac{(k-1)}{(nk-1)} \int_{\Re^+\times\Re^+} \left(\frac{y(x)}{x} - \frac{y(x')}{x'}\right)^2 xx' f(x) f(x') dx dx'$$
$$+ \frac{(n-1)}{n}\frac{nk}{(nk-1)} \sum_{i=1}^{nk} \int_{B_i\times B_i} \left(\frac{y(x)}{x} - \frac{y(x')}{x'}\right)^2 xx' f(x) f(x') dx dx' \Big\}$$
$$V_{P_{RHC}}(t_{RHC}) = \Big\{ \frac{(k-1)}{(nk-1)} \int_0^\infty \left(\frac{y(x)}{x/\mu} - m_y\right)^2 \frac{x}{\mu} f(x) dx$$
$$+ \frac{(n-1)}{2n}\frac{nk}{(nk-1)} \sum_{i=1}^{nk} \int_{B_i\times B_i} \left(\frac{y(x)}{x} - \frac{y(x')}{x'}\right)^2 xx' f(x) f(x') dx dx' \Big\}.$$

Hence the theorem.

In what follows we show that the second term in (2.1) goes to zero as $k$, the number of intervals in each of the $n$ groups, goes to infinity. Observe that

$$V_{ppx}\left(\frac{y(x)}{x/\mu}\right) = \int_0^\infty \left(\frac{y(x)}{x/\mu} - m_y\right)^2 \frac{x}{\mu} f(x) dx \qquad \ldots (2.2)$$
$$= \mu \sum_{i=1}^{nk} \int_{{}_kB_i} \left(\frac{y(x)}{x} - {}_k\theta_i + {}_k\theta_i - \frac{m_y}{\mu}\right)^2 x f(x) dx.$$

Expanding the square and simplifying, we get

$$V_{ppx}\left(\frac{y(x)}{x/\mu}\right) = \mu\left(\sum_{i=1}^{nk} \int_{{}_kB_i} \left(\frac{y(x)}{x} - {}_k\theta_i\right)^2 x f(x) dx + \sum_{i=1}^{nk} \left({}_k\theta_i - \frac{m_y}{\mu}\right)^2 {}_k\mu_i\ {}_kW_i\right) \qquad \ldots (2.3)$$

where for $k \geq 2$ ${}_kB_1, {}_kB_2, \ldots, {}_kB_{nk}$ are the $nk$ intervals;

$${}_kW_i = \int_{{}_kB_i} f(x)dx;\ {}_k\mu_i\ {}_kW_i = \int_{{}_kB_i} x f(x) dx;\ {}_k\theta_i = \frac{\int_{{}_kB_i} y(x) f(x) dx}{\int_{{}_kB_i} x f(x) dx};\ 1 \leq i \leq nk.$$

Lemma 2.1. *Let $y(.)$ be a.e. (Lebesgue) continuous function on $\Re^+$. Let further $g_k(x) = \sum_{i=1}^{nk} {}_k\mu_i \left({}_k\theta_i - \frac{m_y}{\mu}\right)^2 \chi_{{}_kB_i(x)}$ and $g(x) = x\left(\frac{y(x)}{x} - \frac{m_y}{\mu}\right)^2$ where $\chi_B$ denotes the indicator function of the set $B$, then $g_k(x)$ converges to $g(x)$ a.e. (Lebesgue) as $k$ tends to $\infty$.*

*Proof.* Let $x_0$ be any continuity point of $y$. Let ${}_kB_i(x_0)$ denote the interval containing the point $x_0$ at the $k$-th stage, $k = 2, 3, \dots,$. It is clear that for any function $u$ that is continuous at $x_0$ we have $\sup_{x \in {}_kB_i(x_0)} | u(x) - u(x_0) |$ converges to 0. In particular it is true for the functions $y$ and the identity. Therefore

$$| {}_k\mu_i(x_0) - x_0 | = \left| \frac{\int_{{}_kB_i(x_0)} x f(x) dx}{\int_{{}_kB_i(x_0)} f(x) dx} - x_0 \right| \leq \sup_{x \in {}_kB_i(x_0)} | x - x_0 |$$

Thus

$$\lim_{k \to \infty} \frac{{}_k\mu_i(x_0)}{x_0} = 1.$$

Similarly

$$\begin{aligned} \left| {}_k\theta_i(x_0) - \frac{y(x_0)}{x_0} \right| &= \left| \frac{\int_{{}_kB_i(x_0)} y(x) f(x) dx}{\int_{{}_kB_i(x_0)} x f(x) dx} - \frac{y(x_0)}{x_0} \right| \\ &= \left| \frac{x_0 \int_{{}_kB_i(x_0)} \{y(x) - y(x_0)\} f(x) dx - y(x_0) \int_{{}_kB_i(x_0)} (x - x_0) f(x) dx}{x_0 \int_{{}_kB_i(x_0)} x f(x) dx} \right|. \end{aligned}$$

Dividing the numerator as well as the denominator by ${}_kW_i(x_0)$ and simplifying we get

$$\begin{aligned} \left| {}_k\theta_i(x_0) - \frac{y(x_0)}{x_0} \right| &\leq \frac{1}{{}_k\mu_i(x_0)} \left\{ \int_{{}_kB_i(x_0)} | y(x) - y(x_0) | \frac{f(x)}{{}_kW_i(x_0)} dx \right. \\ &\quad \left. + \frac{y(x_0)}{x_0} \int_{{}_kB_i(x_0)} | x - x_0 | \frac{f(x)}{{}_kW_i(x_0)} dx \right\} \\ &\leq \frac{1}{{}_k\mu_i(x_0)} \left\{ \sup_{x \in {}_kB_i(x_0)} | y(x) - y(x_0) | \right. \\ &\quad \left. + \frac{y(x_0)}{x_0} \sup_{x \in {}_kB_i(x_0)} | x - x_0 | \right\} \end{aligned}$$

as ${}_k\mu_i(x_0)$ converges to $x_0$ the above expression goes to 0 as $k$ tends to $\infty$. This proves that $g_k(x_0)$ converges to $g(x_0)$. Hence the lemma.

Lemma 2.2. *For $y(.)$ a.e. (Lebesgue) continuous*

$$\lim_{k \to \infty} \sum_{i=1}^{nk} \int_{{}_kB_i} \left( \frac{y(x)}{x} - {}_k\theta_i \right)^2 x f(x) dx = 0$$

*Proof.* In view of the Lemma 2.1, $\sum_{i=1}^{nk} \left({}_k\theta_i - \frac{m_y}{\mu}\right)^2 {}_k\mu_i \quad {}_kW_i$ converges to $\int_0^\infty \left(\frac{y(x)}{x} - \frac{m_y}{\mu}\right)^2 xf(x)dx$ as $k$ tends to $\infty$.

Now using (2.2) we get that

$$\lim_{k\to\infty} \sum_{i=1}^{nk} \int_{{}_kB_i} \left(\frac{y(x)}{x} - {}_k\theta_i\right)^2 xf(x)dx = 0.$$

We now prove the following theorem.

Theorem 2.3. *The limiting sampling variance of the RHC strategy for a.e. continuous $y(.)$ is given by*

$$\lim_{k\to\infty} V_{P_{RHC}}(t_{RHC}) = \frac{1}{n}\int_0^\infty \left(\frac{y(x)}{x/\mu} - m_y\right)^2 \frac{x}{\mu} f(x)dx. \qquad \ldots (2.4)$$

*Proof.* Note that $\sum_{i=1}^{nk} \int_{{}_kB_i \times {}_kB_i} \left(\frac{y(x)}{x} - \frac{y(x')}{x'}\right)^2 xx'f(x)f(x')dxdx'$

$$\begin{aligned}
&= 2\sum_{i=1}^{nk} \left({}_k\mu_i \quad {}_kW_i \int_{{}_kB_i} \frac{y^2(x)}{x} f(x)dx - \left\{\int_{{}_kB_i} y(x)f(x)dx\right\}^2\right) \\
&= 2\sum_{i=1}^{nk} {}_k\mu_i \quad {}_kW_i \int_{{}_kB_i} \frac{y(x)}{x} - {}_k\theta_i\,)^2\, xf(x)dx \\
&\le 2\left\{\sum_{i=1}^{nk} {}_k\mu_i \quad {}_kW_i\right\}\left\{\sum_{i=1}^{nk}\int_{{}_kB_i} \left(\frac{y(x)}{x} - {}_k\theta_i\right)^2 xf(x)dx\right\} \\
&= 2\mu\sum_{i=1}^{nk} \int_{{}_kB_i} \left(\frac{y(x)}{x} - {}_k\theta_i\right)^2 xf(x)dx.
\end{aligned}$$

Now in view of the above inequality and Lemma 2.2, the second term in (2.1) goes to 0 as $k$ tends to $\infty$. Therefore the limiting variance of the RHC strategy is given by

$$\lim_{k\to\infty} V_{P_{RHC}}(t_{RHC}) = \frac{1}{n}\int_0^\infty \left(\frac{y(x)}{x/\mu} - m_y\right)^2 \frac{x}{\mu} f(x)dx.$$

Theorem 2.4. *For a.e. continuous $y(.)$, the limiting sampling variance of the RHC strategy equals to $V_{ppx}(t_{HT})$.*

*Proof.* It is easy to see that $V_{ppx}(t_{HT}) = V_{ppx}\left(\frac{1}{n}\sum_{i=1}^{n} \frac{y(x_i)}{x_i/\mu}\right) = \frac{1}{n}V_{ppx}\left(\frac{y(x)}{x/\mu}\right)$.
Thus using the equations (2.2) and (2.4) we get the required result.

## 3. Efficiency under the model

In this section we do away with the assumption of continuity of $y(.)$ and evaluate the efficiency of the RHC strategy under the model (1.1). We then compare the RHC strategy with the strategies mentioned earlier. Let us first prove the following lemma that would be useful in computing the limiting value of $M_2(p,t)$ for the RHC strategy.

At the $k$-th stage, as mentioned earlier, we have the percentiles of the variable $x$ such that ${}_kz_0 = 0, {}_kz_1, {}_kz_2, \ldots, {}_kz_{nk-1}, {}_kz_{nk}, = \infty$, etc.

Lemma 3.1. *If the density $f$ satisfies for $u \leq v$,*

$$\int_u^v xf(x)dx = \Psi_1(u,v) + \Psi_2(Prob[u \leq x \leq v]),$$

*where $\Psi_1(u,v) \to 0$ as $u \to \infty, u \leq v$; and $\Psi_2(p) \to 0$ as $p \to 0$. Then*

$$\max_{1\leq i\leq nk} {}_ka_i \to 0 \text{ as } k \to \infty.$$

*Proof.* Let ${}_ka_i = \int_{{}_kz_{i-1}}^{{}_kz_i} xf(x)dx, i = 1,2\ldots,nk$. Given an $\epsilon > 0$, choose $M$ and $k$ such that $\Psi_1(u,v) < \frac{\epsilon}{2}$ for $M \leq u \leq v$ and $\Psi_2\left(\frac{1}{nk}\right) < \frac{\epsilon}{2}$ for $k \geq K$. At the $k$-th stage there exists unique $r_k$ such that ${}_kZ_{r_k-1} \leq M < {}_kZ_{r_k}$. Hence $\max({}_ka_{r_k+1}, {}_ka_{r_k+2}, \ldots,) < \epsilon$

$$\begin{aligned}\text{Note that } {}_ka_{r_k} &= \int_{{}_kz_{r_k}-1}^{{}_kz_{r_k}} xf(x)dx \\ &= \int_{{}_kz_{r_k}-1}^{M} xf(x)dx + \int_M^{{}_kz_{r_k}} xf(x)dx \leq \int_{{}_kz_{r_k}-1}^{M} xf(x)dx + \tfrac{\epsilon}{2}.\end{aligned}$$

Therefore consider $\max\left({}_ka_1, {}_ka_2 \ldots, {}_ka_{r_k-1}, \int_{{}_kz_{r_k}-1}^{M} xf(x)dx\right)$. For convenience we write ${}_kb_i = {}_ka_i, 1 \leq i \leq r_k - 1$, and ${}_kb_{r_k} = \int_{{}_kz_{r_k}-1}^{M} xf(x)dx$.

Now look at $\max({}_kb_1, {}_kb_2, \ldots, {}_kb_{r_k-1}, {}_kb_{r_k})$. This maximum must go to 0 as $k \to \infty$, if not then there exists a $\delta > 0$ and a subsequence $\{k_j, j \geq 1\}$ such that $\max_{1\leq i\leq r_{k_j}} ({}_{k_j}b_i) > \delta \forall j \Rightarrow$ for each $j$ there exists $m_j$ such that ${}_{k_j}b_{m_j} > \delta \Rightarrow$ $\frac{1}{nk_j} \geq \frac{1}{M} {}_{k_j}b_{m_j} > \frac{\delta}{M} > 0$. A contradiction as $\frac{1}{nk_j}$ goes to zero as $j \to \infty$. Hence $\max({}_ka_1, {}_ka_2, \ldots, {}_ka_{r_k-1}, \int_{{}_kz_{r_k}-1}^{M} xf(x)dx)$ goes to zero as $k \to \infty$. Therefore $\max_{1\leq i\leq nk} {}_ka_i \to 0$ as $k \to \infty$.

Theorem 3.1. *For the RHC strategy we have,*

$$M_2(P_{RHC}, t_{RHC}) = \sigma^2 \left\{ \frac{(k-1)\mu^2}{(nk-1)} \frac{\Gamma(\alpha+g-1)}{\Gamma(\alpha+1)} + \frac{(n-1)k}{nk-1} \right.$$

$$\left. \sum_{i=1}^{nk} \left( \int_{B_i} x f(x) dx \right) \left( \int_{B_i} x^{g-1} f(x) dx \right) \right\}. \qquad \ldots (3.1)$$

*Proof.* For a strategy that is both $p$-unbiased and $\xi$-unbiased $M_2(p,t)$ reduces to $E_p V_\xi(t)$. Therefore

$$\begin{aligned} M_2(P_{RHC}, t_{RHC}) &= E_1 E_2 V_\xi \left( \sum_{h=1}^{n} W_h \mu_h \frac{y(x_h)}{x_h} \right) \\ &= E_1 E_2 \left( \sigma^2 \sum_{h=1}^{n} W_h^2 \mu_h^2 x_h^{g-2} \right) \\ &= \sigma^2 E_1 \sum_{h=1}^{n} W_h^2 \mu_h^2 \int_{G_h} x^{g-2} \frac{x}{\mu_h W_h} f(x) dx \\ &= \sigma^2 E_1 \sum_{h=1}^{n} \left( \int_{G_h} x f(x) dx \right) \left( \int_{G_h} x^{g-1} f(x) dx \right). \end{aligned}$$

Now anologous to the proof of Theorem 2.2, we have,

$$
\begin{aligned}
& E_1\Big(\int_{G_h} xf(x)dx\Big)\Big(\int_{G_h} x^{g-1}f(x)dx\Big) \\
&= E_1\left(\Big(\sum_{i=1}^{k}\int_{B_{hi}} xf(x)dx\Big)\Big(\sum_{i=1}^{k}\int_{B_{hi}} x^{g-1}f(x)dx\Big)\right) \\
&= E_1\left(\sum_{i=1}^{k}\Big(\int_{B_{hi}} xf(x)dx\Big)\Big(\int_{B_{hi}} x^{g-1}f(x)dx\Big)\right) \\
&+ E_1\left(\sum_{i\neq j=1}^{k}\Big(\int_{B_{hi}} xf(x)dx\Big)\Big(\int_{h_j} x^{g-1}f(x)dx\Big)\right) \\
&= \left(\frac{1}{n}\sum_{i=1}^{nk}\Big(\int_{B_i} xf(x)dx\Big)\Big(\int_{B_i} x^{g-1}f(x)dx\Big)\right) \\
&+ \left(\frac{k-1}{n(nk-1)}\sum_{i\neq j=1}^{nk}\Big(\int_{B_i} xf(x)dx\Big)\Big(\int_{B_j} x^{g-1}f(x)dx\Big)\right) \\
&= \frac{1}{n}\sum_{i=1}^{nk}\Big(\int_{B_i} xf(x)dx\Big)\Big(\int_{B_i} x^{g-1}f(x)dx\Big) \\
&+ \frac{k-1}{n(nk-1)}\left\{\sum_{i=1}^{nk}\Big(\int_{B_i} xf(x)dx\Big)\sum_{i=1}^{nk}\Big(\int_{B_i} x^{g-1}f(x)dx\Big)\right. \\
&\left. -\sum_{i=1}^{nk}\Big(\int_{B_i} xf(x)dx\Big)\Big(\int_{B_i} x^{g-1}f(x)dx\Big)\right\} \\
&= \frac{k-1}{n(nk-1)}\left\{\int_0^{\infty} xf(x)dx \int_0^{\infty} x^{g-1}f(x)dx\right\} \\
&+ \frac{(n-1)k}{n(nk-1)}\left\{\sum_{i=1}^{nk}\Big(\int_{B_i} xf(x)dx\Big)\Big(\int_{B_i} x^{g-1}f(x)dx\Big)\right\}.
\end{aligned}
$$

Hence, $M_2(P_{RHC}, t_{RHC}) = \sigma^2\Big\{\frac{k-1}{(nk-1)}\Big(\int_0^{\infty} xf(x)dx\Big)\Big(\int_0^{\infty} x^{g-1}f(x)dx\Big)$
$+\frac{(n-1)k}{nk-1}\sum_{i=1}^{nk}\Big(\int_{B_i} xf(x)dx\Big)\Big(\int_{B_i} x^{g-1}f(x)dx\Big)\Big\}$.

Theorem 3.2. *For the RHC strategy the limiting value of* $M_2(p,t)$ *is given by*

$$
\lim_{k\to\infty} M_2(P_{RHC}, t_{RHC}) = \frac{\sigma^2\mu^2}{n}\frac{\Gamma(\alpha+g-1)}{\Gamma(\alpha+1)}. \qquad \ldots(3.2)
$$

*Proof.* Observe that

$$\sum_{i=1}^{nk}\left(\int_{B_i} xf(x)dx\right)\left(\int_{B_i} x^{g-1}f(x)dx\right) \le \left(\max_{1\le i\le nk}\int_{B_i} xf(x)dx\right)\sum_{i=1}^{nk}\left(\int_{B_i} x^{g-1}f(x)dx\right)$$

$$= \frac{\Gamma(\alpha+g-1)}{\Gamma(\alpha)}\left(\max_{1\le i\le nk}\int_{B_i} xf(x)dx\right).$$

Since the gamma density with parameter $\alpha$ satisfies the conditions of Lemma 3.1

$$\max_{1\le i\le nk}\int_{B_i} xf(x)dx \text{ goes to } 0 \text{ as } k\to\infty.$$

Hence the second term in the expression (3.1) goes to 0 as $k\to\infty$. Also $\mu=\alpha$. Therefore $\lim_{k\to\infty} M_2(P_{RHC}, t_{RHC}) = \frac{\sigma^2\mu^2}{n}\frac{\Gamma(\alpha+g-1)}{\Gamma(\alpha+1)}$. Hence the theorem.

We conclude this paper by comparing the RHC strategy with some of the known strategies in the continuous set-up. Cassel *et al.* (1977) in section 7.6 of their book considered certain strategies which were later taken up by Särndal (1980) and Padmawar (1994). The rest of the section deals with the comparison of these and some other strategies with the RHC strategy w.r.t. $M_1(p,t)$ and $M_2(p,t)$. From Särndal (1980) and Padmawar (1994) we have,

$$M_2(ppx, t_{HT}) = \frac{\sigma^2\mu^2}{n}\frac{\Gamma(\alpha+g-1)}{\Gamma(\alpha+1)} \qquad \ldots(3.3)$$

$$M_2(P_g, t_g) = \frac{\sigma^2\mu^2}{n}\frac{\Gamma(\alpha+g-1)}{\Gamma(\alpha+1)} \qquad \ldots(3.4)$$

$$M_2(srs, t_R) = \sigma^2\mu^2\frac{n\Gamma(g+\alpha)/\Gamma(\alpha)}{(g+n\alpha-1)(g+n\alpha-2)} \qquad \ldots(3.5)$$

$$M_2(P_M, t_R) = \sigma^2\mu^2\frac{\Gamma(g+\alpha)/\Gamma(\alpha+1)}{(g+n\alpha-1)}. \qquad \ldots(3.6)$$

Theorem 3.3. *In the limiting sense the RHC strategy is as good as the strategies* $(ppx, t_{HT})$ *and* $(P_g, T_g)$ *w.r.t. either measure of uncertainty* $M_1$ *or* $M_2$ *under the regression model* (1.1).

*Proof.* For any $p$-unbiased as well as $\xi$-unbiased strategies $M_1$ and $M_2$ differ by the quantity $E_\xi(m_y-\beta\mu)^2$ which is clearly independent of any strategy. Therefore it is enough to use the measure of uncertainty $M_2$ for the purpose of comparisons as all three strategies are $p$-unbiased as well as $\xi$-unbiased. In view of (3.2), (3.3) and (3.4) the strategies $(P_{RHC}, t_{RHC})$, $(ppx, t_{HT})$ and $(P_g, t_g)$ are equally efficient.

*Remark* 3.1. In the limiting sense the RHC strategy is as good as the strategies $(ppx, t_{HT})$ and $(P_g, t_g)$ w.r.t. either measure of uncertainty $M_1$ or $M_2$ under the regression model (1.1). This result indicates that from the practical point of view the strategy $(ppx, t_{HT})$ would score over the other two competing strategies $(P_g, t_g)$ and $(P_{RHC}, t_{RHC})$ as the strategy $(P_g, t_g)$ depends on the parameter $g$ of the model (1.1) that may not always be known and $(P_{RHC}, t_{RHC})$ is as good as $(ppx, t_{HT})$ only in the limiting sense.

Using Theorem 3.3 and the results from Padmawar (1994) we finally have the following.

Theorem 3.4. *Under the model* (1.1) *we have, in the limiting sense,*

(a) $M_2(srs, t_R) > M_2(P_{RHC}, t_{RHC})$

(b) *for* $n \geq 2$ *and* $g + n\alpha - 1 > 0$ $M_r(P_M, t_R) \lesseqgtr M_r(P_{RHC}, t_{RHC})$

according as $g \lesseqgtr 1, r = 1, 2.$

(c) *for* $g = 1$ *the strategies* $(P_M, t_R), (ppx, t_{HT}), (P_g, t_g)$ *and* $(P_{RHC}, t_{RHC})$ *are equally efficient w.r.t. either measure of uncertainty.*

*Proof.* The proof follows using the above Theorem 3.3 along with Corollary 2.1, Theorems 2.3 and 2.4 of Padmawar (1994) and the expressions (3.3) through (3.6).

*Remark* 3.2. The parameter $g$ of the model (1.1) may not always be known in practice. If the sampler has to choose between the strategies $(P_{RHC}, t_{RHC})$ and $(P_M, t_R)$ then there is a clear demarcation of the parametric space of $g$. If there are reasons to believe that $g$ is less than unity then $(P_M, t_R)$ would perform better than $(P_{RHC}, t_{RHC})$. On the other hand if the sampler speculates $g$ to be greater than unity then $(P_{RHC}, t_{RHC})$ should be preferred to $(P_M, t_R)$.

*Acknowledgement.* The author thanks Prof. R. L Karandikar for a useful discussion on the construction of the Rao-Hartley-Cochran strategy and the referee for his suggestions.

## References

Cassel, C. M. and Särndal, C. E. (1972). A model for studying robustness of estimators and informativeness of labels in sampling with varying probabilities. *J. R. Statist. Soc.* B, **34**, 279 -289.

– – – – (1974). Evaluation of some sampling strategies for finite populations using a continuous variable framework. *Commun. Statist.*, **3**, 373 -390.

Cassel, C. M., Särndal, C. E. and Wretman, J. H. (1977). *Foundations of Inference in Survey Sampling*, New York : Wiley.

CORDY, C. B. (1993). An extension of the Horvitz-Thompson theorem to point sampling from a continuous universe. *Statistics and Probability Letters.*, **18**, 353 - 362.

PADMAWAR, V. R. (1981). A note on the comparison of certain sampling strategies. *J. R. Statist. Soc.* B, **43**, 321 - 326.

– – –– (1982). Optimal strategies under superpopulation models. *Ph.D. Thesis*, Indian Statistical Institute.

– – –– (1984). Two existence theorems in survey sampling of continuous populations. *Sankhyā* B, **46**, 217 - 227.

– – –– (1994). A note on comparison of estimation strategies in survey sampling of continuoous populations. *(Submitted for publication.)*

RAO, J. N. K., HARTLEY, H. O., and COCHRAN, W. G. (1962). On a simple procedure of unequal probability sampling without replacement. *J. R. Statist. Soc.* B, **24**, 482 - 491.

SÄRNDAL, C. E. (1980). A method for assessing efficiency and bias of estimation strategies in survey sampling. *South African Statist.* J, **14**, 17 - 30,

STAT–MATH DIVISION
INDIAN STATISTICAL INSTITUTE
8TH MILE, MYSORE ROAD
BANGALORE 560 059
INDIA.