Survey Methodology, June 1998
Vol. 24, No. 1, pp. 57-67
Statistics Canada

Confidence Intervals for Domain Parameters When the Domain Sample Size is Random

ROBERT J. CASADY, ALAN H. DORFMAN and SUOJIN WANG[1]

ABSTRACT

Let A be a population domain of interest and assume that the elements of A cannot be identified on the sampling frame and the number of elements in A is not known. Further assume that a sample of fixed size (say n) is selected from the entire frame and the resulting domain sample size (say n_A) is random. The problem addressed is the construction of a confidence interval for a domain parameter such as the domain aggregate $T_A = \sum_{i \in A} x_i$. The usual approach to this problem is to redefine x_i, by setting $x_i = 0$ if $i \notin A$. Thus, the construction of a confidence interval for the domain total is recast as the construction of a confidence interval for a population total which can be addressed (at least asymptotically in n) by normal theory. As an alternative, we condition on n_A and construct confidence intervals which have approximately nominal coverage under certain assumptions regarding the domain population. We evaluate the new approach empirically using artificial populations and data from the Bureau of Labor Statistics (BLS) Occupational Compensation Survey.

KEY WORDS: Bayes method; Conditioning; Establishment surveys; Simple random sampling; Stratification; Survey methods.

1. INTRODUCTION

In sampling from a finite population, we often are interested in the estimation of totals, means, or other quantities, for parts of that population, usually referred to as domains. Such domains are not explicitly listed in the frame, the number of items that will occur in the survey is not known in advance, and often enough, we do not even know the number of their elements in the population. For example, we might sample schoolchildren for certain medical problems, and then wish to know the mean blood pressure of those children who are underweight. The class of underweight children would constitute a domain. The only information we have as to whether or not a child is underweight is likely to be among the sampled children; if so, then this would be a case where the domain is not explicitly listed on the frame.

An essential part of the inference process is the estimation of the precision of our estimators; this is typically given by an estimated standard deviation, coefficient of variation, or confidence interval. The notion of a valid confidence interval underlies whatever measure of precision we use. All confidence intervals have, by construction, a stated "nominal" confidence level. A valid confidence interval is a confidence interval with actual coverage matching the nominal coverage. The actual coverage may be determined theoretically or by empirical work mimicking the practical circumstances in which the confidence interval would be used. If a standard deviation is not such as to give rise to a valid confidence interval, then the standard deviation needs to be regarded as misleading.

In the case of estimates for domains, confidence intervals constructed along traditional lines can lead to serious under-coverage, a fact not always appreciated in the literature. We refer to this as the domain problem. The present paper addresses this problem by a somewhat complex methodology involving Bayesian ideas, which, however, leads to a rather simple practical solution, improving on current methodology. The main change in method lies in replacing the standard normal statistic used in the construction of confidence intervals, with a Student's t-statistic having degrees of freedom that depend on the number and configuration of the domain items in the sample.

We shall focus on domain totals and domain means for the two common cases of simple random sampling and stratified random sampling. In the case of simple random sampling, it turns out that standard methods are satisfactory for the mean; however, for the total, coverage can be lower than nominal but not usually worrisome. For stratified random sampling, confidence intervals for both the mean and the total pose serious difficulties with regard to coverage level. In this case, the new methodology is augmented by use of a well known approximation due to Satterthwaite (1946). Alternate approaches to ours, also using this approximation, may be found in Johnson and Rust (1993) and Kott (1994).

An outline of the paper is as follows: In Section 2, to introduce ideas, we consider the case of the total in simple random sampling, using it to illustrate the standard approach for domain estimation, the coverage problem to which this gives rise, and the approach here taken to rectify the difficulty. Section 3 describes the extension to stratified random sampling. Section 4 states our conclusions.

[1] Robert J. Casady and Alan H. Dorfman, U.S. Bureau of Labor Statistics, 2 Massachusetts Ave. N.E., Washington D.C., 20212-0001, U.S.A.; Suojin Wang, Department of Statistics, Texas A&M University, College Station, TX 77843, U.S.A.

2. THE CASE OF SIMPLE RANDOM SAMPLING

2.1 Standard Method

The standard approach to domain estimation is well described in Särndal, Swensson, and Wretman (1992); Sections 3.3, 5.8, and Chapter 10) (henceforth SSW). Their approach is general. Here we paraphrase it for the case of simple random sampling, and, by mild extension, for stratified random sampling as well, and focus on the domain total.

Let x_i be the value of the characteristic of interest for the i-th ($i = 1, 2, ..., N$) element of the population and let A be a domain of interest. We shall consider only the case where the elements of A cannot be identified on the frame and the number N_A of elements in A is not known; the case where N_A is known is fully treated in SSW. It is assumed that any element of A included in a sample can be identified. The problem is to construct a confidence interval for the domain total, $T_A = \sum_{i \in A} x_i$, based on a sample of n elements selected from the entire frame.

Explicitly (as in SSW, Section 3.3) or implicitly (as in SSW, Section 10.3) the standard approach to this problem is to redefine x_i, by setting $x_i = 0$ if $i \notin A$, which forces the population total $T = \sum_{i=1}^{N} x_i$ to be equal to T_A. Thus, the construction of a confidence interval for the domain total is recast as the construction of a confidence interval for a population total. In what follows it is assumed that the x_i's have been redefined as above. We shall also assume, here and throughout this paper, that n is sufficiently large and n/N sufficiently small that second order terms can be ignored. Define the additional population parameters,

$\bar{X} = T/N$ = population mean,

$S^2 = \sum_{i=1}^{N} (x_i - \bar{X})^2/N$ = population variance, and

$p_A = N_A/N$ = proportion of population in A.

Then

(1) $\hat{T}_A = (N/n) \sum_{i=1}^{n} x_i$, $\bar{x} = \sum_{i=1}^{n} x_i/n = \hat{T}_A/N$, $s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2/(n - 1)$, and $\hat{p}_A = n_A/n$ (where n_A is the number of sample elements in A) are unbiased for the corresponding population parameters,

(2) $E(\hat{T}_A) = T_A$,

(3) $\text{var}(\hat{T}_A) = N^2 S^2/n$,

(4) $\sqrt{n}(\hat{T}_A - T_A)/(NS) \xrightarrow{d} N(0, 1)$, and

(5) s^2 is consistent for S^2.

It follows that $\sqrt{n}(\hat{T}_A - T_A)/(Ns) \xrightarrow{d} N(0, 1)$, so, when n is "sufficiently large", appropriate values from the normal distribution can be used to construct confidence intervals for T_A, as noted by SSW, p. 391.

The proportion of the population in A^c is $1 - p_A$ and $x_i = 0$ for $i \in A^c$; therefore, when p_A is small and the values of the x_i's for $i \in A$ are concentrated away from zero, the convergence in distribution in (4) can be slow. Consequently, the distribution of $\sqrt{n}(\hat{T}_A - T_A)/Ns$ can deviate from normal even for what are usually considered to be moderate to large values of n. The simulation study in Section 2.5 illustrates this.

For the case of stratified random sampling, confidence interval coverage for domain quantities using standard methods can be poor. Dorfman and Valliant (1993) noted the problem in their study of wage distributions for domains consisting of workers in specific occupational groups. Preliminary empirical work by the authors indicated that supposed 95% confidence intervals for total workers and total wages for occupation based domains typically provided only 75% to 85% coverage even for a large total sample size ($n = 353$ establishments). These results are verified as part of the empirical work described in Section 3. Furthermore, their work indicated that the distribution of $\hat{T}_A - T_A$ was strongly dependent on the realized value of n_A, which suggested that some type of "conditional" confidence interval should be considered. It seems desirable to establish methodology for the construction of conditional (on n_A or equivalently $\hat{p}_A$) confidence intervals for T_A, which provide nominal, or near nominal, coverage regardless of the realized value of the domain sample size. Inference conditional on sample size is discussed in SSW, Section 10.4, but only for the case of known N_A; we are concerned throughout this paper with the case of unknown N_A.

2.2 Definitions and Notation

We define the following parameters and estimators:

Domain parameters:

$\mu_A = T_A/N_A$ = domain mean,

$\sigma_A^2 = \sum_{i \in A} (x_i - \mu_A)^2/N_A$ = variance of population elements in A.

Domain estimators:

$\hat{N}_A = \hat{p}_A N$,

$\hat{\mu}_A = \sum_{i=1}^{n_A} x_i/n_A = \hat{T}_A/\hat{N}_A$ (only defined for $n_A \geq 1$), and

$\hat{\sigma}_A^2 = \sum_{i=1}^{n_A} (x_i - \hat{\mu}_A)^2/(n_A - 1)$ (only defined for $n_A \geq 2$).

In what follows it is understood that $n_A \geq 2$ (or equivalently $\hat{p}_A \geq 2/n$) unless specifically stated otherwise. At $n_A = 1$ or 0, it is preferable to supply an "insufficient information" tag, rather than attempt inference. The relationships given below follow directly from the definitions:

$T_A = Np_A\mu_A$ and $\hat{T}_A = N\hat{p}_A\hat{\mu}_A$,

$\bar{X} = p_A\mu_A$ and $\bar{x} = \hat{p}_A\hat{\mu}_A$,

$S^2 = p_A(1 - p_A)\mu_A^2 + p_A\sigma_A^2$

and

$$s^2 = \frac{n}{n-1}\hat{p}_A\left(1 - \hat{p}_A\right)\hat{\mu}_A^2 + \frac{n\hat{p}_A - 1}{n - 1}\hat{\sigma}_A^2. \qquad (1)$$

Also, it is straightforward to verify that

$$\left(\sqrt{n}/N\right)\left(\hat{T}_A - T_A\right) = \sqrt{n}\mu_A\left(\hat{p}_A - p_A\right) + \sqrt{\hat{p}_A}\sigma_A Z, \qquad (2)$$

where $Z = \sqrt{n\hat{p}_A}(\hat{\mu}_A - \mu_A)/\sigma_A$. Thus, conditionally on $\hat{p}_A$, $\hat{T}_A$ is biased for T_A, and if, for example, we assume an underlying normality, and standardize $(\sqrt{n}/N)(\hat{T}_A - T_A)$ by the corresponding conditional variance, we will get a non-central t-distribution with unknown non-centrality parameter proportional to $\sqrt{n}\mu_A(\hat{p}_A - p_A)$, providing little basis for (conditional) sound inference. This is the problem which the discussions in the next sections attempt to address.

We remark that in estimating the mean μ_A by $\hat{\mu}_A$, the bias is zero, and the problem of the preceding paragraph does not arise. This is the reason that, in simple random sampling, standard inference for means is sound, at least when the domain variates are normally distributed.

2.3 General Methodology for Confidence Intervals

Let $\hat{\theta} = (\hat{T}_A - T_A)/s_{\hat{T}_A}$, where $s^2_{\hat{T}_A}$ is an estimator (to be specified) of the (conditional or unconditional) variance of the total. Assume that the form of the conditional (on $\hat{p}_A$) distribution function of $\hat{\theta}$, say $H(\cdot \mid \hat{p}_A; p_A, \mu_A, \sigma^2_A)$, is known where p_A, μ_A and σ^2_A represent unknown parameters. In order to construct a conditional equal tailed $(1 - \alpha) \times 100\%$ confidence interval (CI) for T_A, we define an upper critical value

$$c_u \equiv c_u(\alpha, \hat{p}_A, p_A) = -\inf\left\{x \mid H\left(x \mid \hat{p}_A; p_A\right) \geq \alpha/2\right\} = -H^{-1}\left(\alpha/2, \hat{p}_A; p_A\right)$$

where p_A is considered fixed and the dependence on μ_A and σ^2_A is temporarily suppressed; a lower critical value, say c_ℓ, is defined in a similar manner. A conditional, equal tailed $(1 - \alpha) \times 100\%$ CI for T_A is then given by $\mathrm{CI}(1 - \alpha) = (\ell, u)$, where

$$u = \hat{T}_A + c_u s_{\hat{T}_A} \text{ and } \ell = \hat{T}_A + c_\ell s_{\hat{T}_A}. \qquad (3)$$

At this point the obvious practical problem is that the critical values c_u and c_ℓ depend not only on $\hat{p}_A$ but also on the unknown parameter p_A. One approach to this problem is to take a Bayesian tack and assume the parameter p_A is the realization of a random variable. Adjusting the notation to reflect the assumption that p_A is stochastic, we replace $H(x \mid \hat{p}_A; p_A)$ by $H(x \mid \hat{p}_A, p_A)$ and have that

$$\Pr\left\{\hat{\theta} \leq x \mid \hat{p}_A\right\} = F\left(X \mid \hat{p}_A\right)$$

$$= \frac{1}{h(\hat{p}_A)} \int H\left(x \mid \hat{p}_A, p_A\right) f\left(\hat{p}_A \mid p_A\right) g\left(p_A\right) dp_A, \qquad (4)$$

where $h(\hat{p}_A) = \int f(\hat{p}_A \mid p_A) g(p_A) dp_A$ and $g(p_A)$ is the density of p_A. It should be noted that as a consequence of our sampling scheme the distribution of $n\hat{p}_A$, conditional on p_A, is Binomial (n, p_A) so that $f(\hat{p}_A \mid p_A)$ is known. Under the Bayesian approach, the critical values are $c^*_u \equiv c^*_u(\alpha, \hat{p}_A) = -F^{-1}(\alpha/2 \mid \hat{p}_A)$ and $c^*_\ell \equiv c^*_\ell(\alpha, \hat{p}_A) = -F^{-1}(1 - \alpha/2 \mid \hat{p}_A)$ so the upper and lower limits for a conditional $(1 - \alpha) \times 100\%$ CI for T_A are

$$u = \hat{T}_A + c^*_u s_{\hat{T}_A} \text{ and } \ell = \hat{T}_A + c^*_\ell s_{\hat{T}_A}. \qquad (5)$$

For the purposes of our current research, we assume that the prior distribution $g(p_A)$ is $N(\mu_{p_A}, \sigma^2_{p_A})$ with μ_{p_A} and $\sigma^2_{p_A}$ to be specified, with the understanding that $\sigma^2_{p_A}$ is sufficiently small that p_A lies between 0 and 1 with near certainty. The normality assumption is made for mathematical convenience. It also captures notions we may have of degrees of closeness to, and symmetry about, μ_{p_A}. For an empirical Bayes approach, we use $\mu_{p_A} = \hat{p}_A$; we consider several possible alternatives for $\sigma^2_{p_A}$ discussed in detail below. Our experience indicates that the normality assumption is not crucial; rather, it is primarily a matter of convenience.

2.4 Confidence Intervals Under Normal Assumptions

To proceed further we assume that within the domain A the x_i are distributed $N(\mu_A, \sigma^2_A)$. In practice, this assumption may not be met. Nonetheless, it leads to suggested modifications that will not at any rate give lower coverage of confidence intervals than the standard approach. Combining this assumption with earlier results, in particular equation (2), and ignoring lower order terms, we have

(a) $[\sqrt{n}(\hat{T}_A - T_A)/n \mid \hat{p}_A, p_A]$ is distributed $N(\sqrt{n}\mu_A(\hat{p}_A - p_A), \hat{p}_A\sigma^2_A)$,

(b) $\left[(n\hat{p}_A - 1)\dfrac{\hat{\sigma}^2_A}{\sigma^2_A} \mid \hat{p}_A, p_A\right]$ is distributed $\chi^2(n\hat{p}_A - 1)$, and

(c) the conditional random variable in (b) is stochastically independent of the conditional random variable in (a).

Consider $\hat{\theta}_1 = (\hat{T}_A - T_A)/(N\hat{\sigma}_A\sqrt{\hat{p}_A}/\sqrt{n})$, which utilizes the conditional variance of $\hat{T}_A$ as the standardizing term. It follows immediately from (a), (b) and (c) that, conditional on $(\hat{p}_A, p_A)$ the random variable $\hat{\theta}_1$ is distributed as a non-central t with $n\hat{p}_A - 1 = n_A - 1$ degrees of freedom and non-centrality parameter

$$\lambda = \sqrt{n}\gamma_A(\hat{p}_A - p_A)/\sqrt{\hat{p}_A},$$

with

$$\gamma_A = \mu_A/\sigma_A.$$

Thus, we have specified the conditional distribution function $H(\cdot\,|\,\hat{p}_A, p_A)$ of $\hat{\theta}_1$. As $f(\hat{p}_A\,|\,p_A)$ and $g(p_A)$ have been previously specified, it follows that $F(\cdot\,|\,\hat{p}_A)$ in (4) is well-defined although extremely cumbersome to calculate. The dependence on μ_A and σ_A^2, through γ_A, should be noted.

Although $F(\cdot\,|\,\hat{p}_A)$ as given above can be used to determine the critical values, they are extremely difficult to calculate. A relatively simple approach, given in the next paragraph, provides a close approximation to the critical values. We have verified the closeness of the approximation by computing the exact values for selected cases using large scale simulations.

Adoption of a locally uniform prior on p_A leads to the approximate posterior distribution $p_A \sim N(\hat{p}_A, \mathrm{var}(\hat{p}_A))$ and we could approximate $\mathrm{var}(\hat{p}_A)$ by $\hat{p}_A(1-\hat{p}_A)/n$. We adopt the slightly more flexible prior $p_A \sim N(\mu, \sigma_{p_A}^2)$, and empirically choose $\mu = \hat{p}_A$, with several possibilities for $\sigma_{p_A}^2$ that will be specified below. It follows from Appendix A that $[\lambda\,|\,\hat{p}_A]$ is distributed approximately as a normal with mean zero and variance $\gamma_A^2(1-\hat{p}_A)/(1+\psi_A)$, where

$$\psi_A = \hat{p}_A(1-\hat{p}_A)/n\sigma_{p_A}^2.$$

Then, from the result in Appendix B, conditional on $\hat{p}_A$,

$$\frac{\left(\hat{T}_A - T_A\right)}{\dfrac{N\hat{\sigma}_A\sqrt{\hat{p}_A}}{\sqrt{n}}\sqrt{\dfrac{\gamma_A^2(1-\hat{p}_A)}{1+\psi_A}+1}}$$

is distributed as a central t with $n_A - 1$ degrees of freedom. Let $t_{1-\alpha/2,n_A-1}$ be the $(1-\alpha/2)100\%$ percentile of this distribution. The upper confidence limit u, defined in (5), is given (approximately) by

$$u = \hat{T}_A + N\hat{\sigma}_A\sqrt{\hat{p}_A/n}\times \left(\left(\gamma_A^2\left(1-\hat{p}_A\right)+1+\psi_A\right)\Big/\left(1+\psi_A\right)\right)^{1/2} t_{1-\alpha/2,n_A-1}. \tag{6}$$

As $\hat{\sigma}_A^2$ is conditionally unbiased for σ_A^2 and $\hat{\mu}_A^2 - \hat{\sigma}_A^2/n_A$ is conditionally unbiased for μ_A^2, we use $\hat{\gamma}_A^2 = (\hat{\mu}_A^2 - \hat{\sigma}_A^2/n_A)/\hat{\sigma}_A^2$ to estimate γ_A^2. Substituting $\hat{\gamma}_A^2$ for γ_A^2 in (6) yields

$$\tilde{u} \cong \hat{T}_A + \left(Ns/\sqrt{n}\right)\times \left(\left(1+\frac{\hat{p}_A\hat{\sigma}_A^2\psi_A}{s^2}\right)\Big/\left(1+\psi_A\right)\right)^{1/2} t_{1-\alpha/2,n_A-1} \tag{7}$$

where s^2 is defined in (1).

It remains to choose ψ_A. We note that $\tilde{u}$ is strictly decreasing as ψ_A increases and

$$\tilde{u} \to \hat{T}_A + \frac{Ns}{\sqrt{n}} t_{1-\alpha/2,n_A-1} = \tilde{u} \text{ as } \psi_A \text{ becomes small,}$$

$$\tilde{u} = \hat{T}_A + \frac{Ns}{\sqrt{n}}\left(\frac{1+\hat{p}_A\hat{\sigma}_A^2/s^2}{2}\right)^{1/2} t_{1-\alpha/2,n_A-1} = \tilde{u}_2 \text{ for } \psi_A = 1,$$

and

$$\tilde{u} \to \hat{T}_A + \frac{Ns}{\sqrt{n}}\left(\frac{\sqrt{\hat{p}_A}\hat{\sigma}_A}{s}\right) t_{1-\alpha/2,n_A-1} = \tilde{u}_3$$

$$\text{as } \psi_A \text{ becomes large.} \tag{8}$$

In each case the lower critical value can be dealt with in an analogous manner resulting in three competing confidence intervals; namely, $\mathrm{CI}_i(1-\alpha) = (\tilde{\ell}_i, \tilde{u}_i)$, $i = 1, 2, 3$, with $\tilde{\ell}_i$ defined similarly to $\tilde{u}_i$ in (8) with $t_{1-\alpha/2,n_A-1}$ replaced by $t_{\alpha/2,n_A-1}$. The competing confidence intervals are labeled in order of decreasing length.

The first case is equivalent to assuming that $\sigma_{p_A}^2$ is large relative to $\mathrm{var}(\hat{p}_A)$ and leads to using the usual unconditional variance but with degrees of freedom equal to $n_A - 1$. In most practical problems this seems reasonable since $\sigma_{p_A}^2$ is an unknown constant and $\mathrm{var}(\hat{p}_A)$ is $0(p_A/n)$. The second interval corresponds to adoption of a normal prior as noted above, with $\sigma_{p_A}^2 = \hat{p}_A(1-\hat{p}_A)/n$. The last confidence interval is based on the assumption that p_A is essentially degenerate at $\hat{p}_A$.

2.5 Empirical Study for SRS

We compared the several confidence intervals of Section 2.4 in a small empirical study, using artificial populations, for which the domain variable was normal. In all cases the population size N was 1,000, and the sample size n was 100 or 300. The parameters p_A and γ_A varied from population to population. Letting M_2 be the number of runs with $n_A \geq 2$, we allowed the run size M to vary to give $M_2 = 10{,}000$. Table 1 gives coverage results. CI_0 represents the confidence interval based on the standard normal methodology. The results for CI_2 closely approximated the results for CI_1 and are excluded. The value of M is included to indicate how many trials fell into the "insufficient information" pile, at a given setting of the parameters. Several conclusions seem warranted:

1. Standard confidence intervals using the usual variance estimate and normal quantiles can give low coverage. This occurs for several values of p_A when $\gamma_A = 1/2$ or $\gamma_A = 2$, however, the under-coverage is not too severe if the domain variable is normal. The case where

$\gamma_A = 2$ or takes even larger values is probably more likely in practice. Thus if the domain variable is normal, the use of standard confidence intervals under simple random sampling case is not particularly worrisome.

2. The strictly conditional intervals (*i.e.*, CI_3) using the conditional variance can give abominable coverage, when γ_A is large. That is, confidence intervals based on "large" values of ψ_A gave very poor results.
3. The use of the standard variance estimate but replacing the standard normal quantile with a *t*-quantile having degrees of freedom based on the number of sample units in the domain (*i.e.*, CI_1) gives approximately nominal or conservative coverage regardless of the value of γ_A.

Table 1
Coverage of 95% Confidence Intervals for Domain Total for Artificial Populations with Domain Variate Normally Distributed*

			Coverage		
p_A	n	M	CI_0	CI_1	CI_3
		$y = 1/2$			
.01	100	38774	100.0	100.0	91.2
	300	11773	98.3	100.0	83.2
.02	100	16327	91.1	99.4	95.0
	300	10078	88.6	95.5	93.9
.05	100	10303	88.7	97.8	93.5
	300	10000	92.3	94.4	92.5
.10	100	10001	90.9	94.8	92.5
	300	10000	94.0	95.0	92.3
		$y = 2$			
.01	100	37749	99.9	100.0	83.5
	300	11740	94.4	100.0	89.1
.02	100	16348	99.0	100.0	88.4
	300	10075	91.4	98.9	74.7
.05	100	10312	90.5	99.5	77.6
	300	10000	93.8	95.8	66.6
.10	100	10000	91.7	96.5	67.9
	300	10000	94.0	95.2	65.0

* See Equation (8) and accompanying text for definition of CI_1 and CI_3. CI_0 is the standard normal confidence interval.

As a minor observation on the results, we note the counter-intuitive increases in coverage for smaller p_A and n. We believe this is due to the fact that, at very small values of p_A and n, $\hat{p}_A$ is constrained to be positive, and so cannot deviate much below p_A. Were intervals calculable for $n_A = 0$, there would be a serious drop in coverage in these cases. Note that the coverage rises unexpectedly only where M is large.

3. THE CASE OF STRATIFIED RANDOM SAMPLING

3.1 Definitions and Notation

Assume there are K strata and, where appropriate, terms previously defined have corresponding stratum level definitions. For example, n_k is the sample size and n_{Ak} is the number of sample elements in A for the k-th stratum. Thus, a natural estimator for the domain total $T_A = \sum_{k=1}^{K} \sum_{i\in A} x_{ki} = \sum_{k=1}^{K} N_k \hat{p}_{Ak} \mu_{Ak}$ is

$$\hat{T}_A = \sum_{k\in B_1} \hat{T}_{Ak} = \sum_{k\in B_1} N_k \hat{p}_{Ak} \hat{\mu}_{Ak},$$

where $\hat{p}_{Ak} = n_{Ak}/n_k$, $\hat{\mu}_{Ak} = \sum_{i=1}^{n_{Ak}} x_{ki}/n_{Ak}$ and $B_1 = \{k \mid n_{Ak} \geq 1$ and $1 \leq k \leq K\}$. As $\hat{p}_{Ak} = 0$ for $k \notin B_1$, it is straightforward to verify that

$$E\left[(\hat{T}_A - T_A) \mid \hat{\boldsymbol{p}}_A, \boldsymbol{p}_A\right] = \sum_{k=1}^{K} N_k (\hat{p}_{Ak} - p_{Ak}) \mu_{Ak} \equiv \tilde{\mu}_A \quad (9)$$

and

$$\text{var}\left[(\hat{T}_A - T_A) \mid \hat{\boldsymbol{p}}_A, \boldsymbol{p}_A\right] = \sum_{k\in B_1} N_k^2 \hat{p}_{Ak}^2 \sigma_{Ak}^2 / n_{Ak} =$$

$$\sum_{k\in B_1} N_k^2 \hat{p}_{Ak}^2 \sigma_{Ak}^2 / n_k \equiv \tilde{\sigma}_A^2,$$

where $\hat{\boldsymbol{p}}_A = [\hat{p}_{A1} \hat{p}_{A2} \cdots \hat{p}_{AK}]$, $\boldsymbol{p}_A = [p_{A1} p_{A2} \cdots p_{AK}]$. Thus, as in the simple random sampling case, there is a conditional bias $\tilde{\mu}_A$, which needs to be taken into account.

3.2 A Methodology for Confidence Intervals

The general methodology for confidence intervals of Section 2.3 for simple random sampling holds here as well. One need only reinterpret scalars as vectors; for example, replace $\hat{p}_A$ by $\hat{\boldsymbol{p}}_A = (\hat{p}_{A1}, \ldots, \hat{p}_{AK})'$. In particular, $H(x \mid \hat{\boldsymbol{p}}_A, \boldsymbol{p}_A) = \Pr\{\hat{\theta} \leq x \mid \hat{\boldsymbol{p}}_A, \boldsymbol{p}_A\}$ will be the conditional distribution function of $\hat{\theta} = (\hat{T}_A - T_A)/\hat{\sigma}_A$, where $\hat{\sigma}_A$ is a re-scaling factor to be specified.

Let $B_2 = \{k \mid n_{Ak} \geq 2$ and $1 \leq k \leq K\}$ and, for $k \in B_2$, define $\hat{\sigma}_{Ak}^2 = \sum_{i=1}^{n_{Ak}} (x_{ki} - \hat{\mu}_{Ak})^2/(n_{Ak} - 1)$. Under normality, $(n_{Ak} - 1)\hat{\sigma}_{Ak}^2/\sigma_{Ak}^2 \sim \chi^2(n_{Ak} - 1)$, so if $\{d_k \mid k \in B_2\}$ are non-negative constants with $\sum_{k\in B_2} d_k > 0$, then by the usual Satterthwaite (1946) two moment approximation, the conditional random variable

$$\left[(1/c) \sum_{k\in B_2} d_k (n_{Ak} - 1)(\hat{\sigma}_{Ak}^2/\sigma_{Ak}^2) \mid \hat{\boldsymbol{p}}_A, \boldsymbol{p}_A\right]$$

is distributed approximately as a $\chi^2(v)$, where

$$c = \sum_{k\in B_2} d_k^2 (n_{Ak} - 1) \Big/ \sum_{k\in B_2} d_k (n_{Ak} - 1)$$

and

$$v = \Big(\sum_{k\in B_2} d_k (n_{Ak} - 1)\Big)^2 \Big/ \sum_{k\in B_2} d_k^2 (n_{Ak} - 1).$$

This suggests that we restrict our attention to expressions of the general form

$$\hat{\sigma}_A^2 = \sum_{k\in B_2} d_k (n_{Ak} - 1) \hat{\sigma}_{Ak}^2 / \sigma_{Ak}^2$$

with choice of the d_k to be specified. Note that when $B_1 = B_2$ and $d_k = N_k^2 \hat{p}_{Ak} \sigma_{Ak}^2 / n_k(n_{Ak} - 1)$, $\hat{\sigma}_A^2 = \hat{\tilde{\sigma}}_A^2 \equiv \sum_{k \in B_2} d_k(n_{Ak} - 1) \hat{\sigma}_{Ak}^2 / \sigma_{Ak}^2$ is an unbiased estimator for the conditional variance $\tilde{\sigma}_A^2$. However, as in the simple random sampling case, this estimator will tend to be too small. We use the more general expression to develop a family of t-statistics when we "uncondition" on $\boldsymbol{p}_A$. Each of these will involve unknown parameters, and, as in the simple random sampling case (transition of equation (6) to equation (7)), estimation of these unknowns will be necessary. Thus the net result will be several rival "near t-statistics" which we may then compare empirically.

Because the samples are selected independently from each stratum we have $f(\hat{\boldsymbol{p}}_A | \boldsymbol{p}_A) = \Pi_{k=1}^K f_k(\hat{p}_{Ak} | p_{Ak})$ and, as a consequence of our within stratum sampling scheme, $n_k \hat{p}_{Ak}$ has a binomial distribution $B(n_k, p_{Ak})$. We assume that the $\{p_{Ak} | 1 \le k \le K\}$ are jointly independent so $g(\boldsymbol{p}_A) = \Pi_{k=1}^K g_k(p_{Ak})$ which implies

$$f(\hat{\boldsymbol{p}}_A | \boldsymbol{p}_A) g(\boldsymbol{p}_A) = \prod_{k=1}^K f_k(\hat{p}_{Ak} | p_{Ak}) g_k(p_{Ak})$$

and

$$h(\hat{\boldsymbol{p}}_A) = \prod_{k=1}^K \int f_k(\hat{p}_{Ak} | p_{Ak}) g_k(p_{Ak}) dp_{Ak}.$$

In what follows, we assume that the prior distribution of p_{Ak} is $N(\mu_{p_{Ak}}, \sigma_{p_{Ak}}^2)$ and for the empirical Bayes approach, we use $\mu_{p_{Ak}} = \hat{p}_{Ak}$ and, analogously to the case of simple random sampling, we define

$$\psi_{Ak} = \hat{p}_{Ak}(1 - \hat{p}_{Ak}) / n_k \sigma_{p_{Ak}}^2.$$

It is straightforward to extend the result in Appendix A to the case of stratified random sampling and it then follows that, for $\tilde{\mu}_A$ defined by (9), $[\tilde{\mu}_A / \tilde{\sigma}_A | \hat{\boldsymbol{p}}_A]$ is distributed $N(0, \mathrm{var}(\tilde{\mu}_A | \hat{\boldsymbol{p}}_A) / \tilde{\sigma}_A^2)$, where $\mathrm{var}(\tilde{\mu}_A | \hat{\boldsymbol{p}}_A) = \sum_{k \in B_1} N_k^2 \mu_{Ak}^2 \hat{p}_{Ak}(1 - \hat{p}_{Ak}) / n_k(1 + \psi_{Ak})$. Using the result in Appendix B, it follows that, conditional on $\hat{\boldsymbol{p}}_A$, the random variable

$$\hat{\theta} = \frac{(\hat{T}_A - T_A) \Big/ \sqrt{\mathrm{var}(\tilde{\mu}_A | \hat{\boldsymbol{p}}_A) + \tilde{\sigma}_A^2}}{\sqrt{\hat{\sigma}_A^2 / cv}} =$$

$$\frac{(\hat{T}_A - T_A) \Big/ \sqrt{\mathrm{var}(\tilde{\mu}_A | \hat{\boldsymbol{p}}_A) + \tilde{\sigma}_A^2}}{\sqrt{\sum_{k \in B_1} d_k(n_{Ak} - 1)(\hat{\sigma}_{Ak}^2 / \sigma_{Ak}^2) \Big/ \sum_{k \in B_1} d_k(n_{Ak} - 1)}}$$

is distributed approximately as a central t with v degrees of freedom.

Letting $\Theta = \mathrm{var}(\tilde{\mu}_A | \hat{\boldsymbol{p}}_A) + \tilde{\sigma}_A^2$, with

$$\gamma_{Ak}^2 = \mu_{Ak}^2 / \sigma_{Ak}^2$$

and assuming the ψ_{Ak} are near zero we have

$$\Theta = \sum_{k \in B_1} \frac{N_k^2 \hat{p}_{Ak} \sigma_{Ak}^2}{n_k} (\gamma_{Ak}^2 (1 - \hat{p}_{Ak}) + 1).$$

Thus, the upper bound on the CI would be (approximately)

$$u = \hat{T}_A + \frac{\sqrt{\sum_{k \in B_2} d_k(n_{Ak} - 1)(\hat{\sigma}_{Ak}^2 / \sigma_{Ak}^2)}}{\sqrt{\sum_{k \in B_2} d_k(n_{Ak} - 1)}} \Theta^{1/2} t_v, \qquad (10)$$

where t_v stands for the critical values of the t_v distribution. Unfortunately the bound depends not only on our choice of the d_k, but also on the unknown parameters μ_{Ak} and σ_{Ak}^2.

It is not hard to show that $v \le \sum_{k \in B_2} (n_{Ak} - 1) \equiv v_{\max}$ and, if we set $d_k = 1$ (or any constant for that matter) then $v = v_{\max}$. We refer to $v_{\max}$ specifically as the unweighted degrees of freedom. In this case the upper bound on the CI would be

$$u = \hat{T}_A + \frac{\sqrt{\sum_{k \in B_2} d_k(n_{Ak} - 1)(\hat{\sigma}_{Ak}^2 / \sigma_{Ak}^2)}}{\sqrt{\sum_{k \in B_2} (n_{Ak} - 1)}} \Theta^{1/2} t_{v_{\max}}.$$

Another approach is to attempt to finesse the problem of estimating Θ (at least when $B_1 = B_2$) by a judicious choice of the d_k. To that end let us assume that $B_1 = B_2$ and let

$$d_k = \frac{N_k^2 \hat{p}_{Ak} \sigma_{Ak}^2}{n_k(n_{A_k} - 1)} (\gamma_{Ak}^2 (1 - \hat{p}_{Ak}) + 1)$$

so that $\sum_{k \in B_2} d_k(n_{Ak} - 1) = \Theta$ and Θ cancels out in (10). We then have

$$u = \hat{T}_A + \sqrt{\sum_{k \in B_2} \frac{N_k^2 \hat{p}_{A_k} \hat{\sigma}_{Ak}^2}{n_k} (\gamma_{Ak}^2 (1 - \hat{p}_{Ak}) + 1) t_{v_1}},$$

where v_1 is the degrees of freedom associated with this second choice of the d_k. More generally (*i.e.*, when $B_1 \ne B_2$), we have

$$u = \hat{T}_A + \frac{\sqrt{\sum_{k \in B_2} \frac{N_k^2 \hat{p}_{A_k} \hat{\sigma}_{Ak}^2}{n_k} (\gamma_{Ak}^2 (1 - \hat{p}_{Ak}) + 1)}}{\sqrt{\sum_{k \in B_2} \frac{N_k^2 \hat{p}_{A_k} \sigma_{Ak}^2}{n_k} (\gamma_{Ak}^2 (1 - \hat{p}_{Ak}) + 1)}} \Theta^{1/2} t_{v_1}.$$

In any event, we are still faced with the problem of estimating the population parameters and we have the additional problem of estimating the degrees of freedom.

A third possibility, which we have already mentioned, is to let $d_k = N_k^2 \hat{p}_{Ak} \sigma_{Ak}^2 / n_k(n_{Ak} - 1)$ so that when $B_1 = B_2$, $\hat{\sigma}_A^2 = \hat{\tilde{\sigma}}_A^2 \equiv \sum_{k \in B_2} d_k(n_{Ak} - 1) \hat{\sigma}_{Ak}^2 / \sigma_{Ak}^2$ is a conditionally unbiased estimator for $\tilde{\sigma}_A^2$. In this case we have

$$u = \hat{T}_A + \frac{\sqrt{\sum_{k \in B_2} N_k^2 \hat{p}_k \hat{\sigma}_{Ak}^2 / n_k}}{\sqrt{\sum_{k \in B_2} N_k^2 \hat{p}_k \sigma_{Ak}^2 / n_k}} \Theta^{1/2} t_{v_2},$$

where v_2 is the degrees of freedom associated with this third choice of the d_k. As in the second case, we are faced with the problem of estimating the population parameters and the degrees of freedom.

Now, it should be noted that if we estimate σ_{Ak}^2 with $\hat{\sigma}_{Ak}^2$ for $k \in B_2$ and let $\hat{\Theta}$ be a yet to be specified estimator of Θ then the (estimated) upper bounds above are $u = \hat{T}_A + \hat{\Theta}^{1/2} t_{v_{\max}}$, $u = \hat{T}_A + \hat{\Theta}^{1/2} t_{\hat{v}_1}$ and $u = \hat{T}_A + \hat{\Theta}^{1/2} t_{\hat{v}_2}$ respectively. The degrees of freedom are estimated by substituting estimates of the population parameters into the two respective choices of the d_k. Both $\hat{v}_1$ and $\hat{v}_2$ are smaller than $v_{\max}$, so, for any realized value of $\hat{\Theta}$, the confidence interval using $v_{\max}$ will be the shortest. There is no general relationship between the sizes of $\hat{v}_1$ and $\hat{v}_2$. Empirical evidence indicates that there is little to choose between the second and third approach.

Addressing the problem of estimating Θ, we can write

$$\Theta = \sum_{k \in B_1 - B_2} N_k^2 \hat{p}_{Ak} \left(\mu_{Ak}^2 (1 - \hat{p}_{Ak}) + \sigma_{Ak}^2 \right) \Big/ n_k +$$

$$\sum_{k \in B_2} N_k^2 \hat{p}_{Ak} \left(\mu_{Ak}^2 (1 - \hat{p}_{Ak}) + \sigma_{Ak}^2 \right) \Big/ n_k.$$

For $k \in B_1 - B_2$ the estimator $\hat{\sigma}_{Ak}^2$ is not defined, however, it is straightforward to verify that $(1 - \hat{p}_{Ak}) E[\hat{\mu}_{Ak}^2 \mid n_{Ak}] \le \sigma_{Ak}^2 + \mu_{Ak}^2 (1 - \hat{p}_{Ak}) \le E[\hat{\mu}_{Ak}^2 \mid n_{Ak}]$. It follows that

$$s_a^2 = \sum_{k \in B_1} N_k^2 \hat{p}_{Ak} (1 - \hat{p}_{Ak}) \hat{\mu}_{Ak}^2 / n_k +$$

$$\sum_{k \in B_2} N_k^2 \hat{p}_{Ak} \hat{\sigma}_{Ak}^2 (1 + 1/n_k - 1/n_{Ak}) / n_k$$

will tend to underestimate Θ, and

$$s_b^2 = \sum_{k \in B_1 - B_2} N_k^2 \hat{p}_{Ak} \hat{\mu}_{Ak} \Big/ n_k + \sum_{k \in B_1} N_k^2 \hat{p}_{Ak} (1 - \hat{p}_{Ak}) \hat{\mu}_{Ak}^2 \Big/ n_k +$$

$$\sum_{k \in B_2} N_k^2 \hat{p}_{Ak} \hat{\sigma}_{Ak}^2 (1 + 1/n_k - 1/n_{Ak}) / n_k$$

will tend to overestimate Θ. Clearly, $s_a^2 \le s_b^2$ with equality only when $B_1 = B_2$.

It can also be verified that in the case of stratified sampling, the standard variance estimator for estimated population totals is

$$s_{\text{std}}^2 = \sum_{k \in B_1} N_k^2 s_k^2 \Big/ n_k = \sum_{k \in B_1} N_k^2 \hat{p}_{Ak} (1 - \hat{p}_{Ak}) \hat{\mu}_{Ak}^2 \Big/ (n_k - 1)$$

$$+ \sum_{k \in B_2} N_k^2 \hat{p}_{Ak} \hat{\sigma}_{Ak}^2 (1 - 1/n_{Ak}) / (n_k - 1).$$

This looks like a satisfactory estimator of Θ, if the n_k are not small.

These results imply that CIs of the form $(\hat{T}_A \pm s_b t_{1-\alpha/2, \hat{v}_1})$ will provide the highest level of coverage; but CIs of the form $(\hat{T}_A \pm s_{\text{std}} t_{1-\alpha/2, v_{\max}})$ and even perhaps $(\hat{T}_A \pm s_{\text{std}} t_{1-\alpha/2, \hat{v}_1})$ have obvious computational advantages. Several of these competing forms of CI are evaluated empirically in Section 3.3. These results can easily be extended to ratio estimators by the standard linearization approach.

3.3 Empirical Investigation for Stratified Random Sampling: the BLS Wage Data

With a view to improving estimation of precision on wage data produced by the U.S. Bureau of Labor Statistics, we investigated coverage and interval length in two simulation studies on populations constructed from a test sample of the Occupational Compensation Survey Program (OCSP) conducted in 1991. The OCSP consisted of establishment surveys in several metropolitan areas, aimed at estimating wages levels for a select group of occupations. The surveys were carried out by stratified simple random sampling, with establishments stratified by employment size and industrial classification.

One population (the "Small Population") took the test sample itself as the population, with six non-certainty strata, and one certainty stratum of 12 establishments. Five hundred stratified random samples were taken from this population at sizes $n = 36$ and 60, corresponding to the choices $n_k = 4$ and $n_k = 8$, reflecting relative sample sizes of sampling from the original population. The second population (the "Large Population") was constructed by expanding the sample data through replication (by simple random sampling with replacement, within each Small Population stratum) of establishments to achieve a population the size of the original population; again there were six noncertainty and one certainty strata; for each stratum sample sizes were the same as in the actual sample. Domains are defined by the different occupations of interest; only a fraction of establishments have workers in a particular occupation, and lie in the corresponding domain. Table 2 gives the number of establishments having workers in the selected occupations for the small population.

In both cases sampling was without replacement, so finite population correction factors were included (as appropriate) in the construction of the CIs. Also, the study was limited to a concern with 95% coverage.

Table 2
Number of Establishments in Given Domain (Occupation), by Stratum for Small Population

	stratum							
Occupation	1	2	3	4	5	6	7	total
4021	0	4	11	10	8	10	7	50
1141	0	3	11	7	11	9	7	48
1122	0	3	8	13	14	12	6	56
3180	10	11	5	25	20	4	5	80
2911	0	3	14	2	13	17	7	56
1142	2	8	15	9	15	19	9	77
1180	17	20	5	61	31	3	1	138
1403	12	16	22	28	25	27	9	139
All Estabs	35	35	33	136	66	36	12	353

Small Population: Table 3 gives coverage and median relative interval length for total wages, at two sample sizes $n_k = 4$ and $n_k = 8$, for 8 occupations, and three methods of confidence interval construction: the standard variance estimator, s^2_{std}, with the standard normal z-quantile, the unweighted degrees of freedom $\nu_{\max}$, and the weighted degrees of freedom ν_1. Occupations are ordered by increasing values of the average value, over runs, of the unweighted degrees of freedom. We note:

1) Almost universally, coverage using the standard variance estimator and the standard normal quantiles (infinite *df*) is poor.

2) Coverage for the other interval types is far more satisfactory. In general, the coverage is near the nominal 95%, or slightly conservative, for weighted degrees of freedom; as expected, intervals based on unweighted degrees of freedom tend to yield coverage a few points below those based on weighted degrees of freedom.

3) Two occupations (1122, 4021) yield seriously low coverage for totals even with the improved procedures. Investigation of these particular occupations suggests a strong violation of the normality assumption. In 4021, for example, two units in stratum 5 have a number of workers, and hence total wages, an order of magnitude higher than the other establishments in this stratum and indeed in the population. Furthermore, the wage rate of these two outliers is markedly lower than the great bulk of establishments: with just these two excluded from the population, the overall population average wage would be $9.68/hour; with them in, it is $8.28. Since there are 66 establishments in stratum 5, it is easy for these two establishments to escape being in a sample of size 8; the consequence is a serious overestimate of the mean wage or underestimate of total wage. At the same time, wages for the establishments that are in the sample are relatively homogeneous, so the variance estimate will tend to be too low. The presence of several smaller establishments in the domain contribute to enlarging the degrees of freedom, and so the t-adjustment is unable to compensate fully. It is hard to see how to guard against such a problem short of having prior information, and allotting such outliers to a certainty stratum. Even so, the adjusted intervals are a significant improvement on the naïve normal distribution based interval.

Interval lengths are taken relative to $2 \times z_{.975} \approx 4$ times the root mean square error of $\hat{T}_A$ calculated over runs. We report the median of these standardized lengths (across runs). When the distribution of $\hat{T}_A$ is actually normal, the median length is close to 1.

Table 3
Estimated degrees of freedom, coverage, and relative median length of CIs for total wages of workers in occupation, for the small population

	Four Sample Establishments Per Stratum								Eight Sample Establishments Per Stratum							
Occupation	4021	1141	1122	3180	2911	1142	1180	1403	1141	4021	1122	3180	2911	1142	1180	1403
$df = \nu_{\max}$	1.5	1.6	1.6	2.0	2.3	2.8	4.3	6.1	3.7	3.8	3.9	5.6	6.0	8.0	12.3	16.6
$df = \hat{\nu}_1$	1.3	1.3	1.4	1.5	1.7	1.9	2.3	3.5	2.0	2.3	2.3	3.1	3.5	4.3	5.4	9.7
								Coverage								
$\hat{T}_A \pm s_{\text{std}} z$	.47	.69	.51	.75	.73	.85	.89	.87	.74	.49	.65	.79	.78	.86	.88	.92
$\hat{T}_A \pm s_{\text{std}} t_{\nu_{\max}}$	.89	.92	.93	.99	.95	.96	.97	.92	.87	.65	.75	.89	.86	.90	.90	.94
$\hat{T}_A \pm s_{\text{std}} t_{\hat{\nu}_1}$	.92	.93	.95	.99	.96	.96	.98	.95	.91	.74	.80	.94	.89	.95	.96	.96
								Median Relative Length								
$\hat{T}_A \pm s_{\text{std}} z$	0.53	0.75	0.59	0.70	0.74	0.85	0.90	0.88	0.87	0.63	0.66	0.80	0.83	0.88	0.92	0.96
$\hat{T}_A \pm s_{\text{std}} t_{\nu_{\max}}$	2.65	3.67	2.80	2.60	2.20	1.98	1.50	1.14	1.63	1.09	1.13	1.10	1.10	1.06	1.02	1.04
$\hat{T}_A \pm s_{\text{std}} t_{\hat{\nu}_1}$	3.30	4.32	3.19	3.40	3.08	3.06	2.70	1.58	3.08	2.40	2.38	2.00	1.74	1.38	1.38	1.13

4) The relative interval length of the standard interval tends to be too small, that is, it tends to be less than 1.

5) Interval length among the other variance-degrees of freedom combinations is largest for s_{std}^2 with $\hat{v}_1$, and smallest for s_{std}^2 with $v_{\max}$. These differences can be appreciable; there is a tradeoff between coverage and interval size.

6) For a given interval type, the relative interval length tends to 1 as $v_{\max}$ increases. The conclusions from a study of mean wages are similar.

Large Population: Table 4 gives coverage and interval length for total wages for five interval types, and a wider range of occupations, ordered by average $v_{\max}$. The interval types include the three used previously for the small population. The two new intervals utilize the weighted degrees of freedom together with s_a and s_b respectively. Results are based on 5,000 runs.

1) The results are consistent with those for the Small Population, in terms of the relative coverage and interval sizes of the several interval types. The standard normal is unsatisfactory for many occupations.

2) The coverage for intervals using the weighted degrees of freedom, $\hat{v}_1$, is less than 90% for only a small fraction of cases.

3) There can be marked differences in interval length for the different interval types; however, all ratios of interval length to 4 × root mean square error tend to 1, as $v_{\max}$ gets large.

4) Little difference results from using s_a, s_b, or s_{std} with $t_{\hat{v}_1}$.

Again, the results for mean wages, while differing in detail, lead to the same overall conclusions, and are omitted.

4. SUMMARY AND CONCLUSIONS

From our theoretical investigation and simulation work, we draw the following conclusions:

1. Standard 95% confidence intervals for domain means or totals, when based on the standard normal distribution and standard methods of variance estimation, tend to yield less than actual 95% coverage. The extent of the deviation will vary with domain (occupation in the wage study), but can be quite considerable even when the sample size is large.
2. New nonstandard methods offer a sharp improvement, giving intervals with better coverage, typically at or close to the nominal 95% coverage. These intervals tend to be longer than the standard intervals. The increase in length will vary with domain, and will depend on the particular method for CI construction that is adopted.

Table 4
Estimated degrees of freedom, coverage, and relative median length of CIs for total wages of workers in occupation, for the large population

	Occupation																				
	1718	1604	1802	1716	2911	2052	1332	1141	4021	1232	2853	3020	1122	1142	1714	1514	3180	4030	1063	1403	1180
$df = v_{\max}$	2.97	3.45	4.44	11.9	12.4	13.1	15.3	16.9	16.8	17.3	20.6	24.9	28.0	28.6	29.1	34.8	41.5	59.9	77.6	77.9	128
$df = \hat{v}_1$	2.67	2.34	2.35	5.97	5.90	4.25	11.4	9.00	6.32	15.5	13.5	10.4	15.2	9.67	15.3	18.0	25.2	14.3	27.4	28.5	90.0
Coverage																					
$\hat{T}_A \pm s_{std} z$	.89	.60	.85	.87	.87	.89	.93	.93	.89	.92	.92	.92	.88	.89	.85	.93	.92	.81	.94	.94	.94
$\hat{T}_A \pm s_{std} t_{v_{\max}}$	.96	.83	.94	.89	.88	.91	.95	.95	.91	.94	.94	.93	.88	.90	.86	.93	.92	.81	.95	.94	.95
$\hat{T}_A \pm s_a t_{\hat{v}_1}$	.97	.88	.94	.91	.89	.97	.96	.96	.91	.94	.94	.95	.89	.91	.86	.94	.93	.83	.95	.94	.95
$\hat{T}_A \pm s_{std} t_{\hat{v}_1}$	.97	.89	.94	.92	.90	.97	.96	.91	.94	.94	.95	.89	.89	.91	.86	.94	.93	.83	.95	.95	.95
$\hat{T}_A \pm s_b t_{\hat{v}_1}$	.97	.89	.97	.92	.90	.97	.96	.96	.91	.95	.94	.95	.89	.91	.87	.95	.93	.83	.95	.94	.95
Median Relative Length																					
$\hat{T}_A \pm s_{std} z$	0.99	0.78	0.92	0.97	0.95	0.96	0.99	0.98	0.96	0.97	0.98	.98	0.95	0.96	0.93	0.98	1.00	0.91	1.00	1.00	1.01
$\hat{T}_A \pm s_{std} t_{v_{\max}}$	2.14	1.47	1.40	1.08	1.06	1.06	1.08	1.06	1.04	1.04	1.04	1.03	0.99	1.00	0.98	1.01	1.03	0.93	1.01	1.01	1.02
$\hat{T}_A \pm s_a t_{\hat{v}_1}$	2.32	2.24	2.46	1.37	1.37	1.59	1.12	1.15	1.34	1.05	1.11	1.16	1.04	1.19	1.04	1.04	1.05	1.07	1.09	1.04	1.02
$\hat{T}_A \pm s_{std} t_{\hat{v}_1}$	2.34	2.27	2.48	1.37	1.39	1.60	1.13	1.18	1.34	1.05	1.13	1.18	1.04	1.20	1.04	1.04	1.06	1.07	1.10	1.05	1.02
$\hat{T}_A \pm s_b t_{\hat{v}_1}$	2.47	2.33	2.79	1.39	1.38	1.61	1.14	1.20	1.35	1.07	1.13	1.18	1.04	1.19	1.05	1.05	1.06	1.07	1.10	1.04	1.02

For domains which yield large samples, there will be little difference from standard intervals.

3. The instances where coverage fell below nominal, even using the t-adjusted intervals, may be ascribed to severe violation of the normality assumption for the domain data. Thus the t-adjustment is not a cure-all. Nonetheless, even in such cases there is a good deal of improvement in coverage over the use of the standard normal interval.

4. The key idea behind these intervals is to condition on the amount of information on the particular occupation, which, roughly speaking, is measured in terms of the number of units in the sample that belong to the domain. The fraction of such units within each stratum is unknown, and to handle this fact we put a prior distribution on this unknown, reflective of the degree of our ignorance of it, an idea we borrow from the Bayesians. However, in the final analysis, it is the realized coverage probabilities that determine the merit of the approach.

5. The principal effect of these ideas is the abandonment, for purposes of CI construction, of the standard normal quantiles (± 1.96 for 95% coverage). These are replaced by quantiles from the Student's t-distribution, with degrees of freedom determined from the sample and varying with domain. If because of publication requirements or for other reasons, there is need to report standard deviations rather than confidence intervals, then we recommend reporting an effective standard deviation given by the length of the proposed t-based 95% confidence interval divided by twice 1.96.

6. The standard estimate of variance seems acceptable for estimating the variance, when accompanying the new t-quantile. In most instances this combination should be quite satisfactory, so that the only change from standard methodology will be the introduction of adjusted degrees of freedom. However, in some instances, the alternative standard deviations may improve coverage or reduce the length of confidence intervals.

7. An open question concerns what degree and type of collapsing of strata (if any) should be used in the estimation of variances and of the degrees of freedom for the purpose of confidence interval construction. In general, there will be a tradeoff: as strata are reduced in number, the estimate of variance will tend to increase, but so will the degrees of freedom (reducing the size of $t_{v_{\max}}$ or $t_{\hat{v}_1}$.) The answer to this question may be population specific, and experience from past surveys useful.

ACKNOWLEDGEMENTS

The authors wish to thank the Associate Editor and the three referees for their many helpful suggestions. The research of S. Wang was supported in part by an ASA/NSF/BLS research fellowship, and grants from the National Security Agency (MDA904-96-1-0029), the National Science Foundation (DMS-9504589) and the Texas A&M University Scholarly and Creative Activities Program (95-59). The views are those of the authors and do not necessarily reflect the U.S. Bureau of Labor Statistics policy.

APPENDIX A

From the discussion in Section 2.2 we know that $n\hat{p}_A$ has a binomial distribution $Bin(n, p_A)$, hence, for $\hat{p}_A = 0, 1/n, 2/n, \ldots, 1$,

$$f(\hat{p}_A | p_A) = \frac{\Gamma(n+1) \qquad \Gamma(n+2)}{\Gamma(n+2)\Gamma(n\hat{p}_A + 1)\Gamma(n(1-\hat{p}_A)+1)} \times$$

$$p_A^{(n\hat{p}_A+1)-1}(1-p_A)^{(N(1-\hat{p}_A)+1)-1} = k_{\hat{p}_A}(p_A)/(n+1).$$

For each (fixed) value of $\hat{p}_A$, the function $k_{\hat{p}_A}(p_A)$ is the pdf of a Beta distribution with parameters $\omega_1 = n\hat{p}_A + 1$ and $\omega_2 = n(1-\hat{p}_A) + 1$. As both ω_1 and ω_2 will be larger than unity with high probability (at least in most real world situations), it is reasonable to approximate $k_{\hat{p}_A}(p_A)$ with a normal pdf having equivalent mean and variance, which are approximately $\hat{p}_A$ and $\hat{p}_A(1-\hat{p}_A)/n$ respectively.

Assuming that $p_A \sim N(\mu, \sigma^2)$, it follows that the posterior distribution is

$$h(p_A | \hat{p}_A) = f(\hat{p}_A | p_A) g(p_A) \Big/$$

$$\int_0^1 f(\hat{p}_A | p_A) g(p_A) dp_A \cong ce^{-\frac{1}{2}\left(\frac{(p_A - \hat{p}_A)^2}{\hat{p}_A(1-\hat{p}_A)/n} + \frac{(p_A - \mu)^2}{\sigma^2} \right)},$$

where c is the normalizing constant.

Under the "empirical Bayes" assumption that $\mu = \hat{p}_A$ and $\sigma^2 = \hat{p}_A(1-\hat{p}_A)/n$ we have

$$h(p_A | \hat{p}_A) \cong \frac{1}{\sqrt{2\pi}\sqrt{\hat{p}_A(1-\hat{p}_A)/2n}} e^{-\frac{1}{2}\left(\frac{(p_A - \hat{p}_A)^2}{\hat{p}_A(1-\hat{p}_A)/2n} \right)}.$$

If we drop the specific assumption regarding σ^2, and let $\psi = (\hat{p}_A(1-\hat{p}_A)/n)/\sigma^2$ then $[p_A | \hat{p}_A] \sim N(\hat{p}_A, \hat{p}_A(1-\hat{p}_A)/(1+\psi)n)$.

APPENDIX B

Result: Assume W is distributed $N(0, c^2)$ and, conditional on $W = w$, the random variable T is distributed as a non-central t with v degrees of freedom and non- centrality parameter w. Then, the unconditional distribution of $T/\sqrt{c^2 + 1}$ is central t with v degrees of freedom.

Proof: First notice that T can be written as $T = (X + W)/\sqrt{S^2/v}$, where X is distributed as $N(0, 1)$, S^2 is distributed as χ^2_v, and X, W, and S^2 are mutually independent. Therefore, $X' = (X + W)/\sqrt{1 + c^2}$ is distributed as $N(0, 1)$. As X' and S^2 are independent, it follows by definition that $T' = T/\sqrt{1 + c^2} = X'/\sqrt{S^2/v}$ is distributed as t_v.

REFERENCES

DORFMAN, A., and VALLIANT, R. (1993). Quantile variance estimators in complex surveys. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 866-871.

JOHNSON, E.G., and RUST, K.F. (1993). Effective Degrees of Freedom for Variance Estimates from a Complex Sample Survey. Paper presented at the 1993 Joint Statistical Meetings, San Francisco.

KOTT, P.S. (1994). A hypothesis test of linear regression coefficients with survey data. *Survey Methodology*, 20, 159-164.

SÄRNDAL, C.-E., SWENSSON, B., and WRETMAN, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

SATTERTHWAITE, F. (1946). An approximate distribution of estimates of variance components. *Biometrics*, 2, 110-114.