The Annals of Statistics
1984, Vol. 12, No. 2, 675–684

ADMISSIBLE ESTIMATORS FOR THE TOTAL OF A STRATIFIED POPULATION THAT EMPLOY PRIOR INFORMATION[1]

BY STEPHEN VARDEMAN AND GLEN MEEDEN

Iowa State University

We consider the problem of the estimation of the total of a stratified finite population. For two levels of prior knowledge about the stratification, we provide Bayes and pseudo-Bayes estimators that make use of this prior knowledge in sensible ways. We then note that admissibility results can be established for these estimators using the techniques of Meeden and Ghosh (1982, 1983) and indicate some possible natural extensions of the present work.

1. Introduction. We suppose that a population contains units labeled 1, 2, $\cdots$, N and that y_i is the value of a characteristic attached to unit i. The vector $\mathbf{y} = (y_1, y_2, \cdots, y_N) \in \mathbb{R}^N$ is unknown and of particular interest is the function of $\mathbf{y}$

$$\tau = \sum_{i=1}^{N} y_i.$$

We will further assume that the population is stratified into strata 1, 2, $\cdots$, J and that attached to each unit i is a stratum membership j_i. The stratification may represent a division of the population into groups on the basis of the values of the y_i or on the basis of some other variable(s) possibly, but not necessarily, related to the y_i.

In what follows, we will make a variety of assumptions about one's knowledge concerning $\mathbf{j} = (j_1, j_2, \cdots, j_N)$. The case where $\mathbf{j}$ is completely known and is thus available for use in constructing estimators of τ is the situation of usual stratified sampling. In situations where the j_i are known only for those units sampled, it is common to use the term poststratification and the strata are sometimes called domains of study (see for example, Chapters 2 and 5A of Cochran, 1977). We will use a framework that covers these two cases and others as well. That is, for $s = \{i_1, i_2, \cdots, i_n\}$ and $s^* = \{i_1, i_2, \cdots, i_n, i_{n+1}, \cdots, i_{n^*}\}$ subsets of $\{1, 2, \cdots, N\}$ of size n and n^* respectively (with $s \subset s^*$) we will term the pair (s, s^*) a sample of sizes n and n^*. A probability distribution over those pairs (s, s^*) where $s \neq \phi$, Δ, is called a sampling design, and a function of $(s, s^*, \mathbf{y}, \mathbf{j})$ that depends on $(\mathbf{y}, \mathbf{j})$ only through $\mathbf{y}_s = (y_{i_1}, y_{i_2}, \cdots, y_{i_n})$ and $\mathbf{j}_s = (j_{i_1}, j_{i_2}, \cdots, j_{i_n}, j_{i_{n+1}}, \cdots, j_{i_{n^*}})$ will be termed an estimator of τ. As noted above, the cases where the design is such that $n^* \equiv N$ or $n^* \equiv n$ are well documented in the sampling literature. The intermediate case, where with Δ probability 1, $n < n^* < N$, is the case of so called double sampling for stratification (see Chapter 12 of Cochran, 1977).

Received June 1982; revised October 1983.
[1] This research supported by NSF Grant No. MCS-8202116.
AMS 1980 *subject classifications.* Primary 62D05, 62C15.
Key words and phrases. Finite population sampling, stratification, prior information, admissibility, stepwise Bayes.

Our concern here is with estimators of τ that incorporate prior information about the stratification. In Sections 2 and 3 we introduce two such estimators and give Bayesian and pseudo-Bayesian arguments for their use. In Section 4 we outline how recent stepwise Bayes techniques of Meeden and Ghosh can be used to establish admissibility results for the estimators under squared error loss. Section 5 contains some comments about natural extensions of this work and some closing remarks.

2. Estimation where prior information about stratum memberships is vague. Consider first a situation where a priori one's beliefs about the likely stratum memberships of all units are exchangeable. Suppose further that $\pi_1, \pi_2, \cdots, \pi_J$ (with each $\pi_j \geq 0$ and $\sum_{j=1}^{J} \pi_j = 1$) are prior guesses at the relative sizes of the strata and $\mu_1, \mu_2, \cdots, \mu_J$ are guessed means for the strata. Then for constants $M, M_1, M_2, \cdots, M_J$ each belonging to $[0, \infty]$ an estimator of τ incorporating these a priori values is

$$t = \sum_{i \in s} y_i + \sum_{j=1}^{J} \left[(n_j^* - n_j) + (N - n^*) \left(\frac{M\pi_j + n_j^*}{M + n^*} \right) \right] \left[\frac{M_j}{M_j + n_j} \mu_j + \frac{n_j}{M_j + n_j} \bar{y}_j \right]$$

where n_j and n_j^* are respectively the number of units in s and s^* belonging to stratum j, and $\bar{y}_j$ is the mean of those y_i's attached to units in s belonging to stratum j. When $n_j = 0$ we'll understand $\bar{y}_j$ to be 0 and when both M_j and n_j are 0, we'll take the factor

$$\hat{\mu}_j = \frac{M_j}{M_j + n_j} \mu_j + \frac{n_j}{M_j + n_j} \bar{y}_j$$

to be μ_j.

To motivate this estimator, notice that $\hat{\mu}_j$ is a Bayes-like predictor of any unobserved y_i known to belong to stratum j and that

$$\hat{\pi}_j = \frac{M\pi_j + n_j^*}{M + n^*}$$

is a Bayes-like estimator of the probability that a unit with $i \notin s^*$ belongs to stratum j. t is then obtained from the expression $\tau = \sum y_i$ replacing each y_i having $i \in s^* - s$ with $\hat{\mu}_{j_i}$ and each y_i having $i \notin s^*$ with the $\hat{\pi}_j$ weighted average of the $\hat{\mu}_j$'s. The constants $M, M_1, M_2, \cdots, M_J$ of course control the relative weightings of the prior and sample values, ranging from domination of sample information in "0" cases to domination of the prior values in "∞" cases. The possibility that these constants are all 0 includes in our discussion the classical estimators common in stratified sampling, poststratification and double sampling for stratification. (In this regard, compare t with all M's equal to 0 to the expressions on pages 91, 134 and 328 of Cochran (1977) respectively.)

As for other cases of t that have appeared in the literature, the special case where $J = 1$ has been discussed in detail in Vardeman and Meeden (1983a), while the general J case of t in the instance that $n^* = N$ has been treated by Binder (1982). The original motivation for the present work was provided by the paper of Hidiroglou and Srinath (1981). Their $\hat{Y}_1$ is the special case of t appropriate where Δ is such that $n \equiv n^*$ and $J = 2$, $M = \infty$, $M_1 = 0$, $\pi_1 = 1$, $\pi_2 = 0$ and stratum 1 consists of exactly those units i with $y_i \leq \gamma$, for γ a large positive constant.

For cases of t where M and each M_j and π_j are positive, it is possible to give proper Bayesian derivations for t. One such, appropriate when in fact M and each M_j are in $(0, \infty)$, is to note that t is a version of the conditional mean of τ given the observable $(\mathbf{y}_s, \mathbf{j}_{s^*})$ under a joint distribution for $(\mathbf{y}, \mathbf{j})$ specified as follows. Suppose first that given hyperparameters $\alpha_1, \alpha_2, \cdots, \alpha_J$ with each $\alpha_j \geq 0$ and $\sum \alpha_j = 1$, the entries of $\mathbf{j}$ are iid according to $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_J)$ and that $\boldsymbol{\alpha}$ has a Dirichlet distribution with parameters $M\pi_1, M\pi_2, \cdots, M\pi_J$. (In the terminology of Ferguson (1973), the entries of $\mathbf{j}$ form a sample from a Dirichlet process with parameter measure on $\{1, 2, \cdots, J\}$ defined by the $M\pi_j$'s.) Then suppose that the conditional distribution of $\mathbf{y}$ given $\mathbf{j}$ is made up of J independent factors, those y_i with $j_i = j$ constituting a sample from a Dirichlet process with parameter measure $M_j\beta_j$, where β_j is a probability distribution on $\mathbb{R}$ with $\mu_j = \int x \, d\beta_j(x)$, for each $j = 1, 2, \cdots, J$. (We should remark that in the spirit of Ericson (1969), products of two stage normal priors could, for example, be used at this point as well.)

Proper Bayesian derivations of t in cases where M and/or some of the M_j's are ∞ come about by replacing the corresponding Dirichlet process components above with the (limiting as the corresponding M goes to ∞) distribution of independence. And although cases of t where M or some of the M_j's are 0 don't have Bayesian justifications, one might argue at appropriate points in the above derivation (as in Vardeman and Meeden, 1983a, or earlier Basu, 1971, and Godambe, 1966) that when prior information is nill it is appropriate to use as a pseudo-posterior the empiric distribution of what one observes, and obtain a pseudo-Baysian argument for use of t.

Regardless of whether or not one finds the distributional assumptions for $(\mathbf{y}, \mathbf{j})$ that we used to produce t appealing, we would argue that t makes use of the kind of prior information about $(\mathbf{y}, \mathbf{j})$ that can be available, in a sensible fashion. Even an only partially Bayesian sampler armed with guesses at the relative sizes of his strata and strata means, and a notion of how he wants to weight the various guesses against the sample information, ought to find t intuitively appealing. In addition, we will argue in Section 4 that t possesses some attractive admissibility properties.

3. Estimation where prior information about stratum memberships is less vague. In this section we present a generalization of the estimator t that will be useful in problems where one's prior information about $\mathbf{j}$ is sharper than that assumed in the previous section, in that it is not exchangeable. As motivation for such a discussion, consider a situation where the strata are in fact defined in terms of the y_i themselves, with perhaps, "low", "medium" and "high" strata having been defined in terms of the unknown values of the y_i. Suppose further that (perhaps on the basis of a census of y's at a previous period) the sampler can segment the N units into several groups indicating "likely stratum membership". In the present example, a three group segmentation into "likely low", "likely medium", and "likely high" seems natural, but a two group segmentation into, say, "likely low to medium" and "likely medium to high" groups would also be possible, as would segmentations into more than three groups. In any case, though the sampler might well have the information needed to employ t, it is an inappropriate estimator in that it ignores the prior information implicit in his ability to segment the population. We proceed to define a generalization of t which would be appropriate in circumstances similar to these.

Suppose that in addition to y_i and j_i, there is attached to each unit i a variable k_i taking an integer value from 1 to K and suppose that $\mathbf{k} = (k_1, k_2, \cdots, k_N)$ is completely known. (In the example above, the values k_i would specify the group memberships established by the sampler in his segmentation of the population using his prior information and beliefs as opposed to the actual stratification of the population.) We'll further let N_{kj}, n^*_{kj} and n_{kj} be respectively the number of units in the population, s^*, and s with $k_i = k$ and $j_i = j$ and indicate sums of these with the usual dot notation. Then for each k take $\pi_{k1}, \pi_{k2}, \cdots, \pi_{kJ}$ (with each $\pi_{kj} \geq 0$ and $\sum_{j=1}^{J} \pi_{kj} = 1$) as prior guesses at the relative sizes of $N_{k1}, N_{k2}, \cdots, N_{kJ}$ and again let $\mu_1, \mu_2, \cdots, \mu_J$ be guessed means for the strata. Then for constants $\tilde{M}_1, \tilde{M}_2, \cdots, \tilde{M}_K, M_1, M_2, \cdots, M_J$ each belonging to $[0, \infty]$, our generalization of t is

$$\tilde{t} = \sum_{i\in s} y_i + \sum_{j=1}^{J} \left[(n^*_{\cdot j} - n_{\cdot j}) + \sum_{k=1}^{K} (N_{k\cdot} - n^*_{k\cdot}) \left(\frac{\tilde{M}_k \pi_{kj} + n^*_{kj}}{\tilde{M}_k + n^*_{k\cdot}} \right) \right] \cdot \left[\frac{M_j}{M_j + n_{\cdot j}} \mu_j + \frac{n_{\cdot j}}{M_j + n_{\cdot j}} \bar{y}_j \right],$$

where as before, $\bar{y}_j$ is the mean of those y_i with $i \in s$ and $j_i = j$, we take $\bar{y}_j = 0$ when $n_j = 0$, and

$$\hat{\mu}_j = \frac{M_j}{M_j + n_{\cdot j}} \mu_j + \frac{n_{\cdot j}}{M_j + n_{\cdot j}} \bar{y}_j$$

is understood as μ_j when both M_j and $n_{\cdot j}$ are 0 and

$$\hat{\pi}_{kj} = \frac{\tilde{M}_k \pi_{kj} + n^*_{kj}}{\tilde{M}_k + n^*_{k\cdot}}$$

is understood as π_{kj} when both $\tilde{M}_k$ and $n^*_{k\cdot}$ are 0.

$\tilde{t}$ can be obtained from the expression $\tau = \sum y_i$ by replacing each y_i with $i \in s^* - s$ and $j_i = j$ by $\hat{\mu}_j$ and each y_i with $i \notin s^*$ and $k_i = k$ with the $\hat{\pi}_{kj}$ weighted average of the $\hat{\mu}_j$'s. The M_j and $\tilde{M}_k$ of course govern how strongly the prior stratum means and guessed stratum membership probabilities for the various indices k are weighted against the sample information.

Bayesian and pseudo-Bayesian justifications for $\tilde{t}$ can be made by modifying slightly the derivations outlined in Section 2 for t. The modifications necessary involve only specification of the marginal distribution of $\mathbf{j}$ (and not the conditional distribution of $\mathbf{y}$ given $\mathbf{j}$). Basically, the exchangeable structures assumed in the derivation of t must be replaced by a product of K exchangeable structures, one for each set of indices i with $k_i = 1, 2, \cdots, K$. For example, in the case that each $\tilde{M}_1, \tilde{M}_2, \cdots, \tilde{M}_K$ is in $(0, \infty)$, we would assume that for each $k = 1, 2, \cdots, K$ the j_i corresponding to those i with $k_i = k$ are a sample from a Dirichlet process and that those samples from K Dirichlet processes are independent.

It seems to the authors that instances of $\tilde{t}$ with positive $\tilde{M}_k$'s and/or M_j's should be useful to all but the most adamantly nonBayesian samplers, and that even such individuals should consider

$$\sum_{i\in s} y_i + \sum_{j=1}^{J} \left[(n^*_{\cdot j} - n_{\cdot j}) + \sum_{k=1}^{K} (N_{k\cdot} - n^*_{k\cdot}) \frac{n^*_{kj}}{n^*_{k\cdot}} \right] \bar{y}_j,$$

(the "all M's and $\tilde{M}$'s equal to 0" version of $\tilde{t}$ provided each $n_{\cdot j} > 0$ and each $n^*_{k\cdot} > 0$) to be an attractive way to make use of an ability to segment the population, in the estimation of τ.

The classical properties of t and $\tilde{t}$, such as bias and mean squared error, of course depend on Δ, the sampling design, and appear to us to require case by case consideration as the values of the M's and $\tilde{M}$'s and the sampling design change. One type of weak optimality property that can be established for t and $\tilde{t}$ in some generality is that of admissibility. In the next section we note what the recent techniques of Meeden and Ghosh (1982, 1983) yield in the way of admissibility results for t and $\tilde{t}$.

4. Admissibility results for t and $\tilde{t}$. If $e(s, s^*, \mathbf{y}_s, \mathbf{j}_{s^*})$ is an estimator of τ and one employs the sampling design Δ, then the mean squared error suffered in the estimation of τ is

$$R(e, \Delta, (\mathbf{y}, \mathbf{j})) = \sum_{(s,s^*)} (e(s, s^*, \mathbf{y}_s, \mathbf{j}_{s^*}) - \tau)^2 \Delta((s, s^*)).$$

Various notions of admissibility are possible here, depending upon which arguments of R one wishes to fix and how widely the others are allowed to vary. Two definitions that we will use are the following.

DEFINITION 4.1. If Δ is a design, $\mathscr{A}$ is a subset of $\mathbb{R}^N \times \{1, 2, \cdots, J\}^N$ and there exists no estimator e^* with

$$R(e^*, \Delta^*, (\mathbf{y}, \mathbf{j})) \le R(e, \Delta, (\mathbf{y}, \mathbf{j}))$$

for all $(\mathbf{y}, \mathbf{j}) \in \mathscr{A}$ with strict inequality for some $(\mathbf{y}, \mathbf{j}) \in \mathscr{A}$, we will say that e is admissible for design Δ and parameter set $\mathscr{A}$.

By further allowing Δ to vary over some class of designs, we get the notion of global or uniform admissibility of Joshi (1966) and Godambe (1966). That is, we will make:

DEFINITION 4.2. If $\mathscr{C}$ is a class of designs containing Δ, $\mathscr{A}$ is a subset of

$\mathbb{R}^N \times \{1, 2, \cdots, J\}^N$ and there exists no estimator-design pair (e^*, Δ^*) with $\Delta^* \in \mathscr{G}$ and

$$R(e^*, \Delta^*, (\mathbf{y}, \mathbf{j})) \leq R(e, \Delta, (\mathbf{y}, \mathbf{j}))$$

for all $(\mathbf{y}, \mathbf{j}) \in \mathscr{A}$ with strict inequality for some $(\mathbf{y}, \mathbf{j}) \in \mathscr{A}$, we will say that the pair (e, Δ) is uniformly admissible relative to $\mathscr{G}$ when the parameter set is $\mathscr{A}$.

For each $j = 1, 2, \cdots, J$ let $\mathscr{B}_j$ be a subset of $\mathbb{R}$ and take $\mathscr{B} = (\mathscr{B}_1, \cdots, \mathscr{B}_J)$. In the present context a natural type of $\mathscr{A}$ for use in admissibility considerations is

$$\mathscr{A}_{\mathscr{B}} = \{(\mathbf{y}, \mathbf{j}) \mid \mathbf{j} \in \{1, 2, \cdots, J\}^N \text{ and } y_i \in \mathscr{B}_{j_i} \text{ for all } i\},$$

although for the case of $\tilde{t}$ one might possibly wish to consider some subset of $\mathscr{A}_{\mathscr{B}}$ where the value of k_i might preclude the possibility that $j_i = j$ for one or more j's. Choices of finite $\mathscr{B}_j$'s are especially tractable and, as noted by Meeden and Ghosh (1982, 1983), can lead to admissibility theorems for nonfinite choices of $\mathscr{A}$ as corollaries. For example, if for each $(\mathbf{y}, \mathbf{j}) \in \mathscr{A}$ there is a choice of finite $\mathscr{B}_1, \mathscr{B}_2, \cdots, \mathscr{B}_J$ such that $(\mathbf{y}, \mathbf{j}) \in \mathscr{A}_{\mathscr{B}}$ and an estimator e is admissible (or an estimator-design pair (e, Δ) is uniformly admissible) with parameter set $\mathscr{A}_{\mathscr{B}}$, then e (or (e, Δ)) is admissible (uniformly admissible) when the parameter set is $\mathscr{A}$. In what follows we will thus be concerned primarily with parameter sets of the form $\mathscr{A}_{\mathscr{B}}$ for finite $\mathscr{B}_j$'s, but will also point out where some nonfinite $\mathscr{A}$ admissibility theorems follow as corollaries.

Consider first the problem of establishing the admissibility of t or $\tilde{t}$ for a fixed design Δ and parameter set $\mathscr{A}_{\mathscr{B}}$ with finite $\mathscr{B}_1, \mathscr{B}_2, \cdots, \mathscr{B}_J$. If one is to prove the admissibility of t or $\tilde{t}$, it stands to reason that the prior means $\mu_1, \mu_2, \cdots, \mu_J$ should in some sense be consonant with the parameter set. We will say that $\mu = (\mu_1, \mu_2, \cdots, \mu_J)$ and $\mathscr{B}$ are *compatible* provided for each $j = 1, 2, \cdots, J$ there is a distribution on the set $\mathscr{B}_j$ with mean μ_j. Then with this terminology it is possible to use the characterization of admissible estimators in finite problems that appears as Theorem 1 of Meeden and Ghosh (1981) and prove:

THEOREM 4.1. *For any design Δ, if each $\mathscr{B}_1, \mathscr{B}_2, \cdots, \mathscr{B}_J$ is finite and μ and $\mathscr{B}$ are compatible, then the estimators t and $\tilde{t}$ are admissible for the design Δ and parameter set $\mathscr{A}_{\mathscr{B}}$.*

According to the characterization of Meeden and Ghosh, to prove this theorem, it suffices to specify a partition of $\mathscr{A}_{\mathscr{B}}$ and a sequence of (mutually orthogonal) priors on the elements of this partition and show that the estimator in question is stepwise Bayes with resepct to the sequence. This can be done for both t and its generalization $\tilde{t}$. Unfortunately, many different cases must be considered and the notational burden, which is severe for t, is even worse for $\tilde{t}$. The unpublished Vardeman and Meeden (1982), which is available from the authors, presents a complete proof of the admissibility of t in the case where $n = n^*$ and $\mathscr{B}_1, \cdots, \mathscr{B}_J$ are disjoint. In fact, these restrictions are not necessary and that proof carries over to the present case. The proof for $\tilde{t}$ then comes about as a fairly obvious but notationally unpleasant modification of the proof for t. Here, to illustrate the type of argument that is required, without developing all the notation and different cases needed to prove the theorem in general, we will prove only the admissibility of $\tilde{t}$ where all $\tilde{M}_k$'s and M_j's are 0.

PROOF OF THEOREM 4.1 WHEN ALL M'S ARE 0. Without loss of generality, we will suppose that the units are indexed in such a way that the first $N_{1.}$ units have $k_i = 1$, the next $N_{2.}$ units have $k_i = 2$, etc. and take $T_k = \sum_{\ell=1}^{k} N_{\ell.}$ for $k = 1, 2, \cdots, K$. Then with $\mathscr{B}_j = \{b_{j1}, b_{j2}, \cdots, b_{jp_j}\}$ (p_j is the number of different values in $\mathscr{B}_j$), we proceed to define a partition of $\mathscr{A}_{\mathscr{B}}$. For $k = 1, 2, \cdots, K$, let ℓ_k be the number of different values amongst $\{j_{T_{k-1}+1}, j_{T_{k-1}+2}, \cdots, j_{T_k}\}$ and take $\ell = (\ell_1, \ell_2, \cdots, \ell_K)$. Similarly, for $j = 1, 2, \cdots, J$, let m_j be the number of different elements of $\mathscr{B}_j$ represented amongst $\{y_i \mid j_i = j\}$ and take $\mathbf{m} = (m_1, m_2, \cdots, m_J)$. Then for each pair $(\ell^*, \mathbf{m}^*)$ with ℓ_k^*'s between 1 and K and m_j^*'s respectively between 1 and p_j, we will let $\mathscr{A}_{\ell^*,\mathbf{m}^*}$ stand for the (possibly void) subset of $\mathscr{A}_{\mathscr{B}}$ containing all $(\mathbf{y}, \mathbf{j})$ with $(\ell, \mathbf{m}) = (\ell^*, \mathbf{m}^*)$. Clearly, the nonvoid $\mathscr{A}_{\ell,\mathbf{m}}$ partition $\mathscr{A}_{\mathscr{B}}$.

Next, we must define a distribution on each nonvoid $\mathscr{A}_{\ell,\mathbf{m}}$. We will do this by giving first a distribution for $\mathbf{j}$ and then a conditional distribution for $\mathbf{y}$ given $\mathbf{j}$. Let ϕ_k be a distribution for $(j_{T_{k-1}+1}, j_{T_{k-1}+2}, \cdots, j_{T_k})$ defined in the case $\ell_k = 1$ by $\Phi_k((j, j, \cdots, j)) = \pi_{kj}$ and otherwise by

$$\phi_k((j_{T_{k-1}+1}, \cdots, j_{T_k})) \propto \textstyle\prod_{j=1}^{J} \Gamma(N_{kj}).$$

(Notice that only ℓ_k of the N_{kj} are positive and we'll understand $\Gamma(0)$ to be 1.) Our distribution for $\mathbf{j}$ is then ϕ^{ℓ} defined by $\phi^{\ell} = \prod_{k=1}^{K} \phi_k$. Next, if $\mathbf{y}_j$ is the vector of length $N_{.j}$ whose entries are those y_i with corresponding $j_i = j$, we must define a conditional distribution ρ_j for $\mathbf{y}_j$. In the event $m_j = 1$, we suppose $\theta_{j1}, \cdots, \theta_{jp_j}$ specify a probability distribution on $\mathscr{B}_j$ such that $\sum_{u=1}^{p_j} \theta_{ju} b_{ju} = \mu_j$ and take $\rho_j((b_{ju}, b_{ju}, \cdots, b_{ju})) = \theta_{ju}$. Otherwise we suppose

$$\rho_j(\mathbf{y}_j) \propto \textstyle\prod_{u=1}^{p_j} \Gamma(\text{the number of entries of } \mathbf{y}_j \text{ equal to } b_{ju}).$$

(Notice here that only m_j of the terms in this product are other than $\Gamma(0)$.) Our conditional distribution for $\mathbf{y}$ given $\mathbf{j}$ is $\rho^{\mathbf{m},\mathbf{j}}$ defined by $\rho^{\mathbf{m},\mathbf{j}}(\mathbf{y}) = \prod_{j=1}^{J} \rho_j(\mathbf{y}_j)$, and so we take as our distribution on $\mathscr{A}_{\ell,\mathbf{m}}$, $\lambda_{\ell,\mathbf{m}}$ defined for $(\mathbf{y}, \mathbf{j}) \in \mathscr{A}_{\ell,\mathbf{m}}$ by

$$\lambda_{\ell,\mathbf{m}}(\mathbf{y}, \mathbf{j}) = \phi^{\ell}(\mathbf{j})\rho^{\mathbf{m},\mathbf{j}}(\mathbf{y}).$$

Next we need to specify an ordering of the nonvoid $\mathscr{A}_{\ell,\mathbf{m}}$ and their corresponding $\lambda_{\ell,\mathbf{m}}$. For this it suffices to place the $\mathscr{A}_{\ell,\mathbf{m}}$ in lexicographical order according to their "$K + J$ digit lables" $(\ell, \mathbf{m})$. Then with $\Lambda_{\ell,\mathbf{m}}$ standing for the set of $(s, s^*, \mathbf{y}_s, \mathbf{j}_{s^*})$ possible under Δ and $\lambda_{\ell,\mathbf{m}}$ but not under Δ and any $\lambda_{\ell^*,\mathbf{m}^*}$ standing before $\lambda_{\ell,\mathbf{m}}$ in the ordering, it is possible to verify that $\tilde{t}$ is the $\lambda_{\ell,\mathbf{m}}$ conditional mean of τ given $(s, s^*, \mathbf{y}_s, \mathbf{j}_{s^*})$. As such, for the observable $(s, s^*, \mathbf{y}_s, \mathbf{j}_{s^*})$, $\tilde{t}$ is a $\lambda_{\ell,\mathbf{m}}$ Bayes estimate of τ and thus by Theorem 1 of Meeden and Ghosh (1981), $\tilde{t}$ is admissible for Δ with parameter set $\mathscr{A}_{\mathscr{B}}$. □

Notice now that even in the case that one or more of $\mathscr{B}_1, \cdots, \mathscr{B}_J$ are nonfinite, if $\boldsymbol{\mu}$ and $\mathscr{B}$ are compatible, each $(\mathbf{y}, \mathbf{j})$ belonging to $\mathscr{A}_{\mathscr{B}}$ also belongs to some

$\mathscr{A}_{\mathscr{B}^0}$ where $\mathscr{B}_1^0, \mathscr{B}_2^0, \cdots, \mathscr{B}_J^0$ are finite and μ and $\mathscr{B}^0$ are compatible. Hence, as suggested earlier, Theorem 4.1 provides its own generalization and we have:

COROLLARY 4.1. *For any design Δ, if each $\mathscr{B}_1, \mathscr{B}_2, \cdots, \mathscr{B}_J$ is a subset of $\mathbb{R}$ (finite or infinite), then the estimators t and $\tilde{t}$ are admissible for the design Δ and parameter set $\mathscr{A}_{\mathscr{B}}$.*

Of course, beyond questions of admissibility of t and $\tilde{t}$ for fixed Δ are questions of uniform admissibility for pairs (t, Δ) and $(\tilde{t}, \Delta)$. Meeden and Ghosh (1983) have shown how uniform admissibility problems in finite population sampling can be treated in terms of making an admissible choice between finitely many different possible experiments, and how notions of stepwise Bayesness arise naturally in making such a choice. We proceed to outline how their line of reasoning can be applied here.

First, when contemplating estimator design pairs (t, Δ), notice that a reasonable class of designs to use in uniform admissibility considerations is $\mathscr{L}_{n,n^*}$, the collection where the size of s is fixed at n and the size of s^* is fixed at n^*. (At least on an intuitive basis, designs of fixed sizes (5, 10) ought to be at a disadvantage if one allows comparison to designs of fixed sizes say (50, 100).) The fact is that, because of their symmetric nature, the priors that one would use to prove the t part of Theorem 4.1 produce the same Bayes averages of the mean squared error of t for each degenerate (nonrandom) design $\Delta \in \mathscr{L}_{n,n^*}$. Following the reasoning of Meeden and Ghosh (1983), this observation then leads to:

THEOREM 4.2. *If $\mathscr{B}_1, \mathscr{B}_2, \cdots, \mathscr{B}_J$ are finite, μ and $\mathscr{B}$ are compatible and Δ is any element of $\mathscr{L}_{n,n^*}$, then the estimator-design pair (t, Δ) is uniformly admissible relative to $\mathscr{L}_{n,n^*}$ when the parameter set is $\mathscr{A}_{\mathscr{B}}$.*

Of course, just as Theorem 4.1 immediately gave Corollary 4.1, it is easy to see that the restriction to finite $\mathscr{B}_j$'s may be dropped and the uniform admissibility conclusion will still hold.

The situation of $\tilde{t}$ (for $K > 1$) is only slightly more complicated than that of t. Because $\tilde{t}$ makes use of at least potentially asymmetric prior information about $\mathbf{j}$, it doesn't seem to the authors that the class of designs $\mathscr{L}_{n,n^*}$ is natural for use in uniform admissibility considerations for $\tilde{t}$. Instead, for $\mathbf{n} = (n_{1\cdot}, n_{2\cdot}, \cdots, n_{K\cdot})$ and $\mathbf{n}^* = (n_{1\cdot}^*, n_{2\cdot}^*, \cdots, n_{K\cdot}^*)$ we let $\mathscr{L}_{\mathbf{n},\mathbf{n}^*}$ be the collection of designs producing the fixed vectors of segment sample sizes $\mathbf{n}$ and $\mathbf{n}^*$. Then the priors one would use to prove the $\tilde{t}$ part of Theorem 4.1 produce the same Bayes mean squared errors for each degenerate design in $\mathscr{L}_{\mathbf{n},\mathbf{n}^*}$ and again using the Meeden-Ghosh argument one has:

THEOREM 4.3. *If $\mathscr{B}_1, \mathscr{B}_2, \cdots, \mathscr{B}_J$ are finite, μ and $\mathscr{B}$ are compatible and Δ is any element of $\mathscr{L}_{\mathbf{n},\mathbf{n}^*}$, then the estimator-design pair $(\tilde{t}, \Delta)$ is uniformly admissible relative to $\mathscr{L}_{\mathbf{n},\mathbf{n}^*}$ when the parameter set is $\mathscr{A}_{\mathscr{B}}$.*

And once again, the finiteness requirement on the $\mathscr{B}_j$'s can be dropped and the uniform admissibility conclusion will still hold.

5. Some possible extensions and closing remarks. There are a number of directions in which the present work can easily be extended. For one thing, the ideas of Section 4 of Vardeman and Meeden (1982) and Section 3 of Vardeman and Meeden (1983a) can be used to replace the $\hat{\mu}_j$ in t or $\tilde{t}$ with ratio or difference type estimators of stratum means and still have admissibility, and in some cases uniform admissibility results, follow fairly easily from the Meeden-Ghosh techniques. Indeed, the ideas of Vardeman and Meeden (1983b) could even be employed to replace the $\hat{\mu}_j$ with estimators of stratum means that employ trimming or Winsorization.

Also, though this entire article has been phrased in terms of estimation of τ, the present techniques could clearly be used in the admissible estimation of other functions of $(\mathbf{y}, \mathbf{j})$ such as population variances or differences in stratum means.

Next, we should point out that as t and $\tilde{t}$ can be thought of as arising as the mean of τ based on either legitimate posterior distributions (in cases where all M's are positive) or pseudo posteriors (in the case that some M's are 0) it is possible to think of producing Bayesian or pseudo-Bayesian credible sets from these distributions in addition to or in preference to the point estimates discussed here. (For example, in the case of $\tilde{t}$ where all M's are 0, the pseudo-posterior suggested by our proof of Theorem 4.1 would be described as follows. After observing $(\mathbf{y}_s, \mathbf{j}_{s^*})$, the unobserved entries of $\mathbf{j}$ could conceptually be filled in using K independent Polya urn schemes, where urn k initially contains n^*_{kj} "balls" of type j. Then the unobserved entries of $\mathbf{y}$ could conceptually be provided by drawing from J independent Polya urn schemes, urn j being used to generate the y_j with $j_i = j$ and the initial composition of the urn being determined by the frequency distribution of those y_i with $i \in s$ and $j_i = j$.) Binder (1982) has applied such reasoning to give Bayesian derivations of the usual confidence interval for the mean of a finite population and for the usual confidence interval for the mean of a stratified population.

Finally, it could be argued that admissibility proofs are of limited interest since for a typical decision problem there is always available a large class of admissible decision rules. Until the recent work using the stepwise Bayes technique, most admissibility proofs in finite population sampling were quite difficult and worked only in a limited number of cases (see Godambe and Joshi, 1965, Joshi, 1965, Joshi, 1966, and Godambe, 1969). In this paper, we used the stepwise Bayes technique to prove the admissibility of a wide variety of estimators, some well known and some new. Of more interest, however, is the fact that the technique itself seems helpful in suggesting new estimators that incorporate prior information in a pseudo-Bayesian way.

REFERENCES

BASU, D. (1971). An essay on the logical foundations of survey sampling, part one. *Foundations of Statistical Inference*. Ed. V. P. Godambe and D. A. Sprott. Holt, Toronto.

BINDER, D. A. (1982). Non-parametric Bayesian models for samples from finite populations. *J. Roy. Statist. Soc. B* **44** 388–393.

COCHRAN, W. G. (1977). *Sampling Techniques*. Wiley, New York.

ERICSON, W. A. (1969). Subjective Bayesian models in sampling finite populations (with discussion). *J. Roy. Statist. Soc. B* **31** 195–233.

FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230.

GODAMBE, V. P. (1966). Bayes and empirical Bayes estimation in sampling finite populations (Abstract). *Ann. Math. Statist.* **37** 552.

GODAMBE, V. P. (1969). Admissibility and Bayes estimation in sampling finite populations V. *Ann. Math. Statist.* **40** 672–676.

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

HIDIROGLOU, M. and SRINATH, K. (1981). Some estimators of a population total from simple random samples containing large units. *J. Amer. Statist. Assoc.* **76** 690–695.

JOSHI, V. M. (1965). Admissibility and Bayes estimation in sampling finite populations II and III. *Ann. Math. Statist.* **36** 1723–1742.

JOSHI, V. M. (1966). Admissibility and Bayes estimation in sampling finite populations IV. *Ann. Math. Statist.* **37** 1658–1670.

MEEDEN, G. and GHOSH, M. (1981). Admissibility in finite problems. *Ann. Statist.* **9** 846–852.

MEEDEN, G. and GHOSH, M. (1982). On the admissibility and uniform admissibility of ratio type estimates. *Proceedings of the Golden Jubilee Conference*, Indian Statistical Institute.

MEEDEN, G. and GHOSH, M. (1983). Choosing between experiments: applications to finite population sampling. *Ann. Statist.* **11** 296–305.

VARDEMAN, S. and MEEDEN, G. (1982). The admissibility of some estimators of the population total using sample based weights. Unpublished manuscript.

VARDEMAN, S. and MEEDEN, G. (1983a). Admissibile estimators in finite population sampling employing various types of prior information. *J. Statist. Plann. Inf.* **7** 329–341.

VARDEMAN, S. and MEEDEN, G. (1983b). Admissible estimators of the population total using trimming and Winsorization. *Statist. Probab. Letters* **1** 317–321.

DEPARTMENT OF STATISTICS
IOWA STATE UNIVERSITY
AMES, IOWA 50011