"No" Is the Easiest Answer: Using Calibration to Assess Nonignorable Nonresponse in the 2002 Census of Agriculture

Phillip S. Kott

The 2002 Census of Agriculture adjusted for whole-farm nonresponse by dividing the potential farms on its list into size classes and then weighting the respondents within each class to account for the nonrespondents. Unfortunately, to assign the size-class memberships, a consistent measure of size was needed for all potential farms not just census respondents. Subsequently-collected census information sometimes contradicts these class assignments. By defining indicator variables for census respondents based on "corrected" class assignments, instrumental-variable calibration can be used to construct an alternative set of nonresponse weights. Assuming the response model underpinning these new weights is correct, the bias from using the original set of nonresponse weights can be assessed. The potential bias in the estimated farm count is caused by nonfarms on the list being more likely to respond to the census than farms.

KEY WORDS: Weighting class; Calibration; Response model; Instrumental variable; Measurement-error model.

Phillip S. Kott is Chief Research Statistician, Research and Development Division, National Agricultural Statistics Service, 3251 Old Lee Highway, Room 305, Fairfax, Virginia, 22030. This paper was prepared for the Joint Statistical Meetings, August 2005, Minneapolis, Minnesota.

1. Introduction

Despite its name, it is helpful to think of the 2002 US Census of Agriculture as a survey. It's core was indeed a census of all places capable of producing $1,000 or more of annual agricultural sales – what the National Agricultural Statistics Service (NASS) defined as farms. This core, however, had, to be supplemented to two directions. First, not all entities on the Census Mailing List (CML) maintained by NASS responded to the Census of Agriculture. Second, not all places NASS defined as farms were on the CML.

To compensate for nonresponse and undercoverage, Census records were reweighted. Unlike most surveys, the completed records in the 2002 Census of Agriculture had original weights of 1. These weights were adjusted first to account for the nonresponse and then the undercoverage. How both of these were done is explained in some detail on the NASS web site (Kott, 2004a).

Our attention here will be focused on the adjustment for nonresponse in the 2002 Census of Agriculture. NASS divided the entities receiving Census forms into mutually exclusive "response groups" based on what NASS believed to be each entity's county of operation, its size class as measured by expected sales, and whether or not the entity responded to an agency survey since 1997. Potential farms above a certain expected size or without expected sales information were removed from this categorization and handled separately, as were entities whose forms were returned as undeliverable by the Post Office.

Reweighting within mutually exclusive groups is a well-known and much-used procedure. See, for example, Lohr (1999, pp. 266-267). There are two ways to defend this practice theoretically. One is with a response model, in which every unit in a groups is assumed to be equally likely to respond, irrespective of its Census item values. This approach, known as *quasi-randomization*, treats the response (nonresponse) process as a phase of random sampling.

The alternative defense is to assume a prediction model for each of the item values on the census. The response/nonresponse mechanism is assumed to be ignorable; that is to say, the same prediction model applies to respondents and nonrespondents. The unit values for an item are independently distributed random values with a common mean within each response group. The response groups under theses assumptions are more correctly thought of as prediction-model groups.

Both the response and prediction models have one obvious defect when used in the Census of Agriculture. Groups are defined using expected 2002 sales before enumeration rather than the actual sales reported on the US Census of Agriculture. It is more reasonable to assume similar behavior from farms in the same *actual*-sales group than from entities in the sample *expected*-sales group. An obvious example: two farms in the sample expected-sales group but in different actual-sales groups can have vastly different actual sales, a Census item.

The single most important Census item is whether an entity meets the definition of a farm (the item value is 1 if it does, 0 otherwise). Here, again, actual sales is a much better predictor for the item value than expected sales.

Why then did NASS use expected sales in creating response groups? The answer applies to many survey using reweighting not just the Census of Agriculture. In order to apply reweighting, the entire sample, or in this case the entire categorizable population, must be assigned to response groups. Since NASS only knows the actual sales for respondents, it could not use actual sales in creating groups.

This paper explores an alternative approach to reweighting. Section 2 first puts reweighting for nonresponse into a general quasi-randomization theory of calibration. A mild extension allows the set of variables in the weight equation, the predictors of response/nonresponse, to contain some variables not in set used in the calibration equation. As long as there are the same number of variables in each set, instrumental-variable calibration (Estevao and Särndal, 2000) can be employed. Furthermore, the response-predicting instrumental variables need only be defined for the responding units.

Section 3 explores using the instrumental-variable calibration with 2002 Census-of-Agriculture data. Instrumental-variable calibration turns out to have some undesirable properties in this application. Calibrated weights can be less than 1, and estimated quasi-randomization variances are much higher than with simple reweighting within groups. Still, under the assumption that the model using the actual sales-size groups is correct, estimates based on simple reweighting can be biased downward because potential farms that are not farms are more likely to respond than entities meeting the definition of a farm. “No” is the easiest answer to give on a survey.

Section 4 offers some concluding remarks. It looks at a prediction-model where the distribution of an item value given the calibration variables need not be the same for respondents and nonrespondents; that is to say, the response/nonresponse mechanism may be nonignorable. This situation can be handled by assuming a prediction model where the item values are conditioned on the instrumental variables rather than the calibration variables. A direction for future research is also discussed.

2. A Quasi-Randomization Approach to Calibration

Poststratification

In the quasi-randomization approach to reweighting for nonresponse, each unit k in the original sample (or population) has a positive Poisson probability of responding to the survey, p_k. This probability may be a function of covariates associated with the unit, but whether unit k responds to the survey is independent of whether other units respond.

Suppose we have a sample. Unit k in the sample has original sample weight a_k. In the case of the Census of Agriculture, all $a_k = 1$. Folsom and Singh (2000) assume that the unit-k Poisson probability of response has the form:

$$p_k = p(\mathbf{x}_k\delta), \tag{1}$$

where $\mathbf{x}_k$ is a known row vector of Q covariates, δ an unknown column vector of the same size, and p(z) is a known continuous, monotonic function. The unit Poisson probabilities in equation (1) are assumed independent of each other (hence the name) and of all the Census item values.

When the population total, $\sum_U \mathbf{x}_j$, is known, Folsom and Singh argue that δ in equation (1) can be estimated implicitly by finding a set of *calibration weights* for each j in the respondent subsample, S, satisfying $w_j = a_j / p(\mathbf{x}_j\mathbf{g})$ for all $k \in S$, and the *calibration equation*:

$$\sum_S w_j\mathbf{x}_j = \sum_U \mathbf{x}_j. \qquad (2)$$

Observe that under this framework using the "true" sampling weight, $a_k / p(\mathbf{x}_k\delta)$, for each k in S would come asymptotically close to solving the calibration equation. Consequently, the **g** that satisfies the calibration equation exactly should be a consistent estimator for δ.

It is convenient to think of the calibration adjustment for nonresponse as the factor, $f(\mathbf{x}_k\mathbf{g}) = 1/p(\mathbf{x}_k\mathbf{g})$. This factor when multiplied by the original weight, a_k, results in the calibration weight, w_k When p(t) has the awkward form p(t) = 1/(1 + t) so that f(t) = 1/p(t) is linear, Fuller, Loughin, and Baker (1994) point out that the usual linear calibration weights result. That is because with f(t) = 1 + t, and
$\mathbf{g} = (\sum_S a_j\mathbf{x}_j'\mathbf{x}_j)^{-1}(\sum_U \mathbf{x}_j - \sum_S a_j\mathbf{x}_j)'$,

$$f(\mathbf{x}_k\mathbf{g}) = f(\mathbf{g}'\mathbf{x}_k') = 1 + \mathbf{g}'\mathbf{x}_k' = 1 + (\sum_U \mathbf{x}_j - \sum_S a_j\mathbf{x}_j)(\sum_S a_j\mathbf{x}_j'\mathbf{x}_j)^{-1}\mathbf{x}_k', \qquad (3)$$

and $\sum_S w_j\mathbf{x}_j = \sum_S a_j f(\mathbf{x}_j\mathbf{g})\mathbf{x}_j = \sum_U \mathbf{x}_j$.

Suppose $\mathbf{x}_k$ takes the form of a vector of response group indicators. It's q-th member is 1 when k is in response group q, zero otherwise. Since the groups are mutually exclusive, all functions of $\mathbf{x}_k\delta$ are equivalent to a linear function of $\mathbf{x}_k$ (because $f(\mathbf{x}_k\delta) = f(\sum^Q x_{kq}\delta_q) = \sum^Q x_{kq}f(\delta_q) = \sum^Q x_{kq}\delta_q^*$). Thus, the seemingly awkward model, $p_k = p(\mathbf{x}_k\delta) = (1 + \mathbf{x}_k\delta)^{-1}$ is the same as $p_k = \sum^Q x_{kq}\delta_q^*$.

If all of the $a_k = 1$, as in the 2002 Census of Agriculture, then the calibration equation is solved when each $\delta_q^* = n_q / N_q$, the respondent size within group q divided by the population size within the group. Equivalently, for responding unit k in group q, $w_k = N_q / n_q$. Calibration weighting preserves estimates of the population sizes of potential farms within each group. This type of reweighting is usually given the name "poststratification."

Instrumental-variable calibration

Estevao and Sarndal (2000) point out that, in general, the variables in the weight equations, $w_k = a_k f(\mathbf{x}_k\mathbf{g})$, need not coincide with the variables in the calibration equation (2). One can replace $\mathbf{x}_k$ with another row vector with the same number of components, say $\mathbf{z}_k$. Finding a vector $\mathbf{g}$ that satisfies the slightly-revised calibration equation,

$$\sum_S a_j f(\mathbf{z}_k\mathbf{g})\mathbf{x}_j = \sum_U \mathbf{x}_j. \tag{4}$$

is "simply" a matter of solving Q equations with Q unknowns. Such a solution can often, but not always, be found.

The components of $\mathbf{z}_k$ that are not in $\mathbf{x}_k$ are called *instrumental variables*. Because of this, the vector $\mathbf{z}_k$ is called an *instrumental-variable vector* when it does not exactly coincide with $\mathbf{x}_k$, and weights that satisfy the calibration equation in (4) *instrumental-variable calibration weights*.

Observe that the underlying response model assumption is that the Poisson probability of unit k responding has the form:

$$p_k = p(\mathbf{z}_k\delta) = 1/f(\mathbf{z}_k\delta);$$

that is to say, response is a function of the components of $\mathbf{z}_k$. Unlike the components of $\mathbf{x}_k$, the components of $\mathbf{z}_k$ need only be known for the members of the respondent sample. Furthermore, calibration weights can often be found without the population z-total, $\sum_U \mathbf{z}_j$, being known.

In our empirical analysis of the Census of Agriculture in the following section, we will replace components of $\mathbf{x}_k$ with analogously-defined components that substitute actual 2002 sales for expected sales. Our new $\mathbf{z}_k$, like $\mathbf{x}_k$, will be a vector with a single 1-valued component and Q−1 0-valued components. Consequently, all calibration factors are essentially linear. They can be written as the following generalization of equation (3):

$$f(\mathbf{z}_k\mathbf{g}) = f(\mathbf{g}'\mathbf{z}_k') = 1 + (\textstyle\sum_U \mathbf{x}_j - \sum_S a_j\mathbf{x}_j)(\sum_S a_j\mathbf{z}_j'\mathbf{x}_j)^{-1}\mathbf{z}_k'. \tag{5}$$

Equation (3) is the special case where $\mathbf{z}_k \equiv \mathbf{x}_k$.

Given a calibration estimator $\sum_S w_k y_k$ for $\sum_U y_k$ based on Poisson response probabilities of the form $p_k = p(\mathbf{z}_k'\delta)$, Kott (2004b) argues that under certain conditions a good estimator for the quasi-randomization mean squared error of $\sum_S w_k y_k$ is

$$v = \textstyle\sum_S (w_k^2 - w_k)e_k^2, \tag{6}$$

where

$$e_k = y_k - \mathbf{x}_k(\textstyle\sum_S a_j f'(\mathbf{z}_j\mathbf{g})\mathbf{z}_j'\mathbf{x}_j)^{-1}\sum_S a_j f'(\mathbf{z}_j\mathbf{g})\mathbf{z}_j' y_j, \tag{7}$$

$f(t) = 1/p(t)$, and $f'(t) = df(t)/dt$. In the Census-of-Agriculture context under investigation in the next section, all $a_j = 1$. Moreover, $f(t)$ is linear, so $f'(t) = 1$.

3. An Evaluation of the Nonresponse-Adjustment Methodology in the 2002 Census of Agriculture

As noted in the introduction, NASS removed certain potential farms from the population before creating reweighting groups for the 2002 Census of Agriculture. In the analysis to be presented here, these and other problem entities (like those that turned out to be associated with more than one farms) are removed from the data set beforehand.

NASS actually created response groups within counties. When too few potential farms responded in a putative group, collapsing rules were followed. In order to focus on the repercussions of using expected-sales in response-group formation and to avoid sticky small-group problems, the analysis here forms groups at the state level in

47 states (NASS used whole-farm imputation rather than reweighting in Alaska, Hawaii, and Rhode Island.)

Paralleling the within-county routines actually used by NASS, reweighting is first done within states by creating five mutually exclusive response groups:

X-Group 1: Expected 2002 sales less than $2,500;

X-Group 2: Expected 2002 sales between $2,500 and $9,999;

X-Group 3: Expected 2002 sales between $10,000 and $49,999 and previously reported survey data from 1997 or later;

X-Group 4: Expected 2002 sales greater than or equal to $50,000 and reported survey data from 1997 or later;

X-Group 5: Expected 2002 sales greater than or equal to $10,000 and no reported survey data from 1997 or later.

All the units in the data set belonged to one of these five groups. Mathematically, the covariate $\mathbf{x}_k = (x_{k1}, x_{k2}, ..., x_{k5})$, where x_{kg} =1 when k is in X-Group g and x_{kg} =0 otherwise. Reweighting within states using $\mathbf{x}_k$ in equation (3) can be compared to instrumental-variable calibration weighting using $\mathbf{z}_k$ in equation (5), with

z_{k1} =1 if unit k had reported 2002 sales less than $2,500, 0 otherwise;

z_{k2} =1 if unit k had reported 2002 sales between $2,500 and $9,999, 0 otherwise; .

z_{k3} =1 if unit k had reported 2002 sales between $10,000 and $49,999 and previously reported survey data from 1997 or later, 0 otherwise;.

z_{k4} =1 if unit k had reported 2002 sales greater than or equal to $50,000 and reported survey data from 1997 or later, 0 otherwise;

z_{k5} =1 if unit k had reported 2002 sales greater than or equal to $10,000 and no reported survey data from 1997 or later, 0 otherwise.

The sets of components making up $\mathbf{x}_k$ and $\mathbf{z}_k$ are completely analogously, with expected sales in the x-definitions replaced by actual sales in the z-definitions. We called the original five response groups that defined the components of $\mathbf{x}_k$ the "x-groups." In a parallel fashion, there are also five mutually exclusive groups related to the components of $\mathbf{z}_k$. We call them the "z-groups."

Reweighting using the x-groups is justified if one assumes that all potential farms in the same x-group are equally likely to responds. Instrumental-variable calibration with the $\mathbf{z}_k$ is justified if one assumes that potential farms in the sample z-group are equally likely to respond. The latter assumption is the more reasonable.

One way to summarize the two approaches is to say that the "x-group method" treats the x-groups as both the response groups and the *calibration groups*. The latter are the groups for which the calibration-weighted totals equal the population totals (the calibration equation (2) holds). The "z-group method" treats the z-groups as the response groups while the x-groups remain the calibration groups.

The empirical analysis here focuses on one item: the estimated number of farms (entities capable of having at least $1,000 in agricultural sales in 2002). This is called the "farm count." Summary statistics for the two reweighting methods are contained in Table 1. The biases and relative biases are for the x-groups method assuming the z-group method is (asymptotically) unbiased. Estimated standard errors are computed using the square root of the right hand of equation (6). In these calculations, when the x-group (z-group) method is used, it is assumed to be unbiased.

If the probability of response is really constant within z-groups, but we reweight as if they were constant within x-groups, as NASS did at the county-level for the 2002 Census of Agriculture, then the number of farms (in the categorized population) is undercounted by as estimated 2.3%. There is a mild overcount of farms with annual sales of less that $10,000, but that is overwhelmed by the undercount of farms with more than $10,000 in annual sales, especially among those farms that did not respond to a NASS survey from 1997 to 2002, what we have called Z-Group 5).

Why this happens is pretty obvious. Using the x-group methodology, the average response rate for farms in the Z-Group 5 is estimated to be 80%. Using the z-group methodology it is only 55%. Since all potential farms in this group are farms (they have annual sales above $1,000), the x-group methodology does not adjust their weights enough. Using either methodology, it is clear that the response rate goes down as sales goes up. With the x-group methodology, however, the proper weighting is confounded by potential farms that have higher expected sales than actual sales. Unlike it analogous z-group, not all entities in X-Group 5 (or 4, 3, or 2) are farms.

Observe that the estimated standard errors using the z-group method are always double those of using the x-group method. That may be the price one pays for creating response groups based on information available only from respondents.

Table 2 displays the variability of the average weight within z-groups across states for the two methods as well as the variability of the estimated relative bias from using the x-group method relative to the z-group method. Observe that the minimum value of the weight is less than 1 in Z-Group 2. In all, instrumental-variable calibration results in weights less than unity three out of 5 × 47 = 235 times. This happens in two states for Z-Group 2 and in one state for Z-Group 1, although in the latter the weight rounded to 1. As a result of this, the estimated number of farms is less than the responding number of farms in these groups, an absurdity. By contrast, simple poststratification never allows the calibration weight to be less than 1.

Only one out of 47 states has a smaller estimated farm count using the z-group method. From this, we can infer that the downward bias in the x-group method relative to the z-group method is significant. For the majority of states, the estimated relative bias of the x-group method is in the -1 to -3% range.

Table 1. Summary Statistics on the Two Methods of Reweighting for Nonresponse on the 2002 Census of Agriculture

		Using x-groups			*Using z-groups*				
Z-group	Unadjusted counts	Adjusted counts	response rate	estimated se	Adjusted counts	response rate	estimated se	Bias	Relative bias
1	408243	465059	88	219	446322	91	531	18737	0.040
2	288040	330097	87	197	329755	87	1401	342	0.001
3	256280	297589	86	187	309230	83	7321	-11641	-0.039
4	216022	257410	84	150	268045	81	327	-10635	-0.041
5	53107	66327	80	110	96127	55	344	-29800	-0.449
Total	1221692	1416482	86	256	1449479	84	721	-32997	-0.023

Bias and relative bias are estimates for using the x-group methodology when the z-group methodology is (asymptotically) unbiased.

Table 2. The Variability of the Weights and the Relative Biases Across States

Less than $2,500 in sales: Z-Group 1

Variable	Mean	Min	Lower Quartile	Median	Upper Quartile	Max
x-weight	1.14	1.04	1.12	1.14	1.17	1.21
z-weight	1.10	1.00	1.07	1.10	1.12	1.16
relative bias	0.04	-0.01	0.02	0.03	0.06	0.14

Between $2,500 and $9,999 in sales: Z-Group 2

Variable	Mean	Min	Lower Quartile	Median	Upper Quartile	Max
x-weight	1.15	1.05	1.13	1.15	1.18	1.24
z-weight	1.17	0.94	1.12	1.15	1.22	1.37
relative bias	-0.01	-0.13	-0.05	-0.02	0.02	0.20

Between $10,000 and $49,999 in sales and appearance on a NASS survey since 1997: Z-Group 3

Variable	Mean	Min	Lower Quartile	Median	Upper Quartile	Max
x-weight	1.17	1.06	1.13	1.17	1.20	1.26
z-weight	1.22	1.06	1.16	1.19	1.27	1.53
relative bias	-0.04	-0.19	-0.06	-0.02	-0.01	0.04

Over $50,000 in sales and appearance on a NASS survey since 1997: Z-Group 4

Variable	Mean	Min	Lower Quartile	Median	Upper Quartile	Max
x-weight	1.20	1.06	1.17	1.20	1.24	1.32
z-weight	1.25	1.06	1.19	1.23	1.30	1.52
relative bias	-0.04	-0.15	-0.06	-0.04	-0.02	0.09

Over $10,000 in sales and no appearance on NASS survey since 1997: Z-Group 5

Variable	Mean	Min	Lower Quartile	Median	Upper Quartile	Max
x-weight	1.22	1.05	1.18	1.22	1.26	1.40
z-weight	1.70	1.22	1.43	1.65	1.85	2.72
relative bias	-0.26	-0.52	-0.32	-0.25	-0.17	-0.06

All farms

Variable	Mean	Min	Lower Quartile	Median	Upper Quartile	Max
x-weight	1.16	1.05	1.14	1.16	1.19	1.24
z-weight	1.19	1.05	1.15	1.19	1.22	1.26
relative bias	-0.02	-0.06	-0.03	-0.02	-0.01	0.01

x-weight is the average weight – adjusted count/unadjusted count -- in a state using the x-groups method;
z-weight uses the z-group (instrumental-variable calibration) methodology;
relative bias is the estimated relative bias in a state of the x-group methodology assuming the z-group methodology is (asymptotically) unbiased.

4. Concluding Remarks

The introduction discussed two theoretical justifications for handling nonresponse by reweighting within mutually exclusive groups: response modeling and prediction modeling. In the former, every unit in the same response group was assumed equally likely to respond to the survey (or, in the case analyzed in depth here, the census). In the latter, the unit values for an item of interest were assumed to have a common mean within response groups (which more accurately could be called "prediction-model groups" in this context).

Under either justification the adjusted method was the same: each unit in the original sample (or, in the case here, the categorized population) was assigned to one of the response groups, and the respondents were reweighted so that the sum of the new weights among respondents within groups equaled the sum of the old weights among both respondents and nonrespondents.

In Section 2, a instrumental-variable calibration methodology was developed and justified under a response model in which, again, every unit in a response group was equally likely to respond. Now, however, only the responding units needed to be assigned to a response group. This was made possible because there was an equal number of calibration groups to which every unit in the original sample had been assigned whether or not it responded.

A prediction-model approach

Instrumental-variable calibration can be given a general prediction-model justification. In brief, suppose the prediction model for a item of interest y has the form:

$$y_k = \mathbf{z}_k\beta + \epsilon_k, \tag{8}$$

where $E(\epsilon_k|\mathbf{z}_k) = 0$ for both respondents and nonrespondents. Moreover, suppose the calibration-variable row vector, $\mathbf{x}_k$, has an analogous form:

$$\mathbf{x}_k = \mathbf{z}_k\Gamma + \xi_k, \tag{9}$$

where $E(\xi_k|\mathbf{z}_k) = \mathbf{0}$ for both respondents and nonrespondents.

Equation (9) is often referred to as a *measurement-error* model. Note that although $E(\epsilon_k|\mathbf{z}_k) = 0$ and $E(\xi_k|\mathbf{z}_k) = \mathbf{0}$ are assumed to hold for both respondents and nonrespondents, $E(\epsilon_k|\mathbf{x}_k) = 0$ need *not* be zero for both respondents and nonrespondents. It is for this reason that the response/nonresponse mechanism is called "nonignorable" in the title.

Under the prediction and measurement-error models in equations (8) and (9) respectively, the difference between a calibration estimator, $t = \sum_S w_k y_k$ and its target, $T = \sum_U y_k$, is

$$\begin{aligned} t - T &= \textstyle\sum_S w_k y_k - \sum_U y_k \\ &= \textstyle\sum_S w_k[(\mathbf{x}_k - \xi_k)\Gamma^{-1}\beta + \epsilon_k] - \sum_U [\,\Gamma^{-1}(\mathbf{x}_k - \xi_k)\Gamma^{-1}\beta + \epsilon_k] \\ &= \textstyle\sum_S w_k(\epsilon_k - \xi_k\Gamma^{-1}\beta) - \sum_U(\epsilon_k - \xi_k\Gamma^{-1}\beta), \end{aligned} \tag{10}$$

when the weights are such that the calibration equation, $\sum_S w_k\mathbf{x}_k = \sum_U \mathbf{x}_k$, holds, and Γ is invertible.

Equation (10) appears to support the notion for *any* calibration estimator would be unbiased under the prediction and measurement-error models so long as $\sum_S w_k\mathbf{x}_k = \sum_U \mathbf{x}_k$ holds. The fault with this reasoning is in treating the calibration weights as constants given the $\mathbf{z}_k$. Conventional calibration weights, those not involving instrumental variables, have the form $w_k = a_k f(\mathbf{x}_k\mathbf{g})$. They are functions of the $\mathbf{x}_k$, which are random variables under the measurement-error model.

The weights in instrumental-variable calibration are designed to be functions of the $\mathbf{z}_k$ rather than the $\mathbf{x}_k$. Nevertheless, they too are functions of the $\mathbf{x}_k$ through the calibration-equation-solving $\mathbf{g}$ in $w_k = a_k f(\mathbf{z}_k\mathbf{g})$. Consequently, the instrumental-variable calibration estimator is not strictly prediction-model unbiased given the $\mathbf{z}_k$. Nevertheless, it is a consistent estimator under mild conditions so long as $\mathbf{g}$ converges to a vector with finite components as the respondent sample grows arbitrarily large. Note that although f(.) is used to define $\mathbf{g}$, its inverse, the Poisson response probability p(.), need *not* be correctly specified under this prediction-model justification for instrumental-variable calibration.

If both the ϵ_k and ξ_k are uncorrelated across units given $\mathbf{z}_k$, then under mild conditions, e_k in equation (7) is approximately:

$$
\begin{aligned}
e_k &= y_k - \mathbf{x}_k\left(\textstyle\sum_S a_j f'(\mathbf{z}_j\mathbf{g})\mathbf{z}_j'\mathbf{x}_j\right)^{-1} \textstyle\sum_S a_j f'(\mathbf{z}_j\mathbf{g})\mathbf{z}_j' y_j) \\
&= \mathbf{z}_k\beta + \epsilon_k - (\mathbf{z}_k\Gamma + \xi_k)\left(\textstyle\sum_S a_j f'(\mathbf{z}_j\mathbf{g})\mathbf{z}_j'[\mathbf{z}_k\Gamma + \xi_k]\right)^{-1} \textstyle\sum_S a_j f'(\mathbf{z}_j\mathbf{g})\mathbf{z}_j'(\mathbf{z}_k\beta + \epsilon_k) \\
&\approx \mathbf{z}_k\beta + \epsilon_k - (\mathbf{z}_k\Gamma + \xi_k)\Gamma^{-1}\beta \\
&= \epsilon_k - \xi_k\Gamma^{-1}\beta,
\end{aligned}
$$

which allows a prediction-model interpretation to the mean-squared-error estimator in equation (6).

Caveats and a possible extension

There are two obvious drawbacks to the instrumental-variable calibration methodology laid out in Section 2. The first is that the theory underpinning it is asymptotic (since $\mathbf{g}$ in $w_k = a_k f(\mathbf{z}_k\mathbf{g})$ is a consistent estimator for δ in $p_k = p(\mathbf{z}_k\delta)$). That is not much of a limitation when reweighting for nonresponse in the Census of Agriculture at the state level because both calibration and response group sizes are in the hundreds. It *is* a limitation had we tried to apply the methodology at the county level where some group sizes are small to nonexistent. Since one of the purposes of the Census of Agriculture is to measure farm activity at the county level, this limitation is very problematic. In principle, the limitation appears to be shared by simple poststratification as well, but one can define the fraction of responses in a group as the exact *conditional* probability of response in the group and avoid asymptotics entirely. A group is only too small when its respondent size sinks below one, making estimation impossible.

A more general drawback of instrumental-variable calibration is that the methodology is limited to a response (or prediction-model) vector ($\mathbf{z}_k$) with the same number of components as the calibration vector ($\mathbf{x}_k$). Chang and Kott (2005) develop an approach to calibration where the calibration vector is larger than the response vector. This is made possible because their approach loosens the restriction that the calibration equation holds *exactly*.

Even when the number of components in the response and calibration vectors coincide, there may be no set of weights in the proper form that satisfy the calibration equation exactly. For example, suppose one assumes the probability of response in the 2002 Census of Agriculture was logistic ($p(\mathbf{z}_k\delta) = [1 + \exp(-\mathbf{z}_k\delta)]^{-1}$). Consequently, the calibration weight for each farm has inverse-logistic form ($w_k = 1 + \exp(-\mathbf{z}_k\delta)$) and must exceed 1. In Section 3, we saw that the solution of the 5-equation-5-unknown problems results in some weights less than 1 in three states. It is not hard to show that these solutions are unique. Thus, no set of weights in inverse-logistic form satisfy the calibration equation exactly in these states. Chang and Kott's loosening of the calibration-equation constraint is appealing in this situation.

References

Chang, T. and Kott, P.S. (2005). "Using Calibration to Fit Nonresponse and Noncoverage Models." *ASA Proceedings of the Survey Research Methods Section*, to appear.

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." *ASA Proceedings of the Section on Survey Research Methods*, 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. (2004a). "Collected Technical Notes on Weighting and its Impact on the 2002 Census of Agriculture." http://www.nass.usda.gov/research/reports/COMBINED.pdf

Kott, P.S. (2004b). "Comment on Demnati and Rao." *Survey Methodology*, **30**, 27-28.

Lohr, S. L. (1999). *Sampling Design and Analysis*. Duxbury Press, Pacific Grove, CA.