# Building a Better Delete-a-Group Jackknife for a Calibration Estimator (Like That Based on Data from the ARMS III)

Phillip S. Kott

This note summarizes much of the theory behind the use of the Delete-a-Group (DAG) jackknife with calibrated survey data like that coming from the third phase of the Agricultural Resources and Management Survey (ARMS III). A DAG jackknife employing 15 sets of replicate weights is used for variance estimation with ARMS-III data. Many analyses of this data are conducted by the Economics Research Service and by other agricultural economists outside of the US Department of Agriculture, making a portable variance-estimation technique like the DAG jackknife especially desirable. Some suggestions are offered to improve the present methodology.

KEY WORDS: Sampling Weight; Replicate Weight; Variance; Mean Squared Error.

Phillip S. Kott is Chief Research Statistician, Research and Development Division, National Agricultural Statistics Service, 3251 Old Lee Highway, Room 305, Fairfax, Virginia, 22030. An earlier version of this paper was prepared for the Workshop on Calibration and Estimation in Surveys, October/November 2007, Ottawa, Ontario. The final section ("The ARMS III Application") and Appendix B were *not* in the original.

## Introduction

The National Agricultural Statistics Service (NASS) has increasingly been using a delete-a-group (DAG) jackknife for variance estimation. In surveys where this technique is used, each sampled element k (usually a farm) is given 16 weights: the element's actual sampling weight *after* incorporating all nonresponse and calibration adjustments, $w_k$, and 15 jackknife replicate weights, $w_{k(g)}$, with g = 1, ..., 15.

The estimator, t, for a univariate parameter of interest, T, is computed using the actual (post-adjustment) sampling weights. Fifteen replicate estimators, $t_g$, each computed with its respective set of replicate weights are then calculated. The DAG jackknife variance estimator for t is

$$v_{DAG} = (14/15)\sum^{15} (t - t_{(g)})^2. \qquad (1)$$

NASS recommends that when computing coverage intervals or testing hypotheses, one treat the associated z statistic, $z = (t - T)/\sqrt{v_J}$, as if it had a Student's t distribution with 14 degrees of freedom.

This note summarizes much of the theory behind the use of the DAG jackknife with data from the third phase of the Agricultural Resources and Management Survey (ARMS III). The survey is based on a multi-phase composite sample of farms. There is also an area-frame component to pick up farms not on the NASS list. The ARMS III provides an annual assessment of the economic conditions of US farms.

The respondent sample for the ARMS III is calibrated to targets determined from other NASS surveys. A delete-a-group jackknife employing 15 sets of replicate weights is used for variance estimation. Much of the analysis of ARMS III data is conducted by the Economics Research Service and by other agricultural economists outside of the US Department of Agriculture, making a portable variance-estimation technique like the DAG jackknife especially desirable.

In general, the theory underpinning the use of the DAG jackknife (and all jackknives for that matter) is asymptotic. See Kott (1998).

The focus here will be on the properties of the DAG jackknife for a calibration estimator, with a particular emphasis on model-based properties. If a variance

estimator does not behave well under a reasonable model, then it is unlikely to behave well when that model fails.

**Some Calibration Theory**

Suppose our goal is to estimate a population total, $T_y = \sum_U y_k$, where U denotes the population of N members. The expansion estimator for $T_y$ is

$$t_{y\_E} = \sum_S d_k y_k,$$

where S is the sample of n members, and $d_k$ is the original sampling weight of element k (usually the inverse of the unit's selection probability).

Suppose there is a row vector of auxiliary variables, $\mathbf{x}_k = (x_{1k}, ..., x_{Pk})$, for which the population total(s), $\mathbf{T_x} = \sum_U \mathbf{x}_k$, is known. Deville and Särndal (1992) coined the term "calibration estimator" to describe an estimator of the form

$$t_{y\_CAL} = \sum_S w_k y_k,$$

where the $w_k$ are close to the $d_k$, and the calibration equation,

$$\sum_S w_k \mathbf{x}_k = \sum_U \mathbf{x}_k,$$

holds.

As with the expansion estimator, the same set of calibration weights can be used no matter what the y-variable. Usually, $t_{y\_CAL}$ is nearly randomization unbiased for a sufficiently large sample under mild conditions. See, for example, Kott (2005a).

Observe that $t_{y\_CAL}$ estimates $T_y$ perfectly when $y_k$ equals $\mathbf{x}_k\boldsymbol{\beta}$ exactly for all k. Thus, it is reasonable to expect $t_{y\_CAL}$ to be a good estimator when $y_k$ and $\mathbf{x}_k\boldsymbol{\beta}$ are almost always close. This is formalized by assuming the $y_k$ are random variables satisfying the linear prediction model:

$$y_k = \mathbf{x}_k\boldsymbol{\beta} + \varepsilon_k, \qquad (2)$$

where $E(\varepsilon_k | \mathbf{x}_g, I_g) = 0$ for all $k \in U$; $I_g = 1$ when $g \in S$, 0 otherwise. Under this model, it is easy to see that $t_{y\_CAL}$ is an unbiased estimator for $T_y$ in the sense that

$E_\varepsilon(t_{y\_CAL} - T_y) = 0$.

Most of the calibration weighting in practice involve a variant of least squares, where the calibration weights have linear the form:

$$w_k = d_k(1 + c_k \mathbf{x}_k \mathbf{g}) = d_k(1 + c_k \mathbf{g}' \mathbf{x}_k')$$

for some set of constants $\{c_k\}$, and a vector: $\mathbf{g} = (\sum_S d_j c_j \mathbf{x}_j' \mathbf{x}_j)^{-1} (\mathbf{T_x} - \sum_S d_j \mathbf{x}_j)'$.

This calibration estimator is also a generalized regression (GREG) estimator:

$$\begin{aligned} t_{y\_CAL} &= \sum_S d_k (1 + c_k \mathbf{g}' \mathbf{x}_k') y_k \\ &= \sum_S d_k y_k + \mathbf{g}' \sum_S d_k c_k \mathbf{x}_k' y_k \\ &= t_{y\_E} + (\mathbf{T_x} - \sum_S d_j \mathbf{x}_j)(\sum_S d_j c_j \mathbf{x}_j' \mathbf{x}_j)^{-1} \sum_S d_k c_k \mathbf{x}_k' y_k \\ &= t_{y\_E} + (\mathbf{T_x} - \sum_S d_j \mathbf{x}_j) \mathbf{b}, \end{aligned}$$

where $\mathbf{b} = (\sum_S d_j c_j \mathbf{x}_j' \mathbf{x}_j)^{-1} \sum_S d_k c_k \mathbf{x}_k' y_k$. It is easy to see the $\mathbf{b}$ is an unbiased estimator for $\boldsymbol{\beta}$ under the model.

If the $\varepsilon_k$ in the model are uncorrelated, each with a variance of $\sigma_k^2$, then the model variance of $t_{\_CAL}$ is

$$\begin{aligned} E_\varepsilon[(\sum_S w_k y_k - \sum_U y_k)^2] &= E_\varepsilon[(\sum_S w_k \varepsilon_k - \sum_U \varepsilon_k)^2] \\ &= \sum_S w_k^2 \sigma_k^2 - 2 \sum_S w_k \sigma_k^2 + \sum_U \sigma_k^2 \\ &\approx \sum_S w_k^2 \sigma_k^2 \end{aligned}$$

when most $w_k >> 1$, which we will assume to be the case here.

**The Delete-a-Group Jackknife**

Suppose the (respondent) sample was randomly divided into 15 mutually exclusive and nearly equal groups denoted $S_g$, where g = 1, ..., 15. We call the complement of each group, the *jackknife replicate group* and denote it by $S_{(g)}$.

Observe that

$$\sum_S w_k\varepsilon_k - \sum_{S_{(g)}} (15/14)\, w_k\varepsilon_k = \sum_{S_g} w_k\varepsilon_k - \sum_{S_{(g)}} (1/14)\, w_k\varepsilon_k.$$

A little work reveals that

$$E_M[\sum^{15} (14/15)(\sum_S w_k\varepsilon_k - \sum_{S_{(g)}} (15/14)\, w_k\varepsilon_k)^2] = \sum_S w_k^2\sigma_k^2,$$

the approximate model variance of $t_{\_CAL}$.

Unfortunately, the $\varepsilon_k$ are unknown. We could replace them by the $y_k$ in the above expression if the calibration equation

$$\sum_{S_{(g)}} (15/14)\, w_k\mathbf{x}_k = \sum_U \mathbf{x}_k$$

held for all g. It generally does not. We can, however, replace each $(15/14)w_k$ within $S_{(g)}$ by the corresponding (nonstandard) g'th replicate weight:

$$w_{k(g)} = (15/14)w_k + (\sum_U \mathbf{x}_i - \sum_{S_{(g)}} [15/14] w_i\mathbf{x}_i)[\sum_{S_{(g)}} c_j d_j \mathbf{x}_j' \mathbf{x}_j]^{-1} c_k d_k \mathbf{x}_k' .$$

(Most of the results discussed here hold equally well for a more conventionally defined replicate weight, but the one above is more useful for our purposes.)

Each replicate weight in $S_{(g)}$ is close to its corresponding $(15/14)w_k$. Moreover, as a group, they satisfy the calibration equation since

$$\begin{aligned}\sum_{S_{(g)}} w_{k(g)}\mathbf{x}_k &= \sum_{S_{(g)}} (15/14)\, w_k\mathbf{x}_k + \\ &\qquad (\sum_U \mathbf{x}_i - \sum_{S_{(g)}} [15/14] w_i\mathbf{x}_i)[\sum_{S_{(g)}} c_j d_j \mathbf{x}_j' \mathbf{x}_j]^{-1} \sum_{S_{(g)}} c_k d_k \mathbf{x}_k' \mathbf{x}_k \\ &= \sum_U \mathbf{x}_i.\end{aligned}$$

Defining $w_{k(g)}$ to be 0 when $k \in S_g$, the delete-a-group jackknife variance estimator can be expressed as

$$v_{DAG} = \sum_{g=1}^{15} (14/15)\Big(\sum_{k\in S} w_k y_k - \sum_{k\in S} w_{k(g)} y_k\Big)^2$$

$$= \sum^{15} (14/15)\Big(\sum_S w_k \varepsilon_k - \sum_S w_{k(g)} \varepsilon_k\Big)^2 \qquad (3)$$

$$\approx \sum^{15} (14/15)\Big(\sum_S w_k \varepsilon_k - \sum_{S_{(g)}} [15/14] w_k \varepsilon_k\Big)^2 .$$

Thus, $v_{DAG}$ is a nearly unbiased estimator for the model variance of $t_{y\_CAL}$. It is not hard to see that the model bias, caused by the replication weights not exactly equaling $(15/14)w_k$ for $k \in S_{(g)}$, is usually upward. There is no theoretical guarantee of that, however.

Let us now turn to the randomization-based properties of the DAG jackknife under two particular sampling designs. Kott (2006a) shows that under Poisson sampling $v_{DAG}$ is almost unbiased for the randomization mean squared error of $t_{y\_CAL}$ when the population residual, $e_k = y_k - \mathbf{x}_k(\sum_U c_j \mathbf{x}_j' \mathbf{x}_j)^{-1} \sum_U c_j \mathbf{x}_j' y_j$, is such that $\sum_U e_k = 0$. The key is that both terms of

$$\sum_S w_k e_k - \sum_{S_{(g)}} (15/14)\, w_k e_k$$

need to estimate $\sum_U e_k = 0$. For that to happen $c_k$ must be expressible as $\mathbf{x}_k\boldsymbol{\lambda}$ for some column vector $\boldsymbol{\lambda}$. This will often rule out the randomization-optimal estimator where $c_k = d_k - 1$. Although not always a good estimator for the randomization mean squared error of $t_{y\_CAL}$ under Poisson sampling, the DAG jackknife remains a nearly unbiased estimator for the model variance so long as the $w_k$ are all large.

Let us now look at the expansion estimator under stratified simple random sampling before any calibration or nonresponse adjustment. Ignoring finite population correction, it has a randomization variance of

$$V_{rand} = \sum_{h=1}^{H} (N_h^2/n_h) \sum_{k\in U_h} e_k^2/(N_h - 1) \approx \sum_{k\in U} w_k e_k^2,$$

where $e_k = y_k - \sum_{i\in U_h} y_j / N_h$, and $w_k = N_h / n_k$ for $k \in U_h$.

Although incalculable (because the $e_k$ are unknown),

$v = \sum_S w^2 e_k^2$

is an unbiased estimator for $\sum_U w_k e_k^2$, while v itself is approximated by the DAG jackknife in equation (1) with $w_{k(g)} = N_h / n_{h(g)}$ for $k \in S_h$ so long as $N_h / n_{h(g)}$ is approximately $(15/14)N_h / n_h$. This last approximation requires large-enough $n_h$ since each $(14/15)n_h$ must approximate $n_{h(g)}$. A good rule of thumb is that all $n_h$ should exceed 5. Kott (2001) proposes an extended DAG jackknife for designs with stratum sample sizes smaller than that.

When the calibration occurs *after* the latest phase of a multi-phase sample, one can construct the following replicate weights:

$$w_{k(g)} = (d_{k(g)}/d_k)w_k + \sum_U \mathbf{x}_i - \sum_{S_{(g)}} [d_{i(g)}/d_i]\, w_i \mathbf{x}_i)[\ \sum_{S_{(g)}} c_j d_j \mathbf{x}_j\, '\, \mathbf{x}_j\ ]^{-1} c_k d_k \mathbf{x}_k',$$

where $d_k$ is the sampling weight before calibration, while $d_{k(g)}$ is the replicate weight before calibration. When $d_{k(g)} / d_k \approx 15/14$, the approximate model unbiasedness of the DAG jackknife holds, but there is an additional tendency for the bias to be upward. If the DAG jackknife is almost randomization unbiased before calibration, it remains so after.

**Nonresponse and Coverage Adjustment**

We can use the same prediction model to justify both calibration weighting and the DAG jackknife when the sample is subject to nonresponse or the frame to coverage errors. For handling coverage errors, the true (without coverage-error) $\mathbf{T_x}$ is assumed known.

An alternative justification for calibration treats unit response or frame coverage as another phase of Poisson sampling. This is called quasi-randomization or quasi-random response (coverage) modeling. The theory supporting the use of the DAG jackknife is analogous to the theory in Kott (2006b) for the conventional stratified jackknife.

**Can We Build a Better DAG Jackknife (for the ARMS III)?**

*Should NASS increase the number of replicates?*

Having more than 15 replicates would be helpful for multivariate testing and for increasing the precision of variance estimates. On the other hand, when there are fewer replicates, there are more units within each, which is more compatible with the invocation of asymptotic normality. Setting G = 30 seems a reasonable compromise.

C*an we remove the upward bias in the DAG jackknife?*

It is not hard to see that replacing

$$v_{DAG} = \sum^{G} [(G-1)/G](\sum_S w_k y_k - \sum_S w_{k(g)} y_k)^2$$

with

$$v_{DAG}^{*} = \sum^{G} \frac{\sum_{S_g} w_k^2}{\sum_S (w_k - w_{k(g)})^2} (\sum_S w_k y_k - \sum_S w_{k(g)} y_k)^2 \qquad (4)$$

will remove the bias of the DAG under the model in equation (2) when the $\varepsilon_k$ are uncorrelated and have a common variance. Note that the asymptotic randomization-based properties of the DAG jackknife are unaffected by this substitution.

We can similarly remove the bias under an alternative working model – say, that the element variances are proportional to $f_k$ – by replacing

$$\frac{\sum_{S_g} w_k^2}{\sum_S (w_k - w_{k(g)})^2} \quad \text{with} \quad \frac{\sum_{S_g} w_k^2 f_k}{\sum_S (w_k - w_{k(g)})^2 f_k}$$

in equation (4). So long as the $\varepsilon_k$ are uncorrelated, the (bias-adjusted) DAG jackknife remains *nearly* model unbiased even when the working-model speculation about the relative sizes of the $\sigma_k^2$ is wrong.

Although we have so far only discussed totals, the DAG jackknife is also effective when t is component of a regression coefficient of the form:

$$\mathbf{b} = (\sum_S w_j \mathbf{h}_j \mathbf{z}_j')^{-1} \sum_S w_k \mathbf{h}_k y_k$$

for column vectors $\mathbf{h}_j$ and $\mathbf{z}_j$ (in a common example both are equal to $\mathbf{x}_k'$). To compute the DAG jackknife we replace $w_k$ and $w_{k(g)}$ with the analogous components of

$$\boldsymbol{\omega}_k = (\sum_S w_j \mathbf{h}_j \mathbf{z}_j')^{-1} w_k \mathbf{h}_k \text{ and}$$

$$\boldsymbol{\omega}_{k(g)} = (\sum_S w_{j(g)} \mathbf{h}_j \mathbf{z}_j')^{-1} w_{k(g)} \mathbf{h}_k .$$

For the special case of a domain mean, $\mathbf{z}_j$ is the scalar 1, while $\mathbf{h}_j$ is 1 when j is in the domain and 0 otherwise.

In Appendix A, a working model is invoked to compute a theoretically better measure for the effective degrees of freedom of the DAG jackknife than G – 1. It remains an open question how relevant removing the model bias and better measuring the model variance of the DAG jackknife are in the face of inevitable model failure.

*What can we do about replicate-weight outliers*?

Ideally, a replicate weight should be close to $G/[G-1]w_k$. What if it were not? We could truncate troublesome replicate weights or even remove troublesome target variables from the replicate calibration. Surprisingly, this may tend to cause an *upward* bias in variance estimation.

To see why, look at the model expectation of DAG jackknife expressed in equation (3). The term

$$\sum_{g=1}^{15} (14/15) (\sum_S w_k \mathbf{x}_k \boldsymbol{\beta} - \sum_S w_{k(g)} \mathbf{x}_k \boldsymbol{\beta})^2 = \sum (14/15)[(\sum_S w_k \mathbf{x}_k - \sum_S w_{k(g)} \mathbf{x}_k) \boldsymbol{\beta}]^2$$

is missing from the right hand side of the equation because the $w_{k(g)}$, like the $w_k$, are calibrated on all the components of $\mathbf{x}_k$. When that calibration is not enforced (due, for example, to truncating the replicate weights), the term above adds a positive model bias to the DAG jackknife. This size of the bias is directly related to the effectiveness of calibration in reducing the variance/mse of the estimator itself (through the absolute sizes of the components of $\boldsymbol{\beta}$).

Since the DAG jackknife both before and after truncating replicate weights – or after removing calibration targets when computing the replicate weights – has a

tendency to be biased upward, computing the DAG both ways and taking the lesser value is a reasonable policy.

**The ARMS III Application**

The actual survey design used for the ARMS surveys varies from year to year. In most years, the ARMS begins with a stratified simple random sample drawn from the NASS list frame. This *screening sample* is selected in late summer to determine which potential farms on the list frame are in business. The sample is also often used to determine whether which farms are engaged in particular enterprises of interest. Those enterprises could be corn and hogs in one year, wheat in another, and nothing in a third. A subsample of the farms engaged in each enterprise of interest serves as a component of the ARMS III.

Although the stratum sample sizes almost always exceed five in the screening survey described above, the same cannot always be said about the accompanying June area sample used to identify farms missing from the NASS list frame. Rather than using an extended DAG jackknife to handle such situations, NASS treats the larger area-frame land-use strata as the design strata in DAG variance estimation (the actual design strata for the survey are the geographically-determined substrata). When one or more land-use strata contain five of fewer sampled segments, the resulting DAG jackknife has, if anything, a tendency to overestimate variances. Since such sampled segments usually contain no farms, this tendency is more theoretical that real.

In years where there are one or more enterprise samples, they are combined with a general subsample of the screening sample to form the list-component of the ARMS-III sample. The factor(s) used in weighting the composite list sample is (are) used in constructing the replicate weights.

The list sample is combined with farms identified by the area-frame sample as not-on-the-list (NOL). Weights for this final ARMS-III sample are calibrated at the region level (with many large states serving as regions) to meet outside targets determined from other NASS surveys. These targets are treated as known constants by NASS for

variance-estimation purposes. Kott (2005b) describes how to incorporate targets with their own variances into a DAG jackknife.

NASS uses a truncated version of linear calibration to create ARMS-III sampling weights. First, the $c_k$ are set at $1 - 1/d_k$, where $d_k$ is the weight before calibration. After an initial round of calibration, weights that would otherwise be less than 1 are set to 1, and weights deemed too high are set back to $d_k$. Farms having these truncated weights are removed before a second round of linear calibration is conducted. The process of truncating too-small and too-large weights and then removing the associated farms from the calibration is repeated as necessary.

In recent years, replicate weights have been truncated at zero (to aid users employing software packages that cannot handle negative weights). In addition, some targets used in calibration have been dropped when constructing replicate weights on a region-by-region basis.

Appendix B describes a method for constructing a "stratified" jackknife from 2006 ARMS-III replicate weights. This jackknife allows the sophisticated user to conduct multivariate tests in certain situations where the DAG jackknife cannot be used.

**Appendix A: Effective Degrees of Freedom**

For simplicity, let G = 15 and assume $w_{k(g)} \approx (15/14)w_k$ for $k \in S_{(g)}$ (which holds in the limit as S becomes asymptotically large), Consequently,

$$
\begin{aligned}
t - t_{(g)} &= \sum_{k \in S_g} w_k \varepsilon_k + \sum_{k \in S_{(g)}} (w_k - w_{k(r)})\, \varepsilon_k, \\
&\approx \sum_{k \in S_g} w_k \varepsilon_k - \sum_{k \in S_{(g)}} w_k \varepsilon_k / 14, \\
&= (15/14)\Big( \sum_{k \in S_g} w_k \varepsilon_k - \sum_{k \in S} w_k \varepsilon_k / 15 \Big),
\end{aligned}
$$

which is dominated by the first summation. This lead to the following *ad hoc* approximation:

$$
t - t_{(g)} \approx (15/14)\Big( \sum_{k \in S_g} w_k \varepsilon_k - \sum_{r=1}^{G} \sum_{k \in S_r} w_k \varepsilon_k / 15 \Big),
$$

where the dominant term is correct, but the other term relies on $w_{k(g)} \approx (15/14)w_k$.

From this last expression, we have the following approximation for $v_{DAG}$:

$$v_{approx} = (15/14)^2 \{ \sum^{15} u_g^2 - ( \sum^{15} u_g)^2/15 \},$$

where $u_g = \sum_{S_g} w_k \varepsilon_k$.

If we assume the $u_g$ are approximately normal random variables, then following the logic in Kott (1994) eq. (6), we can see that $v_{approx}$ has roughly the multiple of a chi-squared distribution with

$$\Delta = ( \sum_{g=1}^{15} q_g)^2 / \{ \sum_{g=1}^{15} q_g^2 + \sum_{g' \neq g} q_g q_{g'} /196\}.$$

degrees of freedom, where $q_g$ is a multiple of the model variance of $u_g$. If we conjecture that $\sigma_k^2 \propto f_k$, then $q_g = \sum_{S_g} w_k^2 f_k$ . Consequently, the z statistic, $z = (t - T)/\sqrt{v}$, has roughly a Student's t distribution with $\Delta$ degrees of freedom.

Even if the model holds, $\sigma_k^2 \propto f_k$, and $w_{k(g)} \approx (15/14)w_k$, when the sample size is small (less than, say, 15 x 15 = 225) assuming the $u_r$ are roughly normal is tantamount to assuming that the $\varepsilon_k$ are roughly normal or, at least, each has a fourth moment approximately equal to $3\sigma_k^4$. This may be unreasonable in many applications. Nevertheless, assuming the z statistic has $\Delta$ degrees of freedom is more reasonable than assuming it has 14 degrees of freedom.

## Appendix B: Creating a "Stratified" Jackknife for the 2006 ARMS III

Estimating variances for a vector of coefficients in a large model can be problematic when a DAG jackknife has only 15 replicates. To see why, let **t** be a d-vector of estimated coefficients. Its DAG jackknife variance estimator, $\mathbf{v}_{DAG}(\mathbf{t})$, is the $d \times d$ matrix computed by replacing $(t - t_{(g)})^2$ in equation (1) with $(\mathbf{t} - \mathbf{t}_{(g)})(\mathbf{t} - \mathbf{t}_{(g)})'$ ) (itself a $d \times d$ matrix).

A test statistic for the null hypothesis $\mathbf{t} = \mathbf{0}$ is $F = \mathbf{t}'\left[\mathbf{v}_{DJ}(\mathbf{t})\right]^{-1}\mathbf{t}/d$, which has approximately an F(d, 15 – d) distribution under the hypothesis. This statistic only exists when d is less than 15. In fact, when $d \geq 15$, $\mathbf{v}_{DJ}(\mathbf{t})$ will not be invertible.

In some circumstances, a user will be able to construct a "stratified" jackknife variance estimator for a vector like **t** that is invertible when the DAG is not. This appendix sketches how.

The 2006 ARMS III (respondent) sample can be divided into H = 20 mutually exclusive calibration regions or "strata." These strata consist of 15 large states and 5 "rump" collections of smaller states. We let $T_h$ denote that part of the 2006 sample from stratum h.

For the stratified jackknife, we can create 300 ($20 \times 15$) *stratified replicate weights* for each sample farm with the formula:

$$w_{k(hg)} = \begin{Bmatrix} w_{k(g)} & \text{when } k \in T_h \\ w_k & \text{otherwise} \end{Bmatrix} \tag{B1}$$

For any estimated parameter vector **t** computed with the calibration weights, a user can also compute 300 *stratified replicate estimates*, each denoted $\mathbf{t}_{(hg)}$, using the appropriate set of stratified replicate weights, and then estimate the variance of **t** with

$$\mathbf{v}_{SJ}(t) = (14/15)\sum_{h=1}^{H}\sum_{g=1}^{15}(\mathbf{t} - \mathbf{t}_{(hg)})(\mathbf{t} - \mathbf{t}_{(hg)})'. \tag{B2}$$

Observe that if **t** is computed from data deriving from a subset S′ of the sample, and the subset only contains farms in H′ < 20 strata, then a user could repl ace H in the first summation of the right-hand side of equation (B2) by H′ without changing the result. In fact, if the data was wholly contained within a single stratum, then the stratified jackknife in equation (B2) would be identical to the DAG jackknife.

When evaluating a large model involving data from a number of strata, the stratified jackknife variance estimator for **t** in equation (B2) may be invertible when that

analogous DAG jackknife is not. The test statistic $\chi^2 = \mathbf{t}'\left[\mathbf{v}_{SJ}(\mathbf{t})\right]^{-1}\mathbf{t}$ will then be asymptotically chi-squared with 14 degrees of freedom under the null hypothesis that $\mathbf{t} = \mathbf{0}$. The alternative *ad hoc* statistic $F = \chi^2/d$ will be roughly $F(d, 14H' + 1 - d)$ (note: replace $H'$ with $H$ when the sample contains farms from every stratum). This statistic is *ad hoc* because, among other things, the strata are not necessarily of nearly equal size. It is nonetheless preferable to use F in place of $\chi^2$ when $14H' + 1 - d$ is not large.

When the sample contains data from all 20 strata, some users may find computing the 300 replicate estimates necessary for determining $\mathbf{v}_{SJ}(\mathbf{t})$ in equation (B2) burdensome. A simplified version of the stratified jackknife can be constructed that collapses the 20 strata into a more manageable number of *variance strata*. Stratified replicate weights, replicate estimates, and test statistics can be computed accordingly. For example, one could collapse the 15 state strata into five mutually exclusive variance strata, each containing three states, while placing the states in the five rump strata into a sixth variance strata. Each farm would then have 90 ($6 \times 15$) replicate weights computed using equation (B1) with h redefined as a variance stratum and $T_h$ as the subsample containing farms in h. The new replicate weights could then be used to compute 90 replicates of $\mathbf{t}$. Armed with these replicate estimates, one could them employ equation (B2) to estimate a variance matrix with h redefined as above and H replaced by 6.

Finally, note that in the spirit of the "better" DAG jackknife of equation (4), we could replace the stratified jackknife in equation (B2) with

$$\mathbf{v}_{SJ}(t) = \sum_{h=1}^{H} \frac{\sum_{S_g \cap T_h} w_k^2}{\sum_{S \cap T_h} (w_k - w_{k(hg)})^2} \sum_{g=1}^{15} (\mathbf{t} - \mathbf{t}_{(hg)})(\mathbf{t} - \mathbf{t}_{(hg)})'.$$

**References**


Kott, P.S. (1994). "Hypothesis Testing of Linear Regression Coefficients with Survey Data," *Survey Methodology*, 159-164.

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*, 521-526.

Kott, P.S. (2005a). "Randomization-Assisted Model-Based Survey Sampling," *Journal of Statistical Planning and Inference*, 263-277.

Kott, P.S. (2005b). "Using the Delete-a-Group Jackknife in an Economic Survey of US Farms," Presented at the 55th Session of the International Statistical Institute, Sydney, Australia (CD-Rom).

Kott, P.S. (2006a). "Delete-a-Group Variance Estimation for the General Regression Estimator under Poisson Sampling," *Journal of Official Statistics*, 759-767.

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