ESTIMATION OF PROPORTIONS IN A FINITE POPULATION

Richard Valliant, Bureau of Labor Statistics

1. INTRODUCTION

The linear least squares prediction approach to finite population sampling theory has been developed as an alternative to the traditional design-based approach. In prediction theory the actual values for each unit in the finite population are treated as realizations of random variables. The joint probability law of these random variables is given by a superpopulation model which links sample and nonsample units and forms the basis for inference. Previous work in prediction theory for finite populations has examined models in which the expected value of each random variable is a linear combination of a set of unknown parameters. In many applications a linear model is too restrictive and a nonlinear model of some type will provide a better approximation to reality.

Virtually all of the work in prediction theory has involved the estimation of totals or means of quantitative variables. A natural situation in which a nonlinear model may be appropriate is that of estimating the proportion of units or total number of units in a population which have a particular characteristic. Often some relevant auxiliary information for each unit is available which can be used in model construction.

Section 2 presents an estimator of the total under a nonlinear Bernoulli model in which the random variables are independent. Conditions are given under which the estimator of the total is asymptotically normal and consistent. Estimation of the variance of the estimated total under the Bernoulli model is discussed and further conditions are given for the consistency of the variance estimator. An approximate jackknife variance estimator is also presented as an alternative. Section 3 describes the results of an empirical study. A finite population of hospitals in the United States is used and the total numbers of hospitals offering certain types of services are estimated. Whether a hospital offers a service is modelled as the realization of a Bernoulli random variable whose expected value depends on the size of the hospital. The results confirm that nonlinear models can be usefully employed in finite population inferences and that the asymptotic properties of the estimated total do describe the moderate sample size situation with useful accuracy.

2. BASIC RESULTS UNDER A NONLINEAR BERNOULLI MODEL

A finite population of N units will be considered with the units labeled 1, 2, ..., N. Associated with unit i are a random variable y_i and a vector of known auxiliary variables $\underset{\sim}{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iq})'$. The random variables y_i, i = 1, 2, ..., N are assumed to be independent. The population vector of random variables and matrix of auxiliaries will be written as $\underset{\sim}{y} = (y_1, \ldots, y_N)'$ and $\underset{\sim}{X} = (\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_N)'$. When a sample of n units is selected from the population of N, $\underset{\sim}{y}$ and $\underset{\sim}{X}$ can be reordered and partitioned into sections referring to the sample units, denoted by s, and the nonsample units, denoted by r: $\underset{\sim}{y} = (\underset{\sim}{y}_s', \underset{\sim}{y}_r')'$ and $\underset{\sim}{X} = (\underset{\sim}{X}_s', \underset{\sim}{X}_r')'$ where $\underset{\sim}{y}_s$ is nx1, $\underset{\sim}{y}_r$ is (N-n) x 1, $\underset{\sim}{X}_s$ is nxq, and $\underset{\sim}{X}_r$ is (N-n) x q. Denote the nxn diagonal covariance matrix associated with the sample units by $\underset{\sim}{V}_{ss}$ and the corresponding diagonal (N-n) x (N-n) matrix for the nonsample units by $\underset{\sim}{V}_{rr}$. The model to be considered asserts that $E(y_i) = f(\underset{\sim}{x}_i; \underset{\sim}{\theta})$ and $Var(y_i) = f(\underset{\sim}{x}_i; \underset{\sim}{\theta})[1-f(\underset{\sim}{x}_i; \underset{\sim}{\theta})]$ where $\underset{\sim}{\theta}$ is a px1 vector of unknown constants and $f(\cdot;\cdot)$ has at least three partial derivative with respect to $\underset{\sim}{\theta}$ for all $\underset{\sim}{\theta}$ in the parameter space and for all N of the $\underset{\sim}{x}_i$. The function $f(\cdot;\cdot)$ is in general a nonlinear function of the elements $\underset{\sim}{\theta}$. The vector of expected values of $\underset{\sim}{y}$ will be denoted by $\underset{\sim}{f}(\underset{\sim}{\theta}) = (f(\underset{\sim}{x}_1; \underset{\sim}{\theta}), \ldots, f(\underset{\sim}{x}_N; \underset{\sim}{\theta}))'$ and its sample and nonsample components by $\underset{\sim}{f}_s(\underset{\sim}{\theta})$ and $\underset{\sim}{f}_r(\underset{\sim}{\theta})$. The full specification of the model is then

$$E(\underset{\sim}{y}) = \underset{\sim}{f}(\underset{\sim}{\theta}) = [\underset{\sim}{f}_s(\underset{\sim}{\theta})', \underset{\sim}{f}_r(\underset{\sim}{\theta})']'$$

$$Var(\underset{\sim}{y}) = \underset{\sim}{V} = \begin{bmatrix} \underset{\sim}{V}_{ss} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{V}_{rr} \end{bmatrix} \qquad (1)$$

In addition, we will need the vectors and matrices of partial derivatives defined by $\underset{\sim}{z}_i(\underset{\sim}{\theta}) = [\partial f(\underset{\sim}{x}_i; \underset{\sim}{\theta})/\partial\theta_1, \ldots, \partial f(\underset{\sim}{x}_i; \underset{\sim}{\theta})/\partial\theta_p]'$ for i=1, 2, ..., N and $\underset{\sim}{F}(\underset{\sim}{\theta}) = [\underset{\sim}{z}_1(\underset{\sim}{\theta}), \ldots, \underset{\sim}{z}_N(\underset{\sim}{\theta})]' = [\underset{\sim}{F}_s(\underset{\sim}{\theta})', \underset{\sim}{F}_r(\underset{\sim}{\theta})']'$ where $\underset{\sim}{F}_s(\underset{\sim}{\theta})$ is the nxp matrix of partial derivatives for the sample units and $\underset{\sim}{F}_r(\underset{\sim}{\theta})$ is the (N-n) x p matrix of partials for the nonsample units. In much of the following, the argument $\underset{\sim}{\theta}$ will be suppressed in $\underset{\sim}{z}_i(\underset{\sim}{\theta})$, $\underset{\sim}{F}_s(\underset{\sim}{\theta})$, and $\underset{\sim}{F}_r(\underset{\sim}{\theta})$, for compactness of notation.

For a given sample the population total $T = \Sigma_1^N y_i$ can be written as $T = T_s + T_r$ where T_s is the total for units in the sample and T_r is the total for the nonsample units. After the vector $\underset{\sim}{y}_s$ is observed, the problem of estimating T is equivalent to the problem of predicting the sum T_r for the unobserved random variables. When the parameter $\underset{\sim}{\theta}$ is known the best linear unbiased (BLU) predictor of T is obtained by adding to the observed T_s the BLU predictor of

T_r. This result is stated explicitly in the following theorem which is a simple consequence of a standard result in linear prediction theory (Bibby and Toutenburg 1979, Ch. 5). The type of linear estimator considered by the theorem has the form $\hat{T}^* = \underset{\sim}{\alpha}_1' \underset{\sim}{y}_s + \alpha_0$ where $\underset{\sim}{\alpha}_1$ is an n x 1 vector of constants and α_0 is a scalar.

Theorem 1:
Under model (1) with $\underset{\sim}{\theta}$ known, among linear estimators $\hat{T}^*$ satisfying $E(\hat{T}^*-T) = 0$, the error-variance $E(\hat{T}^*-T)^2$ is minimized by $\hat{T}_o = T_s + \underset{\sim}{1}_r' \underset{\sim}{f}_r(\underset{\sim}{\theta})$ where $\underset{\sim}{1}_r$ is an (N-n) - vector of 1's.

When $\underset{\sim}{\theta}$ is unknown an estimator must be used. The standard estimator of $\underset{\sim}{\theta}$ in a nonlinear regression problem is obtained by generalized least squares (GLS). The GLS estimator (GLSE) is the value of $\underset{\sim}{\theta}$ which minimizes $(\underset{\sim}{y}_s - \underset{\sim}{f}_s(\underset{\sim}{\theta}))' \underset{\sim}{V}_{ss}^{-1} (\underset{\sim}{y}_s - \underset{\sim}{f}(\underset{\sim}{\theta}))$. Under model (1) the GLSE is equivalent to the maximum likelihood estimator (MLE) in the sense that both can be obtained by solving the same set of equations using the method of iterative reweighted least squares (Nelder and Wedderburn 1972). Bradley and Gart (1962) have given conditions under which the MLE under model (1) will be consistent and asymptotically normal.

Motivated by Theorem 1 the estimator of the total we consider here is $\hat{T} = T_s + \underset{\sim}{1}_r' \underset{\sim}{f}_r(\hat{\underset{\sim}{\theta}})$ where $\hat{\underset{\sim}{\theta}}$ is the MLE. This estimator is directly analogous to the BLU predictor of T under a linear model with $E(\underset{\sim}{y}) = \underset{\sim}{X}\underset{\sim}{\beta}$ where $\underset{\sim}{\beta}$ is a q-vector of regression coefficients and with $\underset{\sim}{V}$ a known diagonal covariance matrix. Under that linear model the BLU predictor is $T_L = T_s + \underset{\sim}{1}_r' \underset{\sim}{X}_r \hat{\underset{\sim}{\beta}}$ with $\hat{\underset{\sim}{\beta}} = (\underset{\sim}{X}_s' \underset{\sim}{V}_{ss}^{-1} \underset{\sim}{X}_s)^{-1} \underset{\sim}{X}_s' \underset{\sim}{V}_{ss}^{-1} \underset{\sim}{y}_s$ (Royall 1976). In the special case of $\underset{\sim}{f}(\underset{\sim}{\theta}) = \underset{\sim}{X}\underset{\sim}{\theta}$ and $\underset{\sim}{V}$ known the estimator $\hat{T}$ equals $\hat{T}_L$.

A delta method approximation to the prediction variance of $\hat{T}$ is

$$V_{\hat{T}} = \underset{\sim}{1}_r' \underset{\sim}{F}_r (\underset{\sim}{F}_s' \underset{\sim}{V}_{ss}^{-1} \underset{\sim}{F}_s)^{-1} \underset{\sim}{F}_r' \underset{\sim}{1}_r + \underset{\sim}{1}_r' \underset{\sim}{V}_{rr} \underset{\sim}{1}_r \qquad (2)$$

where each element in the matrices in (2) is evaluated at $\underset{\sim}{\theta}$ (Valliant 1984). The factor $(\underset{\sim}{F}_s' \underset{\sim}{V}_{ss}^{-1} \underset{\sim}{F}_s)^{-1}$ is the inverse of the information matrix which is the usual estimator of the covariance matrix of the MLE $\hat{\underset{\sim}{\theta}}$. Under some reasonable conditions the second term on the righthand side of (2), which is the variance of the sum of the nonsample units, will be negligible compared to the first in large samples. The natural estimator of the prediction variance of $\hat{T}$ is obtained by evaluating (2) at $\hat{\underset{\sim}{\theta}}$ and will be denoted by $v_{\hat{T}}$.

Under the regularity conditions of Theorems 2 and 3 below, $\hat{T}$ properly standardized is asymptotically normal and both $\hat{T}/N$ and $v_{\hat{T}}$ are consistent. Proofs are given in Valliant (1984). Below we denote the normal probability law with mean μ and variance σ^2 by $N(\mu, \sigma^2)$.

Theorem 2:
Suppose the model given by (1) holds. If, as N and $n \to \infty$, $f = n/N \to 0$ and the following conditions hold:

(i) $\underset{\sim}{F}(\hat{\underset{\sim}{\theta}})$ exists and is nonzero at $\hat{\underset{\sim}{\theta}} = \underset{\sim}{\theta}$,

(ii) $\underset{\sim}{1}_r' \underset{\sim}{F}_r/(N-n) \to \underset{\sim}{\lambda}_o'$, a nonzero vector of constants,

(iii) $\underset{\sim}{F}_s' \underset{\sim}{V}_{ss}^{-1} \underset{\sim}{F}_s / n \to \underset{\sim}{A}_o$, a positive definite matrix,

(iv) $\underset{\sim}{1}_r' \underset{\sim}{V}_{rr} \underset{\sim}{1}_r/(N-n) \to \bar{v}_r$, a positive constant,

(v) for every $\underset{\sim}{\theta}^*$ in a neighborhood of $\underset{\sim}{\theta}$ there exist real valued functions $h(\underset{\sim}{x}_i)$ such that $\left|\partial^2 f(x_i;\theta)/\partial\theta_j \partial\theta_k\right| \le h(\underset{\sim}{x}_i)$ and $\sum_r h(\underset{\sim}{x}_i)/(N-n)$ converges to a positive constant for j,k = 1,2,...,p, and

(vi) $\hat{\underset{\sim}{\theta}}$ is an estimator of $\underset{\sim}{\theta}$ such that $\sqrt{n}(\hat{\underset{\sim}{\theta}}-\underset{\sim}{\theta}) \xrightarrow{d} N(\underset{\sim}{0}, \underset{\sim}{A}_o^{-1})$,

then $(\hat{T}-T)/\sqrt{V_{\hat{T}}} \xrightarrow{d} N(0,1)$ where "$\xrightarrow{d}$" means "converges in distribution."

The weak consistency of the estimated proportion $\hat{T}/N$ follows directly from Theorem 2. The asymptotic variance of $(\hat{T}-T)/N$ is $\underset{\sim}{1}_r' \underset{\sim}{F}_r (\underset{\sim}{F}_s' \underset{\sim}{V}_{ss} \underset{\sim}{F}_s)^{-1} \underset{\sim}{F}_r' \underset{\sim}{1}_r / N^2$. This variance converges to zero as n, $N \to \infty$ by conditions (ii) and (iii) of Theorem 2. The result follows from Chebyshev's inequality.

Before stating the conditions for consistency of $v_{\hat{T}}$, we define the (N-n) x p matrix for the nonsample units $\underset{\sim}{G}_{rm} = [\partial^2 f(\underset{\sim}{x}_i;\underset{\sim}{\theta})/\partial\theta_j \partial\theta_m]$, $i \varepsilon r$ and j=1, 2,...,p, the corresponding n x p matrix $\underset{\sim}{G}_{sm}$ for the sample units, then n x n matrix $\underset{\sim}{D}_{sm} = \partial \underset{\sim}{V}_{ss}^{-1}/\partial\theta_m$ for the sample units, and the (N-n) x (N-n) matrix $\underset{\sim}{E}_{rm} = \partial \underset{\sim}{V}_{rr}/\partial\theta_m$ for the nonsample units. Each of the matrices is defined for m = 1,2,...,p and is associated with $\partial v_{\hat{T}}/\partial\theta_m$.

Theorem 3:
Suppose model (1) holds. If as n, $N \to \infty$, $f \to 0$, conditions (i)-(iv) of Theorem 2 and conditions (i) and (ii) below hold, then $v_{\hat{T}}/V_{\hat{T}} \xrightarrow{p} 1$ where "$\xrightarrow{p}$" means "converges in probability."

(i) $\hat{\underset{\sim}{\theta}} \xrightarrow{p} \underset{\sim}{\theta}$

(ii) For every θ^* in a neighborhood of $\underset{\sim}{\theta}$ each element of the following quantities is bounded in absolute value when evaluated at θ^*:

$$1_r'F_r/(N-n), 1_r'G_{rm}/(N-n), n(F_s'V_{ss}^{-1}F_s)^{-1},$$

$$G_{sm}'V_{ss}^{-1}F_s/n, F_s'D_{sm}F_s/n, \text{ and}$$

$$1_r'E_{rm}1_r/(N-n) \text{ for } m = 1,2,\ldots,p.$$

An important consequence of Theorems 2 and 3 is that the standardized form of $\hat{T}$ is still asymptotically normal when the approximate variance of $\hat{T}$ is replaced by its sample estimator, i.e., under the conditions of Theorems 2 and 3 $(\hat{T} - T)/\sqrt{v_{\hat{T}}} \xrightarrow{d} N(0,1)$ as $n \to \infty$.

The following simple example illustrates the use of the two theorems. The example involves a single auxiliary variable x which is a common situation in finite population estimation. If x is bounded, the sufficient conditions of Bradley and Gart (1962) for the asymptotic normality and consistency of the MLE $\hat{\theta}$ can be verified though the steps are somewhat lengthy and are not given here. In the following define $\bar{x}_r = \Sigma_r x_i/(N-n)$, $\bar{x}_{r2} = \Sigma_r x_i^2/(N-n)$, $\bar{x}_{s3} = \Sigma_s x_i^3/n$, and $f(x_i;\theta) = p_i$.

Example: Two parameter logistic model.

Let $y_1, y_2, \ldots, y_n$ be independent Bernoulli random variables with $p_i = [1 + \exp(-\theta_o - \theta_1 x_i)]^{-1}$ and suppose $x_i \geq 1$ for all i = 1, 2, .., N. The quantities covered by conditions (i)-(iv) of Theorem 2 which must converge are $1_r'F_r/(N-n) = [\Sigma_r p_i q_i, \Sigma_r p_i q_i x_i]/(N-n)$,

$$F_s'V_{ss}^{-1}F_s/n = \frac{1}{n}\begin{bmatrix} \Sigma_s p_i q_i & \Sigma_s p_i q_i x_i \\ & \Sigma_s p_i q_i x_i^2 \end{bmatrix},$$

and $1_r'V_{rr}1_r/(N-n) = \Sigma_r p_i q_i/(N-n)$.

The second partial derivatives of p_i are $\partial^2 p_i/\partial\theta_o^2 = p_i q_i(1-2p_i)$, $\partial^2 p_i/\partial\theta_1^2 = p_i q_i x_i^2(1-2p_i)$ and $\partial^2 p_i/\partial\theta_o\partial\theta_1 = p_i q_i x_i(1-2p_i)$, all of which are bounded in absolute value by x_i^2. Condition (v) of Theorem 2 will be satisfied if $\bar{x}_{r2}$ converges. Turning to Theorem 3, the elements $1_r'F_r^*/(N-n)$ are bounded by $\bar{x}_r$. The elements of $n(F_s^{*\prime}V_{ss}^{*-1}F_s^*)^{-1}$ are bounded if $\bar{x}_{r2}$ converges and the determinant of $F_s^{*\prime}V_{ss}^*F_s^*$, $(\Sigma_s p_i^*q_i^*x_i^2)(\Sigma_s p_i^*q_i^*) - (\Sigma_s p_i^*q_i^*x_i)^2$, is nonzero where the * superscript indicates the elements are evaluated at θ^*. The determinant is zero in the trivial case in which x is constant or p is always zero or one but in general is nonzero. Letting

$\Sigma_r p_i^*q_i^*(1-2p_i^*)/(N-n) = g_{1r}$, $\Sigma_r p_i^*q_i^*x_i(1-2p_i^*)/(N-n) = g_{2r}$, $\Sigma_r p_i^*q_i^*x_i^2(1-2p_i^*)/(N-n) = g_{3r}$, g_{1s}, g_{2s} and g_{3s} be the analogous sample averages, and $\Sigma_s p_i^*q_i^*x_i^3(1-2p_i^*)/n = g_{4s}$, the other quantities in condition (ii) of Theorem 3 are

$$1_r'G_{ro}^*/(N-n) = [g_{1r}, g_{2r}],\ 1_r'G_{r1}^*/(N-n) = [g_{2r}, g_{3r}]$$

$$G_{so}^{*\prime}V_{ss}^{*-1}F_s^*/n = -F_s^{*\prime}D_{so}^*F_s^*/n = \begin{bmatrix} g_{1s} & g_{2s} \\ g_{2s} & g_{3s} \end{bmatrix},$$

$$G_{s1}^*V_{ss}^{*-1}F_s^*/n = -F_s^*D_{s1}^*F_s^*/n = \begin{bmatrix} g_{2s} & g_{3s} \\ g_{3s} & g_{4s} \end{bmatrix},$$

$$1_r'E_{ro}^*1_r/(N-n) = g_{1r},\ 1_r'E_{r1}^*1_r/(N-n) = g_{2r}.$$

The terms g_{1r}, g_{2r}, and g_{3r} are bounded if $\bar{x}_{r2}$ converges while the terms g_{1s} through g_{4s} are bounded if $\bar{x}_{s3}$ converges. Thus, reasonable behavior of moments involving the auxiliary variable x are sufficient to assure asymptotic normality of $\hat{T}$ and consistency of $v_{\hat{T}}$.

In this as in other prediction theory investigations, a key issue is how well the estimators of the total and variance perform when the model fails. In linear model analyses variance estimators dictated by the model may be extremely nonrobust when the model is wrong (Royall and Cumberland 1978, 1981a, 1981b). The same kind of problem can be expected under nonlinear models. An alternative variance estimator we will mention briefly is an approximate jackknife based on a covariance matrix estimator suggested by Fox, Hinkley, and Larntz (1980). When applied to the problem of estimating the prediction variance of $\hat{T}$, the approximate jackknife is

$$v_J = 1_r'F_r V_J F_r'1_r$$

where

$$V_J = A^{-1}\{F_s'V_{ss}^{-1}D_{ss}^2V_{ss}^{-1}F_s - \tfrac{1}{n}F_s'V_{ss}^{-1}D_{ss}1_s1_s'D_{ss}V_{ss}^{-1}F_s\}A^{-1},$$

$A = F_s'V_{ss}^{-1}F_s$, and

$D_{ss} = \text{diag}\{r_i/(1-k_i)\}$, $i \in s$.

The remaining terms are the sample residual $r_i = y_i - \hat{p}_i$, $k_i = z_i'A^{-1}z_i/\hat{p}_i\hat{q}_i$, where $\hat{p}_i = f(x_i;\hat{\theta})$, and 1_s which is an n-vector of 1's. All terms are evaluated at $\hat{\theta}$. The matrix V_J is a linear approximation to the jackknife estimator of the covariance of $\hat{\theta}$. Computationally, this estimator is much less demanding then the full jackknife. The estimator v_J requires only one nonlinear fit while the full jackknife requires n + 1 fits —one for the full sample and one each for n subsamples.

3. EMPIRICAL RESULTS

The finite population used for the empirical study is a subset of the hospitals in the United States in 1980. The hospitals were taken from those covered in the 1980 Annual Survey of Hospitals conducted by the American Hospital Association. The subset used consisted of 5050 general medical and surgical hospitals which provided complete responses to questions on the types of services each offered and on the number of inpatient beds each had. Results for three services are reported here:(1) psychiatric acute care, (2) pediatric intensive care, and (3) neonatal intensive care. The percentages of hospitals offering services 1, 2, and 3 are 33.2, 26.7, and 16.2 as shown in Table 1. The percentage offering each service also depends on the hospital bed size (x) as illustrated by Figures 1-3 which are plots of the percentages within bed size classes versus the midpoints of the size classes.

Bernoulli models were identified which fit the scatter plots in Figures 1-3 reasonably well based on analyses of residuals and on formal tests of significance using the error sum of squares from each fitted model and which produced acceptable estimates of the total for each service when the estimated expected values were summed over the entire population. The particicular forms adopted for the function $f(\cdot;\cdot)$ were $1-\exp(-\theta x)$ for service 1 and $[1+\exp(-\theta_0-\theta_1\sqrt{x})]^{1}$ for services 2 and 3. The fitted curves based on the entire population of hospitals are shown as solid lines in Figures 1-3.

To test the theory for $\hat{T}$ and $v_{\hat{T}}$, 1000 simple random samples of size 32 were selected. The validity of the theory in section 2 does not require that simple random sampling or any other probability sampling plan be used. Simple random sampling was used because it provides a wide variety of samples to test the theory. The size of 32 was chosen as being large enough to test asymptotic properties but small enough to expose anomalies that might not appear in very large samples. For each sample, $\underset{\sim}{\theta}$ was estimated in each model by iterative reweighted least squares. The convergence criterion used was that the maximum relative change in $\hat{\underset{\sim}{\theta}}$ between successive iterations be .01 or less with a maximum of 5 iterations allowed per sample to limit computation time. Samples for which $\Sigma_s y = 0$ or n were eliminated since computations could not be carried through in those instances.

For each of the samples which were retained the estimated total $\hat{T}$, the estimation error $\hat{T} - T$, the squared error $(\hat{T} - T)^2$, the estimated variances $v_{\hat{T}}$ and v_J, and the standardized error $(\hat{T} - T)/\sqrt{v}$ with $v = v_{\hat{T}}$ and v_J were computed. Table 1 gives the average error of $\hat{T}$, defined as $AE(\hat{T}) = \Sigma_1^S (\hat{T} - T)/S$, and the average relative error $AE(\hat{T})/T$ where S is the number of samples retained for a model. The table also gives the square roots of the averages of $(\hat{T} - T)^2, v_{\hat{T}}$, and v_J along with the ratios $\{\Sigma v/\Sigma(\hat{T} - T)^2\}^{1/2}$, where $v = v_{\hat{T}}$ or v_J. The relative error of $\hat{T}$ ranges from -.030 to .008 in accordance with the theoretical asymptotic unbiasedness of $\hat{T}$.

Results for $v_{\hat{T}}$ are somewhat less satisfactory. The variance estimator underestimates the empirical mean square error in each case with the ratio of the root of the average $v_{\hat{T}}$ to the root MSE ranging from .93 to .94. The approximate jackknife v_J is more conservative than $v_{\hat{T}}$, producing ratios of .96, 1.03, and 1.06.

Table 1 also gives empirical coverage properties of 95 percent confidence intervals for the total T. If the empirical distribution of the standardized error (SZE) is approximately standard normal, then about 95 percent of the SZE's would be less than 1.96 in absolute value. When $v_{\hat{T}}$ is used to standardize, the empirical coverage percentages are 92.5, 91.1, and 88.3 for the three services. The observed undercoverage is due in part to $v_{\hat{T}}$'s being on the average an underestimate of the MSE. When v_J is used to standardize, coverage percentages improve to 92.9, 92.6, and 91.8.

4. CONCLUSION

The empirical results illustrate that a nonlinear Bernoulli model can usefully describe the dependence of a zero-one variable on an auxiliary variable in a finite population. The nonlinear models adopted here do produce approximately unbiased estimators of totals. However, the variance estimator $v_{\hat{T}}$ underestimates the mean square error of $\hat{T}$ in this study. The underestimation may be due to the relatively small sample size used in the study and to departures from the adopted models. The approximate jackknife variance estimator appears to be a reasonable alternative to $v_{\hat{T}}$ which produces more conservative variance estimators and confidence intervals that have actual coverage probabilities closer to the nominal probabilities.

The models which were appropriate for the population examined here were monotone in the auxiliary variable x, but other nonmonotone relationships are certainly conceivable. In such cases acceptable models may be difficult to formulate. An alternative prediction approach to fitting a single model is to approximate the nonlinear relationship by a series of piecewise linear functions. Separate ratio or linear regression estimation might then be used as suggested in Royall and Herson (1973).

REFERENCES

Bibby, John and Toutenburg, Helge (1979), Prediction and Improved Estimation in Linear Models, New York: John Wiley and Sons.

Bradley, Ralph A. and Gart, John J. (1962), "The Asymptotic Properties of ML Estimators When Sampling from Associated Populations," Biometrika, 49, 205-214.

Fox, Terry, Hinkley, David, and Larntz, Kinley (1980), "Jackknifing in Nonlinear Regression," Technometrics, 22, 29-33.

Nelder, John A. and Wedderburn, R.W.M. (1972), "Generalized Linear Models," Journal of the Royal Statistical Society, A 135, 370-384.

Royall, Richard M. (1976), "The Linear Least Squares Prediction Approach to Two-Stage Sampling," Journal of the American Statistical Association, 71, 657-664.

Royall, Richard M. and Cumberland, William G. (1978), "Variance Estimation in Finite Population Sampling," Journal of the American Statistical Association, 73, 351-358.

---------- (1981a), "An Empirical Study of the Ratio Estimator and Estimators of Its Variance," Journal of the American Statistical Association, 76, 66-77.

---------- (1981b), "The Finite Population Linear Regression Estimator and Estimators of Its Variance --- An Empirical Study," Journal of the American Statistical Association, 76, 66-77.

Royall, Richard M. and Herson, J. (1973), "Robust Estimation in Finite Populations II: Stratification on a Size Variable," Journal of the American Statistical Association, 68, 890-893.

Valliant, Richard (1984), "Nonlinear Prediction Theory and the Estimation of Proportions in a Finite Population," submitted for publication.

Table 1. Empirical results for the bias of $\hat{T}$, estimators of the standard error of $\hat{T}$, and 95 percent confidence intervals for T.

Summary Item	Psychiatric Acute Care	Pediatric Intensive Care	Neonatal Intensive Care
1. Percent offering service	33.2	26.7	16.2
2. No. of samples	1000	1000	993
3. T	1675	1346	816
4. Avg.$(\hat{T} - T)$	14	-41	-1
5. $MSE^{1/2}$	378	386	296
6. $[\text{Avg } v_{\hat{T}}]^{1/2}$	355	364	276
7. $[\text{Avg } v_J]^{1/2}$	362	396	313
8. Relative error of $\hat{T}$	.008	-.030	-.002
9. $[\text{Avg } v_{\hat{T}} /MSE]^{1/2}$	.94	.94	.93
10. $[\text{Avg } v_J/MSE]^{1/2}$	.96	1.03	1.06
Percentage SZE <1.96			
11. using $v_{\hat{T}}$	92.5	91.1	88.3
12. using v_J	92.9	92.6	91.8

NOTE: Avg. $(\hat{T} - T) = \Sigma(\hat{T} - T)/S$, $MSE = \Sigma(\hat{T} - T)^2/S$, Avg. $v = \Sigma v/S$ for $v = v_{\hat{T}}$ and v_J, (relative error $\hat{T}$) $= \Sigma(\hat{T} - T)/(ST)$, where S = no. of samples and summations are over the samples used for each service.

Figure 1. Proportion of hospitals providing psychiatric acute care by bed size.

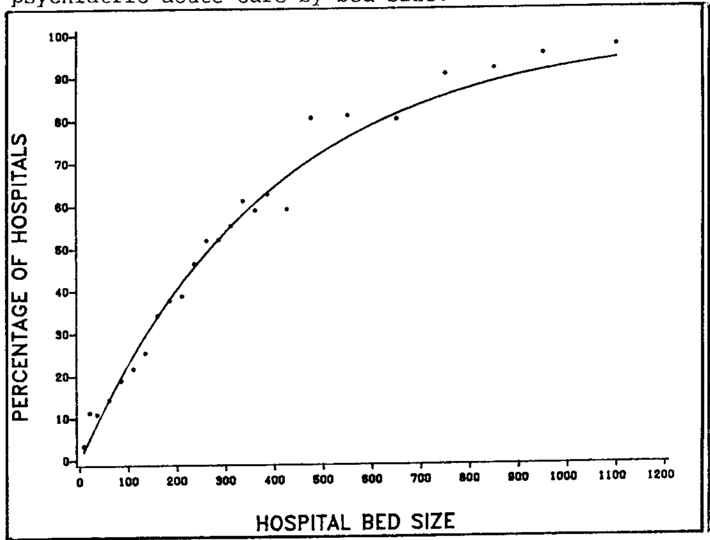

Figure 2. Proportion of hospitals providing pediatric intensive care by bed size.

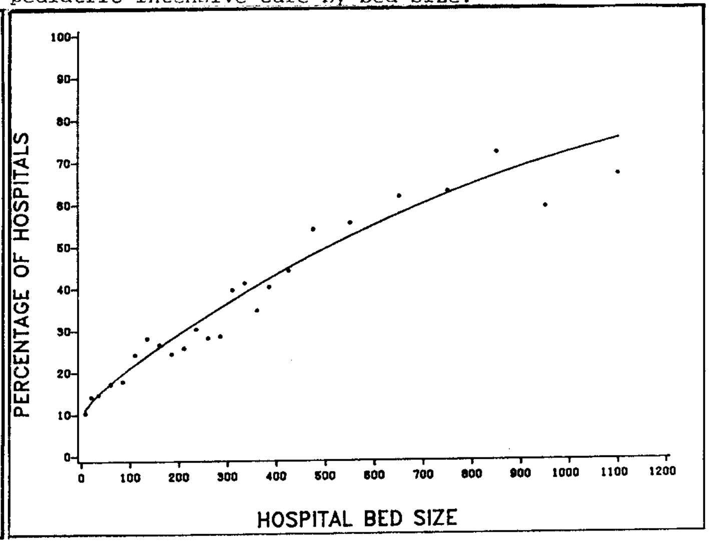

Figure 3. Proportion of hospitals providing neonatal intensive care by bed size.