USING THE DELETE-A-GROUP JACKKNIFE VARIANCE ESTIMATOR IN NASS SURVEYS, by Phillip S. Kott, Research Division, National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, DC, March 1998, NASS Research Report, RD-98-01 (Revised July 2001).

ABSTRACT

The National Agricultural Statistics Service (NASS) plans to use estimation strategies of increasing complexity in the future and will need to estimate the variances resulting from those strategies. This report describes a relatively simple method of variance/mean squared error estimation, the delete-a-group jackknife, that can be used meaningfully in a remarkably broad range of settings employing complex estimation strategies. The text describes a number of applications of the method in abstract terms. It goes on to shows how the delete-a-group jackknife has been applied to some recent NASS surveys.

KEY WORDS

Calibrated weight; Multi-phase sampling; Poisson sampling; Ratio-adjusted weight; Replicate weight; Restricted regression; Systematic probability sampling.

ACKNOWLEDGEMENTS

The author would like to thank Susan Hicks, John Amrhein, Matt Fetter, Carl Scott, and Don McKinnie for their help in making the use of the delete-a-group jackknife at NASS a reality. Special thanks go to Bill Iwig, Tapan Nayak, and Susan Hicks (again) for constructive comments on earlier drafts of this report.

TABLE OF CONTENTS

ABSTRACT i

SUMMARY iii

INTRODUCTION 1

WHY USE THE DELETE-A-GROUP JACKKNIFE? 2

COMPUTING A DELETE-A-GROUP JACKKNIFE: AN OVERVIEW 3

PARTICULAR CASES (IGNORING FINITE POPULATION CORRECTION) 4
- *Stratified Simple Random Sampling* 4
- *Stratified Probability Sampling* 4
- *Restricted Regression (A Form of Calibration)* 5
- *Ratio-Adjusted Weights (Another Form of Calibration)* 6
- *NASS Applications of Ratio-Adjusted Weighting* 6
- *Composite Estimators* 7

SINGLE-PHASE POISSON SAMPLING AND FINITE POPULATION CORRECTION 8
- *A Calibrated Estimator for a Total* 8
- *An Estimator for a Ratio* 9
- *Some Explanations and Extensions* 9

SUMMARY OF NASS USES (SO FAR) 10

A DIGRESSION ON MODEL-BASED INFERENCE 11

CONCLUDING REMARKS 13

JULY 2001 UPDATE 13

REFERENCES 14

APPENDIX A: Justifying the Delete-a-Group Jackknife Under a Single-Phase, Stratified, Sampling Design 15
APPENDIX B: Justifying the Delete-a-Group Jackknife for a Restricted-Regression Estimator Under a Two-Phase Sample 17
APPENDIX C: Justifying the Delete-a-Group Jackknife for Certain Composite Estimators in the ARMS 19
APPENDIX D: Justifying the Delete-a-Group Jackknife with Finite Population Correction for a Single-Phase Poisson Sample 21

SUMMARY

Historically, NASS has employed mostly expansion estimators and ratios of expansion estimators based on stratified, simple random samples when computing indications of agricultural activity. This is changing due in no small part to increasing demands on the agency to make more efficient use of the information it collects. Fortunately, parallel increases in computing power are allowing NASS to use more sophisticated estimation strategies involving multi-phase sampling designs and calibration estimators. For example, the 1996 Agricultural Resource Management Study (ARMS) used a multi-phase sampling design (a first-phase sample is randomly drawn, then a second-phase subsample is randomly drawn from the first-phase sample, and so forth). Ratio adjustments of the initial inverse-probability sampling weights capture relevant information from the ARMS screening phase about farms *not* selected for a particular survey module.

This report shows how different variations of a delete-a-group jackknife can be used to estimate the variances (or, more precisely, the mean squared errors) of a variety of estimation strategies. Many of these are strategies are currently in use by NASS.

The delete-a-group jackknife is simple to use once appropriate replicate weights are constructed. By contrast, the “linearization” methods traditionally used by NASS for estimating variances can be exceedingly complicated and cumbersome when applied to complex estimators strategies. The advantage of the delete-a-group jackknife over the traditional, delete-one-primary-sampling-unit-at-a-time jackknife (see Rust 1985) is that the number of needed replicate weights per sample record is kept manageable.

A disadvantage of the delete-a-group jackknife over the delete-one jackknife is that it requires the first-phase stratum sample sizes to be large – at least five sample units per stratum. Otherwise, the delete-a-group jackknife will be overly conservative; that is, higher, on average, than the true variance it is measuring. As a result, when this jackknife is applied to estimators from the NASS area frame, it will be biased upward.

Like the delete-one jackknife, the delete-a-group jackknife is a nearly unbiased estimator of variance only when the first-phase sampling fractions are small – no more than 1/5 for most records. Otherwise, the delete-a-group jackknife tends to be biased upward. This bias is likely to be ignorable in most NASS applications. For the 1996 VCUS, however, it was so great that the delete-a-group jackknife has to be modified. A potential modification is discussed in the text. It is useful, but has a striking limitation: One set of replicate weights is needed when estimating the variances of totals and another when estimating the variances of ratios.

INTRODUCTION

This report addresses the construction of delete-a-group jackknife variance estimators for a variety of estimation strategies (an estimation strategy is a sampling design paired with an estimator). The emphasis will be on computational formulae, which will be rendered in fairly abstract form. Relevant theoretical comments will be made where appropriate, but most proofs are left for the appendices.

The sampling designs with which we will be dealing may have any number of phases. At each phase, one of the following selection schemes is assumed to be used:

1) stratified simple random sampling without replacement,

2) systematic probability sampling (usually called systematic probability proportional to size sampling; here we want to de-emphasize the "size" measure),

3) the converse of systematic probability sampling (what remains in a frame after a systematic probability sample has been removed), or

4) Poisson sampling (in which each element is given its own selection probability, and the sampling of one element has no impact on whether another gets selected).

All stratum samples are assumed to be large (contain at least five sampling units). Violation of this assumption in the first-phase of sampling can cause the delete-a-group jackknife to be biased upward. This is shown in Appendix A.

NASS currently incorporates two types of calibration in its estimators and does not plan to use any other types in the near future. "Calibration" is a general term for a sampling-weight adjustment that forces the estimates of certain item totals based on the sample at one phase of sampling to equal the same totals based on a previous phase or frame (control) data.

Ratio adjustments, the most common form of calibration, were used repeatedly in the 1996 Agricultural Resource Management Study (ARMS). Restricted regression, another population form of calibration, was used in both the 1997 Minnesota pilot Quarterly Agriculture Survey (QAS) and the second-phase of the 1996 Vegetable Chemical Use Survey (VCUS). Only these forms of calibration are discussed in the text.

Most of the results in this report are supported with randomization-based (design-based) analyses. As a consequence, all estimators of population parameters are assumed to be randomization consistent (i.e., have small randomization mean squared errors and even smaller randomization biases). A brief discussion of the model-based properties of the delete-a-group jackknife is reserved for a separate section.

The concise term "variance estimation" will be used throughout the text in place of the more cumbersome "mean squared error estimation." It should be understood, however, when the delete-a-group jackknife is a good estimator for the variance of a randomization-consistent estimator, it is also a good estimator for its mean squared error.

For our purposes, the term "nearly unbiased" will mean that the bias of the estimator in question is an ignorably small fraction of its mean squared error. The term "biased" will be used to mean "not (necessarily) nearly unbiased."

When first-phase stratum sample sizes are large, the delete-a-group jackknife is appropriate (has only a small potential for bias) whenever the conventional, randomization-based, delete-one jackknife is. Kott and Stukel (1997) have extended the use of the latter jackknife to two-phase estimators with calibration in the second phase. This report relies heavily on their results. Here, however, systematic probability can be used in the second design phase, even though Kott and Stukel only treated strategies featuring stratified simple random sampling in the second phase.

In most applications of the delete-a-group jackknife at NASS the need for finite population correction (fpc) is ignored. One section of the text discusses a number of those applications.

A subsequent section takes up variance estimation of totals and ratios when proper fpc is a concern. Strictly speaking, the variant of the delete-a-group jackknife that captures fpc requires single-phase Poisson sampling to be nearly unbiased. Nevertheless, the practical application can be broader, as we shall see.

WHY USE THE DELETE-A-GROUP JACKKNIFE?

The delete-a-group jackknife is simple to compute once appropriate replicate weights are constructed. The so-called "linearization" methods traditionally used by NASS for estimating variances can be very cumbersome when applied to estimators based on multi-phase designs like the 1996 Vegetable Chemical Use Survey (VCUS) 1998) and components of the 1996 ARMS (Kott and Fetter 1997). Estimators using calibrated weights based on restricted regression, like those calculated for the 1997 Minnesota pilot Quarterly Agriculture Survey (QAS), pose even greater practical problems for linearization variance methods (a multivariate regression coefficient would need to be estimated for every item of interest).

It is also a relatively simple matter to apply the delete-a-group jackknife to the composite estimators associated with the ARMS. With 1996 survey data, for example, results from the Phase II Corn Production Practices Report (PPR) were composited with results from the Phase II Corn-for-Grain Production Practices and Costs Report (PPCR). In addition, results from the Phase III Cost and Returns Report (CRR) stand-alone (based on respondents that were *not* in the Phase II PPCR sample) were composited with results from the Phase III CRR follow-on (based on respondents that were).

The advantage of the delete-a-group jackknife over the traditional, delete-one-primary-sampling-unit-at-a-time jackknife (see Rust 1985) is that the number of needed replicate weights per sample record is kept manageable. A common practice with the delete-one jackknife for handling this problem is to group primary sampling units (PSU's) into variance PSU's. This practice reduces the number of replicate weights needed per record – there is one for every variance PSU. Nevertheless, NASS would need at least 15 replicate weights per record to compute variances for state estimators. This would result in national variance estimates employing several hundred replicate weights per record.

COMPUTING A DELETE-A-GROUP JACKKNIFE: AN OVERVIEW

Suppose we have a sampling design with any number of phases and a randomization-consistent estimator, t, we wish to apply to the resultant sample. To compute a delete-a-group jackknife variance estimator for t, we first divide the first-phase sample – both respondents and non-respondents – into R (jackknife) groups. Currently, R is 15 in NASS applications. Consequently, we will assume R = 15 in the text. By setting R at 15, we lengthen the traditional, normality-based 95% confidence interval by ten percent. To see why this is so, observe that the ratio of the t-value at 0.975 for a Student's t distribution with 14 degrees of freedom and the normal z-value at 0.975 is approximately 1.1.

Suppose we have a survey which may have multiple phases. Let F be the sample selected at the first phase of the sampling process. The first-phase sample units may be composed of distinct elements (e.g., farms) or it may consist of clusters of elements (e.g., area segments). Many survey designs feature a single phase of sample selection.

The delete-a-group jackknife begins by dividing the first-phase sample F, into 15 groups. This can be done as follows: order F in an appropriate manner (discussed below); select the first, sixteenth, thirty-first, ... units for the first group; select the second, seventeenth, thirty-second, ..., units for the second group; continue until all 15 groups are created. Unless the number of units in F is divisible by 15 (which is unlikely), the groups will not all be of the same size.

Ordering in an "appropriate manner" depends on the context. If F was drawn using stratified random sampling, then order the sample so that units in the same stratum are listed together (i.e., contiguously). If samples were drawn using Poisson sampling, order the sample units randomly.

Let S denote the final respondent sample used to compute t, and let w_i denote the sampling weight for element i in S. The elements in S may be the same as the sample units in F or they may be a subsample of those units. The elements in S may also have a different nature than the original sample units in F; for example, they may be farms as opposed to area segments or fields as opposed to farms. In all such cases, however, each element in S must be contained within an original sample unit in F in a clearly defined way. Let e_i be the original sampling weight of the unit containing i (which may be i itself); that is, e_i is the inverse of the unit's first-phase probability of selection.

Let S_r denote that part of the final sample originating in first-phase sample units assigned to group r. The *jackknife replicate* $S_{(r)}$ is the whole final sample S with S_r removed. We similarly define $F_{(r)}$ as the set of first-phase sample units not in r.

We need to create 15 sets of *replicate weights* $\{w_{i(r)}\}$, one for each r, in the following manner: $w_{i(r)} = 0$ for all elements in S_r; for other elements, $w_{i(r)}$ will be close to $(15/14)w_i$ but adjusted to satisfy calibration constraints similar to those satisfied by w_i (exactly how to do this in a number of situations is the subject matter of the following section). *Observe that a $w_{i(r)}$-value has been assigned to every element in S including those in S_r.*

Now t is an estimate based on the sample S calculated using the set of weights, $\{w_i\}$. Let $t_{(r)}$ be the same estimate but with the member of $\{w_{i(r)}\}$ replacing $\{w_i\}$. The delete-a-group jackknife variance estimator for t is

$$v_J = (14/15) \sum^{15} (t_{(r)} - t)^2. \quad (1)$$

PARTICULAR CASES (IGNORING FINITE POPULATION CORRECTION)

In this section, we see how the delete-a-group jackknife can be fruitfully applied in a number of estimation strategies where fpc may be ignored; that is, when the first-phase selection probabilities are all small (say less than or equal to 1/5).

One sampling design *not* discussed in detail subsequently is stratified multi-*stage* sampling, in which subsampling within each primary (first-stage) sampling unit is conducted independently of subsampling in other primary sampling units. When the first stage of sampling has ignorably small selection probabilities, the conventional variance estimator for a stratified multi-stage sample looks exactly like that for a stratified single-stage cluster sample with estimated totals for primary sampling units used in place of actual values. As a result, when a delete-a-group jackknife is appropriate for an estimator based on a stratified single-stage sample, it is appropriate for an estimator based on a stratified multi-stage sample.

Stratified Simple Random Sampling

Suppose we have a single-phase stratified simple random sample without any nonresponse (handling nonresponse will be discussed later). The original and final sampling weight for a unit i in stratum h is $e_i = w_i = N_h / n_h$, where N_h is the population size of stratum h and n_h is its sample size.

Let us now consider the r'th set of replicate weights. For a unit i in $S_{(r)}$ and stratum h, $e_{i(r)} = (15/14)N_h / n_h$. By contrast, the appropriate final r'th replicate weight for unit i *recognizes the calibration equations inherent in the direct expansion estimator* (i.e., $N_h = \sum_{j\in S(r)\cap h} w_{j(r)}$ for all h). It is $w_{i(r)} = N_h / n_{h(r)} = (n_h / n_{h(r)}) e_i$, where $n_{h(r)}$ is the number of sample units in both $S_{(r)}$ and h. Observe that $e_{i(r)} = w_{i(r)}$ only when n_h is divisible by 15.

Stratified Systematic Probability Sampling

Suppose we have a single-phase, stratified systematic probability sample. The original and final sampling weight for a unit i in stratum h is $e_i = w_i = M_h / (n_h m_i)$, where m_i is the measure of size of unit i in stratum h, M_h is the sum of the m_j across all units in stratum h, and n_h is the stratum sample size.

Analogous to the simple random sampling case, the appropriate final r'th replicate weight for element i recognizes the calibration equations inherent in the Horvitz-Thompson expansion estimator (i.e., $M_h = \sum_{j\in S(r)\cap h} w_{j(r)} m_j$ for all h). It is $w_{i(r)} = (n_h / n_{h(r)}) e_i$, where $n_{h(r)}$ is the number of sample units in both $S_{(r)}$ and h.

Stratified simple random sampling can be viewed as equivalent to a special case of systematic probability sampling from randomly-order lists (one in which m_i is constant within strata). Appendix A provides some theoretical justification for using the delete-a-group jackknife as described above with a stratified, single-phase systematic probability sampling design under certain conditions. One of those conditions is that the systematic samples be drawn from *randomly*-ordered lists. Variance estimation can be problematic when systematic samples are drawn from *purposefully*-ordered lists.

Purposefully-ordered lists can reduce the variance in estimators based on systematic samples. Unfortunately, the reduction in variance due to a well-designed ordering usually can not be measured in an effective manner.

Restricted Regression
(A Form of Calibration)

There are many versions of restricted regression. Below is a description of a method similar to what was used in the 1996 VCUS and 1997 Minnesota pilot QAS. The version presented here will likely be used in the future.

Suppose, for exposition purposes, there are two sampling phases. Suppose further that the second phase sample is calibrated to a row vector of totals, η, based on estimates from the first-phase sample or determined from the frame itself.

Let f_j be the weight for element j after the first phase of sampling, and let p_j be the element's selection probability in the second sampling phase. In the absence of non-response (again, nonresponse will be dealt with later) in the second sampling phase, a general form of the calibrated weight for j under restricted regression is

$$w_j = f_j/p_j + (\eta^* - \textstyle\sum_{i \in S^*} [f_i/p_i]\mathbf{x}_i) (\sum_{i \in S^*} [f_i/p_i]\mathbf{x}_i'\mathbf{x}_i)^{-1} [f_j/p_j]\mathbf{x}_j' \qquad (2)$$

for $i \in S^*$, and a predetermined value otherwise (chosen so that w_j is not too small or too far from f_j/p_j), where S is the second-phase sample, S* a subset containing almost all the elements of S, $\mathbf{x}_i$ is a row vector of covariates whose sum across all elements in the population is either η or has been previously estimated to be η – that is, $\eta = \sum_F f_i\mathbf{x}_i$, where F denotes the elements in the first-phase sample; finally, $\eta^* = \eta - \sum_{S-S^*} w_i\mathbf{x}_i$.

Let $f_{j(r)}$ be the r'th jackknife replicate weight for unit j after the first sampling phase. The r'th jackknife replicate weight for element j is 0 when $j \in S_r$; otherwise, it is

$$w_{j(r)} = w_j[f_{j(r)}/f_j] + (\eta_{(r)} - \textstyle\sum_{i \in S(r)} w_i[f_{i(r)}/f_i]\mathbf{x}_i) (\sum_{i \in S(r)} w_i[f_{i(r)}/f_i]\mathbf{x}_i'\mathbf{x}_i)^{-1} w_j [f_{j(r)}/f_j]\mathbf{x}_j', \qquad (3)$$

where $\eta_{(r)} = \eta$ when η has been determined from frame; $\eta_{(r)} = \sum_F f_{i(r)}\mathbf{x}_i$ when η has been estimated from the first-phase sample.

Equation (3) is *not* the standard way to construct jackknife replicate weights. The expression $w_k[f_{k(r)}/f_k]$ has been used in place of the more common $f_{k(r)}/p_k$, with which it is nearly equal (because $w_k \approx f_k/p_k$). Equation (3)'s strength is that it forces the replicate weights (for elements not in group r) to be fairly close to the associated calibrated weights. This appears to reduce the upward bias that unexpected differences between the two can cause. It should be noted that any such upward bias is small; in fact, it is asymptotically ignorable. We live, however, in a finite world.

Restricted-regression as described above can be done at any phase of sampling. At the t'th phase, f_i in equation (2) becomes the weight for element i at the t-1'th phase and p_i the element's conditional selection probability at the t'th phase. For a single-phase restricted-regression estimator, we can set all $p_i = 1$ in equation (2).

When the phase of sampling calibrated in this manner contains more than a single stratum, the jackknife can have an upward bias (see Appendix B). In addition, for a single-phase Poisson sample, $\mathbf{x}_i\lambda = 1$ must hold for some λ (see the section on Poisson sampling and Appendix D).

Ratio-Adjusted Weights (Another Form of Calibration)

Consider, again, a two-phase sample with f_i and p_i as above. A very common form of calibration occurs when a vector of covariates for element i, $\mathbf{x}_i$, is defined in such a way that only one component of the vector is non-zero for each i. That is to say, the elements are categorized into G mutually exclusive calibration (or ratio-adjustment) groups, and $x_{ig} > 0$ only when element i is in group g; otherwise, $x_{ig} = 0$.

Under that structure, a ratio-adjusted weight for an element j in group g is

$$w_j = \eta_g \left(\sum_{i \in S} [f_i / p_i] x_{ig} \right)^{-1} [f_j / p_j], \qquad (4)$$

and $\eta = (\eta_1, ..., \eta_G)$. Similarly, the corresponding replicate weight is 0 for $j \in S_r$, and

$$w_{j(r)} = \eta_{g(r)} \left(\sum_{i \in S(r)} f_{i(r)} / p_i] x_{ig} \right)^{-1} [f_{j(r)} / p_j] \qquad (5)$$

otherwise, where $\eta_{(r)} = (\eta_{1(r)}, ..., \eta_{G(r)})$.

If the second-phase sample is stratified, and more than one of these strata are contained within a calibration group, then the jackknife can have an upward bias (see Appendix B). When the second-phase sample is unstratified or the second-phase strata and ratio-adjustment groups coincide, the delete-a-group jackknife is nearly unbiased. In the 1996 ARMS and 1996 VCUS, second-(and later-)phase sampling was unstratified.

Extensions of these results to estimation strategies with $t > 2$ phases are straight-forward; the f_i in equation (4) and $f_{i(r)}$ in equation (5) become the weight and replicate weight at the t-1'th phase. For a single-phase sample, we can set all the p_i equal to 1 in both equations (4) and (5).

The establishment of the appropriateness of the delete-a-group jackknife for ratio-adjusted estimators parallels that of restricted-regression estimators, which is outlined in Appendix B.

NASS Applications of Ratio-Adjusted Weighting

One way to handle nonresponse is to treat the set of responding elements (at any phase of the design) as a stratified simple random subsample of the selected sample. This was essentially what was done in the first-phase of the 1996 VCUS. All the original sample elements (respondents and nonrespondents) were assigned to jackknife replicates, and nonresponse was treated as a second phase of sampling. The "second-phase" strata and calibration groups coincided with the original stratum definitions, and x_{ig} was set equal to 1 when i was in group g (0, otherwise). Since f_i was equal for all i in the same stratum, and p_i was likewise identical for each respondent i in the stratum, w_i in equation (4) collapsed to the population size in the stratum containing i divided by the respondent sample size in that stratum. Equation (5) collapsed similarly.

In the 1996 ARMS, a stratified simple random screening sample of farms was subsampled sequentially for several mutually exclusive survey modules (see Kott and Fetter 1997). Farms were selected for the Phase II Soybean PPR in Nebraska, for example, using an additional five phases of sampling (to be selected for this module, a respondent farm from the screening sampled had to avoid being subsampled for one of the four modules preceding it). Each of these phases employed unstratified systematic probability sampling from a purposefully-ordered list (the theory in Appendix B is assumes a randomly-ordered list; if anything, purposeful ordering should reduce mean squared errors and contribute an upward bias to the delete-a-group jackknife). Finally, a field was randomly selected from each sampled soybean farm.

The separate-ratio estimator in equation (4) was used twice in Phase II Soybean PPR estimates. It was used to ratio adjust the weights for the screening-sample respondents to the frame total-value-of-sales within every screening stratum (notice that response/nonresponse on the screening survey is treated here implicitly as another phase of sampling). In addition, the soybean field sample was divided into three size groups. Here, η_g was the total soybean acres in calibration (size) group g as estimated from the screening sample with the weights described above, and p_i was the product of the six (conditional) probabilities: the probabilities that the farm containing field i was *not* selected for the four modules preceding soybeans, the probability that this farm was selected for the production practices module, and the probability that field i was subsampled from the farm.

We treated the fields from which we collected Phase II PPR information as if they were a stratified simple random subsample of the selected fields, where the three calibration groups served as strata. This had no practical effect on the calculation of the p_i (observe that if all the p_i in a group are multiplied by the same factor, the computed weights in equations (4) and (5) are unchanged).

Composite Estimators

Consider a set of C distinct samples, each of which can be used to estimate a common target value. Let S denote the combined sample, and $w_i^{(c)}$ denote the weight for element i in original sample c. If i is not in sample c, set $w_i^{(c)} = 0$. A composite estimator t uses the set of weights $\{w_i\}$, where each $w_i = \sum^C \lambda_c w_i^{(c)}$ and $\sum \lambda_c = 1$.

To estimate the variance of t, we can create 15 sets of replicate weights for every $w_i^{(c)}$ and denote each by $\{w_{i(r)}^{(c)}\}$. We then estimate the $t_{(r)}$ using $w_{i(r)} = \sum^C \lambda_c w_{i(r)}^{(c)}$ and compute v_J using equation (1).

Composite estimation was used, for example, to combine the Phase III Beef and Corn-for-Grain CRR follow-on samples in the 1996 ARMS with the Phase III CRR stand-alone sample. First the two enterprise CRR samples were composited and then this combined sample was composited with the other CRR sample (see Kott and Fetter 1997).

Samples being combined need *not* correspond to identical target populations. For example, the population of list farms with corn for grain in 1996 is not the same as the population of list farms with ten weaned calves in 1996 (the Beef CRR population). When combining CRR samples, we also combine target populations; in this case, to the set of all list farms with *either* grain corn or ten weaned calves in 1996. Only those sample farms having both corn for grain and at least 10 weaned calves are assigned composite weights as described above. Other farms in the combined sample retain their pre-composite weights.

Appendix C shows why the delete-a-group jackknife works for the composite estimators used in the ARMS in which the components were separate modules based on the same screening sampling. Composite estimation was also used in the ARMS to combine independently drawn samples like the Phase II Soybean PPR sample and the National Resource Inventory sample. Here, like a conventional jackknife, when the delete-a-group jackknife is appropriate for each independent component, it is also appropriate for any linear combination of the components.

SINGLE-PHASE POISSON SAMPLING AND FINITE POPULATION CORRECTION

In this section, we restrict our attention – at first – to a single-phase Poisson sample of elements. Let π_j be the selection probability of element j. We assume there is no nonresponse.

The versions of the delete-a-group jackknife developed in this section *will* contain finite population corrections. The versions are different for an estimator of a total and the estimator of a ratio. This is a reflection of the fact that *a simple formula like equation (1) does NOT work for all smooth transformations of calibrated expansion estimators when finite population correction is an issue* (note: a "smooth" transformation has first, second, and third derivatives; most statistics of interest are smooth transformations of expansion estimations, the major exception being percentiles).

A Calibrated Estimator for a Total

Suppose we have a calibrated estimator for a total, $t = \sum_S w_j y_j$, where

$$w_j = 1/\pi_j + (\eta^* - \textstyle\sum_{i \in S^*} [1/\pi_i]\mathbf{x}_i)(\sum_{i \in S^*} [1/\pi_i]\mathbf{x}_i'\mathbf{x}_i)^{-1} [1/\pi_j]\mathbf{x}_j' \qquad (6)$$

for $j \in S^*$, and a predetermined value otherwise (chosen so that $w_j \geq 1$ and, perhaps, not too far from $1/\pi_j$), S is the sample, S* a subset containing almost all the elements of S, $\mathbf{x}_i$ is a row vector of covariates whose sum across all elements in the population is η, and $\eta^* = \eta - \sum_{S-S^*} w_i\mathbf{x}_i$. There must also be a vector λ such that $\mathbf{x}_j\lambda = \sqrt{(1 - \pi_j)}$ for all j (that is to say, either a component of $\mathbf{x}_j$ or a linear combination of components must equal $\sqrt{(1 - \pi_j)}$). Since we are dealing with a single-phase sample, (6) is simply equation (2) with $1/\pi_k$ replacing f_k / p_k (i.e., f_k in equation (2) is 1, while p_k is π_k).

To estimate the variance of t, we use equation (1) but replace t with $t^{(v)} = \sum_S w_j^{(v)} y_j$, and $t_{(r)}$ with $t_{(r)}^{(v)} = \sum_S w_{j(r)}^{(v)} y_j$, where

$$w_j^{(v)} = w_j\sqrt{(1 - 1/w_j)}, \qquad (7)$$

and

$$w_{j(r)}^{(v)} = w_j^{(v)}\{1 + (\textstyle\sum_S w_i^{(v)}\mathbf{x}_i - \sum_{S(r)} w_i^{(v)}\mathbf{x}_i)(\sum_{S(r)} w_i^{(v)}\mathbf{x}_i'\mathbf{x}_i)^{-1}\mathbf{x}_j'\} \qquad (8)$$

when $j \in S_{(r)}$ and 0 otherwise. Appendix D outlines why this works.

Observe that $w_j^{(v)} \approx w_j\sqrt{(1 - \pi_j)}$, so that $w_j^{(v)} \approx w_j$ when the selection probability for element j is ignorably small. When all element selection probabilities are very small, there is little difference between this delete-a-group jackknife for a total estimator with finite population correction, $v_{J(fpcT)}$, and the standard delete-a-group jackknife, v_J. Moreover, the rather odd assumption that there exists a λ such that $\mathbf{x}_j\lambda = \sqrt{(1 - \pi_j)}$ becomes close to the more standard assumption that either a component of $\mathbf{x}_j$ or a linear combination of components is a constant (i.e., $\mathbf{x}_j\lambda = 1$ for some λ).

In fact, if we were to ignore finite population correction (which we can do for most surveys, but not VCUS), we could estimate the variance of t with equation (1), replacing equation (8) with

$$w_{j(r)} = w_j\{1 + (\sum_S w_i\mathbf{x}_i - \sum_{S(r)} w_i\mathbf{x}_i)(\sum_{S(r)} w_i\mathbf{x}_i'\mathbf{x}_i)^{-1}\mathbf{x}_j'\} \qquad (8')$$

when $j \in S_{(r)}$ and 0 otherwise as long as $\mathbf{x}_j\lambda = 1$ for some λ. This is what we did for the 1997 Minnesota pilot QAS (see Bailey and Kott 1977).

An Estimator for a Ratio

Suppose t_R is an estimator for a ratio of the form, $t_R = \sum_S w_jy_j / \sum_S w_jz_j$, where w_j is calibrated as above. One can estimate the variance of t with

$$v_{J(fpcR)} = (\sum_S w_j^{(v)}z_j / \sum_S w_jz_j)^2 (14/15)\sum^{15} (t_{R(r)}^{(v)} - t_R^{(v)})^2, \qquad (9)$$

where $t_R^{(v)} = \sum_S w_j^{(v)}y_j / \sum_S w_j^{(v)}z_j$, and $t_{R(r)}^{(v)} = \sum_S w_{j(r)}^{(v)}y_j / \sum_S w_{j(r)}^{(v)}z_j$. This assumes $\mathbf{x}_j\lambda = \sqrt{(1 - \pi_j)}$ for some λ. Even without this assumption holding, in fact, even without calibration, $v_{J(fpcR)}$ will likely be a reasonable variance estimator; as we shall see.

Alternatively, we could estimate the variance of t_R ignoring finite population correction with equation (1). We need not assume that $\mathbf{x}_j\lambda = 1$ for some λ**.** *In fact, the w_i need not even be calibrated in this case* (to see why, observe that multiplying all the weights in t_R by a fixed constant so that $\sum_S w_j$ equals the population size has no effect on the estimator; consequently, all ratio estimators are effectively calibrated on $x_j = 1$).

Some Explanations and Extensions

Consider a single-phase element sample that is not necessarily Poisson. Suppose we wish to estimate the variance of $t = \sum_S w_jy_j$, where the w_j satisfy equation (6). Let $u_k = y_k - \mathbf{x}_k(\sum_U \mathbf{x}_i'\mathbf{x}_i)^{-1}\sum_U \mathbf{x}_i'y_i$, where U denotes the population. The variance of t is approximately

$$V = \sum_U u_k^2(1 - \pi_k)/\pi_k + \sum_{U(k \neq i)} u_ku_i(\pi_{ki} - \pi_k\pi_i)/(\pi_k\pi_i). \qquad (10)$$

Under Poisson sampling the joint selection probability of k and i, π_{ki}, is equal to the product $\pi_k\pi_i$, and so V collapses to $\sum_U u_k^2(1 - \pi_k)/\pi_k$. This can, in principle, be estimated by $\sum_S u_k^2(1 - \pi_k)/\pi_k^2$, which is approximately equal to $\sum_S[w_ku_k\sqrt{(1 - 1/w_k)}]^2$, which is what $v_{J(fpcT)}$ estimates (see Appendix D).

A similar argument can be made for the ratio estimator, $t_R = \sum_S w_jy_j/\sum_S w_jz_j$, except that now V becomes approximately

$$V^* = (\sum_S w_jz_j)^{-2}[\sum_U (u_k^*)^2(1 - \pi_k)/\pi_k + \sum_{U(k \neq i)} u_k^*u_i^*(\pi_{ki} - \pi_k\pi_i)/(\pi_k\pi_i)],$$

where

$$u_k^* = u_k^+ - \mathbf{x}_k(\sum_P \mathbf{x}_i'\mathbf{x}_i)^{-1}\sum_P \mathbf{x}_i'u_i^+, \text{ and}$$

$$u_k^+ = y_k - (\sum_U y_i / \sum_U z_i)\, z_k.$$

Under Poisson sampling, V^* collapses to $(\sum_S w_jz_j)^{-2}[\sum_U (u_k^*)^2(1 - \pi_k)/\pi_k]$. If we tried to compute a delete-a-group jackknife with equation (1) replacing t by $t_R^{(v)}$ and $t_{(r)}$ by $t_{R(r)}^{(v)}$, we would get a reasonable estimator for $(\sum_S w_j^{(v)}z_j)^{-2}[\sum_U (u_k^*)^2(1 - \pi_k)/\pi_k]$ rather than V^*, hence the factor $(\sum_S w_j^{(v)}z_j / \sum_S w_jz_j)^2$ on the right hand side of equation (9).

This factor is unnecessary if finite population correction is ignored. In fact, since $\sum_U u_i^+ = 0$ (simplifying the proof in Appendix D), the weights need not be calibrated for the delete-a-group jackknife variance estimator for t_R to be nearly unbiased.

Calibrated estimators of totals were computed in the 1997 Minnesota pilot QAS. Sampling was not exactly Poisson due to the need to combine some samples and subsample others (see Bailey and Kott 1997). Nevertheless, it is not unreasonable to assume that $\sum_{U(k\neq i)} u_k u_i(\pi_{ki} - \pi_k\pi_i)/(\pi_k\pi_i)$ in the right hand side of equation (10) is roughly zero and then – ignoring finite population correction – use v_J to estimate variances.

It is of interest to note that for systematic probability sampling from an *purposefully-ordered* list, π_{ki} will often be zero when i and k are listed sequentially in the ordering. If u_k and u_i in equation (10) tend to have the same sign when i and k are listed together, then it is likely that $\sum_{U(k\neq i)} u_k u_i(\pi_{ki} - \pi_k\pi_i)/(\pi_k\pi_i)$ will be negative – reducing the variance of t. The delete-a-group jackknife does not capture this variance-reducing phenomenon, however. That is why it was claimed earlier that the use of systematic unequal probability sampling from an ordered list will, if anything, bias the delete-a-group jackknife upward. This presupposes that elements (or units) listed together in the ordering are in some sense similar.

Remember the delete-a-group jackknife for an estimated total with finite population correction, $v_{J(fpcT)}$, is only appropriate when there is no nonresponse. Still, computing $v_{J(fpcT)}$ and v_J using imputed values in place of real ones can provide a means of evaluating the impact of high selection probabilities on variance. There is one additional caveat. When one does not require there exists a λ such that $\mathbf{x}_j\lambda = \downarrow(1 - \pi_j)$ for all j, then $v_{J(fpcT)}$ may be biased downward. This possibility is likely to be remote in practice (see Appendix D).

We could have used $v_{J(fpcR)}$ to estimate variances from the 1996 VCUS. In theory, this might not be appropriate since the calibration in that survey was to first-phase totals rather than to control totals. Moreover, we did not require that there be a vector λ such that $\mathbf{x}_j\lambda = \downarrow(1 - \pi_j)$ for all j. It is unlikely that either failing would cause much bias in $v_{J(fpcR)}$. This is because calibration does little to reduce the variance of t in the VCUS. Moreover, it is likely that $\sum_S w_i^{(v)}u_i^*$ (or $\sum_S w_i^{(v)}u_i^+$ when $v_{J(fpcR)}$ is used with an un-calibrated VCUS estimator) is close to zero even when there is no λ such that $\mathbf{x}_j\lambda = \downarrow(1 - \pi_j)$. Appendix D explains why $\sum_S w_i^{(v)}u_i^*$ must be near zero.

SUMMARY OF NASS USES (SO FAR)

The delete-a-group jackknife was used to estimate variances for the 1996 ARMS, 1996 VCUS, and 1997 Minnesota pilot QAS. It has also been used in some of NASS's foreign consulting work, but that is beyond the scope of this report.

Bailey and Kott (1997) describes the sample design used in the Minnesota pilot QAS. Since NASS imputes for nonresponse on the QAS, there was essentially a single-phase sample in Minnesota. Equation (2), with all p_j set equal to 1 and f_j computed as described in Bailey and Kott, was used to generate most of the calibration weights. The vector $\mathbf{x}_i$ had 20 components including a constant term.

When a w_j calculated with equation (2) would have been less than 1, farm j was removed from S*, and w_j was set equal to 1. Sampled farms were randomly assigned to jackknife groups, and replicate weights were calculated using the more conventional

$$
\begin{aligned}
w_{j(r)} &= [15/14]f_j + (\eta_{(r)} - \textstyle\sum_{i\in S(r)} [15/14]f_i\mathbf{x}_i)(\textstyle\sum_{i\in S(r)} [15/14]f_i\mathbf{x}_i'\mathbf{x}_i)^{-1}[15/14]f_j\mathbf{x}_j' \\
&= f_j + (\eta_{(r)} - \textstyle\sum_{i\in S(r)} f_i\mathbf{x}_i)(\textstyle\sum_{i\in S(r)} f_i\mathbf{x}_i'\mathbf{x}_i)^{-1} f_j\mathbf{x}_j'
\end{aligned}
$$

instead of equation (3). This was because the advantages of using the latter equation was not clear at the time.

The 1996 VCUS had a two-phase design. The first phase was stratified simple random sampling. Sampled units were randomly assigned to jackknife groups within first-phase strata. Nonresponse to the first design phase of the VCUS was treated as an additional phase of stratified simple random sampling where the strata were the same as the first-phase.

The second design phase of the VCUS used systematic unequal probability sampling. Nonresponse to this phase was treated as an additional phase of simple random sample. Equation (2) was used to compute calibrated weights. The "first-phase" weight, f_j, was actually the population size of the first-phase stratum containing i divided by the number of first-phase usables in the stratum; η was a vector of estimated planted-acre totals for in-scope vegetables based on the first-phase sample adjusted for nonresponse; p_i was the second-phase probability of selection multiplied by the number of second-phase usables and divided by the number of second-phase sample farms. Replicate weights were computed using

$$w_{j(r)} = w_j + (\eta_{(r)} - \sum_{i\in S(r)} w_i \mathbf{x}_i) (\sum_{i\in S(r)} w_i \mathbf{x}_i' \mathbf{x}_i)^{-1} w_j \mathbf{x}_j',$$

which turned out to have slightly better empirical properties (less negative values; values closer to w_j) in this context than those produced by equation (3) for some reason .

The many uses of ratio-adjustment and composite estimation in list-based estimates from the 1996 ARMS are discussed thoroughly in Kott and Fetter (1997). The original screening sample was randomly allocated into jackknife groups on a stratum-by-stratum basis. The text provides some details for a couple of examples. See Kott and Fetter for more.

The delete-a-group jackknife was also used for the non-overlap (area) portion of the 1996 ARMS for economic statistics. The area design had effectively three-phases: 1) a stratified simple random sample of area segments, 2) a restratified simple random subsample of farms; and 3) a stratified (using the first-phase strata) simple random subsample of respondents. Using the delete-a-group jackknife in this context treats the three-phase sample as if it were a three-*stage* sample. As a result, the variance estimator can be biased upward (see Kott 1990). The problem here is that the second-phase sample is not calibrated in any way.

There is an additional source of upward bias in the delete-a-group jackknife applied to the 1996 ARMS non-overlap sample. Some area substrata have very small samples sizes (less than five areas segments). Collapsing substrata into land-use strata helped some, but on occasion even land-use strata had small sample sizes. Appendix A shows why this can cause a bias in the delete-a-group jackknife.

A DIGRESSION ON MODEL-BASED INFERENCE

The delete-a-group jackknife can be applied to estimate variances in a reasonable fashion under a variety of complex estimation strategies. Both the text so far and the appendices rely exclusively on the principles of randomization-based inference. As a result of this, we were forced to assume two number of questionable or erroneous assumptions:

1) systematic probability sampling is conducted by NASS from randomly-ordered lists, and

2) farm in the same ratio-adjustment (calibration) group are equally likely to respond to a survey.

These assumptions would not be necessary if we replace them by the model assumptions behind calibration; namely;

$$y_i = \mathbf{x}_i\beta + \epsilon_i,$$

where the ϵ_i have zero mean, bounded variances, and are uncorrelated – at least across first-phase sampling units.

For example, consider the difference between the calibration estimator,
$t = \sum_S w_i y_i$, and its target, $T = \sum_U y_i$:

$$t - T = \sum_S w_i y_i - \sum_U y_i = \sum_S w_i \epsilon_i - \sum_U \epsilon_i = \sum_U (w_i I_i - 1)\epsilon_i,$$

where $I_i = 1$ when $i \in S$, and $I_i = 0$ otherwise. Now

$$(t - T)^2 = \sum_U (w_i I_i - 1)^2 \epsilon_i^2 + \sum_{U(i \neq k)} (w_i I_i - 1)(w_k I_k - 1)\epsilon_i \epsilon_k.$$

If ϵ_i and ϵ_k are uncorrelated, the model variance of t as an estimator for T is

$$E_\epsilon[(t - T)^2] = \sum_U (w_i^2 I_i - 2w_i I_i + 1)E(\epsilon_i^2) = \sum_S (w_i^2 - w_i)E(\epsilon_i^2) - [\sum_S w_i E(\epsilon_i^2) - \sum_U E(\epsilon_i^2)]$$

no matter what the sampling design. Moreover, if $E(\epsilon_i^2) = \mathbf{x}_i\gamma$ for some γ, then the model variance of t as an estimator for T collapses further to $\sum_S (w_i^2 - w_i)E(\epsilon_i^2)$, which is what the delete-a-group jackknife for a total *with* finite population correction estimates. (Note: even if $E(\epsilon_i^2)$ does not equal $\mathbf{x}_i\gamma$ for some γ, we know that $\sum_S w_i E(\epsilon_i^2) \approx \sum_S (1/\pi_i)E(\epsilon_i^2) \approx \sum_U E(\epsilon_i^2)$ for randomization-based reasons).

Similar arguments can be made for calibration in the second (or later) phase of sampling. Kott (1997) contains a treatment of this topic for the conventional delete-one-primary-sampling-unit jackknife. The interested reader may also want to look at the expression for Var_{2d} in Appendix B and replace each u_k with ϵ_k. Similar sub-stitutions can profitably be made in Appendices C and D as well.

In the real world, models fail, which is the reason NASS insists on using randomization-consistent estimators where possible. The impact of model failure is typically greater on bias than on variance. This is because model failure is usually small and subtle but can nonetheless lead to a bias in a non-randomization-consistent estimator that is not asymptotically ignorable. Once the potential for asymptotic bias is removed by using a randomization-consistent estimator, a model can often be safely invoked to estimate variance.

The situation can be reversed when ratio-adjustment is used (in part) to handle nonresponse as in the 1996 ARMS and VCUS. The model assumption that the expected value for an unknown y-value is a fixed multiple of a known x-value within a ratio-adjustment group is usually more reasonable than the quasi-randomization assumption that all farms in the group are equally likely to be survey respondents. In this situation, the assumption of the linear model offers some protection against a systematic bias in an estimated value due to the failure of the quasi-randomization assumption.

CONCLUDING REMARKS

The delete-a-group jackknife is remarkably simple to compute once appropriate replicate weights are determined. We have seen how this variance estimation method can be meaningfully applied to a number of complex estimation strategies. These include the 1996 ARMS (with multiple phases and ratio adjustments), the 1997 Minnesota pilot QAS (restricted regression and Poisson sampling), and the 1996 VCUS (two phases, calibration of the second phase to the first, and finite population correction problems).

Like any variance estimator, the delete-a-group jackknife is not necessarily nearly unbiased when any phase of the sample is drawn systematically from a *purposefully*-ordered list (as is the case in latter phases of the ARMS). If anything, however, the delete-a-group jackknife will usually be conservative (biased upward) in this circumstance. In addition, the delete-a-group jackknife requires the following to be nearly unbiased:

1) results from each phase of a survey – including the first phase – be calibrated for some key items of interest on results from either an earlier same phase or the frame (for example, the estimated number of farm names on the list frame is often forced to equal the actual number of farm names on the list frame); and

2) the sample size of at every follow-on phase of a non-nested (not multi-stage) multi-phase design be large (contain at least five sample units per stratum at that phase).

These are not difficult requirements, and NASS need keep them in mind when developing estimators in the future.

A disadvantage of the delete-a-group jackknife over potential competitors is that it requires the first-phase stratum sample sizes to be large (at least five sample units per stratum). Otherwise, the delete-a-group jackknife can be overly conservative. As a result, when this jackknife is applied to estimators from the NASS area frame – as it was with the non-overlap component of the 1996 Phase III CRR, it has an upward bias. NASS needs to assess how big a problem this constitutes in practice.

JULY 2001 UPDATE

NASS made the Minnesota QAS pilot operational in all states in 2000. A slightly different form of restricted regression is used. Variances are estimated ignoring finite population correction. For more details, see Kott and Bailey (2000).

REFERENCES

Bailey, J.T. and Kott, P.S. (1997). An Application of Multiple List Frame Sampling for Multi-Purpose Surveys. *ASA Proceedings of the Section on Survey Research Methods*, 496-500.

Harley, H.O., and Rao, J.N.K. (1962). Sampling with Unequal Probabilities and Without Replacement. *The Annals of Mathematical Statistics*, 350-74.

Kott, P.S. (1990). Variance Estimation when a First Phase Area Sample is Restratified. *Survey Methodology*, 99-104.

Kott, P.S. (1997). A (Partially) Model-based Look at Jackknife Variance Estimation with Two-Phase Samples, *Statistical Society of Canada Proceeding of the Survey Methods Section*, 37-45.

Kott, P.S. and Bailey J.T. (2000). The Theory and Practice of Maximal Brewer Selection with Poisson PRN Sampling. *ICES II: The Second International Conference on Establishment Surveys*, 269-78.

Kott, P.S. and Fetter, M. (1999). A Multi-Phase Design to Co-ordinate Surveys and Limit Response Burden. *ASA Proceedings of the Section on Survey Research Methods*, 33-7.

Kott, P.S. and Stukel, D.M. (1997). Can the Jackknife Be Used With a Two-Phase Sample? *Survey Methodology*, 81-9.

Rust, Keith (1985). Variance Estimation for Complex Estimators in Sample Surveys. *Journal of Official Statistics*, 1, 381-397.

APPENDIX A: Justifying the Delete-a-Group Jackknife Under a Single-Phase, Stratified Sampling Design

Suppose we have a probability sample design with H strata and n_h sampled units within each stratum h. Let us assume that the sample was selected without replacement but the selection probabilities are all so small, and the joint selection probabilities are such, that using the with-replacement variance estimator is appropriate (this rules out systematic sampling from a purposefully-ordered lists). In particular, let us assume that the estimator itself can be written in the form:

$$t = \sum_{h=1}^{H} \sum_{j=1}^{n_h} t_{hj}.$$

Let $q_{hj} = t_{hj} - \sum^{n_h} t_{hg}/n_h$. The randomization variance of t is $Var(t) = \sum^{H} Var(\sum t_{h+})$, where $t_{h+} = \sum_j t_{hj}$. Now $Var(t_{h+})$ can be estimated in a (nearly) unbiased fashion by

$$var(t_{h+}) = (n_h/[n_h - 1]) \sum_{j=1}^{n_h} q_{hj}^{2}$$

("nearly" because we are ignoring finite population correction).

In order to estimate Var(t) with a delete-a-group jackknife as suggested in the text, we first order the strata in some fashion and then order the units within each stratum randomly. The sample is partitioned into R (i.e., 15) systematic samples using the resulting ordered list. Let S_r denote one such systematic sample, S_{hr} the set of n_{hr} units in both S_r and stratum h, and $S_{h(r)}$ the set of $n_{h(r)}$ units in stratum h and *not* in r.

The jackknife replicate estimator $t_{(r)}$ is

$$t_{(r)} = \sum_{h=1}^{H} (n_h/n_{h(r)}) \sum_{j \in S_{h(r)}} t_{hj}.$$

Now

$$t_{(r)} - t = \sum_{h=1}^{H} [(n_h/n_{h(r)}) \sum_{j \in S_{h(r)}} t_{hj} - t_{h+}].$$

Treating each $S_{h(r)}$ as a simple random subsample of the sample in stratum h, we have

$$E_2[(t_{(r)} - t)^2] = \sum^{H} Var_2([n_h/n_{h(r)}] \sum_{j \in S_{h(r)}} t_{hj})$$

$$= \sum^{H} (n_h^2/n_{h(r)})[1 - (n_{h(r)}/n_h)] \sum^{n_h} q_{hj}^2/(n_h - 1)$$
$$= \sum^{H} (n_h/[n_h - 1])(n_{hr}/n_{h(r)}) \sum^{n_h} q_{hj}^2$$
$$= \sum^{H} (n_{hr}/n_{h(r)}) \text{ var}(t_{h+}),$$

where E_2 denotes expectation with respect to the subsampling.

Observe that for strata where $n_h < R$, $n_{hr}/n_{h(r)}$ is either zero because there are no units in both r and h or $n_{hr}/n_{h(r)}$ is $1/(n_h - 1)$ because there is one unit in both r and h. Since the latter situation occurs in exactly n_h replicates, $\sum^R n_{hr}/n_{h(r)} = n_h/(n_h - 1)$.

For strata where $n_h \geq R$, $n_{hr}/n_{h(r)} = O(1/R)$ and $\sum^R n_{hr}/n_{h(r)} \approx 1 + O(1/R)$. (Technical note: $z = O(1/R)$ means $\lim_{R\to\infty} R|z|$ is a constant). In fact, when n_h/R is an integer, $n_{hr}/n_{h(r)}$ exactly equals $1/(R-1)$, and $\sum^R n_{hr}/n_{h(r)} = R/[R-1]$.

Since Var(t) can itself be estimated in an approximately unbiased fashion by var(t) = $\sum^H (n_h/[n_h - 1]) \sum_j q_{hj}^2$, it is not difficult to see that the delete-a-group jackknife variance estimator, $v_J = ([R-1]/R) \sum^R (t_{(r)} - t)^2$ is approximately unbiased for var(t) and thus for Var(t) when all strata are such that $n_h \geq R$ and is biased upward otherwise. Moreover, the relative upward bias is bounded by $([R-1]/R)\min_h \{1/(n_h - 1)\}$.

APPENDIX B: Justifying the Delete-a-Group Jackknife for a Restricted-Regression Estimator Under a Two-Phase Sample (Including When the "Second" Phase is a String of Phases)

Consider the estimator, $t = \sum_S w_i y_i$, where w_i have the same definition as in equation (2), and y_i is a value of interest for element i. For ease of exposition, we will let f_i be the inverse of the first-phase selection probability of element i, which we assume to be large for all i. It is a simple matter to use induction to cover the situation where f_i is itself the result of several phases and calibrations.

We also assume that neither phase sample is Poisson. Poisson sampling in the first phase is discussed in Appendix D. Poisson sampling in a later phase is identical to an additional *stage* of sampling because Poisson sampling is independent from one selection to the next.

Let $\mathbf{B} = (\sum_U \mathbf{x}_i'\mathbf{x}_i)^{-1} \sum_U \mathbf{x}_i' y_i$, where U is the set of all elements in the population, and $u_i = y_i - \mathbf{x}_i\mathbf{B}$. Equation (2) allows us to rewrite t as $t = \eta\mathbf{B} + \sum_S w_i u_i$, and the variance of t as $\text{Var}(t) \approx \text{Var}(\eta\mathbf{B}) + \text{Var}(\sum_S w_i u_i) + 2\text{Cov}(\eta\mathbf{B}, \sum_S w_i u_i)$. Now η has no variance when it comes from the frame, and Var(t) collapses to $\text{Var}(\sum_S w_i u_i)$.

Let O and o define asymptotic orders ($z = O(m)$ means $\lim_{m\to\infty} |z|/m$ is a constant; $z = o(m)$ means $\lim_{m\to\infty} |z|/m = 0$). We assume that equation (2) holds for almost all elements in the sample (i.e., it fails at most $o_P(m)$ times, where m is the size of S). As a result, $\sum_S w_i u_i \approx \sum_S (f_i/p_i)u_i$ under mild conditions, we assume to hold (this is because, treating each (f_i/p_i) as O(1), $(\eta^* - \sum_{i\in S^*} [f_i/p_i]\mathbf{x}_i)(\sum_{i\in S^*} [f_i/p_i]\mathbf{x}_i'\mathbf{x}_i)^{-1}\sum_{S^*}[f_j/p_j]\mathbf{x}_j'u_j = O_P(\sqrt{m})O_P(1/m)O_P(\sqrt{m})$ is ignorably small compared to $t = O_P(m)$; note that the equality $\sum_U \mathbf{x}_j'u_j = 0$ has a role in making this contention viable). Thus, $\text{Var}(\sum_S w_i u_i)$ is approximately the variance of a double-expansion estimator, $\sum_S (f_i/p_i)u_i$. Assuming the second-phase samples within each second-phase stratum are large, results in Kott (1990, p. 103) show the single-phase variance estimator with estimated primary sampling unit values put in place of actual values will over-estimate the variance of a double expansion estimator unless the sum of the $f_i u_i$ in the second-phase strata before subsampling are equal to zero (note that since both m_d and n_h in equation (B) of Kott are large, only $e_{d..}^2$ matters).

Kott assumed stratified simple random sampling in both phases, but extensions to stratified systematic probability sampling from randomly-ordered lists are straight-forward. For the first-phase sample all the f_i must be large (as in the simple random sampling case), so that the with-replacement expression for variance is appropriate. For the second-phase sample, the selection probabilities and population must be such that the approximation $p_{ik} \approx (m_d - 1)p_i p_k/m_d$ holds (see Hartley and Rao 1962), where p_{ik} is the second-phase joint selection probability of two elements, i and k, from second-phase stratum d, and m_d is the number of sampled elements in that stratum. In most NASS applications, the second-phase of selection is unstratified, which is equivalent to d being all the elements in the first-phase sample.

The second-phase variance of $\sum_S w_i u_i$ originating from second-phase stratum d can be expressed by (we are assuming m_d is large)

$$
\begin{aligned}
Var_{2d} &= \sum (f_i u_i)^2(1-p_i)/p_i + \sum (f_i u_i)(f_k u_k)(p_{ik} - p_i p_k)/(p_i p_k) \\
&\approx \sum (f_i u_i)^2(1-p_i)/p_i - \sum (f_i u_i)(f_k u_k) p_i p_k / m_d \\
&= \sum (f_i u_i)^2(1 - [(m_d - 1)/m_d] p_i)/p_i - (\sum f_i u_i)^2/m_d \\
&\approx \sum (f_i u_i)^2(1-p_i)/p_i - (\sum f_i u_i)^2/m_d
\end{aligned}
$$

where the summations are over all elements in second-phase stratum d *before* the second-phase of sampling takes place. This is analogous to equation (3) in Kott.

Note that $Var_{2d} \le \sum (f_i u_i)^2(1-p_i)/p_i$, where equality holds only when $\sum f_i u_i = 0$. This turns out to be the reason (not directly proven here) why the delete-a-group jackknife over-estimates the variance of t when the sum of the $f_i u_i$ within all second-phase strata are not approximately zero. Observe that given *any* column vector λ of the same dimension as $\mathbf{x}_i'$, $\sum_F \lambda'\mathbf{x}_i' f_i u_i \approx \sum_U \lambda'\mathbf{x}_i' u_i = \sum_U \lambda'\mathbf{x}_i'[y_i - \mathbf{x}_i(\sum_U \mathbf{x}_i'\mathbf{x}_i)^{-1} \sum_U \mathbf{x}_i' y_i] = 0$. Since $\sum_F \lambda'\mathbf{x}_i' f_i u_i \approx 0$ for any λ, when there exists a λ_d for every second-phase stratum d such that $\mathbf{x}_i \lambda_d = \lambda'_d \mathbf{x}_i'$ equals 1 when i is in d and 0 otherwise, then the sum of the $f_i u_i$ in any second-phase strata before subsampling is approximately (asymptotically) zero.

Applying the weights defined by equation (3) to t, we get $t_{(r)} \approx \eta_{(r)}\mathbf{B} + \sum_{S(r)} [f_{j(r)}/f_j] w_j u_j \approx \sum_{S(r)} [f_{j(r)}/p_j] u_j$. The second part of the first s approximation makes use of the facts that the components of $(\eta_{(r)} - \sum_{i \in S(r)} [f_{j(r)}/f_j] w_i \mathbf{x}_i)$ are $O_P(m/R)$, while the diagonal components of $\sum_{i \in S(r)} [f_{i(r)}/f_i] w_i \mathbf{x}_i' \mathbf{x}_i$ are $O_P(m)$ under mild conditions.

When $\eta = \eta_{(r)}$ is a frame value, $t_{(r)} - t \approx \sum_{S(r)} [f_{j(r)}/p_j] u_j - \sum_S [f_j/p_j] u_j$. It is straight-forward to show that the delete-a-group jackknife estimates $Var(t) = Var(\sum_S w_i u_i) \approx Var(\sum_S [f_j/p_j] u_j)$ fairly well with the possibility of being upwardly biased when the sum of the $f_i u_i$ before subsampling in one or more of the second-phase strata is not equal to zero.

When $\eta = \sum_F f_i \mathbf{x}_i$, then $\eta\mathbf{B} = \sum_F f_i \mathbf{x}_i \mathbf{B}$, and t can be rewritten as $t = \sum_F f_i y_i + (\sum_S w_i u_i - \sum_F f_i u_i) \approx \sum_F f_i y_i + (\sum_S [f_i/p_i] u_i - \sum_F f_i u_i) = \sum_F f_i (y_i + \{[I_i/p_i] - 1\} u_i)$, where $I_i = 1$ when i is in S and zero otherwise . In a similar fashion, $t_{(r)} \approx \sum_F f_{i(r)} (y_i + \{[I_i/p_i] - 1\} u_i)$. It is straight-forward to show that the delete-a-group jackknife estimates the conventional multi-*stage* variance estimator ignoring fpc at the first stage, which in turn estimates Var(t) fairly well but has the possibility of being upwardly biased when the sum of the $f_i u_i$ before subsampling in one or more of the second-phase strata is not equal to zero (see Kott and Stukel [1997] for some missing details).

Extension of the above result to a sample design where the "second-phase" sample is itself the result of a string of phases, all within the same second-phase strata, is a simple matter. We need only assume that $p_{ik}^* \approx \alpha p_i^* p_k^*$ where p_j^* (p_{ik}^*) denotes the appropriate product of conditional (joint) selection probabilities, and $\alpha = 1 - O(1/m_d)$. Appendix C has more on the sequence-of-sampling-phases methodology used in the ARMS design.

APPENDIX C: Justifying the Delete-a-Group Jackknife for Certain Composite Estimators in the ARMS

We restrict attention in this appendix to the unusual composite estimators used with the ARMS surveys.

To show that the delete-a-group jackknife works for a composite estimator like the Phase III Corn-for-Grain/Beef CRR described in the text, one needs to show that it works when estimating a total for:

1) the intersection of the two original target populations (list farms with grain corn and at least 10 weaned calves),

2) each of the two "rump" populations that contain elements in one population but not the other, and,

3) the union of the two rumps and the intersection.

The delete-a-group jackknife works for estimates of a rump total because it works for domains (by defining item values within one target population as zero for farms outside the domain), and it works for estimated totals in the union – assuming it works for estimated totals in the intersection – because it works for functions of estimators (like the sum of the totals in the two rumps and the intersection). We discuss intersections below.

Let us call the two samples we are compositing A and B. In principle, we can estimate an item total in the intersection of the target population using either sample. Let $t^C = \sum_C w_i^C y_i$ be the estimated total calculated using sample C (= A or B), and let $t = \lambda t^A + (1-\lambda)t^B$ be the composite total. Note the y_i is defined to be zero for farms outside the intersection.

The ARMS samples are drawn sequentially to avoid overlap using an unstratified variant of systematic unequal probability sampling at every phase after the initial screening phase. Let π_i^t be the probability of selecting farm i for sample t given that it is available for sampling after sample t-1 is drawn. Let t=1 denote the first sample drawn after the screening sample. Finally, let $p_i^t = (1 - \pi_i^1) \cdots (1 - \pi_i^{t-1})\pi_i^t$. Note that $\pi_i^s = 0$ when farm i is not in the target population for sample s. Without loss of generality, we will assume sample A was selected before B, and that A, B, and their intersection are of size O(m).

Using arguments similar to those in the previous appendix, we can see that

$$t^C \approx \sum_F f_i y_i + (\sum_C [f_i/p_i^C]u_i^C - \sum_F f_i u_i^C),$$

where $u_i^C = y_i - (\sum y_k/\sum x_k^C)x_i^C$, the summations being over the farms in the population that are in the same calibration group as i when computing t^C, and x_k^C is the x-value of farm k when computing t^C. Observe that the first-phase sample applies to both A and B since there is one ARMS screening sample for all purposes. By contrast x and u-values as well as calibration group memberships may differ across samples for the same i. This happens when the 1996 Phase II Corn-for-Grain PPCR is composited with the Phase II Corn PPR; x_i^A is corn-for-grain acres for farm i, while x_i^B is general corn acres.

We can now see that

$$\text{Var}(t) = \text{Var}_1(\textstyle\sum_F f_i y_i) + E_1\{\text{Var}_2(\sum_A [f_i /p_i^A]u_i^A) + \text{Var}_2(\sum_B [f_i /p_i^B]u_i^B) + \\ 2\text{Cov}_2(\sum_A [f_i /p_i^A]u_i^A, \sum_B [f_i /p_i^B]u_i^B)\}.$$

Now

$$\text{Cov}_2(\textstyle\sum_A [f_i /p_i^A]u_i^A, \sum_B [f_i /p_i^B]u_i^B) = \sum_{i\in F(A)} \sum_{k\in F(B)} (f_i u_i^A)(f_k u_k^B)\{[(p_i^A{}_k^B)/(p_i^A p_k^B)] - 1\},$$

where F(C) denotes that part of the first-phase sample in the target population for sample C, and $p_i^A{}_k^B = \pi_{i*k*}^{1} \cdots \pi_{i*k*}^{A-1}\pi_{ik*}^{A}(1 - \pi_k^{A+1}) \cdots (1 - \pi_k^{B-1})\pi_k$, when π_{i*k*}^{t} is the probability of selecting neither farm i nor k for sample t providing both are available after sample t-1, and π_{ik*}^{A} is the probability of selecting farm i but not k for sample A given that both are available after A-1. We will assume that the design is such that given $k\neq i$, $p_i^A{}_k^B \approx (1 + \alpha_k)p_i^A p_k^B$, where $\alpha_k = O(1/m)$ (if $\pi_{i*j*}^{s}/\pi_{i*}^{s}\pi_{j*}^{s} \approx 1$ for all s < A, then the assumption is equivalent to $\pi_{ik*}^{A} = \pi_{k*}^{A}$ Prob(i chosen for A|k not chosen for A) $= \pi_{k*}^{A}\pi_i^A(1 + \alpha_k)$). Since $\sum_{F(B)} f_i u_i^B \approx 0$ and $p_i^A{}_i^B = 0$, summing the left hand size of the last expression for Cov_2 over i yields a term of order 1/m. Summing then over k yields a term of order 1. Since $\text{Var}_1(t^A)$ is O(m), the covariance term is asymptotically ignorable.

It is now not hard to show using arguments developed here and in the previous appendices that the delete-a-group jackknife is unbiased for t under conditions we assume to hold.

APPENDIX D: Justifying the Delete-a-Group Jackknife with Finite Population Correction for a Single-Phase Poisson Sample

Suppose we have a calibrated estimator for a total, $t = \sum_S w_j y_j$, where the w_j satisfy equation (6):

$$w_j = 1/\pi_j + (\eta^* - \sum_{i\in S^*} [1/\pi_i]\mathbf{x}_i)(\sum_{i\in S^*} [1/\pi_i]\mathbf{x}_i'\mathbf{x}_i)^{-1} [1/\pi_j]\mathbf{x}_j' \qquad (6)$$

for $j \in S^*$, and a predetermined value otherwise. In addition, there is a vector λ such that $\mathbf{x}_j\lambda = \sqrt{(1 - \pi_j)}$ for all j.

Let $\mathbf{B} = (\sum_U \mathbf{x}_i'\mathbf{x}_i)^{-1} \sum_U \mathbf{x}_i'y_i$, where U is the set of all elements in the population, and $u_i = y_i - \mathbf{x}_i\mathbf{B}$. Now $t = \sum_S w_j y_j = \sum_S w_j(\mathbf{x}_j\mathbf{B} + u_j) = \eta + \sum_S w_j u_j$. Consequently, $Var(t) = Var(\sum_S w_j u_j) \approx Var(\sum_S u_j/\pi_j)$ under conditions we assume to hold (see Appendix B).

If the sample is Poisson, we have $Var(t) \approx \sum_U u_j^2(1 - \pi_j)/\pi_j$, which, in principle, can be estimated in a nearly unbiased fashion by $var(t) = \sum_S u_j^2(1 - \pi_j)/\pi_j^2 \approx \sum_S w_j^2(1 - 1/w_j)u_j^2 = \sum_S (w_j^{(v)})^2 u_j^2$, where $w_j^{(v)} = w_j \sqrt{(1 - 1/w_j)}$ (see equation (7)).

Using the definition of $w_{j(r)}^{(v)}$ in equation (8), we have $t_{(r)}^{(v)} = \sum_S w_{j(r)}^{(v)} y_j \approx \sum_S w_j^{(v)}\mathbf{x}_j\mathbf{B} + \sum_{S(r)} w_{jr}^{(v)}u_j$. Consequently, $t_{(r)}^{(v)} - t^{(v)} = \sum_{S(r)} w_{jr}^{(v)}u_j - \sum_S w_j^{(v)}u_j$. Now $\sum_S w_i^{(v)}u_i/\sqrt{n} = \sum_U (1 - 1/w_i)^{1/2}u_i/\sqrt{n} \approx \sum_U (1 - \pi_i)^{1/2}u_i \sqrt{n} = \sum_U \lambda'\mathbf{x}_i'u_i/\sqrt{n} = 0$, since $\lambda'\mathbf{x}_j' = \mathbf{x}_j\lambda = \sqrt{(1 - \pi_j)}$ for some λ, while $\sum_U \mathbf{x}_i'u_i = 0$. Similarly, $\sum_S w_j^{(v)}u_j/\sqrt{n} \approx 0$.

The the $n_{(r)}$ members of $S_{(r)}$ can be viewed as a simple random subsample of the n members of S. Since $n/n_{(r)} = 1 + O_P(1/R)$, $t_{(r)}^{(v)} - t^{(v)} \approx (n/n_{(r)})\sum_{S(r)} w_{j(r)}^{(v)}u_j$. Using arguments similar to ones made in Appendix A, we have $E_2[(t_{(r)}^{(v)} - t^{(v)})^2] \approx (n/n_{(r)})(1 - n_{(r)}/n)[\sum_S (w_j^{(v)})^2u_j^2 - (\sum_S w_j^{(v)}u_j)^2/n]$. So, $E_2[(t_{(r)}^{(v)} - t^{(v)})^2] \approx (n/n_{(r)})(1 - n_{(r)}/n) \sum_S (w_j^{(v)})^2u_j^2 = (n_r/n_{(r)}) \sum_S (w_j^{(v)})^2u_j^2$, where n_r is the size of S_r. From here it is easy to see that $v_{J(fpcT)}$ is nearly unbiased for $\sum_S (w_j^{(v)})^2u_j^2$, which is turn is nearly unbiased for Var(t).

Observe that when $\sum_U (1 - \pi_i)^{1/2}u_i \neq 0$, the jackknife is biased, although the bias depends on the size of the π_i. In practice, this may be of little importance because if we felt that $\sum_U (1 - \pi_i)^{1/2}u_i/\sqrt{n}$ had an absolute value far from zero, we would include $\sqrt{(1 - \pi_j)}$ as a component of $\mathbf{x}_j$.