Journal of Official Statistics, Vol. 22, No. 4, 2006, pp. 759–767

Delete-a-Group Variance Estimation for the General Regression Estimator Under Poisson Sampling

Phillip S. Kott[1]

Calibration makes the generalized regression (GREG) estimator under Poisson sampling practical when there are a number of items of interest. The question becomes how to estimate mean squared errors in an equally practical manner. When all the selection probabilities are small, and the GREG is expressible in projection form, an appropriately defined delete-a-group jackknife variance estimator can have desirable asymptotic properties making it a useful measure in many applications.

Key words: Asymptotic; finite population correction; model variance; randomization mean squared error.

1. Introduction

Poisson sampling is perhaps the simplest form of unequal probability selection. Its application often leads to inefficient estimation, which is why it is not more widely used. As noted by Särndal (1996), however, when combined with a regression-type estimator the advantages of Poisson sampling can be realized. This is one reason why the U.S. National Agricultural Statistics Service (NASS) has overhauled its major crop survey program and adopted Poisson sampling. Another is the usefulness of Poisson sampling in coordinating surveys. For that, see Kott and Bailey (2000).

This article reviews the theory supporting Poisson sampling coupled with a generalized regression (GREG) estimator. It then shows when a delete-a-group jackknife variance estimator may be used with this estimation strategy. Section 2 introduces the strategy, and Section 3 discusses its properties. Section 4 discusses delete-a-group jackknife variance estimation, while Section 5 investigates three modest empirical examples, the first based on real data. Section 6 provides some concluding remarks. New proofs have been relegated to the Appendix.

2. Background

Suppose we want to estimate a population (U) total, $T = \sum_U y_k$ based on a sample (S) of y-values. If the probability that population unit k is in the sample is $\pi_k > 0$, then the expansion estimator for T is $t = \sum_S y_k/\pi_k$. Another useful way to render t is as $t = \sum_U y_k I_k/\pi_k$, where I_k is a random variable equal to 1 when $k \in S$ and 0 otherwise. The simple expansion estimator is a randomization-unbiased estimator of T; that is,

[1]National Agricultural Statistics Service, Fairfax, VA 20230, U.S.A. Email: pkott@nass.usda.gov

$E_p(t) = T$, where the subscript p denotes the expectation with respect to the I_k (this is a convention; the p derives from "probability").

Under Poisson sampling (see for example, Särndal, Swensson, and Wretman 1992, pp. 85–87), each unit k is sampled independently of every other unit in the population. The randomization variance of t is

$$Var_p(t) = \sum_{k \in U} (y_k/\pi_k)^2(\pi_k - \pi_k^2) = \sum_{k \in U} (y_k^2/\pi_k)(1 - \pi_k)$$

which has a simple unbiased estimator:

$$var_p(t) = \sum_{k \in S} (y_k/\pi_k)^2(1 - \pi_k)$$

Using Poisson sampling with t, can lead to a larger-than-necessary randomization variance because the sample size is random. This problem vanishes when Poisson sampling is coupled with the GREG estimator:

$$t_R = t + \left(\sum_{k \in U} \mathbf{x}_k - \sum_{k \in S} \pi_k^{-1}\mathbf{x}_k\right)\left(\sum_{k \in S} c_k \pi_k^{-1}\mathbf{x}_k'\mathbf{x}_k\right)^{-1} \sum_{k \in S} c_k \pi_k^{-1}\mathbf{x}_k' y_k \tag{1}$$

where $\mathbf{x}_k = (\mathbf{x}_{k1}, \ldots, \mathbf{x}_{kQ})$ is a row vector of values known for all S, c_k is a constant, $\sum_U \mathbf{x}_k$ is known, and $\sum_S c_k \pi_k^{-1}\mathbf{x}_k'\mathbf{x}_k$ is invertible. Some authors force c_k to be related to a parameter of an assumed model in their definition of the GREG estimator (see for example, Montanari and Ranalli 2002). That is not the case here.

The GREG estimator can be rewritten as $t_R = \sum_S w_k y_k$, where w_k is the regression weight of k:

$$w_k = \pi_k^{-1} + \left(\sum_{i \in U} \mathbf{x}_i - \sum_{i \in S} \pi_i^{-1}\mathbf{x}_i\right)\left(\sum_{i \in S} c_i \pi_i^{-1}\mathbf{x}_k'\mathbf{x}_k\right)^{-1} c_k \pi_k^{-1}\mathbf{x}_k' \tag{2}$$

It is well-known (and easy to see) that the w_k satisfy the *calibration equation*: $\sum_S w_k \mathbf{x}_k = \sum_U \mathbf{x}_k$ (Deville and Särndal, 1992).

3. Properties of the Estimation Strategy

The GREG estimator in Equation (1) has well-known randomization-based and model-based properties under mild conditions. We will review them briefly below.

3.1. Randomization-based Properties

The randomization-based properties of t_R are asymptotic (we use the more accurate modifier "randomization" in place of the often-used "design"). That is to say, they depend on the expected sample size, say n^*, being large. A sufficient condition for an estimation strategy (an estimator coupled with a sampling design) to be randomization consistent is that its relative mean squared error should approach 0 as n^* grows arbitrarily large.

In what follows, we assume $N^{-1}(\sum_U c_k \mathbf{x}_k' \mathbf{x}_k)$ is invertible, where N is the size of U. Let $\mathbf{B} = (\sum_U c_k \mathbf{x}_k' \mathbf{x}_k)^{-1} \sum_U c_k \mathbf{x}_k' y_k$, and $e_k = y_k - \mathbf{x}_k \mathbf{B}$, so that $\sum_U c_i \mathbf{x}_i' e_i = 0$. We assume further that the population values are such that $\sum_S c_i \pi_i^{-1} \mathbf{x}_i' e_i$ and $\sum_S \pi_k^{-1} \mathbf{x}_k - \sum_U \mathbf{x}_k$ are $O_P(N/\sqrt{n^*})$. We can express the error of t_R as $t_R - T = \sum_S e_i/\pi_i - \sum_U e_i + O_p(N/n^*)$. The derivation of this equality relies on the equality $\sum_U c_i \mathbf{x}_i' e_i = 0$ rather than, as often asserted in the literature, the asymptotic identity of each w_i with its corresponding $1/\pi_i$. See Kott (2004).

3.2. *Model-based Properties*

Suppose the y_k were random variables that satisfied the following model:

$$y_k = \mathbf{x}_k \boldsymbol{\beta} + \varepsilon_k \tag{3}$$

where $\boldsymbol{\beta}$ is an unknown column vector, $E(\varepsilon_k | \mathbf{x}_k, I_k) = E(\varepsilon_k \varepsilon_g | \mathbf{x}_k, \mathbf{x}_g, I_k, I_g) = 0$ for $k \neq g$, and $E(\varepsilon_k^2 | I_k) = \sigma_k^2$. The σ_k^2 *need not be known.*

It is easy to see that as long as the regression weights satisfy the calibration equation, $\sum_S w_k \mathbf{x}_k = \sum_U \mathbf{x}_k$, t_R will be model unbiased in the sense that $E_\varepsilon(t_R - T) = 0$. Moreover, its model variance is $E_\varepsilon[(t_R - T)^2] = E_\varepsilon[(\sum_S w_i \varepsilon_i - \sum_U \varepsilon_i)^2] = \sum_S w_i^2 \sigma_i^2 - 2\sum_S w_i \sigma_i^2 + \sum_U \sigma_i^2 \approx \sum_S w_i^2 \sigma_i^2 - \sum_S w_i \sigma_i^2$. The final near equality is exact when σ_i^2 has the form $\mathbf{x}_i \mathbf{h}$ for a non-necessarily-specified vector $\mathbf{h}$. See Kott (2004) for an alternative justification.

3.3. *Simultaneous Variance Estimation*

Särndal (1996) proposed the following estimator for both the model variance and randomization mean squared error of t_R :

$$v_S = \sum_{i \in S} w_i^2 (1 - \pi_i) r_i^2 \tag{4}$$

where $r_i = y_i - \mathbf{x}_i \mathbf{b}$, and $\mathbf{b} = (\sum_S c_k \pi_k^{-1} \mathbf{x}_k' \mathbf{x}_k)^{-1} \sum_S c_k \pi_k^{-1} \mathbf{x}_k' y_k$. When the π_i are ignorably small so that almost all $w_i^2 >> w_i$, v_S is nearly equal to $v_0 = \sum_S (w_i r_i)^2$.

4. Delete-a-Group Variance Estimation

Many surveys have multiple variables of interest. The problem with v_S in Equation (4) is that is requires r_k to be calculated separately for each such variable, even when a common regressor vector, $\mathbf{x}_k$, is employed. That is one reason why a *delete-a-group jackknife variance estimator* can prove helpful in practice. The term can be found in Kott (2001), while the variance estimator itself in some form has long been used, not always with theoretical justification. A NASS research report, Kott (1998), discusses a wide variety of uses for the delete-a-group jackknife.

In this section, we assume that all the π_k are ignorably small for variance estimation purposes. This means the model variance of t_R, $E_\varepsilon[(t_R - T)^2] \approx \sum_S w_i^2 \sigma_i^2 - \sum_S w_i \sigma_i^2$, is approximately $V_0 = \sum_S w_i^2 \sigma_i^2$.

Let the Poisson sample be randomly divided into G mutally exclusive replicate groups, denoted $S_1, S_2, \ldots, S_G$ (some groups can have one more member than others). The complement of each S_g is called the *jackknife replicate group* $S_{(g)} = S - S_g$. A set of replicate weights is computed for each replicate group. For the gth set: $w_{i(g)} = 0$ when $i \in S_g$; and

$$w_{i(g)} = [G/(G-1)]w_i + \left(\sum_{k \in U} \mathbf{x}_k - \sum_{k \in S_{(g)}} [G/(G-1)] w_k \mathbf{x}_k \right) \left(\sum_{k \in S_{(g)}} c_k \pi_k^{-1} \mathbf{x}_k' \mathbf{x}_k \right)^{-1} c_i \pi_i^{-1} \mathbf{x}_i' \tag{5}$$

otherwise. The $w_{i(g)}$ have been computed to be reasonably close to the corresponding $[G/(G-1)]w_i$ for $i \in S(g)$ and to satisfy the calibration equation $\sum_S w_{k(g)} \mathbf{x}_k = \sum_U \mathbf{x}_k$ for all g.

The delete-a-group variance estimator for t_R is:

$$v_J = (G - 1/G) \sum_{g=1}^{G} \left(\sum_{i \in S} w_{i(g)} y_i - t_R \right)^2 \tag{6}$$

which WESVAR (Westat, 1997) calls JK1. In the appendix, we show that v_J is an asymptotically model unbiased estimator for $V_0 = \sum_S w_i^2 \sigma_i^2$ and asymptotically indistinguishable from $v_0 = \sum_S (w_i r_i)^2$ when $c_i = 1/(\mathbf{x}_i \boldsymbol{\gamma})$ for some vector $\boldsymbol{\gamma}$. It can have a slight upward bias otherwise. The condition that $c_i = 1/(\mathbf{x}_i \boldsymbol{\gamma})$ for some vector $\boldsymbol{\gamma}$ assures t_R can be put into *projection form*: $t_R = (\sum_U \mathbf{x}_k)\mathbf{b}_c$, where $\mathbf{b}_c = (\sum_S c_k \pi_k^{-1} \mathbf{x}_k' \mathbf{x}_k)^{-1} \times \sum_S c_k \pi_k^{-1} \mathbf{x}_k' y_k$. Särndal, Swensson, and Wretman (1992, p. 231) make a similar point.

5. A Modest Empirical Example

Farm data from the 1997 Census of Agriculture in Puerto Rico were used to study the delete-a-group jackknife's applicability given a GREG estimator under a Poisson sample. The cash from coffee production for farm k was used as y_k in Equation (1), while $\mathbf{x}_k$ was the total land on farm k, a scalar value. Each farm's selection probability was set at $\pi_k = 200\sqrt{\mathbf{x}_k} / \sum_U \sqrt{\mathbf{x}_i}$, where U was the set of 19,951 records in the population. That meant π_k was proportional to $\sqrt{\mathbf{x}_k}$, and the average sample size was 200.

Two estimators were considered. One had $c_k = 1/\mathbf{x}_k$, and collapsed into the standard *ratio estimator*: $t_{rat} = (\sum_U \mathbf{x}_k) b_{rat}$, where $b_{rat} = [\sum_S (y_k/\pi_k) / \sum_S (\mathbf{x}_k/\pi_k)]$. The other, the *optimal estimator* (see Rao 1994), had $c_k = (1-\pi_k)/\pi_k$, so that $t_{opt} = \sum_S (y_k/\pi_k) + [\sum_U \mathbf{x}_k - \sum_S (\mathbf{x}_k/\pi_k)] b_{opt}$, where $b_{opt} = \sum_S y_k (1-\pi_k) / [\sum_S \mathbf{x}_k (1-\pi_k)]$. The optimal estimator gets its name from $b^* = \sum_U y_k (1-\pi_k)\pi_k / [\sum_U \mathbf{x}_k (1-\pi_k)\pi_k] = \{Var_p[\sum_S (\mathbf{x}_k/\pi_k)]\}^{-1} Cov_p[\sum_S (\mathbf{x}_k/\pi_k), \sum_S (y_k/\pi_k)]$, the probability limit of b_{opt}. The choice $b = b^*$ minimizes the randomization variance among all estimators of the form: $t = \sum_S (y_k/\pi_k) + [\sum_U \mathbf{x}_k - \sum_S (\mathbf{x}_k/\pi_k)] b$. Montanari and Ranalli (2002) provide a penetrating discussion of the relationships between the GREG and optimal estimators in a more general context.

Returning to the empirical example at hand, the ratio estimator is already in projection form. The optimal estimator, by contrast, cannot be put in projection form.

The use of 15 replicate groups was investigated since that is what NASS uses in its applications. The results of 10,000 simulations are summarized in Table 1. Both estimators appear to be virtually unbiased and have small (empirical) relative mean squared errors. The optimal estimator is slightly more efficient. This small difference in efficiency appears to be real since, over the 10,000 simulations, the value of $[(t_R - T)/T]^2$ has a standard deviation of roughly 0.02 for either estimator.

Most of the variance/mean-squared-error (MSE) estimators displayed appear to have small absolute biases (less than 5%). The biggest exception is the delete-a-group jackknife under the optimal estimator, which has a negative bias of 7.8%, in contrast to the very slight positive bias in v_0 (0.2%). This negative bias is what our theory predicts when the model, $y_x = \beta \mathbf{x}_k + \varepsilon_k$, where $E(\varepsilon_k|\mathbf{x}_k) = 0$, fails. In contrast to that, the theory predicts that v_S should be (asymptotically) unbiased, and the remaining variance/MSE estimators slightly biased upward since not all the π_k are ignorably small.

Nominal two-sided 95% coverage intervals were computed for the two jackknives using 14 degrees of freedom since they were based on only 15 replicates. The other coverage intervals were computed using the 60 degrees of freedom corresponding to the *ad hoc* use of two-standard-error intervals. The delete-a-group jackknife covered a bit worse than v_0 for both estimators. It covered slightly better for the ratio estimator than for the optimal estimator. All coverages were between 90 and 95%, indicating a slight imperfection in the asymptotic theory.

A second and third set of 1,600 simulations were conducted to try to uncover the origin of the imperfection. In the second, each y_k was generated as a multiple of farm land, x_k plus $\sqrt{x_k}$ times an independent and identically distributed normal error term. All the variance/MSE measures had an absolute relative bias of less than 2%, with v_S having an absolute relative bias of less than 0.5% for both estimators. The nominal 95% coverages were all between 94.8 and 95.4%, suggesting the nonnormality in the Census of Agriculture data was a key source of the modest undercoverage in the original set of simulations. Note that in this set of simulations, the model fit the data exactly. As a consequence, the delete-a-group jackknife produced a good measure of the relative squared error for the optimal estimator, as our theory predicted.

In the third set of simulations, coffee sales were first regressed on an intercept and farm land using all the Census of Agriculture data. The two coefficients, call them α and β, were

Table 1. Variance summaries for the Puerto Rico data (10,500 simulations) T normalized to equal 1

	The ratio estimator	The optimal estimator
Averages over the simulations:		
RE $= t - T$	.0043	.0026
MSE $= (t - T)^2$	.0355	.0334
$(v_0 - \text{MSE})/\text{MSE}$	$-.0115$	.0023
$(v_S - \text{MSE})/\text{MSE}$	$-.0132$	$-.0181$
$(v_J - \text{MSE})/\text{MSE}$	$-.0340$	$-.0781$
Nominal 95% coverage rates:		
$v_{0\,(60\,df)}$	93.2	92.2
$v_{S\,(60\,df)}$	92.9	92.0
$v_{J\,(14\,df)}$	92.7	91.2

both significantly positive at the .001 level, but the R^2 was less than 0.03. The fitted sales values, $\alpha + \beta \mathbf{x}_k$, were then used as the y_k in computing the ratio and optimal estimators from 1,600 Poisson samples. The errors in the ratio and optimal estimators for T resulted fully from the failure of the intercept-free linear model implicit in their construction. Although the intercept is a design-balanced variable (Montanari and Ranalli 2002) for the optimal estimator under stratified simple random sampling, it is not one under the unequal-probability Poisson design used in the simulations.

The optimal estimator lived up to its name and was more efficient than the ratio estimator. The empirical relative mean squared error of the optimal estimator was .0046 as opposed to .0073 for the ratio.

The delete-a-group jackknife had a negative bias of 27.9% for the optimal estimator, but only 0.1% for the ratio, which conformed to our theory. The variance/MSE estimator v_0 had a slight positive bias of 2.6 and 1.8% for the optimal and ratio estimators, respectively.

A surprise came in the nominal 95% coverage intervals, where using v_0 lead to an average coverage of 89.9% for the optimal estimator, but 95.7% for the ratio. The delete-a-group jackknife covered the ratio estimator well, at 94.9%. That is gratifying, but the relatively poor coverage of v_0 for the optimal estimator is a bit mystifying (the delete-a-group jackknife covered even worse, at 84.1%). This poor coverage may result from $t_{opt} - T \approx \alpha[\sum_S 1/\pi_k - N]$ being far from normally distributed.

6. Concluding Remarks

When both n and N are large, but N is so large that all the π_k are ignorably small for variance estimation purposes, the delete-a-group jackknife variance estimator (v_J in Equation (6)) can be used to estimate both the model variance and randomization mean squared error of the GREG estimator (t_R in Equation (1)). The asymptotic unbiasedness of the latter requires an additional assumption: $c_i = 1/(\mathbf{x}_i \boldsymbol{\gamma})$ for some vector $\boldsymbol{\gamma}$, which means the estimator can be put into projection form. This usually rules out the asymptotically efficient optimal estimator. If $c_i = 1/(\mathbf{x}_i \boldsymbol{\gamma})$ and the π_k are not all ignorably small, then the delete-a-group jackknife can be asymptotically biased upward.

Had the sample been drawn with probability-proportional-to-π_k with replacement, the π_k, redefined to be the expected number of times k is selected for the sample, need not be small. Moreover, a quick look at Equation (A.4) in the appendix (translated to allow the same unit to be in S more than once) reveals that the delete-a-group jackknife is an asymptotically unbiased estimator for the randomization mean squared error of t_R whether or not the GREG is expressible in projection form.

Finally, the construction of the jackknife replicate weights for $i \in S_{(g)}$ in Equation (5) was nonstandard. The interested reader can verify that using the formulation:

$$\begin{aligned} W_{i(g)} =& [G/(G-1)]\pi_i^{-1} + \left(\sum_{k \in U} \mathbf{x}_k - \sum_{k \in S_{(g)}} [G/(G-1)]\pi_k^{-1} \mathbf{x}_k \right) \\ & \times \left(\sum_{k \in S_{(g)}} c_k [G/(G-1)]\pi_k^{-1} \mathbf{x}_k' \mathbf{x}_k \right)^{-1} c_i [G/(G-1)]\pi_i^{-1} \mathbf{x}_i' \end{aligned} \tag{5'}$$

does not change the asymptotic results.

Appendix: The Asymptotic Properties of v_J

The delete-a-group jackknife in Equation (6) can be re-expressed as

$$v_J = (G - 1/G)]\sum_{g=1}^{G}\left(\sum_{i\in S} w_{i(g)}u_i - \sum_{i\in S} w_i u_i\right)^2$$

where u_i may be either ε_i or e_i depending on whether we are interested in model-based or randomization-based properties.

We assume for this appendix that all the π_i are $O(n^*/N)$, so each w_i is $O_p(N/n^*)$. Without loss of generality, we assume each n/G equals an integer, d. To do otherwise complicates the derivation of the subsequent formulae without adding insight. We also assume that $(n^*/N^2)V_0$ and $(n^*/N^2)E_P(v_S)$ each converge to a positive constant and that $\sum_U c_k \mathbf{x}_k' \mathbf{x}_k/N$ converges to a positive definite matrix as n^* grows arbitrarily large.

The sets S_g and $S_{(g)}$ can be viewed as simple random subsamples of S. With this in mind, we will assume that $\sum_{S_g} w_i\mathbf{x}_i - \sum_S w_i \mathbf{x}_i/G$ is $O_P([N/n^*]\sqrt{d})$. Since $dG = n$, either d or G (or both) must grow arbitrarily large, in probability, with n^*. Remembering that $\sum_S w_i\mathbf{x}_i = \sum_U \mathbf{x}_i$, we have:

$$\begin{aligned}\sum_{i\in S} w_{i(g)}u_i - \sum_{i\in S} w_i u_i &= [G/(G-1)]\left\{-\left(\sum_{i\in S_g} w_i u_i - \sum_{i\in S} w_i u_i/G\right)\right.\\ &\quad \left. + \left(\sum_{i\in S_g} w_i\mathbf{x}_i - \sum_{i\in S} w_i\mathbf{x}_i/G\right)\left(\sum_{i\in S_{(g)}} c_k\pi_k^{-1}\mathbf{x}_k'\mathbf{x}_k\right)^{-1}\sum_{i\in S_{(g)}} c_k\pi_k^{-1}\mathbf{x}_k' u_i\right\}\\ &= -[G/(G-1)]\left\{\sum_{i\in S_g} w_i u_i - \sum_{i\in S} w_i u_i/G\right\}\\ &\quad + O_{(1\times q)p}(\sqrt{d}/[n^* - d])\sum_{i\in S_{(g)}} c_i\pi_i^{-1}\mathbf{x}_i' u_i \end{aligned} \tag{A.1}$$

Consequently,

$$\begin{aligned} E_\varepsilon\left[\left(\sum_{i\in S} w_{i(g)}\varepsilon_i - \sum_{i\in S} w_i\varepsilon_i\right)^2\right] &= \sum_{i\in S_g} w_i^2\sigma_i^2(1 - [2/G])\\ &\quad + \sum_{i\in S} w_i^2\sigma_i^2/G^2 + O_p([N/n^*]^2 d/[n^* - d])\\ &= \sum w_i^2\sigma_i^2(1 - [2/G]) + \sum w_i^2\sigma_i^2/G^2\\ &\quad + O_p([N/n^*]^2/[G-1]) \end{aligned}$$

From this, and Equation (6), we can see that the delete-a-group is an asymptotically model unbiased estimator for $V_0 = \sum_S w_i^2 \sigma_i^2$:

$$E_\varepsilon(v_J) = \sum_{g=1}^{G} \sum_{i \in S_g} w_i^2 \sigma_i^2 + O_p([N/n^*]^2)$$

Establishing the asymptotic randomization-based properties of v_J is a bit more difficult. From Equation (A.1):

$$\begin{aligned} \sum_{i \in S} w_{i(g)} e_i - \sum_{i \in S} w_i e_i &= -[G/(G-1)] \left\{ \sum_{i \in S_g} w_i e_i - \sum_{i \in S} w_i e_i / G \right\} \\ &\quad + O_{(1 \times q)p}(\sqrt{d/[n^* - d]}) \sum_{S(g)} c_i \pi_i^{-1} \mathbf{x}_i' e_i \\ &= -[G/(G-1)] \left\{ \sum_{i \in S_g} w_i e_i - \sum_{i \in S} w_i e_i / G \right\} \\ &\quad + O_p([N/n^*] \sqrt{d/[n^* - d]}) \\ &= -[G/(G-1)] \left\{ \sum w_i e_i - \sum w_i e_i / G \right\} \\ &\quad + O_p([N/n^*] / \sqrt{[G-1]}) \end{aligned} \quad \text{(A.2)}$$

We can combine Equations (6) and (A.2). Thus

$$v_J = [G/(G-1)] \sum_{g=1}^{G} \left\{ \sum_{i \in S_g} w_i e_i - \sum_{i \in S} w_i e_i / G \right\}^2 + O_p([N/n^*]^2) \quad \text{(A.3)}$$

We now turn our attention to the randomization expectation of v_J under the random subsampling of sample S in creating S_g. Note that $E_2\{(\sum_{Sg} w_i e_i / d - \sum_S w_i e_i / n)^2\} = \{(1 - [d/n])/d\}\{\sum_S (w_i e_i)^2 - (\sum_S w_i e_i)^2 / n\}/(n-1)$, where the subscript 2 refers to the subsampling. As a result,

$$E_2(v_J^2) = \left\{ \sum_{i \in S} (w_i e_i)^2 - \left(\sum_{i \in S} w_i e_i \right)^2 / n \right\} + O_p([N/n^*]^2) \quad \text{(A.4)}$$

We need an additional assumption; namely, $c_i = 1/(\mathbf{x}_i \boldsymbol{\gamma})$ for some vector $\boldsymbol{\gamma}$. Under this assumption, $\sum_U e_i = \sum_U y' \mathbf{x}_i' c_i e_i = y' \sum_U c_i \mathbf{x}_i' e_i = 0$. From which we can conclude $E_2(v_J) = \sum_S (w_i e_i)^2 + O_p([N/n^*]^2)$, which is asymptotically indistinguishable from $v_0 = \sum_S (w_i r_i)^2$.

From the derivation of Equation (A.4), we see that when $c_i \neq 1/(\mathbf{x}_i\boldsymbol{\gamma})$, so that $\sum_S w_i e_i \neq O_p(N/\sqrt{n^*})$, v_J can have an downward bias as an estimator of the randomization mean squared error of t_R.

7. References

Deville, J-C. and Särndal, C-E. (1992). Calibration Estimators in Survey Sampling. Journal of the American Statistical Association, 87, 376–382.

Kott, P.S. (1998). Using the Delete-a-Group Jackknife Variance Estimator in NASS Surveys, RD Research Report No. RD-98-01. Washington, DC: National Agricultural Statistics Service.

Kott, P.S. (2001). The Delete-a-group Jackknife. Journal of Official Statistics, 17, 521–526.

Kott, P.S. (2004). Randomization-assisted Model-based Survey Sampling. Journal of Statistical Planning and Inference, 129, 263–277.

Kott, P.S. and Bailey, J.T. (2000). The Theory and Practice of Maximal Brewer Selection. Proceedings of the Second International Conference on Establishment Surveys, Invited papers, 269–278.

Montanari, G.E. and Ranalli, M.G. (2002). Asymptotically Efficient Generalised Regression Estimators. Journal of Official Statistics, 18, 577–589.

Rao, J.N.K. (1994). Estimating Totals and Distribution Functions Using Auxiliary Information at the Estimations Stage. Journal of Official Statistics, 25, 1–21.

Särndal, C.-E. (1996). Efficient Estimators With Simple Variance in Unequal Probability Sampling. Journal of the American Statistical Association, 91, 1289–1300.

Särndal, C.-E., Swensson, B., and Wretman, J. (1992). Model Assisted Survey Sampling. New York: Springer.

Westat, Inc. (1997). A User's Guide to WesVarPC®, Version 2.1. Rockville, MD: Westat.

Received November 2002
Revised September 2005