Evaluating the Asymptotic Limits of the Delete-a-Group Jackknife for Model Analyses

Phillip S. Kott and Steven T. Garren

The delete-a-group jackknife can be effectively used when estimating the variances of statistics based on a large sample. The theory supporting its use is asymptotic, however. Consequently, analysts have questioned its effectiveness when estimating parameters for a small domain computed using only a fraction of the large sample at hand. We investigate this issue empirically by focusing on heavily poststratified estimators for a population mean and a simple regression coefficient, where the poststratification takes place at the full-sample level. Samples are chosen using differentially-weighted Poisson sampling. The bias and stability of delete-a-group jackknife employing either 15 or 30 replicates are evaluated and compared with the behavior of linearization variance estimators.

KEY WORDS: Calibrated weight; Domain; Ignorable; Linearization variance estimator; Model parameter; Relative empirical bias.

Phillip S. Kott is Chief Research Statistician, Research and Development Division, National Agricultural Statistics Service, 3251 Old Lee Highway, Room 305, Fairfax, Virginia, 22030. Steven T. Garren is Professor, Department of Mathematics and Statistics, MSC 1911, James Madison University, Harrisonburg, Virginia 22807.

1. Introduction

The National Agricultural Statistics Service (NASS) has increasingly been using calibration to produce parameter estimates and a delete-a-group (DAG) jackknife to measure the precision of these estimates. In surveys where the DAG jackknife is used, each sample element k is given R+1 weights: the element's sampling weight *after* incorporating all nonresponse and calibration adjustments, w_k, and R jackknife replicate weights, $w_{k(r)}$, with $r = 1, ..., R$.

NASS usually sets R at 15 or 30. The former produces variance estimators for univariate statistics with 14 nominal degrees of freedom and thus only a modest fattening of coverage intervals (the t-value for a two-sided 95% coverage interval is 2.145, not much larger than 1.96 under infinite degrees of freedom). Unfortunately, for constructing multivariate test statistics, more replicates may be needed, which is why the agency sets $R = 30$ for some surveys. Most NASS surveys have thousands on primary sampling units (individual farms), rendering delete-one jackknives impractical.

Be that as it may, we do not claim here that the DAG jackknife is theoretically superior to other variance-estimation methods. Rather, our goal is to investigate an empirical *limitation* of the DAG jackknife because that is the method NASS uses.

The theory underpinning the use of the DAG jackknife – and all jackknives for that matter – is asymptotic. See Kott (1998; 2001). We are interested here in evaluating the limitations of the asymptotics. In particular, we will be concerned with how well the DAG jackknife methodology works for parameter estimators defined within a domain when the (respondent) sample size in that domain is small. This is an issue of particular concern to analysts working with data from the third phase of the Agricultural Resources and Management Survey (ARMS-III; see USDA 2007), NASS's principal survey of the economic condition of US farms.

Complicating matters is that the weights for the ARMS-III sample are heavily calibrated. This means initial element sample weights, inverses of the element selection probabilities (perhaps partially adjusted for nonresponse and/or coverage errors), are adjusted so that the

sample-weighted sums of certain benchmark (calibration) variables equal totals derived from outside sources.

In the analyses presented here, we will restrict our attention to a Poisson sample without nonresponse. This is the simplest sample design with variable sample weights. After reviewing the theory for a more general version of linear calibration, our empirical investigations will be confined to perhaps the simplest form of calibrated-weighting: poststratification. By focusing on this relatively simple setup (Poisson sampling with poststratification), we hope to shed light on the particular issue of the usefulness of the DAG jackknife methodology – and the alternative linearization methodology – for a parameter estimate within a domain when the estimator's weights are calibrated to benchmark totals at a higher level of aggregation than the domain.

A well-known limitation of the DAG jackknife is that it ignores the impact of large sampling fractions on finite-population variances. This is of little import to most analysts of ARMS-III data because these analysts are less interested in finite-population parameter estimates than in estimating the parameters of the models generating the finite population under investigation. This subject, as well as other aspects of the theory, is explored in Section 2. Section 3 lays out the framework for the empirical investigaion, the results of which are reported in Section 4. Section 5 offers some concluding remarks.

2. Some Theory

2.1. Preliminaries

Let a_k be the initial sample weight for element k. Let $\mathbf{z}_k = (z_{k1}, ..., z_{kP})$ denote a row vector of calibration variables associated with k, for which the population total(s), $\mathbf{T_z}$, is known. Most of the calibration weighting in practice involves a variant of least squares, where the calibrated weights have the linear form:

$$w_k = a_k + (\mathbf{T_z} - \sum_{j \in S} a_j \mathbf{z}_j)[\sum_{j \in S} a_j c_j \mathbf{z}_j' \mathbf{z}_j)]^{-1} a_k c_k \mathbf{z}_k'$$

for some set of constants $\{c_k\}$, where S denotes the (respondent) sample. By design, $\sum_S w_k \mathbf{z}_k = \mathbf{T_z}$.

The c_k are often chosen to restrict the range of the w_k. A more general linear form is discussed in Estevao and Särndal (2000).

To simplify matters, we assume here a Poisson sample without nonresponse. The a_k are inverses of the element selection probabilities, π_k. We further assume the c_k are all equal to 1, and there is a vector $\boldsymbol{\lambda}$ such that $\mathbf{z}_k\boldsymbol{\lambda} = 1$ for all $k \in S$ (*e.g.*, one of the components of $\mathbf{z}_k$ is always 1). As a result of these assumptions, the calibrated weights can be rendered:

$$w_k = \mathbf{T}_\mathbf{z}\left(\sum_{j\in S} a_j \mathbf{z}_j{}'\mathbf{z}_j\right)^{-1} a_k \mathbf{z}_k{}'. \tag{1}$$

(To see why replace $\Sigma_S\, a_j\mathbf{z}_j$ in $\mathbf{T}_\mathbf{z} - \Sigma_S\, a_j\mathbf{z}_j$ by $\Sigma_S\, a_j\boldsymbol{\lambda}'\,\mathbf{z}_j'\,\mathbf{z}_j = \boldsymbol{\lambda}'\Sigma_S\, a_j\mathbf{z}_j'\,\mathbf{z}_j$.) This also allows the DAG jackknife to have certain desirable properties (see Kott 2006a).

To compute DAG jackknife replicate weights, the sample is randomly ordered and then systematically divided into R mutually exclusive groups. The complements of the groups are the *replicate groups*, denoted $S_{(1)}, \ldots, S_{(R)}$. Each $S_{(r)}$ contains roughly $(R-1)/R$ of the sample. One way to compute the replicate weights is with

$$w_{k(r)} = \tfrac{R}{R-1} w_k + \left(\mathbf{T}_\mathbf{z} - \sum_{j\in S_{(r)}} \tfrac{R}{R-1} w_j\mathbf{z}_j\right)\left[\sum_{j\in S_{(r)}} a_j \mathbf{z}_j{}'\mathbf{z}_j\right]^{-1} a_k\mathbf{z}_k' \tag{2}$$

when $k \in S_{(r)}$, and 0 otherwise. (See Kott 2006b.) By design, $\sum_{S_{(r)}} w_{k(r)}\mathbf{z}_k = \mathbf{T}_\mathbf{z}$. If we replaced the a_j and a_k in equation (2) by their near equalities w_j and w_k, we could write

$$w_{k(r)} = \mathbf{T}_\mathbf{x}\left(\sum_{S_{(r)}} w_j\mathbf{z}_j{}'\mathbf{z}_j\right)^{-1} w_k\mathbf{z}_k{}'.$$

2.2. A Parameter Estimate

We will be interested in a (vector) parameter estimate of the form:

$$\mathbf{b} = \left(\sum_{j\in S} w_j\mathbf{h}_j{}'\mathbf{x}_j\right)^{-1} \sum_{j\in S} w_j\mathbf{h}_j{}' y_j\,, \tag{3}$$

where $\mathbf{h}_j$ and $\mathbf{x}_j$ are row vectors of the same length ($\mathbf{x}_j$ may or may not have components in common with $\mathbf{z}_j$). When $\mathbf{h}_j = \mathbf{x}_j$ has more than one component, $\mathbf{b}$ is a sample-weighted regression

coefficent. When $\mathbf{h}_j = 1$ and $\mathbf{x}_j = x_j$ are scalars, $\mathbf{b} = b$ is a sample-weighted ratio. When, in addition, $x_j = 1$, b is a sample-weighted mean.

The DAG jackknife (matrix) variance estimator for $\mathbf{b}$ is

$$\mathbf{V}_J = \frac{R-1}{R} \sum_{r=1}^{R} (\mathbf{b} - \mathbf{b}_{(r)})(\mathbf{b} - \mathbf{b}_{(r)})', \tag{4}$$

where $\mathbf{b}_{(r)} = (\sum_S w_{j(r)} \mathbf{h}_j' \mathbf{x}_j)^{-1} \sum_S w_{j(r)} \mathbf{h}_j' y_j$. Note that we have yet to specify exactly what $\mathbf{b}$ is estimating, making it difficult to judge how good a job $\mathbf{V}_J$ does at measuring its accuracy.

If the goal of $\mathbf{b}$ is to estimate the limit of $\mathbf{B} = (\sum_U \mathbf{h}_j' \mathbf{x}_j)^{-1} \sum_U \mathbf{h}_j' y_j$ as the population U grows arbitrarily large, then the jackknife can be shown to be an asymptotically unbiased estimator for the variance matrix of $\mathbf{b}$ under mild conditions we assume to hold. In particular, we assume conditions are such that both $\mathbf{B}$ and its limit, call it $\mathbf{B}^*$, exist.

Sample selection is essentially two-phased in this framework. The population can be viewed as a simple random sample drawn from an infinite conceptual population. This is followed by the actual Poisson selection of the sample. Effectively, we have a Poisson sample from the infinite population, where the orginal sampling weights, the a_k, reflect the relative sizes of the inverses of the sample-selection probabilities.

We are interested in estimating the limit of **B,** as opposed to the finite population parameter itself, because we are looking for insights into the underlying model generating the population values. This is what interests most analysts studying the ARMS-III. Ideally, the underlying model is linear and can be expressed in this following two-part form:

$$y_k = \mathbf{x}_k \boldsymbol{\beta} + \varepsilon_k, \tag{5.1}$$

$$\text{with } \mathrm{E}(\varepsilon_k | \{ \mathbf{x}_j, \mathbf{z}_j, \mathbf{h}_j, I_j;\ j \in U\}) = 0, \tag{5.2}$$

where $I_j = 1$ when j is in the sample, 0 otherwise. The ε_k are uncorrelated and have bounded variances, σ_k^2. Under this model, the probability limit of $\mathbf{B}$ is $\boldsymbol{\beta}$.

Although it is often instructive to evaluate variance estimators under the linear model in both parts of equation (5), the DAG jackknife has been designed to work (under mild conditions)

whether or not the model, as specified, holds. For example, equation (5.2) effectively specifies that the design is ignorable since the expectation of ε_k is zero regardless of which elements are selected for the sample. In practice, the sample design may not be ignorable. Still, the model in equation (5.1) may hold with $E(\varepsilon_k|\{ \mathbf{x}_j, \mathbf{h}_j;\ j \in U\}) = 0$. The probability limit of **B** remains β in this case.

An even weaker formulation is possible. Observe the **B** has been defined so that $\sum_U (\mathbf{h}_k' [y_k - \mathbf{x}_k\mathbf{B}]) = \mathbf{0}$. Although many would argue that the following is not really a linear model at all, the way **B** is defined suggests that if the equation (5.1) holds with only $E(\mathbf{h}_k'\varepsilon_k) = \mathbf{0}$, then the probability limit of **B** remains β. This formulation is called the "extended linear model" in Kott (2007).

2.3. Domain Estimates

The asymptotics supporting the use of $\mathbf{V}_J$ (with or without the model in equation (5)) require both the expected sample size (recall the sample is Poisson so its size is random) and R to be large. We will be concerned in the next several sections with domain estimates of the form: $\mathbf{b}_d = (\sum_S d_j w_j \mathbf{h}_j'\mathbf{x}_j)^{-1}\sum_S d_j w_j \mathbf{h}_j' y_j$, where $d_j = 1$ when element j is in the domain of interest, 0 otherwise. Notice that if we redefine $\mathbf{h}_j$ as $d_j\mathbf{h}_j$, then $\mathbf{b}_d$ has exactly the same form as **b** in equation (3). Viewed this way, the realized sample sizes for $\mathbf{b}_d$ and the original **b** are exactly the same! Nevertheless, it seems intuitive that when the expected overall sample size is in the hundreds but the sample size within the domain is less than, say, 30, the asymptotics supporting **b** might not support $\mathbf{b}_d$. (Although the sample within a domain is independently drawn with Poisson sampling, the domain estimator in our setup is computed using calibration weights that depend on the entire sample.)

There is theory behind this intuition. For the asymptotics to work, statistics like the components of $\sum_S w_j\mathbf{h}_j'\mathbf{x}_j$ within the nonlinear expression $(\sum_S w_j\mathbf{h}_j'\mathbf{x}_j)^{-1}$ need to have small relative variances when the sample size is large. If most of the sample values of a component are zero, then that may not be the case. The "mild conditions" we cavalierly added to our requirements for $\mathbf{V}_J$ to be asymptotically unbiased may be violated.

2.4. Why the DAG Jackknife Works (Asymptotically)

We now take a temporary, but useful, digression. An alternative way to estimate the variance of **b** is through linearization (see, for example, Demnati and Rao, 2004). Let

$$\mathbf{U}_k = (\sum_{j \in S} w_j \mathbf{h}_j \mathbf{x}_j')^{-1} \mathbf{h}_k (y_k - \mathbf{x}_k \mathbf{B}^*), \text{ and}$$

$$\mathbf{u}_k = (\sum_{j \in S} w_j \mathbf{h}_j \mathbf{x}_j')^{-1} \mathbf{h}_k (y_k - \mathbf{x}_k \mathbf{b}). \tag{6}$$

Then $\mathbf{b} - \mathbf{B}^*$ can ideally be rendered as $\mathbf{\Delta} = \Sigma_S w_k \mathbf{U}_k$. Of course, $\mathbf{U}_k$ is unknown. It will ultimately be replaced by $\mathbf{u}_k$. For now, however, assume it is known. An *idealized linearization variance estimator* for **b** is

$$\mathbf{V}_{IL}(\mathbf{b}) = \sum_{k \in S} w_k^2 [\mathbf{U}_k - \mathbf{z}_k (\sum_{j \in U} \mathbf{z}_j \mathbf{z}_j')^{-1} \sum_{j \in U} \mathbf{z}_j \mathbf{U}_j][\mathbf{U}_k - \mathbf{z}_k (\sum_{j \in U} \mathbf{z}_j \mathbf{z}_j')^{-1} \sum_{j \in U} \mathbf{z}_j \mathbf{U}_j]'. \tag{7}$$

Often, **b** is treated as an estimator for **B**, and w_k^2 in the above equation is replaced by $a_k^2(1 - \pi_k)$. The $1 - \pi_k$ disappears when **b** estimates **B***. Note that $\mathbf{U}_k - \mathbf{z}_k (\sum_U \mathbf{z}_j \mathbf{z}_j')^{-1} \sum_U \mathbf{z}_j \mathbf{U}_j$ serves as the population regression residual (of the component of $\mathbf{U}_k$ on $\mathbf{z}_k$) due to the calibration. Why we put w_k^2 in the above equation rather than the asymptotically equivalent a_k^2 will be made clear presently.

Observe that if the linear model in equation (5) holds, and the population is large enough both for the distinction between **B** and β to be ignored and for $(\Sigma_U \mathbf{z}_j' \mathbf{z}_j)^{-1} \Sigma_U \mathbf{z}_j' \mathbf{U}_j$ to be effectively equal to a matrix of zeros, then $\mathbf{V}_{IL}(\mathbf{b})$ is $\sum_S w_k^2 E(\mathbf{U}_k \mathbf{U}_k') =$ $(\sum_S w_j \mathbf{h}_j \mathbf{x}_j')^{-1} \sum_S w_k^2 \mathbf{h}_k E(\varepsilon_k \varepsilon_k') \mathbf{h}_k (\sum_S w_j \mathbf{x}_j \mathbf{h}_j)^{-1}$, which is an unbiased estimator for the variance of **b** under the linear model *no matter what the sample size*. An actual linearization estimator for **b**, like

$$\mathbf{V}_L(\mathbf{b}) = \sum_{k\in S} w_k^2[\mathbf{u}_k - \mathbf{z}_k(\sum_{j\in S} a_j\mathbf{z}_j'\mathbf{z}_j)^{-1}\sum_{j\in S} a_j\mathbf{z}_j'\mathbf{u}_j][\mathbf{u}_k - \mathbf{z}_k(\sum_{j\in S} a_j\mathbf{z}_j'\mathbf{z}_j)^{-1}\sum_{j\in S} a_j\mathbf{z}_j'\mathbf{u}_j]', \qquad (8)$$

must rely on information available in the sample and thus needs a large-enough sample size. It should be realized, however, that the potential scarcity of nonzero $\mathbf{x}_j$ when estimating a domain-specific parameter has no impact on the size of $\Sigma_S a_j\mathbf{z}_j'\mathbf{z}_j$. The number of nonzero $\mathbf{x}_j$ does have an effect on $\Sigma_S w_j\mathbf{h}_j'\mathbf{x}_j$ in $\mathbf{u}_k$. Moreover, even under the model in equation (5), which treats $\Sigma_S w_j\mathbf{h}_j'\mathbf{x}_j$ as a constant, the number of nonzero $\mathbf{x}_j$ affects $\mathbf{b}$.

Let us now turn to the DAG jackknife in equation (4). Observe that under the model in equation (5),

$$\mathbf{b} - \mathbf{b}_{(r)} = (\mathbf{b} - \boldsymbol{\beta}) - (\mathbf{b}_{(r)} - \boldsymbol{\beta})$$

$$= (\sum_{j\in S} w_j\mathbf{h}_j'\mathbf{x}_j)^{-1}\sum_{j\in S} w_j\mathbf{h}_j'\varepsilon_j - (\sum_{j\in S_{(r)}} w_{j(r)}\mathbf{h}_j'\mathbf{x}_j)^{-1}\sum_{j\in S_{(r)}} w_{j(r)}\mathbf{h}_j'\varepsilon_j$$

or

$$\mathbf{b} - \mathbf{b}_{(r)} = (\sum_{j\in S} w_j\mathbf{h}_j'\mathbf{x}_j)^{-1}\left[\sum_{j\in S} w_j\mathbf{h}_j'\varepsilon_j - \sum_{j\in S_{(r)}} \tfrac{R}{R-1} w_j\mathbf{h}_j'\varepsilon_j\right] + \qquad (9)$$

$$(\sum_{j\in S} w_j\mathbf{h}_j'\mathbf{x}_j)^{-1}\sum_{j\in S_{(r)}} \tfrac{R}{R-1} w_j\mathbf{h}_j'\varepsilon_j - (\sum_{j\in S_{(r)}} w_{j(r)}\mathbf{h}_j'\mathbf{x}_j)^{-1}\sum_{j\in S_{(r)}} w_{j(r)}\mathbf{h}_j'\varepsilon_j.$$

It takes some work, but the second line on the right-hand side of equation (9) can be shown to be asymptotically dominated by the first line under mild conditions (which can be dubious for domain estimates). This is true even when the model fails and ε_j is replaced by $y_k - \mathbf{x}_k\mathbf{B}^*$. Plugging only the first line into the right-hand side of equation (4), it is not hard to show that the result would be an unbiased estimator the variance of $\mathbf{b}$ under the model in equation (5). This unbiasedness is only asymptotic when the model fails, and $\sum_S w_j\mathbf{h}_j'\mathbf{x}_j$ cannot be viewed as fixed.

3. Setting Up an Empirical Investigation

The simulations discussed in the next section assume a simple form of calibration: poststratication. The population is divided into P mutually exclusive classes, and $\mathbf{z}_k$ in equation

(1) is a row vector of class-indicators. That is to say, $z_{kp} = 1$ when k is in class p, 0 otherwise. Letting N_p be the population size of class p, and S_p the part of the sample is class p (which we assume is not empty) the calibrated weight for a sampled element in class p is

$$w_k = \frac{N_p}{\sum_{j \in S_p} a_j} a_k . \tag{10}$$

It is a simple matter to derive equation (10) from (1).

The r-th replicate weight for a sample element in class p can be derived from equation (2). It is 0 for k not in $S_{(r)}$, and

$$w_{k(r)} = \frac{N_p}{\sum_{j \in S_p \bigcap S_{(r)}} a_j} a_k \tag{11}$$

otherwise.

One estimator we will investigate is the sample-weighted domain mean:

$$\bar{y}_{dS} = \frac{\sum_{k \in S} w_k d_k y_k}{\sum_{k \in S} w_k d_k}, \tag{12}$$

in which $\mathbf{h}_k$ in equation (3) is equal to the scalar d_k (an indicator of domain membership) and $\mathbf{x}_k$ is the scalar 1. The other is the simple domain-specific weighted simple regression coefficient:

$$b_d = \frac{\sum_{k \in S} w_k d_k \left(x_k - \bar{x}_{dS} \right) \left(y_k - \bar{y}_{dS} \right)}{\sum_{k \in S} w_k d_k \left(x_k - \bar{x}_{dS} \right)^2}, \tag{13}$$

which is the second component of $\mathbf{b}$ in equation (3) when $\mathbf{x}_k = (1 \;\; x_k)$, and $\mathbf{h}_k = d_k \mathbf{x}_k$. There are alternative ways to define the variables in equation (3) to produce $\bar{y}_{dS}$ and b_d. One such was discussed in the previous section. We will also be interested in the "degenerate" case where all the $d_k = 1$, and $\bar{y}_{dS}$ and b_d are the whole-sample weighted means and weighted simple regression

coefficient, respectively.

The R replicate estimates for $\bar{y}_{dS}$ and b_d can be calculated by substituting $w_{k(r)}$ for w_k to compute each $\bar{y}_{dS(r)}$ and then substituting $w_{k(r)}$ for w_k, $\bar{y}_{dS(r)}$ for $\bar{y}_{dS}$, and $\bar{x}_{dS(r)}$ for $\bar{x}_{dS}$ to compute each $b_{d(r)}$. The DAG jackknife in equation (4) has the simplified scalar form:

$$v_J = \frac{R-1}{R}\sum_{r=1}^{R}(b - b_{(r)})^2. \tag{14}$$

The idealized linearization and linearization variance estimators in equations (7) and (8) are not so simply rendered. For $\bar{y}_{dS}$, **B** becomes the scalar B = $\bar{y}_{dU} = \sum_U d_k y_k / \sum_U d_k$, so that $\mathbf{U}_k = U_k = (\sum_S w_j d_j)^{-1} d_k (y_k - \bar{y}_{dU})$ and $\mathbf{u}_k = (\sum_S w_j d_j)^{-1} d_k (y_k - \bar{y}_{dS})$. Note that both are zero when k is not in the domain. Plugging into equations (7) and (8), we get

$$v_{IL}(\bar{y}_{dS}) = \frac{\sum_{p=1}^{P}\sum_{k\in S_p} w_k^2 \left[d_k \left(y_k - \bar{y}_{dU} - \frac{\sum_{j=1}^{N_p} d_j \left(y_j - \bar{y}_{dU} \right)}{N_p} \right) \right]^2}{\left(\sum_{k\in S} w_k d_k \right)^2}, \text{ and}$$

$$v_L(\bar{y}_{dS}) = \frac{\sum_{p=1}^{P}\sum_{k\in S_p} w_k^2 \left[d_k \left(y_k - \bar{y}_{dS} - \frac{\sum_{j\in S_p} w_j d_j \left(y_j - \bar{y}_{dS} \right)}{N_p} \right) \right]^2}{\left(\sum_{k\in S} w_k d_k \right)^2}. \tag{15}$$

For b_d as an estimator for the limit of

$$B_d = \frac{\sum_{k \in U} d_k \;\; x_k - \bar{x}_{dU} \;\;\; y_k - \bar{y}_{dU}}{\sum_{k \in U} d_k \;\; x_k - \bar{x}_{dU} \;\;^2},$$

it helps to first redefine $\mathbf{x}_k$ as $(1 \;\; x_k - \bar{x}_{dS})$, with $\mathbf{h}_k = d_k \mathbf{x}_k$ redefined accordingly, so that $\Sigma_S \, w_j \mathbf{h}_j' \mathbf{x}_j$ is diagonal. The scalars U_k and u_k become

$$U_k = \frac{d_k \;\; x_k - \bar{x}_{dS} \;\; e_k}{\sum_{j \in S} w_j d_j \;\; x_j - \bar{x}_{dS} \;\;^2}, \quad \text{and}$$

$$u_k = \frac{d_k \;\; x_k - \bar{x}_{dS} \;\; r_k}{\sum_{j \in S} w_j d_j \;\; x_j - \bar{x}_{dS} \;\;^2},$$

where

$e_k = \; y_k - \bar{y}_{dU} \; - \; x_k - \bar{x}_{dU} \;\; B_d$ is the population residual (for the regression coefficient), and

$r_k = \; y_k - \bar{y}_{dS} \; - \; x_k - \bar{x}_{dS} \;\; b_d$ is the sample residual.

Note that U_k and u_k are again zero when k is not in the domain.

We can now conclude

$$v_{IL}(b_d) = \frac{\sum_{k \in S_p} w_k^2 \left[d_k \;\; x_k - \bar{x}_{dS} \;\; e_k - \frac{\sum_{j=1}^{N_p} d_j \;\; x_j - \bar{x}_{dS} \;\; e_j}{N_p} \right]^2}{\left[\sum_{j \in S} w_j d_j \;\; x_j - \bar{x}_{dS} \;\;^2 \right]^2}, \quad \text{and}$$

$$v_L(b_d) = \frac{\sum_{k \in S_p} w_k^2 \left[d_k \;\; x_k - \bar{x}_{dS} \;\; r_k - \frac{\sum_{j \in S_p} w_j d_j \;\; x_j - \bar{x}_{dS} \;\; r_j}{N_p} \right]^2}{\left[\sum_{j \in S} w_j d_j \;\; x_j - \bar{x}_{dS} \;\;^2 \right]^2}. \qquad (16)$$

It was partly in response to the complicated nature of the equations (15) and (16) that NASS decided to use the DAG jackknife rather than linearization for the ARMS-III. In the next section, we also evaluate simplified versions of each:

$$v_{SL}(\bar{y}_{dS}) = \frac{\sum_{p=1}^{P} \sum_{k \in S_p} w_k{}^2 \left[d_k \;\; y_k - \bar{y}_{dS} \right]^2}{\left(\sum_{k \in S} w_k d_k \right)^2}, \tag{17}$$

and

$$v_{SL}(b_d) = \frac{\sum_{k \in S_p} w_k^2 \left[d_k \;\; x_k - \bar{x}_{dS} \;\; r_k \right]^2}{\left[\sum_{j \in S} w_j d_j \;\; x_j - \bar{x}_{dS} \right]^2}. \tag{18}$$

These simplified versions effectively assume there is no gain (reduction in variance) from poststratification.

4. A Simulation Study

We began our simulation study with an ARMS-III respondent sample of 986 farms in California. Our original plan was to use this sample and its final weights to generate a population.

Each farm in the sample had associated with it a frame value based on previous sales data. We called this value x_k. Classes were created by partitioning the x-values in 22 intervals, where the smallest interval was [0, 10 000), the largest interval was [750 000, ∞), and 20 intervals of equal width were spaced between 10,000 and 750,000.

We assigned a fraction of the 986 farms to domains of interest systematically. One such domain cotained 5% of the population. A second 10%. A third 20%.

Each farm in the sample also had a final weight associated with it, which we integerized and labeled a_k . At this point, each sampled farm had attached to it an x-value, an a-value, a class identifier, and three yes/no domain identifiers. We reproduced each sampled farm and its

attachments 10,000 times.

Our original idea was also to include survey-reported sales as the *y*-value for each of the 986 sampled farms and to create a fixed population of size $N = \Sigma^{9,860,000} a_k$. That is to say, the *y*, *x*, class identifier, and domain identifiers for each sampled farm *k* would be replicated $10{,}000a_k$ times in the population. This would create a very large population with the same moments of *y* and *x* as the *a*-weighted sample. Independent samples could then be drawn from the putative population by giving each element replicated from *k* a Poisson selection probability of $1/(10{,}000a_k)$. The expected size for each sample would be 986.

Alas, no matter how large we made the simulated population, we found the results unsettling. This was because there could only be 986 possible realizations of the *y*-variable. Even if these *y*-values were orginally generated from a normal distribution, the roughly 49 that would fall into the smallest domain of interest could (and sometimes did) behave very idiosyncratically. Consequently, we decided that we needed to generate the *y*-values for each putative population unit directy from a model.

We used two models to generate the *y*-values. Both had the form:

$$y_k = \beta_0 + \beta_1 x_k^{\alpha} + \beta_2 \log(a_k) + \varepsilon_k, \qquad (19)$$

where the ε_k were independent draws from a $N(0, 100^2)$ distribution, $\beta_0 = 50$, and $\beta_1 = 2$. For one of the models, labeled *Model 1*, we set $\alpha = 1$, and $\beta_2 = 0$. It is a simple linear model under an ignorable sampling mechanism. For the other, labeled *Model 2*, we set $\alpha = 1.1$, and $\beta_2 = 100$.

Ten thousand simulated samples were effectively drawn from the putative population with *y*-values generated by one of the two versions of equation (19) in the following maner. A farm in the original sample was associated with a particular *x*-value, class and domain identifiers, and with a_k *y*-values generated from equation (19) with certain settings. Each *y*-value, together with its associated *x*-value, class identifier, and domain identifiers, was given an independent $1/a_k$ probability of being selected into a simulated sample. As a result, the estimated size for each simulated sample was 986. We expected 49.3 farms to be in each 5%-domain sample, 98.6 in each 10%-domain sample, and 197.2 in each 20%-domain sample.

Estimated means and simple regression coefficients were calculated from the simulated samples using equations (12) and (13) respectively.

The targets of the estimated means and simple regression coefficients were parameters of a conceptual infinite population. In the text, such parameters were labelled (when scalars) B*. We computed analogous and near-identical large-population B-values thusly. We generated 986,000 *y*-values under the respective versions of equation (19); 10,000 for each original farm *k*. Such a *y*-value, together with an associated *x*-value, class identifier, and domain identifiers, was repeated a_k times. The mean *y*-value and the slope the linear regression of the y_k on the x_k were then computed for this simulated population and for the three designated domains of the population.

Table 1 displays the relative empirical biases from using alternative methods for estimating the mean squared error of *b* (which could be either either $\bar{y}_{dS}$ and b_d) as an estimator for B. These relative empirical biases are computed using

$$R = \frac{\sum_{t=1}^{10,000} v_t - \sum_{t=1}^{10,000} (b_t - B)^2}{\sum_{t=1}^{10,000} (b_t - B)^2},$$

where b_t and v_t are computations of the parameter estimate and its estimated variance based on the t^{th} simulated sample. The estimated standard errors on these statistics tended to be between 0.015 and 0.02.

The empirical variance as a fraction of empirical mean squared error was always over 96% for every *b* with an estimated mean squared error on the table. Consequently, whether we treat the DAG jackknife and its linearization counterparts as estimators of variance or mean-squared-error makes little practical difference.

As the table shows, the empirical biases from using the DAG jackknife in equation (14) are all positive, while the biases from using the full linearization estimator in equations (15) (for the mean) and (16) (for the simple regression coefficient) are almost all negative. Both tend to get worse, in absolute terms, as the domain sample size decreases. This happens whether

estimating the mean squared error of a domain mean or a simple regression coefficient or whether generating the *y*-values with *Model 1* or *Model 2*.

When estimating means, the relative empirical biases are always under 10% in absolute terms using either the full linearization variance estimator or the DAG jackknife with 15 or 30 replicates. Using the simplified linearization estimator in equation (17), however, appears badly biased for the full-population mean under either model. This variance estimator gets better as the domain sample size gets smaller. It is reasonable to conclude that the effect on mean estimation of poststratification (which was done at the full-sample level) becomes less powerful the smaller the domain of interest.

Estimating the mean squared error of the full-sample simple regression coefficient using the simplified linearization in equation (18) works well under *Model 1* because the poststratification is irrelevant in the context where the *y*-values are generated by a linear form of equation (19), and the $e_k = y_k - \bar{y}_{dU} - x_k - \bar{x}_{dU} B_d$ are uncorrelated with the a_k.

We do not obseve much difference between the full and simplified linearization variance estimators for the full-sample simple regression coefficient under *Model 2*. The impact of poststriatification appears to be overwhelmed by the correlation between the e_k and the a_k in this context.

For the 5% domain (domain sample sizes of around 50), none of the variance estimators for the estimated simple regression coefficient have relative empiricial biases of less than10% in absolute terms under either *Model 1* or *Model 2*. The two jackknives work much better for the 10% domain (domain sample sizes of around 100), however, as do the two linearization estimators under *Model 1*.

Table 2 displays the coefficients of variation for the various variance esitmators. A 30-replicate DAG jackknife has more stability (a small coefficient of variation) than a 15-replicate version. Linearization is more stable than either jackknife. Stability decreases with the size of the domain sample. It is less for the variance estimator of the simple regression coefficient than the mean.

Table 1
Relative Biases of Alternative Estimators for Mean Squared Error

Domain Proportion of Sample	Estimated Mean: DAG Jackknife R =15	Estimated Mean: DAG Jackknife R = 30	Estimated Mean: Linearization Full	Estimated Mean: Linearization Simplified	Estimated Regression Coefficient: DAG Jackknife R =15	Estimated Regression Coefficient: DAG Jackknife R = 30	Estimated Regression Coefficient: Linearization Full	Estimated Regression Coefficient: Linearization Simplified
Model 1								
5%	0.076	0.078	-0.069	0.041	0.126	0.116	-0.241	-0.232
10%	0.026	0.010	-0.041	0.078	0.086	0.093	-0.036	-0.024
20%	0.018	0.002	-0.029	0.151	0.040	0.016	-0.040	-0.012
100%	0.016	0.032	0.000	3.717	0.061	0.035	-0.010	-0.007
Model 2								
5%	0.059	0.066	-0.075	0.019	0.248	0.249	-0.195	-0.182
10%	0.038	0.049	-0.015	0.088	0.048	0.047	-0.176	-0.180
20%	0.024	0.022	-0.013	0.137	0.048	0.005	-0.133	-0.112
100%	0.018	0.019	0.004	2.347	0.069	0.099	-0.123	-0.165

Table 2
Coefficients of Variation of Alternative Estimators for Mean Squared Error

Domain Proportion of Sample	Estimated Mean: DAG Jackknife R =15	Estimated Mean: DAG Jackknife R = 30	Estimated Mean: Linearization Full	Estimated Mean: Linearization Simplified	Estimated Regression Coefficient: DAG Jackknife R =15	Estimated Regression Coefficient: DAG Jackknife R = 30	Estimated Regression Coefficient: Linearization Full	Estimated Regression Coefficient: Linearization Simplified
Model 1								
5%	0.76	0.68	0.59	0.60	0.90	0.83	0.56	0.56
10%	0.55	0.46	0.36	0.36	0.55	0.48	0.36	0.36
20%	0.47	0.37	0.25	0.24	0.47	0.39	0.27	0.27
100%	0.50	0.42	0.32	0.16	0.62	0.56	0.46	0.47
Model 2								
5%	0.76	0.70	0.60	0.63	1.13	1.09	0.70	0.69
10%	0.56	0.47	0.38	0.36	0.76	0.69	0.51	0.49
20%	0.47	0.37	0.26	0.25	0.62	0.55	0.43	0.42
100%	0.52	0.43	0.34	0.18	0.66	0.54	0.36	0.36

5. Concluding Remarks

We are hesitant to make overly bold claims from the results of our limited empirical study. Nevertheless, we were pleased to see that using a delete-a-group jackknife with as few as 15 replicates on a heavily calibrated sample, one containing 22 poststrata, produced reasonable and conservative variance measures for an estimated mean based on samples containing as few as 50 domain members. Variance measures for an estimated simple regression coefficient did not behave as well until domain samples were roughly twice as large. They did, however, remain competitive with more complicated linearization-based alternatives. These alternatives were more stable but also consistently underestimated true mean squared errors.

It seems to us that the DAG jackknife is a reliable variance-estimation tool for simple ratios like the population mean with domain sample sizes in the 50 and above range. On the other hand, we would not be comfortable using the DAG jackknife for estimating the variance of regression coefficients with less than 100 in-scope sample units. This discomfort extends to all "model-free" variance-estimation methods. When sample sizes get too small, we strongly suspect one needs to assume a model and estimate variances using a technique appropriate for that model.

References

Demnati, A. and Rao, J.N.K. (2004). Linearization Variance Estimators for Survey Data. Survey Methodology, 30, 4-13.

Estevao, V. and Särndal, C.E. (2000). A Functional Form Approach to Calibration. Journal of Official Statistics, 16, 379-399.

Kott, P.S. (1998). Using the Delete-a-Group Jackknife Variance Estimator in NASS Surveys, NASS Research Report 98-01 (revised 2001). Reproduced at http://www.nass.usda.gov/research/reports/RRGJ7.pdf.

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

Kott, P.S. (2006a). Delete-a-group variance Eestimation for the General Regression Estimator under Poisson Sampling. Journal of Official Statistics, 22, 759-767.

Kott, P.S. (2006b). Using Calibration Weighting to Adjust for Nonresponse and Coverage Errors. Survey Methodology, 32, 133-142.

Kott, P.S., (2007). Clarifying Some Issues in the Regression Analysis of Survey Data. Survey Research Methods, 1, 11-18.

USDA (2007). http://www.ers.usda.gov/DATA/Arms/GlobalDocumentation.htm.