EVALUATING ALTERNATIVE CALIBRATION SCHEMES FOR AN ECONOMIC SURVEY WITH LARGE NONREPONSE

Chadd Crouse and Phil. Kott
National Agricultural Statistics Service

KEY WORDS: Nonresponse; Quasi-randomization model; Coverage adjustment; Delete-a-group jackknife.

1. Introduction

Government statistical agencies conduct a number of surveys covering the same population. Inevitably estimates for what appears to be the same parameter varies from survey to survey. This problem can be exacerbated when an agency, like the National Agricultural Statistics Service (NASS), publishes what it considers official values for some aggregates.

The Cost and Returns Report (CRR) is one part of the multi-faceted Agricultural Report and Management Study (ARMS). It is the principal tool of the US Department of Agriculture (USDA) for determining the economic conditions under which farms operated in a given year. NASS conducts the survey, but the microdata is shared with and extensively analyzed by its sister USDA agency, the Economic Research Service (ERS).

Although based on a probability survey, the CRR is subject to a large amount of nonresponse by government standards. Moreover, because the questionnaires it uses are relatively long and administered via personal interview, CRR sample sizes are smaller than those of other NASS surveys. For these two reasons, smaller sample sizes and larger nonresponse rates, CRR samples (and the samples of its predecessor, the Farm Costs and Returns Survey) have historically been reweighted to match official NASS numbers of farms by region within five size classes. Beginning with the 2002 CRR, a much more ambitious system of reweighting was used.

This paper begins by describing how the NASS used truncated linear calibration to reweight the 2002 CRR and the delete-a-group (d-a-g) jackknife to estimate the variances of the resulting estimators. It goes on to describe how d-a-g jackknife methodology was employed to investigate, *i*, the effects of replacing the (last phase of the) current nonresponse-adjustment scheme with linear calibration, *ii*, the effects of accounting for the area-frame component of the CRR with linear calibration of the list-frame sample only, *iii*, the effects of using a calibration routine based on a more reasonable log function in place of a linear one.

Section 2 describes the CRR survey itself and the calibration routines used with the 2002 CRR. Section 3 provides some theory for calibration weighting and the delete-a-group jackknife. Section 4 highlights the results of our investigations, while Section 5 contains a discussion.

2. Calibrating the 2002 CRR

2.1 The Cost and Returns Report

The 2002 had both a list and area component. A clustered area sample was used to capture that portion of the farm population not on the NASS list of farms. This is referred to as the nonoverlap sample.

The list sample was the result of a multiphase selection process. Every year ERS chooses from between one to three agricultural commodities on which to focus a Production Practices and Cost Report (PPCR). In 2002, soybeans was the PPCR crop. CRR sample selection began with a stratified screening sample. The screening instrument was used to determine whether the selected farm was in scope and whether it had planted soybeans. One subsample of the screening sample was chosen to be in the PPCR. A second distinct subsample was chosen to be part of the CRR-proper. The respondent PPCR and CRR-proper samples were then composited and added to the area-based nonoverlap sample to form the entire CRR sample.

Samples were selected by state, but statistics were estimated by region. Table 1 shows the respondent sample and estimated population size by region and typology. Of the 12,391 farms in the respondent sample, slightly less than 10% (1,180) came from the non-overlap sample, while slightly more than 15% (2,027) came from the PPCR.

The regions are defined on the bottom of table. Typology is a designation of particular interest to ERS. The typologies are used in estimation but not in sampling. They played a central role in our empirical analyses and are defined below.

Limited-resource farms (*Typology 1*). Any small farm with: (1) gross sales less than $100,000, (2) total farm assets less than $150,000, and (3) total operator household income less than $20,000. Limited-resource farmers may report farming, a nonfarm occupation, or retirement as their major occupation.

Retirement farms (*Typology 2*). Small farms whose operators report they are retired (excludes limited-resource farms operated by retired farmers).

Residential/lifestyle farms (*Typology 3*). Small farms whose operators report they had a major occupation other than farming (excludes limited-resource farms with operators reporting a nonfarm major occupation).

Farming occupation/low-sales (*Typology 4*). Small farms with sales less than $100,000 whose operators report farming as their major occupation (excludes limited-resource farms whose operators report farming as their major occupation).

Farming occupation/high-sales (*Typology 5*). Small farms with sales between $100,000 and $249,999 whose operators report farming as their major occupation.

Large family farms (*Typology 6*). Farms with sales between $250,000 and $499,999.

Very large family farms (*Typology 7*). Farms with sales of $500,000 or more.

Nonfamily farms (*Typology 8*). Farms organized as nonfamily corporations or cooperatives, as well as farms operated by hired managers.

2.2 Calibration

After adjusting the sample for nonresponse at the state level, compositing the PPCR and CRR-proper samples, and adding the nonoverlap component, NASS reweighted each CRR farm so that the weighted totals for 16 target variables within each region equaled official values determined outside of the CRR. The US sums for these target are displayed in Table 2, as are the estimated US totals before calibration.

Since we wanted to investigate whether calibration can be used to adjust for nonresponse (in the last phase of sampling) and/or in place of the nonoverlap portion of the sample, totals are also given for estimates computed without nonresponse adjustment, without the NOL sample's contribution, and without either. Percentages are calculated with respect to the targets.

We were interested in assessing whether calibration could replace the present method of nonresponse adjustment because that method, based on reweighting within size classes, required determining and loading a size value for every sampled farms, respondent and nonrespondent, and assuring that respondent samples within classes were adequately large. Not only is this process cumbersome, it is likely inadequate.

Calibration may be both simpler operationally and theoretically preferable, since it allows nonresponse to be a function of more variables than size, although the linear form of calibration is awkward.

Using calibration rather than the nonoverlap sample to account for farms missing from NASS's list sampling frame of farms has two perceived advantages. One, the usually-small farms from the area sample are often in the CRR sample for five years, the length of their rotation in the area sample. Moreover, they are enumerated as part of many other NASS nonoverlap samples. This is deemed a particularly unfair burden. Two, since the sampling weights of nonoverlap farms are large, there is concern that their presence in the CRR sample may cause unfortunate increases in the variance of estimates that count each farm equally.

Truncated linear calibration was used to meet the 16 targets. Calibrated weights were not allowed to go below 1, because it was felt each respondent should at least represent itself. A generalization of the algorithm used is described in the next section.

Section 3. Some Theory

3.1 Calibration Weighting

Linear calibration weights can be put in the form:

$$w_{i(reg)} = a_i(1 + \mathbf{z}_i\mathbf{h}), \qquad (1)$$

where $\{a_i\}$ is the set of original weights – the inverses of the element selection probabilities, perhaps after some adjustments, $\mathbf{z}_i = (z_{i1}, ..., z_{iP})$ is a row P-vector of instrument variables with $z_{i1} = 1$,

$$\mathbf{h} = (\textstyle\sum_S a_k\mathbf{z}_k'\mathbf{x}_k)^{-1}(\sum_U \mathbf{x}_k - \sum_S a_k\mathbf{x}_k)',$$

$\mathbf{x}_i$ is a row P-vector of control variables, S is the sample, and U is the population. See, for example, Estevao and Särndal (2002). In most applications, including the 2002 CRR, $\mathbf{z}_i \equiv \mathbf{x}_i$.

When there is no nonresponse or coverage errors, the goal of calibration is mean-squared-error (mse) reduction. Calibration assures weighted sample balance on the calibration variables. It forces $\sum_S w_{i(reg)}\mathbf{x}_i$ to equal $\sum_U \mathbf{x}_i$, where before the two vectors were only close in some sense.

Although not originally designed for it, linear calibration can also be used to adjust for nonresponse (or frame-undercoverage) bias. In that context, we are implicitly assuming that the probability of i responding (or being on the sampling frame) is

$$p_i \approx a_i/w_{i(reg)} = 1/(1 + \mathbf{z}_i\mathbf{h}). \qquad (2)$$

The near equality in equation (2) comes from

$w_{i(reg)} \approx a_i/p_i$.

Treating nonresponse (or undercoverage) as another phase of random sampling is called "quasi-randomization modeling." Fuller et al. (1994) explicitly use a truncated form of equation (2) to model response to a one-day food-intake survey.

Quasi-randomization also can be used to model duplications in the frame or previous over-adjustments for nonresponse by letting p_i denote the expected number of times i is on the (previously-adjusted) sampling frame. Within such a framework, the calibration in the 2002 CRR respondent sample can be thought of as implicitly correcting biases in the previous phases of nonresponse adjustment.

A more general form for the model in equation (2) is

$$p_i = \begin{cases} 1/f(\mathbf{z}_i\lambda) & \text{when } b_L \le f(\mathbf{z}_i\lambda) \le b_U \\ 1/f(b_L) & \text{when } f(\mathbf{z}_i\lambda) < b_L \\ 1/f(b_U) & \text{when } f(\mathbf{z}_i\lambda) > b_U \end{cases} \qquad (3)$$

where (for convenience) f is monotonic and twice differentiable within $b_L \le f(\mathbf{z}_i\lambda) \le b_U$. Equation (3) leads to *general* calibration weights of the form:

$$w_{i(gen)} = \begin{cases} a_i f(\mathbf{z}_i\mathbf{h}) & \text{when } b_L \le f(\mathbf{z}_i\mathbf{h}) \le b_U \\ a_i f(b_L) & \text{when } f(\mathbf{z}_i\mathbf{h}) < b_L \\ a_i f(b_U) & \text{when } f(\mathbf{z}_i\mathbf{h}) > b_U \end{cases} \qquad (4)$$

where the vector $\mathbf{h}$ is chosen so that the calibration equation, $\sum_U \mathbf{x}_i = \sum_S w_{i(gen)}\mathbf{x}_i$, holds.

The following iterative process often finds a solution when one exists. Let

$$\mathbf{h}^{(r+1)} = \mathbf{h}^{(r)} + [\sum_S c_i^{(r)} a_i f_1(\mathbf{z}_i\mathbf{h}^{(r)})\, \mathbf{z}_i'\mathbf{x}_i]^{-1} (\sum_U \mathbf{x}_i - \sum_S w_{i(gen)}^{(r)} \mathbf{x}_i)',$$

where $f_1(y) = \partial f(y)/\partial y$,

$w_{i(gen)}^{(r)} = a_i f(\mathbf{z}_i\mathbf{h}^{(r)})$ and $c_i^{(r)} = 1$ when $b_L \le f(\mathbf{z}_i\mathbf{h}^{(r)}) \le b_U$

$w_{i(gen)}^{(r)} = a_i f(b_L)$ and $c_i^{(r)} = 0$ when $f(\mathbf{z}_i\mathbf{h}^{(r)}) < b_L$

$w_{i(gen)}^{(r)} = a_i f(b_U)$ and $c_i^{(r)} = 0$ when $f(\mathbf{z}_i\mathbf{h}^{(r)}) > b_U$,

$c_i^{(0)} = 1$ and $\mathbf{h}^{(0)} = \mathbf{0}$.

Continue until an R is reached such that

$$|(\sum_U x_{ip} - \sum_S w_{i(gen)}^{(R)} x_{ip}) / \sum_U x_{ip}| \le .001 \quad \text{for all p},$$

then set $w_{i(gen)} = w_{i(gen)}^{(R)}$, $c_i = c_i^{(R)}$, = and $\mathbf{h} = \mathbf{h}^{(R)}$.

It is possible to allow b_L and b_U in equations (3) and (4) to vary across elements, but we did not do that here for simplicity. In actual practice, NASS used f(q) = q, and set b_U at ∞ and $b_{L(i)}$ at 1/a.

In linear calibration, we have $b_L = -\infty$, $b_U = \infty$, and f(y) = 1 + y, so that $f_1(y) = 1$. No iteration is necessary. In *log calibration*, $b_L = 0$, $b_U = \infty$, and f(y) = exp(y), so $f_1(y) = \exp(y)$. The name derives from the equality log{f(y)} = y. Also, when $p_i = 1/f(\mathbf{z}_i\mathbf{h})$, $\log(p_i) = -\mathbf{z}_i\mathbf{h}$. Another name for this calibration method is "generalized raking" since it effectively generalizes the popular raking (iterative proportional fitting) routine.

Log calibration has conceptual appeal, because p is a nonnegative function. Many view logistic calibration, where $p_i = [1 + \exp(\mathbf{z}_i\lambda)]^{-1}$, as even more appealing, since p_i is contained within (0, 1). Unfortunately, logistic calibration has the practical limitation that f(.) can not be less than 1, rendering much calibration infeasible.

3.2. Mean Squared Errors

In this subsection we focus on mse estimation under a quasi-random model for nonresponse. Mse estimation under a quasi-random model for coverage error is similar. For simplicity, we assume no pre-calibration adjustment of the weights, one phase of response modeling, and no bounds (b_L or b_U).

Suppose we are estimating $T = \sum_U y_i$ with $t = \sum_S w_i y_i$, where $w_i = a_i f(\mathbf{z}_i\mathbf{h})$, a_i is the inverse of the selection probability for unit i, $1/f(\mathbf{z}_i\lambda)$ is the Poisson probability that i is a respondent, f(.) is a known, monotonic, and twice differentiable function, and $\mathbf{h}$ is chosen so that $\sum_U \mathbf{x}_i = \sum_S w_i\mathbf{x}_i$. Observe that under mild conditions we assume to hold

$$\begin{aligned} t - T &= \sum_S w_i y_i - \sum_U y_i \\ &= \sum_S w_i u_i - \sum_U u_i \\ &= \sum_S a_i f(\mathbf{z}_i\mathbf{h}) u_i - \sum_U u_i \\ &= \sum_S a_i f(\mathbf{z}_i\lambda) u_i + \sum_S a_i f_1(\theta_i)(\mathbf{h} - \lambda)'\mathbf{z}_i' u_i - \sum_U u_i \\ &\approx \sum_S a_i f(\mathbf{z}_i\lambda) u_i - \sum_U u_i, \end{aligned} \qquad (5)$$

where $u_i = y_i - \mathbf{x}_i(\sum_U f_1(\mathbf{z}_k\lambda)\mathbf{z}_k'\mathbf{x}_k]^{-1} \sum_U f_1(\mathbf{z}_k\lambda)\mathbf{z}_k' y_k$, and θ_i is between $\mathbf{z}_i\lambda$ and $\mathbf{z}_i\mathbf{h}$.

From equation (5), it is easy to see that the quasi-randomization mse of t is approximately equal to the randomization mse of $\sum_S a_i f(\mathbf{z}_i\lambda) u_i$. To estimate the mse of t, we need to replace $f(\mathbf{z}_i\lambda)$ by $f(\mathbf{z}_i\mathbf{h})$ and u_i by its sample analogue:

$$e_i = y_i - \mathbf{x}_i(\sum_S a_k f_1(\mathbf{z}_k\mathbf{h})\mathbf{z}_k'\mathbf{x}_k]^{-1} \sum_S a_k f_1(\mathbf{z}_k\mathbf{h})\mathbf{z}_k' y_k.$$

3.3 The Delete-a-Group Jackknife

NASS creates a set of 15 jackknife replicate weights for every farm in the 2002 CRR respondent sample so that delete-a-group jackknife mean-squared-error (mse) methodology could be used. Farms selected for the screening sample were sorted by stratum and then systematically assigned to one of 15 replicate groups. Sampled area segments (clusters of area-sample farms) were likewise assigned to the 15 groups. A replicate is the complement of a group.

Let S_g denote one of the 15 replicate groups (g = 1, ..., 15), and $S_{(g)} = S - S_g$, one of the 15 replicates. Let $\{w_i\}$ denote the set of calibration weights. The replicate weight we used in the analysis for each $k \in S_g$ is $w_{k(g)} = 0$; for each $k \in S_{(g)}$, it was

$$w_{k(g)} = (15/14)w_k + \Big(\sum_{i \in U} \mathbf{x}_i - \sum_{i \in S_{(g)}} [15/14]w_i\mathbf{x}_i \Big) \times \Big[\sum_{i \in S_{(g)}} f_1(\mathbf{z}_i\mathbf{h})c_i a_i \mathbf{z}_i'\mathbf{x}_i \Big]^{-1} f_1(\mathbf{z}_k\mathbf{h})\, c_k a_k \mathbf{z}_k' \qquad (6)$$

The formula for a replicate weight in equation (6) ignores adjustments made to a_i before calibration. When such adjustments were made, and one has replicate weights of the form $a_{k(g)}$ attached to each k, a better formula for the g'th replicate weight for $k \in S$ would be

$$w_{k(g)} = (a_{k(g)}/a_k)w_k + \Big(\sum_{i \in U} \mathbf{x}_i - \sum_{i \in S} [a_{i(g)}/a_i]w_i\mathbf{x}_i \Big) \times \Big[\sum_{i \in S} f_1(\mathbf{z}_i\mathbf{h})c_i a_{i(g)} \mathbf{z}_i'\mathbf{x}_i \Big]^{-1} f_1(\mathbf{z}_k\mathbf{h})c_k a_{k(g)}\mathbf{z}_k' \qquad (7)$$

(Note that with u_k are defined for equation (5), $\sum_S w_{k(g)}u_k \approx \sum_S a_{k(g)}(w_k/a_k)e_k$; see Kott, 2004, for a deeper discussion). This collapses into equation (6) when $a_{k(g)}/a_k = 15/14$ when $k \in S_{(g)}$ and zero otherwise.

We used the suboptimal replicate formula for operational reasons.

3.3 Comparing Two Statistics Based on Different Sets of Weights

Suppose we want to compare two calibrated estimates of a total:

$t = \sum_S w_k y_k$ and $t' = \sum_S w_k' y_k$,

where S is the combined (if necessary) sample with, for example, w_k (w_k') set equal to zero when k is not in the sample used to compute t (t'). For now, we treat the two sets of calibration weights as equally valid.

Given d-a-g jackknife replicate weights for each set of calibration weights, a
z-score for the difference between t and t' is

$$z = (t - t')/\surd v, \qquad (8)$$

where

$$v = [(R - 1)/R] \sum^R (d_{(r)} - d)^2, \qquad (9)$$
$$d = t - t' \qquad (10)$$
$$d_{(r)} = t_{(r)} - t_{(r)}' \qquad (11)$$
$$t_{(r)} = \sum_S w_{k(r)}y_k, \qquad (12)$$
$$t_{(r)}' = \sum_S w_{k(r)}'y_k, \qquad (13)$$

and $w_{k(r)}$ ($w_{k(r)}'$) refers to the rth jackknife replicate weight associated with w_k (w_k'). Under the null hypothesis, z has a Student's t distribution with 14 degrees of freedom.

For a comparison of estimators for a ratio:

$t = \sum_S w_k y_k / \sum_S w_k z_k$, $t' = \sum_S w_k' y_k / \sum_S w_k' z_k$,
$t_{(r)} = \sum_S w_{k(r)} y_k / \sum_S w_{k(r)} z_k$, and
$t_{(r)}' = \sum_S w_{k(r)}' y_k / \sum_S w_{k(r)}' z_k$.

Equations (9), (10), and (11) still apply under the null hypothesis.

Suppose we again want to compare $t = \sum_S w_k y_k$ with $t' = \sum_S w_k' y_k$, but now only the first set of calibration weights, the w_k, are treated as valid. Given d-a-g jackknife replicate weights associated with w_k, a z-score for the difference between t and t' is as before:

$z = (t - t')/\surd v$,

with $v = [(R - 1)/R] \sum^R (d_{(r)} - d)^2$, $d = t - t'$, $d_{(r)} = t_{(r)} - t_{(r)}'$, and $t_{(r)} = \sum_S w_{k(r)} y_k$.

But now, $t_{(r)}' = \sum_S w_{k(r)}[w_k'/w_k]y_k$. A similar procedure can be used to compare ratios.

4. The Results

4.1 Methodology

Since our d-a-g jackknives have only 14 degrees of freedom, testing many estimates simultaneously would be problematic. Instead, we analyzed z-scores for a single variable at a time within one of 80 mutually exclusive domains – the ten regions crossed with the eight typologies. Estimates for a parameter across the regions were independent because samples were drawn and calibrated independently. The same cannot be said about estimates across typologies. Nevertheless, we hope little is lost by treating them as independent.

The three parameters we investigated were farms, total expenditures, and total expenditures per farm. Estimating expenditures is one of the foci of the CRR.

For the three investigations to be described, we to treated the two alternative compared as if they were both valid since that was the null hypothesis. It turns out that treating only one of the alternatives as valid did not meaningfully change any of the results.

4.2 Nonresponse Adjustment

We first investigated whether estimates for farms, expenditures, and expenditures per farm would be significantly different had the truncated-linear calibration begun with the sample weights before the last round of nonresponse adjustment, that is, if we had let the truncated-linear-calibration routine implicitly adjust for (the last phase) of nonresponse. Table 4 displays some of the statistics from PROC UNIVARIATE's run on the z–scores described last section treating both sets of weights as equally valid.

In the table, t is the original calibration estimate and t' the calibrated estimate without the explicit nonresponse adjustment. There appears to be no systematic differences between the two sets of estimates across the 80 domains. The table reports tests performed on Studentized differences. The variance of each difference is computed with the d-a-g jackknife, which recognizes that the sample is the same for both estimates. The median of the actual percent differences between the two sets of estimates was less 0.22% in absolute value for each of the three variables. Over 90% of the absolute percentage differences for each variance were less than 2%.

To compare the relative efficiency of the two sets of estimates, we looked at the difference in the logs of their respective coefficients of variation (cv). By taking logs, we made this relative-efficiency measure symmetric. Moreover, the difference in the logs of the relatives variances is exactly twice size the size of this measure.

The version of calibration without explicit nonresponse adjustment appears to have a tendency of being more efficient (i.e., have smaller cv's). This tendency is or – is not – statistically significant depending on the test and the variable.

4.3 Removing the Non-overlap Sample

Next we considered whether removing the non-overlap sample and replacing it with truncated-linear calibration yielded significantly different results from calibrating the sample with the non-overlap component. The ambiguous results can be seen in Table 4. Further statistical analyses, not reported here, suggested that some of the typologies were significantly affected by removing the non-overlap sample. It seems, for example, that retired farms tend to be on NASS's list sampling frame in greater numbers than their 16 calibration-variable values alone would indicate.

That being said, the median of the actual relative differences was quite small, under 1.2% in absolute value for each variable. In addition, over 90% of the absolute relative differences for each were under 25%. By contrast, the median of the cv's of the estimates computed with the original calibration weights ranged from 9 to 16% across the three variables.

Even if removing the non-overlap sample did not create a bias – and it may have in some domains – there was an apparent tendency to lose efficiency in the estimates of total expenditures and expenditures per farm for removing this part of the sample. Only farms showed an (insignificant) tendency to be more efficient without the non-overlap sample.

4.4 Log Calibration

We compared the log and truncated linear methods of calibration assuming the log method without truncation was valid. For both, we began the calibration routines with the nonresponse-adjusted sampling weights and the entire respondent sample.

For one region, iteration did not find a solution for **h**. For that region, the value of the w_i after 30 iterations were used as the a_i in a linear calibration, which did not need to be truncated. The theoretical assumption was that iteration based on $f(y) = \exp(y)$ brought us close to the true values for the p_i, so that linear calibration simply provided what was needed for complete weighted-sample balance on the calibration variables.

Using the log version of calibration (and the two-step process in that one region) seemed to produce significantly different numbers of farms across the domains (see Table 3), but the actual relative differences were quite small. All three median relative differences were less than 0.02% in absolute value, and 90% of the absolute relative differences were less than 2% for each variable.

What is really surprising is that the log version appeared to have a tendency to be more efficient for all three variables We speculate that this is because log calibration uses weighted-regression that are closer to those of Rao's (1994) optimal estimator. It also strongly suggests that the nonresponse adjustments in 2002 CRR were "corrected" by calibration. Otherwise, linear and log calibration would have been asymptotically identical.

5. Discussion

Our empirical analyses suggest that NASS should seriously consider using calibration in place of the last phase of nonresponse adjustment. In contrast to that, although it may be possible for NASS to use calibration to adjust for farms missing from its list-sampling frame in place of the non-overlap component of the area sample, a prudent move would be first to create calibration targets that better capture differences among the typologies.

Most promising of all is the prospect of changing the calibration routine for truncated-linear to modified-log calibration (modified when the log routine fails to converge). Research will be undertaken on more complex forms of quasi-random modeling like Folsom and Singh's generalized exponential model (2002). Also in need of exploration is the possibility of using a different $\mathbf{z}_i$ than $\mathbf{x}_i$, perhaps, for example, replacing the total-vegetable variable in the former vector for a indicator variable for vegetables.

References

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

Folsom., R.E. and Singh, A.C. (2000). "The Generalized Exponential Model for Sampling Weight Calibration for Extreme Values, nonresponse, and Poststratification." In *Proceedings of the Section on Survey Research Methods*, American Statistical Association, Washington DC, 598-603.

Fuller, W.A., Loughin, M.M., and Baker, H.D. (1994). "Regression Weighting for the 1987-88 National Food Consumption Survey." *Survey Methodology*, **20**, 75-85.

Kott, P.S. (2004). "Using Calibration Weighting to Adjust for Nonresponse and Coverage Errors." Under review by *Survey Methodology*.

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

Table 1. Population and Sample Sizes for the 2002

Population Estimates

Region	Typology 1	2	3	4	5	6	7	8	Total
59	4248	26987	67823	38061	11162	6399	5566	2460	162706
62	19132	70878	119935	58843	7881	6033	6240	1329	290271
63	7940	29902	59434	33529	4854	4656	8268	2414	150996
64	7975	40649	81820	58109	21126	9848	7567	2384	229478
67	13342	57925	150026	95904	40418	21813	17015	12159	408601
68	6577	21015	38071	24160	4334	4206	6283	609	105254
69	9203	21263	50956	46983	25773	11895	6436	3968	176478
70	20721	83098	150000	75610	11932	7050	6673	4256	359340
75	3799	26547	48770	40419	12607	4259	3940	3809	144150
76	12333	25558	50377	32803	10021	6925	9352	7470	143738
Total	94170	403822	817213	504421	150109	83084	77339	40857	2171013

Sample Sizes

Region	Typology 1	2	3	4	5	6	7	8	Total
59	16	81	135	159	115	99	190	29	824
62	40	183	295	226	136	142	331	34	1387
63	24	111	195	147	62	78	324	61	1002
64	19	105	207	252	200	151	344	24	1302
67	35	155	382	369	363	307	460	63	2134
68	30	65	111	109	98	131	319	28	891
69	39	90	172	301	316	218	204	49	1389
70	34	180	299	213	113	86	166	26	1117
75	22	87	179	250	185	117	245	48	1133
76	8	120	171	190	128	125	370	100	1212
Total	267	1177	146	2216	1716	1454	2953	462	12391

59 = CT, DE, ME, MD, MA, NH, NJ, NY, PA, RI, VT
62 = KY, NC, TN, VA, WV
63 = AL, FL, GA, SC
64 = MI, MN, WI
67 = IL, IN, IA, MO, OH
68 = AR, LA, MS
69 = KS, NE, ND, SD
70 = OK, TX
75 = AZ, CO, ID, MT, NV, NM, UT, WY
76 = CA, OR, WA

Table 2. Calibration Targets and Uncalibrated Estimates at the US level

Variable	Calibration Target	Estimate (uncalibrated)
Farms with Sales < $10K	1,169,410	1,145,913 98%
Farms with 10K ≤ Sales< 50K	458,960	509,503 111%
Farms with 50 K≤ Sales< 100K	176,530	166,065 94%
Farms with 100K≤ Sales< 250K	189,240	177,759 94%
Farms with 250K≤ Sales< 500K	89,124	86,213 97%
Farms with 500K ≤ Sales < 1M	42,788	48,725 114%
Farms with Sales ≥ 1M	26,148	36,833 141%
Total Fruit Acres	3,965,010	6,364,472 161%
Vegetable Acres	4,389,020	3,782,758 86%
Value of Production	18,811,783,000	22,689,329,858 121%
Cattle	95,944,000	106,078,581 111%
Corn Acres	69,313,000	69,721,222 101%
Cotton Acres	12,412,000	11,088,322 89%
Number of Farms	2,152,200	2,171,013 101%
Hogs and Pigs	58,917,400	49,319,025 84%
Soybeans Acres	72,160,000	73,092,493 101%
Wheat Acres	45,817,000	45,719,216 100%

Table 3. Statistical Comparisons

(t is the current method; t' is the alternative)

Variable	Mean	Median	Probability of Student's |t|	Sign Test	Ranked Sign Test
Removing the Nonresponse Adjustment					
$(t - t')/\sqrt{v}$					
Farms	0.034	0.043	0.687	0.911	0.996
Expend./Farm	−0.058	−0.069	0.454	0.434	0.435
Expenditures	−0.034	−0.056	0.709	0.576	0.472
log[(cv(t')] − log[cv(t)]					
Farms	−0.034	−0.027	0.016	0.005	0.005
Expend./Farm	−0.020	−0.026	0.090	0.093	0.047
Expenditures	−0.022	−0.017	0.082	0.057	0.023
Removing the Non-overlap Sample					
$(t - t')/\sqrt{v}$					
Farms	−0.059	−0.038	0.614	0.738	0.718
Expend./Farm	−0.137	−0.124	0.206	0.018	0.124
Expenditures	−0.224	−0.139	0.053	0.434	0.137
log[(cv(t')] − log[cv(t)]					
Farms	−0.041	0.023	0.271	0.314	0.816
Expend./Farm	0.090	0.108	0.015	0.000	0.002
Expenditures	0.087	0.088	0.005	0.018	0.001
Using log calibration					
$(t - t')/\sqrt{v}$					
Farms	0.268	0.129	0.027	0.314	0.023
Expend./Farm	−0.139	−0.016	0.357	0.738	0.604
Expenditures	0.026	0.047	0.851	0.576	0.348
log[(cv(t')] − log[cv(t)]					
Farms	−0.144	−0.025	0.108	0.005	0.001
Expend./Farm	−1.854	−0.074	0.000	0.000	0.000
Expenditures	−0.182	−0.043	0.000	0.000	0.000