The Annals of Statistics
1983, Vol. 11, No. 1, 275–285

PROPERTIES OF ESTIMATORS OF QUADRATIC FINITE POPULATION FUNCTIONS: THE BATCH APPROACH[1]

BY T. P. LIU AND M. E. THOMPSON

University of Waterloo

Polynomial finite population functions can be expressed as totals over derived populations of *batches*, or ordered sequences of units from the original population. This paper extends the results of Godambe and Godambe and Joshi on nonexistence of best unbiased estimators and admissibility of the Horvitz-Thompson estimator to the real batch total case. The admissibility results are only partly extendible; an example is given to show that Horvitz-Thompson type estimators of the form $\sum \sum b_{ij}(y_i - y_j)^2/\pi_{ij}$ need not be admissible.

1. Introduction. In the theory of finite population estimation, most of the emphasis has always been placed on the problem of estimating the population mean or total of a real characteristic. This relatively tractable problem is sufficiently general to illustrate most of the features of the foundational theory (see, for example, the book of Cassel, Sarndal and Wretman, 1977). It is also of great practical importance in all areas where sampling is applied; and even the estimation of complex population functions can often be reduced to a linear problem by Taylor series approximation (see Woodruff, 1971, and Kish and Frankel, 1974).

At the same time the estimation of quadratic or higher order polynomial population functions can also be important. A population variance, or the variance of a linear estimator, are two obvious examples. These population functions, although they may be relatively complex, have the advantage of being expressible as *totals* of certain generalized characteristics over a derived finite population (Hanurav, 1966). In the derived finite population, the "units" are combinations of units of the original population; and these combinations, the "units" of the derived population, are what will here be termed *batches*. For example,

$$\phi(\mathbf{y}) = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} y_i y_j$$

is a quadratic polynomial relative to the population $P = \{1, \cdots, N\}$, but may be expressed as

$$\sum_\alpha z_\alpha,$$

where α ranges over all pairs ij with $i < j$, and $z_{ij} = y_i y_j$. Thus $\phi(\mathbf{y})$ is a *batch total* over the collection of batches $A = \{(i, j): i < j\}$.

In this paper we investigate the extent to which some of the results of Godambe (1955) and Godambe and Joshi (1965) for population totals can be applied to more general real valued batch totals. Particular attention will be paid to the estimation of functions having the forms

$$T_1 = \sum_{i=1}^{N} a_i y_i, \quad A = \{i\}, \; z_i = a_i y_i; \tag{1.1}$$

Received February 1980; revised June 1982.

[1] The main results of this paper appear in the first author's Ph.D. dissertation, written at the University of Waterloo.

AMS 1980 *subject classification.* Primary, 62D05.

Key words and phrases. Admissibility, complex sampling estimation, sampling, surveys, unbiased minimum variance estimation.

(1.2) $$T_2 = \sum_{i=1}^N \sum_{j=i}^N c_{ij} y_i y_j, \quad A = \{(i, j) : i \le j\}, \; z_{ij} = c_{ij} y_i y_j;$$

(1.3) $$T_3 = \sum_{i=1}^{N-1} \sum_{j=i+1}^N b_{ij}(y_i - y_j)^2, \quad A = \{(i, j) : i < j\}, \; z_{ij} = b_{ij}(y_i - y_j)^2;$$

(1.4) $$T_4 = \sum_i \sum_j b_{ij}(x_i - x_j)(y_i - y_j), \quad A = \{(i, j) : i < j\}, \; z_{ij} = b_{ij}(x_i - x_j)(y_i - y_j);$$

where x and y are real variates. We shall prove that no best unbiased estimators exist for these, even in the classes of estimators which are linear in the functions y_i, $y_i y_j$, $(y_i - y_j)^2$ and $(x_i - x_j)(y_i - y_j)$ respectively. However, the generalized Horvitz-Thompson estimators of T_1, T_2 and T_4 are admissible among unbiased estimators under very general conditions, to be specified. This is not the case for T_3, of which the usual population variance is a special case, and an example will be described in Section 5 in which the sampling design is such that the generalized HT estimator for T_3 is actually inadmissible. This example suggests that there may be many cases of standard estimators of polynomial population functions which are not generally admissible; a related result, which is easier to prove, shows that the Yates-Grundy estimator for the variance of the HT estimator of the ordinary population total may be inadmissible (Biyani, 1980).

REMARK 1.1. The *bilinear* function T_4 and its corresponding HT estimator $\hat{T}_{4\pi}$ are defined on the *product space* $\mathbb{R}^N \times \mathbb{R}^N$, but the *quadratic* function T_3 and its corresponding HT estimator $\hat{T}_{3\pi}$ are defined on the *diagonal line* D of $\mathbb{R}^N \times \mathbb{R}^N$, (i.e. the subset having $x_i = y_i$, $i = 1, 2, \cdots, N$). When we consider properties such as bestness or admissibility of the estimators for T_4 we must check the whole space $\mathbb{R}^N \times \mathbb{R}^N$, but for the estimators for T_3 we only need check the subset D. That is, the *admissibility* of the estimator $\hat{T}_{4\pi}$ and the *inadmissibility* of the estimator $\hat{T}_{3\pi}$ are compatible, since when we consider the restricted subset D under a specified sampling design there may be an estimator of T_4 superior to $\hat{T}_{4\pi}$ (i.e. an estimator of T_3 superior to $\hat{T}_{3\pi}$).

In Section 2 we shall introduce the notions of batch totals and batch based estimators formally, and discuss conditions for the existence of unbiased estimators. Section 3 deals with the non-existence of optimal unbiased estimators, even in the restricted class of homogeneous linear batch based estimators of batch total functions. There follows a brief discussion of the possibility of generalizing the elegant non-existence proof of Lanke (1973) for estimators of the ordinary population total.

Lanke's proof depends on the Horvitz-Thompson (HT) estimator being admissible. Section 4 is devoted to the formulation of conditions under which the generalized (batch based) HT estimator can be proved to be admissible, in the classes of unbiased estimators or linear unbiased estimators for a batch total.

2. Notation. Let $P = \{1, 2, \cdots, N\}$ denote the *finite population* under study, and suppose that with each *unit* i is associated a multivariate characteristic value

$$w_i = (x_i, y_i, \cdots, z_i, u_i)$$

of dimension L. The *population matrix* $\mathbf{w}$ is of dimension $N \times L$, and has w_i as its ith row. It may also be written in terms of its columns $\mathbf{x}$, $\mathbf{y}$ etc., as

$$\mathbf{w} = (\mathbf{x}, \mathbf{y}, \cdots, \mathbf{z}, \mathbf{u}).$$

Let W denote the space of all $\mathbf{w}$ which are possible before sampling in a given problem, a subspace of the space of all $N \times L$ matrices with real entries.

A *sample* s is a subset of P, and a *sampling design* p is a probability function on $S = \{s : s \subseteq P\}$, the collection of all possible samples. Given a sampling design p, let $S^+(p)$ denote the collection of all samples s such that $p(s) > 0$.

A batch α is defined as a finite ordered sequence of units in P, not necessarily distinct. Thus α is representable in the form

$$\alpha = ij \cdots k$$

with $i, j, \cdots, k$ belonging to P. The *singleton batch* $\alpha = i$ can be identified with the unit i in P.

Let $s(\alpha)$ denote the set of distinct units in the batch α. We say that the sample s *covers* α (notation: $s \ni \alpha$) if $s(\alpha) \subseteq s$. For a given sampling design p and batch α the quantity

$$\pi_\alpha = \textstyle\sum_{s:s\ni\alpha} p(s) \tag{2.1}$$

is called the *inclusion probability* of the batch α. Note that π_α is always invariant under permutations of α, e.g. $\pi_{ij} = \pi_{ji}$.

Now let A denote a finite collection of batches. The *size of sample s relative to A* is denoted by $n_A(s)$ and defined as the number of batches $\alpha \in A$ which are covered by s (i.e. for which $s \ni \alpha$). For example, since P is the collection of all singleton batches i, $n_P(s)$ would be the usual sample size, namely the number of distinct units in the sample.

We shall consider mainly estimation of population functions of the form

$$G_A(\mathbf{w}) = \textstyle\sum_{\alpha\in A} G_\alpha(\mathbf{w}), \tag{2.2}$$

where $A = \{\alpha\}$ is a finite collection of batches, and $G_\alpha(\mathbf{w})$ is real valued and depends on $\mathbf{w}$ only through those w_i for which i appears in the sequence α. Such a function $G_A(\mathbf{w})$ will be referred to as a *batch total.*

REMARK 2.1. For example, each one of the population functions T_1, T_2, T_3 and T_4 is a batch total over its batch collection A defined in (1.1)–(1.4). For (1.1)–(1.3), $\mathbf{w} = \mathbf{y}$ and $G_\alpha(\mathbf{w}) = a_i y_i,\ c_{ij} y_i y_j,\ b_{ij}(y_i - y_j)^2$ respectively; for (1.4), $\mathbf{w} = (\mathbf{x}, \mathbf{y})$ and $G_\alpha(\mathbf{w}) = b_{ij}(x_i - x_j)(y_i - y_j)$.

An *estimator* for $G_A(\mathbf{w})$ will as usual be a real valued function $g(s, \mathbf{w})$ which depends on $\mathbf{w}$ only through those w_i for which $i \in s$. We shall restrict attention mainly to estimators which are *p-unbiased,* namely estimators g such that

$$E_p g(s, \mathbf{w}) = \textstyle\sum_{s\in\mathscr{S}} p(s) g(s, \mathbf{w}) = G_A(\mathbf{w})$$

and

$$\mathrm{MSE}_p g(s, \mathbf{w}) = \mathrm{Var}_p g(s, \mathbf{w}) = \textstyle\sum_{s\in\mathscr{S}} p(s) g^2(s, \mathbf{w}) - G_A^2(\mathbf{w}).$$

An estimator $g(s, \mathbf{w})$ for $G_A(\mathbf{w})$ will be called *A-linear* if it takes the form

$$g(s, \mathbf{w}) = \textstyle\sum_{\alpha:s\ni\alpha} \lambda_\alpha(s) G_\alpha(\mathbf{w}), \tag{2.3}$$

where α is restricted to A. We introduce

(2.4) $\mathscr{G}_\lambda$ = the class of all p-unbiased A-linear estimators for $G_A(\mathbf{w})$.

An example of an estimator in $\mathscr{G}_\lambda$ is the *generalized Horvitz-Thompson* (HT) estimator

$$g_\pi(s, \mathbf{w}) = \textstyle\sum_{\alpha:s\ni\alpha} \dfrac{G_\alpha(\mathbf{w})}{\pi_\alpha} \tag{2.5}$$

if $\pi_\alpha > 0$ for every $\alpha \in A$. For the population functions $T_1 - T_4$ of (1.1)–(1.4) these are respectively

$$\hat{T}_{1\pi} = \textstyle\sum_{i\in s} a_i y_i/\pi_i, \tag{2.6}$$

$$\hat{T}_{2\pi} = \sum\sum_{\substack{i,j\in s\\ i\leq j}} c_{ij} y_i y_j/\pi_{ij}, \tag{2.7}$$

$$\hat{T}_{3\pi} = \sum\sum_{\substack{i,j\in s\\ i<j}} b_{ij}(y_i - y_j)^2/\pi_{ij}, \tag{2.8}$$

$$\hat{T}_{4\pi} = \sum\sum_{\substack{i,j\in s\\ i<j}} b_{ij}(x_i - x_j)(y_i - y_j)/\pi_{ij}. \tag{2.9}$$

(When $\alpha = i$, π_α is π_i; when $\alpha = (i, j)$, π_α is π_{ij}.)

More generally, an estimator $g(s, \mathbf{w})$ will be called *A-based* if $g(s, \mathbf{w})$ depends only on $G_\alpha(\mathbf{w})$ for those α covered by s. For example, an A-based but not A-linear estimator for T_3 of (1.3), where $A = \{(i, j) : i < j\}$, $\mathbf{w} = \mathbf{y}$ and $G_{ij}(\mathbf{w}) = b_{ij}(y_i - y_j)^2$, might take the "trimmed" form

$$\sum\sum_{\substack{i,j \in s \\ i<j}} \beta_{ij}(s, \mathbf{w})(y_i - y_j)^2,$$

where

$$\beta_{ij}(s, \mathbf{w}) = \begin{cases} 0 & \text{if } \left| (y_i - y_j)^2 - \dfrac{2}{n(n-1)} \sum\sum_{\substack{i,j \in s \\ i<j}} (y_i - y_j)^2 \right| > M \\ \beta_{ij} & \text{otherwise} \end{cases}$$

for appropriate constants β_{ij}.

It should be noted that besides being the natural generalization of linear estimators for batch total estimation, A-linear estimators are appealing because they tend to have properties similar to those of the estimand. For example, the generalized HT estimator

$$\sum\sum_{\substack{i,j \in s \\ i<j}} \frac{(y_i - y_j)^2}{\pi_{ij}} \tag{2.10}$$

for

$$\textstyle\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (y_i - y_j)^2 \tag{2.11}$$

is always non-negative, and is 0 when the estimand (2.11) is 0. It may be noted that if the estimand (2.11) is written in the bilinear form

$$\textstyle\sum_{i=1}^{N} y_i^2(N-1) - 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} y_i y_j,$$

the linear batch-based HT estimator is

$$(N-1)\left(\textstyle\sum_{i \in s} y_i^2/\pi_i\right) - 2 \sum\sum_{\substack{i,j \in s \\ i<j}} y_i y_j/\pi_{ij};$$

this estimator is both non-negative for all $\mathbf{y} \in R^N$ and 0 when $y_i =$ constant only if it is equal to (2.10) (Vijayan, 1975).

A sufficient condition for the existence of A-based p-unbiased estimators of $G_A(\mathbf{w})$ is that π_α should be positive for each α for which $G_\alpha(\mathbf{w})$ is not identically constant. Conversely, we may show that this condition is necessary if each non-constant $G_\alpha(\mathbf{w})$ can be varied (by suitable choice of $\mathbf{w}$) independently of the others. For simplicity in what follows, we shall assume that for all given sampling designs, $\pi_\alpha > 0$ for all $\alpha \in A$.

3. Non-existence of p-best A-linear estimators. In this section we shall discuss conditions for A-linear estimators to belong to $\mathscr{G}_\lambda$ of (2.4), and will prove two theorems on the non-existence of minimum variance p-unbiased estimators of $G_A(\mathbf{w})$. Use will be made of two practically non-restrictive conditions in the statements of the results:

CONDITION L. There is a subset $\{\mathbf{w}^\beta : \beta \in B\}$ of W of cardinality equal to the number of batches α in A such that the rows of the matrix GW whose $(\alpha, \beta)^{\text{th}}$ element is $G_\alpha(\mathbf{w}^\beta)$ are linearly independent.

CONDITION N. There exist $s \in S^+(p)$ and $\mathbf{w}, \mathbf{w}' \in W$ such that $w_i = w_i'$ for all $i \in s$ and $G_A(\mathbf{w}) \neq G_A(\mathbf{w}')$.

It is easy to construct situations for which these are not satisfied; but it is also easy to show in the cases of $T_1 - T_4$ in (1.1) $-$ (1.4) that if the constants a_i, b_{ij}, c_{ij} are non-zero, and W is large enough that each x_i or y_i may be 1 or 0 independently of the others, then L is satisfied, and N is satisfied if the design is not a census.

The property of p-unbiasedness for A-linear estimators imposes constraints on the coefficients $\lambda_\alpha(s)$, as described in the following lemma.

LEMMA 3.1. *Suppose Condition L holds. Then an A-linear estimator*

$$g_\lambda(s, \mathbf{w}) = \textstyle\sum_{\alpha:s\ni\alpha} \lambda_\alpha(s) G_\alpha(\mathbf{w})$$

belongs to $\mathscr{G}_\lambda$ *of* (2.4) *if and only if* $\sum_{s:s\ni\alpha} p(s)\lambda_\alpha(s) = 1$ *for every* α.

PROOF. The proof follows easily from the observation that

$$E_p g_\lambda(s, \mathbf{w}) = \textstyle\sum_\alpha \{\sum_{s:s\ni\alpha} p(s)\lambda_\alpha(s)\} G_\alpha(\mathbf{w}),$$

and that for p-unbiasedness of g_λ, this must equal $\sum_\alpha G_\alpha(\mathbf{w})$. □

Given a sampling design p, the collection $S^+(p)$ will be called a *disjoint cover* of A if each $\alpha \in A$ is covered by one and only one sample $s \in S^+(p)$. Lemma 2.1 applied in this particular case proves uniqueness of a p-unbiased A-linear estimator.

COROLLARY 3.1. *If Condition* L *holds and* $S^+(p)$ *is a disjoint cover of* A, *then the unique estimator in* $\mathscr{G}_\lambda$ *is given by*

$$\textstyle\sum_{\alpha:s\ni\alpha} G_\alpha(\mathbf{w})/p(s) = g_\pi(s, \mathbf{w}),$$

the generalized HT *estimator.*

REMARK 3.1. For examples, the $S^+(p)$ of a unicluster sampling design (Cassel et al, 1977, page 67) is a disjoint cover of the batch collection $A = P = \{i\}$, and the $S^+(p)$ of a fixed sample size *two* sampling design is a disjoint cover of the batch collection $A = \{(i, j): i \neq j,\ i, j \in P\}$.

The first non-existence result we prove is for general A-based estimators, and is an analogue of Godambe's (1955) result for estimators of the population total; the proof follows that of Basu (1971).

Let $\mathscr{G}$ be a class of estimators of $G_A(\mathbf{w})$. We say that g^* is *p-best* in $\mathscr{G}$ if $g^* \in \mathscr{G}$ and

$$E_p\{g^* - G_A(\mathbf{w})\}^2 \leq E_p\{g - G_A(\mathbf{w})\}^2$$

for all $g \in \mathscr{G}$ and $\mathbf{w} \in W$; and the import of Theorem 3.1 below is that p-best p-unbiased A-based estimators can be found only in exceptional circumstances.

LEMMA 3.2. *Suppose Condition N holds. Then there is no estimator* $g(s, \mathbf{w})$ *for* $G_A(\mathbf{w})$ *which has zero mean square error for all* $\mathbf{w} \in W$.

THEOREM 3.1. *If Condition N holds then there is no p-best p-unbiased A-based estimator for* $G_A(\mathbf{w})$.

PROOF. It need only be shown that for each $\mathbf{w}_0 \in W$ a p-unbiased A-based estimator $g(s, \mathbf{w})$ exists such that

$$\mathrm{Var}_p g(s, \mathbf{w}_0) = 0.$$

Such an estimator is given by

$$G_A(\mathbf{w}_0) + \textstyle\sum_{\alpha:s\ni\alpha} \{G_\alpha(\mathbf{w}) - G_\alpha(\mathbf{w}_0)\}/\pi_\alpha. \quad \square \tag{3.1}$$

COROLLARY 3.2. *If Condition N holds for the cases of $T_1 - T_4$ of (1.1) – (1.4), there is no p-best p-unbiased A-based estimator for any of the population functions $T_1 - T_4$.*

Since the estimator (3.1) contains a constant term, it is not A-linear by our definition. Thus we may ask whether there might exist a p-best p-unbiased A-linear estimator for $G_A(\mathbf{w})$. Again, while such an estimator does exist under Condition L when $S^+(p)$ is a disjoint cover (Corollary 3.1), the answer is generally in the negative.

THEOREM 3.2. *Suppose that $S^+(p)$ is not a disjoint cover of A, and that Condition L holds. Then there is no p-best estimator in the class $\mathscr{G}_\lambda$ of (2.4).*

PROOF. Assume there exists a p-best estimator

$$g_\lambda(s, \mathbf{w}) = \textstyle\sum_{\alpha:s\ni\alpha} \lambda_\alpha(s) G_\alpha(\mathbf{w}).$$

Since

$$\textstyle\sum_{s:s\ni\alpha} p(s)\lambda_\alpha(s) = 1 \tag{3.2}$$

for all $\alpha \in A$, the coefficients $\lambda_\alpha(s)$, $s \in S^+(p)$, must satisfy the system of equations

$$\frac{\partial}{\partial\lambda_\alpha(s)}[\mathrm{Var}_p g_\lambda(s, \mathbf{w}) - 2\textstyle\sum_{\alpha\in A} r_\alpha\{\sum_{s:s\ni\alpha} p(s)\lambda_\alpha(s)\}] = 0, \tag{3.3}$$

where r_α is a Lagrange multiplier and the variance $\mathrm{Var}_p g_\lambda(s, \mathbf{w})$ is given by

$$\mathrm{Var}_p g_\lambda(s, \mathbf{w}) = \textstyle\sum_s p(s) g_\lambda^2(s, \mathbf{w}) - G_A^2(\mathbf{w}).$$

If $p(s) > 0$ and $s \ni \alpha$ then (3.3) implies

$$G_\alpha(\mathbf{w}) g_\lambda(s, \mathbf{w}) = r_\alpha. \tag{3.4}$$

If there are two samples $s_1, s_2 \in S^+(p)$ which cover α, from (3.4) we have

$$G_\alpha(\mathbf{w}) g_\lambda(s_1, \mathbf{w}) = G_\alpha(\mathbf{w}) g_\lambda(s_2, \mathbf{w}) \tag{3.5}$$

for all $\mathbf{w} \in W$; and when $G_\alpha(\mathbf{w})$ is not zero, from (3.5) we have

$$\textstyle\sum_{\beta:s_1\ni\beta} \lambda_\beta(s_1) G_\beta(\mathbf{w}) = \sum_{\beta:s_2\ni\beta} \lambda_\beta(s_2) G_\beta(\mathbf{w}).$$

That is,

$$\textstyle\sum_{\beta:s_1\ni\beta, s_1\cap s_2\not\ni\beta} \lambda_\beta(s_1) G_\beta(\mathbf{w}) + \sum_{\beta:s_1\cap s_2\ni\beta} \{\lambda_\beta(s_1) - \lambda_\beta(s_2)\} G_\beta(\mathbf{w})$$
$$- \textstyle\sum_{\beta:s_2\ni\beta, s_1\cap s_2\not\ni\beta} \lambda_\beta(s_2) G_\beta(\mathbf{w}) = 0$$

for all $\mathbf{w} \in W$ such that $G_\alpha(\mathbf{w}) \neq 0$ for some α covered by $s_1 \cap s_2$. Since by Condition L the subspace of vectors

$$\{G_\beta(\mathbf{w}),\ s_1 \ni \beta \text{ or } s_2 \ni \beta,\ G_\alpha(\mathbf{w}) \neq 0 \text{ for some } \alpha \text{ covered by } s_1 \cap s_2\}$$

will always contain a set of linearly independent vectors of cardinality at least equal to $n_A(s_1) + n_A(s_2) - n_A(s_1 \cap s_2)$, then

$$\lambda_\beta(s_1) = \lambda_\beta(s_2) \text{ for all } \beta \text{ covered by } s_1 \cap s_2,$$

$\lambda_\beta(s_1) = 0$ if $s_2 \not\ni \beta$, and

$$\lambda_\beta(s_2) = 0 \text{ if } s_1 \not\ni \beta.$$

This implies that

$$\lambda_\beta(s) = \lambda_\beta$$

for all $s \not\ni \beta$, and that

$$\lambda_\beta(s) = \lambda_\beta = 0$$

whenever β is covered by a sample which intersects with a sample not covering β. Since p is not a disjoint cover this condition is satisfied for some β; however, if $\lambda_\beta = 0$ for some β, constraints (3.2) cannot be satisfied, and this implies the theorem. □

COROLLARY 3.3. *In the cases of $T_1 - T_4$ of* (1.1) – (1.4) *suppose that $S^+(p)$ is not a disjoint cover of A. (For $T_2 - T_4$ this means in particular that $S^+(p)$ must include samples of size at least* 3.) *If also Condition L holds, then there is no p-best p-unbiased A-linear estimator for any of $T_1 - T_4$.*

Lanke (1973) has given an alternative proof of Theorem 3.3 for the case $A = P$, but it is possible to generalize his argument only partially. The difficulty, as will be seen in subsequent sections, is that we cannot always show the generalized HT estimator $g_\pi(s, \mathbf{w})$ to be admissible, as it certainly is (for sufficiently extensive W) in the special case where $A = \{i\}$, $G_i(\mathbf{w}) = y_i$.

4. p-admissibility. We begin by defining p-admissibility formally.

DEFINITION 1. Given a sampling design p, a p-unbiased estimator g is *p-superior* to another p-unbiased estimator g^* if and only if

$$\mathrm{Var}_p g^*(s, \mathbf{w}) \geq \mathrm{Var}_p g(s, \mathbf{w})$$

for all $\mathbf{w} \in W$, with strict inequality for some $\mathbf{w} \in W$.

DEFINITION 2. Given a sampling design p, an estimator g^* is *p-admissible* in a class $\{g\}$ of p-unbiased estimators if there does not exist another estimator g in $\{g\}$ which is p-superior to g^*.

The following general admissibility proof for A-based estimators is patterned upon an argument of Godambe and Joshi (1965). In the statement of the theorem,

$$W_k(A)$$

denotes the set of $\mathbf{w} \in W$ such that $G_\alpha(\mathbf{w}) \neq 0$ for exactly k of the batches $\alpha \in A$. Also, for a given $\mathbf{w}$, $S_j^* \subset S^+(p)$ will consist of those $s \in S^+(p)$ for which $G_\alpha(\mathbf{w}) \neq 0$ for exactly j of the batches α covered by s. The proof is by induction, and the kth induction step requires that the following "chain condition" be satisfied:

CONDITION C. For each $\mathbf{w}^* \in W_k(A)$ and each $s \in S_j^*$, $0 \leq j < k$, we can find a $\mathbf{w} \in W_\ell(A)$ for some $j \leq \ell < k$ such that $G_\alpha(\mathbf{w}) = G_\alpha(\mathbf{w}^*)$ for all batches $\alpha \in s$.

Condition C is generally satisfied in the situation of T_1 of (1.1), for if $y^* = (y_1, \cdots, y_N)$ from $\mathbf{w}^* \in W_k(A)$ is such that (say) $y_1, \cdots, y_k$ are the only non-zero components, and s contains only units $\{1, \cdots, j\}$ from $\{1, \cdots, k\}$ then $\mathbf{y}$ obtained from $\mathbf{y}^*$ by replacing y_k by 0 will cause the new $\mathbf{w}$ to belong to $W_\ell(A)$ where $\ell = k - 1$; and clearly $G_i(\mathbf{w}) = G_i(\mathbf{w}^*)$ for all $i \in s$, since $k \notin s$. Condition C is also generally satisfied in the situation of T_2 of (1.2). It should be noted that in this case $W_k(A)$ is non-empty only for $k = 1$, $k = 3$ (when two y_i's are non-zero), $k = 6$, etc. However, Condition C is not satisfied in general in the situations of T_3 and T_4.

THEOREM 4.1. *Let a sampling design p be given. Suppose that $W_0(A)$ is non-empty, and that the "chain condition" C is satisfied for all k. Let $g^*(s, \mathbf{w})$ be an (A-based) estimator which depends on s and* $\mathbf{w}$ *only through $G_\alpha(\mathbf{w}) \neq 0$ for $\alpha \in s$. Then g^* is p-*

admissible in the class of all p-unbiased A-based estimators for $E_p g^*(s, \mathbf{w})$. (*Here* $E_p g^*(s, \mathbf{w})$ *is not necessarily a batch total function.*)

PROOF. If the theorem is not true, there exists another p-unbiased A-based estimator g for $E_p g^*(s, \mathbf{w})$ which is p-superior to g^*. Let

$$g(s, \mathbf{w}) = g^*(s, \mathbf{w}) + h(s, \mathbf{w}).$$

Then we have

$$\sum_s p(s) h(s, \mathbf{w}) = 0 \tag{4.1}$$

for all $\mathbf{w} \in W$, and it follows from the p-superiority of g that

$$\sum_s p(s) h^2(s, \mathbf{w}) \le -2 \sum_s p(s) h(s, \mathbf{w}) g^*(s, \mathbf{w}) \tag{4.2}$$

for all $\mathbf{w} \in W$.

Now assume that $W_{k+1}(A)$ is not empty, and that $p(s)h(s, \mathbf{w})$ has been proved to be 0 for all $s \in S$ and $\mathbf{w} \in \cup_{j=0}^{k} W_j(A)$. If $\mathbf{w}^* \in W_{k+1}(A)$ then (4.1) implies that

$$\sum_{j=0}^{k+1} \sum_{s \in S_j^*} p(s) h(s, \mathbf{w}^*) = 0, \tag{4.3}$$

and (4.2) such that

$$\sum_{j=0}^{k+1} \sum_{s \in S_j^*} p(s) h^2(s, \mathbf{w}^*) \le -2 \sum_{j=0}^{k+1} \sum_{s \in S_j^*} p(s) h(s, \mathbf{w}^*) g^*(s, \mathbf{w}^*). \tag{4.4}$$

Now if $s \in S_j^*$ for some $0 \le j \le k$ we have $G_\alpha(\mathbf{w}^*) \ne 0$ for exactly j batches $\alpha \in s$, and by (C) there exists ℓ satisfying $j \le \ell \le k$ and a $\mathbf{w} \in W_\ell(A)$, such that $G_\alpha(\mathbf{w}) = G_\alpha(\mathbf{w}^*)$ for all $\alpha \in s$. Thus we have $g^*(s, \mathbf{w}) = g^*(s, \mathbf{w}^*)$ and $h(s, \mathbf{w}) = h(s, \mathbf{w}^*)$, and since by hypothesis $p(s)h(s, \mathbf{w}) = 0$ for $s \in S$ and $\mathbf{w} \in \cup_{j=0}^{k} W_j(A)$, it follows that $p(s)h(s, \mathbf{w}^*) = 0$. Hence $p(s)h(s, \mathbf{w}^*) = 0$ for all $s \in \cup_{j=0}^{k} S_j^*$. Thus from (4.3) and (4.4), we have

$$\sum_{s \in S_{k+1}^*} p(s) h(s, \mathbf{w}^*) = 0 \tag{4.5}$$

and

$$\sum_{s \in S_{k+1}^*} p(s) h^2(s, \mathbf{w}^*) \le -2 \sum_{s \in S_{k+1}^*} p(s) h(s, \mathbf{w}^*) g^*(s, \mathbf{w}^*). \tag{4.6}$$

However, in (4.6) $g^*(s, \mathbf{w}^*)$ for all $s \in S^*_{k+1}$ is constant, so that

$$\sum_{s \in S_{k+1}^*} p(s) h^2(s, \mathbf{w}^*) \le - g^*(s, \mathbf{w}) \sum_{s \in S_{k+1}^*} p(s) h(s, \mathbf{w}^*) = 0,$$

and $p(s)h(s, \mathbf{w}^*) = 0$ for all $s \in S_{k+1}$, hence for all $s \in S$.

Since clearly $p(s)h(s, \mathbf{w}) = 0$ for all $s \in S$ and $\mathbf{w} \in W_0(A)$, by induction we must have $p(s)h(s, \mathbf{w}) = 0$ for all $s \in S$ and $\mathbf{w} \in W$, and the theorem is proved. □

Theorem 4.1 is immediately applicable to the generalized HT estimator.

COROLLARY 4.1. *Given a sampling design p, suppose that Condition* C *is satisfied. Then the generalized* HT *estimator* g_π *is a p-admissible estimator in the class of all p-unbiased estimators for* $G_A(\mathbf{w})$.

COROLLARY 4.2. *Given a sampling design p, suppose W is sufficiently extensive (e.g. suppose each* y_i *may be 0 or not independently of the others). Then*

(i) $\hat{T}_{1\pi}$ *of* (2.6) *is p-admissible in the class of all p-unbiased estimators for* T_1 *of* (1.1)

(ii) $\hat{T}_{2\pi}$ *of* (2.9) *is p-admissible in the class of all p-unbiased A-based estimators for* T_2 *of* (1.2).

It is easy to see that the role of the 0 batch value in Theorem 4.1 can be assumed by any number d. For if we replace 0 by d in the chain condition and the definitions of $W_k(A)$, S_j^*, the proof remains valid.

The chain condition C is, as we have seen, rather restrictive. However, to prove p-admissibility of g_π in the smaller class of p-unbiased A-linear estimators we may use an analogue which is much easier to verify, involving $W_1(A)$.

THEOREM 4.2. *Suppose that $W_0(A)$ is non-empty and that for each $\alpha \in A$ the set $W_1(A)$ contains at least one* $\mathbf{w}$ *for which*

$$G_\alpha(\mathbf{w}) \neq 0$$

and

$$G_\beta(\mathbf{w}) = 0, \beta \neq \alpha.$$

Then given a sampling design p, $g_\pi(s, \mathbf{w})$ is p-admissible in the class of all p-unbiased A-linear estimators for $G_A(\mathbf{w})$.

PROOF. If g is A-linear, p-unbiased and p-superior to g_π, then $h(s, \mathbf{w}) = g(s, \mathbf{w}) - g^*(s, \mathbf{w})$ takes a linear form

$$h(s, \mathbf{w}) = \textstyle\sum_{\alpha\in s} \mu(s, \alpha) G_\alpha(\mathbf{w}).$$

The arguments of Theorem 4.1 can be used to show that $p(s)h(s, \mathbf{w}) = 0$ for every $\mathbf{w} \in W_1(A)$. Thus from the hypothesis on $W_1(A)$

$$\mu(s, \alpha) = 0$$

for every $s \in S^+(p)$, $\alpha \in s$. Consequently $p(s)h(s, \mathbf{w}) = 0$ for every $\mathbf{w} \in W$ and $s \in S^+(p)$, and the theorem is proved. □

From Theorems 4.1 and 4.2 we have the following.

COROLLARY 4.3. *Given a sampling design p, suppose W is sufficiently extensive. Then $\hat{T}_{1\pi}$, $\hat{T}_{2\pi}$ and $\hat{T}_{4\pi}$ are respectively p-admissible in the classes of all p-unbiased A-linear estimators for T_1, T_2 and T_4.*

PROOF. The results for $\hat{T}_{1\pi}$ and $\hat{T}_{2\pi}$ follow from Corollary 4.2. For $\hat{T}_{4\pi}$, we note that to make

$$(x_i - x_j)(y_i - y_j) \neq 0 \Leftrightarrow (i, j) = (i_0, j_0)$$

we need only ensure that

$$x_i \neq 0 \Leftrightarrow i = i_0$$

and

$$y_j \neq 0 \Leftrightarrow j = j_0. \; \square$$

Although the chain condition C is not satisfied for T_3, there is a relatively simple sufficient condition for $\hat{T}_{3\pi}(\mathbf{w})$ to be p-admissible among p-unbiased estimators. That is, we may note that

$$T_3(\mathbf{w}) = \textstyle\sum_{i=1}^{N} \sum_{j=i}^{N} c_{ij} y_i y_j,$$

where $c_{ij} = -2b_{ij}$ if $i < j$ and $c_{ii} = \sum_{j=i+1}^{N} b_{ij} + \sum_{j=1}^{i-1} b_{ji}$. Also

$$\hat{T}_{3\pi}(\mathbf{w}) = \sum_{\substack{i,j\in s\\ i<j}}\sum (-2\, b_{ij}/\pi_{ij}) y_i y_j + \textstyle\sum_{i\in s} y_i^2 \left(\sum_{j\in s, j>i} b_{ij}/\pi_{ij} + \sum_{j\in s, j<i} b_{ij}/\pi_{ij}\right).$$

Defining $b_{ij} = b_{ji}$ if $i > j$, we can see that if

(4.7) $$\textstyle\sum_{j\neq i, j\in s} b_{ij} = \pi_i \sum_{j\neq i, j\in s} b_{ji}/\pi_{ij}$$

for every i and $s \in S^+(p)$ containing i, then $\hat{T}_{3\pi}$ has the form of $\hat{T}_{2\pi}$, and is p-admissible.

The estimator $\hat{T}_{4\pi}(\mathbf{w})$ will also be p-admissible under (4.7), which is clearly highly restrictive. It is satisfied if the design and the estimators $\hat{T}_{3\pi}$ are symmetric with respect to certain groups of permutations of $\{1, \cdots, N\}$, namely those for which $S^+(p)$ is *separably transformable* in the sense of Liu (1981). When $b_{ij} \equiv 1$, (4.7) is satisfied for simple random sampling, stratified random sampling, and two stage simple random sampling with PSU's of constant size.

5. Inadmissibility of $\hat{T}_{3\pi}$. The following example shows that $\hat{T}_{3\pi}$ of (2.8) need not be admissible in general. We shall consider here estimators of

$$\sum_{\substack{i,j=1\\ i<j}}^{N}\sum^{N} (y_i - y_j)^2,$$

taking $b_{ij} = 1$, which is equivalent to the usual population variance, but it will be clear that such examples can be constructed for any choice of b_{ij}.

Let $N = 4$, and $S^+(p)$ consist of samples of size 3. Denote the sample probabilities by

$$a = p(\{1, 2, 3\}) > 0, \quad b = p(\{1, 2, 4\}) > 0,$$
$$c = p(\{1, 3, 4\}) > 0, \quad d = p(\{2, 3, 4\}) > 0.$$

Let

$$\hat{T}_{3\pi} = \sum_{\substack{i,j\in s,\\ i<j}}\sum (y_i - y_j)^2/\pi_{ij},$$

and consider a competing A-based estimator $\hat{T}_{3\lambda}$ given by

$$\hat{T}_{3\lambda} = \hat{T}_{3\pi} + h,$$

where h is given by Table 1.

TABLE 1.

Sample s	Estimator h
$\{1, 2, 3\}$	$\frac{2\varepsilon}{a}(y_1 - y_2)^2 + \frac{2\varepsilon}{a}(y_1 - y_3)^2 + \frac{2\varepsilon}{a}(y_2 - y_3)^2$
$\{1, 2, 4\}$	$-\frac{2\varepsilon}{b}(y_1 - y_2)^2 + \frac{\varepsilon}{b}(y_1 - y_4)^2 + \frac{\varepsilon}{b}(y_2 - y_4)^2$
$\{1, 3, 4\}$	$-\frac{2\varepsilon}{c}(y_1 - y_3)^2 - \frac{\varepsilon}{c}(y_1 - y_4)^2 + \frac{\varepsilon}{c}(y_3 - y_4)^2$
$\{2, 3, 4\}$	$-\frac{2\varepsilon}{d}(y_2 - y_3)^2 - \frac{\varepsilon}{d}(y_2 - y_4)^2 - \frac{\varepsilon}{d}(y_3 - y_4)^2.$

Letting $\eta = w_1 - w_2$, $\xi = w_2 - w_3$ and $\tau = w_3 - w_4$, we obtain $\mathrm{Var}_p(\hat{T}_{3\lambda}) - \mathrm{Var}_p(\hat{T}_{3\pi}) = A\eta^4 + B\xi^4 + C\tau^4 + D\eta^3\xi + E\eta^3\tau + F\eta\xi^3 + G\xi^3\tau + I\xi\tau^3 + H\eta\tau^3 + J\eta^2\xi^2 + K\eta^2\tau^2 + L\xi^2\tau^2 + M\eta^2\xi\tau + N\eta\xi^2\tau + O\eta\xi\tau^2$ where the coefficients $A, \cdots, O$ are functions of a, b, c, d and ε. This homogeneous fourth degree polynomial in η, ξ, τ can be shown to have zero as its maximal value (attained at the origin) when

$$a = 0.75,\ b = 0.15,\ c = 0.07,\ d = 0.03,\ \varepsilon = 0.01;$$

in such a case, $\hat{T}_{3\lambda}$ is then superior to $\hat{T}_{3\pi}$, and $\hat{T}_{3\pi}$ is not admissible. In this particular example, $\hat{T}_{3\lambda}$ is everywhere nonnegative. Further details of the proof are available on request.

REFERENCES

BASU, D. (1971). An essay on the logical foundations of survey sampling Part I. Godambe, V. P. and Sprott, D. A. (editors): *Foundations of Statistical Inference.* Holt, Rinehart and Winston, Toronto.

BIYANI, S. H. (1980). On inadmissibility of the Yates-Grundy variance estimator in unequal probability sampling. *J. Amer. Statist. Assoc.* **75** 709–712.

CASSEL, C. M., SÄRNDAL, C. E. and WRETMAN, J. H. (1977). *Foundations of Inference in Survey Sampling.* Wiley-Interscience, New York.

GODAMBE, V. P. (1955). A unified theory of sampling from finite populations. *J. Roy. Statist. Soc. B* **17** 269–278.

GODAMBE, V. P. and JOSHI, V. M. (1965). Admissibility and Bayes estimation in sampling finite populations I. *Ann. Math. Statist.* **36** 1707–1722.

HANURAV, T. V. (1966). Some aspects of unified sampling theory. *Sankhyā Ser. A* **28** 175–204.

KISH, L. and FRANKEL, M. R. (1974). Inference from complex samples. *J. Roy. Statist. Soc. B* **36** 1–37.

LANKE, J. (1973). On UMV-estimators in survey-sampling. *Metrika* **20** 196–202.

LIU, T. P. (1981). A general completeness theorem in sampling theory. *J. Roy. Statist. Soc. B*, to appear.

LIU, T. P. (1979). Optimum estimation for multivariate finite populations. Ph.D. dissertation, University of Waterloo, Waterloo, Ontario.

VIJAYAN, K. (1975). On estimating the variance in unequal probability sampling. *J. Amer. Statist. Assoc.* **70** 713–716.

WOODRUFF, R. S. (1971). A simple method for approximating the variance of a complicated estimator. *J. Amer. Statist. Assoc.* **66** 411–414.

DEPARTMENT OF SYSTEMS DESIGN
UNIVERSITY OF WATERLOO
WATERLOO, CANADA N2L 3G1

DEPARTMENT OF STATISTICS
UNIVERSITY OF WATERLOO
WATERLOO, CANADA N2L 3G1