Using Calibration Weighting to Adjust for Nonresponse Under a Plausible Model (with full appendices)

Ted Chang

Department of Statistics, University of Virginia,

Charlottesville VA 22904-4135 USA

tcc8v@virginia.edu

Phillip S Kott

National Agricultural Statistical Service,

Fairfax VA 22030-1504 USA

Phil_Kott@nass.usda.gov

Summary

Calibration forces the weighted estimates of certain variables to match known or alternatively estimated population totals called benchmarks. It can be used to correct for sample-survey nonresponse or for coverage error resulting from frame undercoverage or unit duplication. The quasi-randomization theory supporting its use in nonresponse adjustment treats response as an additional phase of random sampling. The functional form of a quasi-random response model is assumed to be known, its parameter values estimated implicitly through the creation of calibration weights. Unfortunately, calibration depends upon known benchmark totals

while the variables in a plausible model for survey response are not necessarily the same as the benchmark variables. Moreover, it may be prudent to keep the number of explanatory variables in a response model small. We will address using calibration to adjust for nonresponse when the explanatory model variables and benchmark variables are allowed to differ as long as the number of benchmark variables is at least as great as the number of model variables. Data from National Agricultural Statistical Service's 2002 Census of Agriculture and simulations based upon that data will be used to illustrate alternative adjustments for nonresponse. The paper concludes with some remarks about extension of the methodology to adjustment for coverage error.

Some key words: Benchmark; Consistency; Coverage model; Back-link function; Quasi-randomization; Response model.

1. Introduction

Calibration weighting ensures that sample-estimated totals of certain calibration or benchmark variables match previously determined population totals. Two related special cases of calibration, poststratification and weighting-class adjustment (see Lohr 1999, pp. 266-267) are used extensively to adjust for survey nonresponse, a subject of growing interest as the response rates in both government and private surveys decline.

Oh and Scheuren (1983) provide a theoretical justification for weighting-class adjustment by treating response as a second phase of sample selection. In this quasi-randomization (or quasi-design-based) framework, each sampled unit within the same weighting class has an equal and independent probability of selection into the respondent subsample. That probability is estimated implicitly in the weighting process. The prefix "quasi" is added to "randomization" to emphasize that inference is not model free, but depends on an assumed response model. Like all models, the response model can fail. Unfortunately, any method for handling

nonresponse requires some form of modelling, at least implicitly.

More complex nonresponse adjusting calibration schemes are proposed in Folsom (1991), Fuller, Louglin, and Baker (1994), and Kott (2006). In each the probability of response is assumed to be a known "back-link" function $p(x_i^T \beta_*)$ of an unknown (but estimatable) linear combination of explanatory model variables x_i. The back-link function is the back transformation of the link function in a generalized linear model. See, for example, McCullagh and Nelder (1989). What we have called the "back-link" is sometimes called the "inverse link" in the generalized-linear-model literature.

In Fuller et al., the back-link function has the form $p(\eta) = 1/(1 + \eta)$. This allows calibration to have its conventional linear form. Lundström and Särndal (1999) also proposes using calibration in conventional linear form to adjust for nonresponse but without specifying a back-link function.

Folsom proposes more plausible functions for the modeling of response than $p(\eta) = 1/(1 + \eta)$. One such is the logistic: $p(\eta) = [1 + exp(-\eta)]^{-1}$. In addition, raking is shown to be a form of calibration weighting with a back-link function of the form $p(\eta) = exp(\eta)$. A follow-up, Folsom and Singh (2000), proposes a class of reasonable back-link functions.

In both Fuller et al. and Folsom the explanatory model variables used to estimate implicitly the probabilities of response are the same as the calibration variables for which one has benchmark totals. In Kott that is no longer the case. This extension does not require that model-variable totals be known. For example, one can separate respondents into response groups in an analogue to poststratification based on their survey responses. Still, Kott assumes the number of explanatory model and benchmark variables are equal. Moreover, that paper does not demonstrate the practicality of its approach with data.

Särndal and Lundström (2005) also treats the case where the explanatory model and benchmark variables can differ in definition but not in number. In addition, it allows some of the benchmark totals to be calculable from the sample before

nonresponse. The back-link function is not specified. Moreover, the authors do not appear to notice that calibration is possible when model-variable values are known only for the respondents.

We will show how to use calibration to adjust for nonresponse when the number of benchmark variables is at least as great as the number of explanatory model variables. As in Kott, the values for the explanatory variables need only be known for respondents. In Section 2 we will introduce our notation and motivate our approach to calibration, which is discussed in more detail in Sections 3 and 4.

In section 5 we turn to the estimation of a total for a vector of variables of interest that typically do not include the benchmark variables (since they are either already known or previously estimated) but may include some of the model variables. We show how to measure both the additional asymptotic variance due to the nonresponse in a calibration-weighted estimator and the full asymptotic variance of the estimator itself. All variances in the text are determined either with respect to the randomization mechanism used to select the sample, the response model generating the subset of sample respondents, or both.

Section 6 contains applications of our methodology to nonresponse adjustment for the 2002 Census of Agriculture. We show here how the probability of a farm's responding to the census can be assumed to be a function of its survey-reported sales rather than its expected sales before enumeration, as is currently assumed in practice. In section 7 we report on simulations from an artificial population constructed from the respondents of the Census of Agriculture. These sections also discuss the implications of a misspecified model.

Section 8 provides some concluding remarks. Among them is a brief discussion about extending the findings of the preceding sections to adjusting for coverage errors, a topic of increasing interest for surveys based on incomplete frames such as telephone and internet surveys.

2. Notation and motivation

Suppose z_i is a P-vector of calibration or benchmark variables for the i-th population unit, and x_i a Q-vector of explanatory model variables, hereafter called simply "model variables." We will assume that the probability of i responding (when the unit is selected for the sample) is $p(x_i^T\beta_*)$ for some vector parameter β_*, where $g(\eta) = 1/p(\eta)$ is a known and everywhere monotonic and twice differentiable function, such as $g(\eta) = (1 + exp(-\eta))$. In this quasi-randomization framework, the response probability does not depend, given the x_i value, upon whether the unit is in the sample or the y_i or z_i values per se. Unlike most treatments, however, we allow the possibility that y_i is a component of x_i.

If β_* were also known, then an expansion estimator for the vector of totals of the benchmark variables would be

$$\widehat{t}_z(\beta_*) = \sum_{i\in\mathcal{R}} \frac{d_i}{p(x_i^T\beta_*)} z_i, \tag{1}$$

where d_i is the sampling weight in the absence of nonresponse, and $\mathcal{R}$ is the set of respondents.

If T_z is a vector of calibration target values consisting of known, or previously estimated, population totals, then β_* could be estimated from the data using the calibration equation(s)

$$T_z = \sum_{i\in\mathcal{R}} \frac{d_i}{p(x_i^T\widehat{\beta})} z_i. \tag{2}$$

If the number P of benchmark variables equals the number Q of model variables, equations (2) will usually be sufficient to determine $\widehat{\beta}$. On the other hand, if $P < Q$, $\widehat{\beta}$ will be underdetermined by (2).

Realize, however, that even were β_* known, it is unlikely that $\widehat{t}_z(\beta_*)$ would equal T_z exactly due to sampling variability. The vectors $\widehat{t}_z(\beta_*)$ and T_z should nonetheless be close. With this in mind, we suggest that (1), and its child (2), be viewed heuristically as nonlinear regression-type equations

$$T_z = \widehat{t}_z(\beta) + \epsilon \tag{3}$$

where $\widehat{t}_z(\beta)$ is defined as in equation (1) with β_* replaced by β, and ϵ is a P-vector of

"errors". In the nonlinear regression paradigm, it is desirable that $P > Q$. Indeed, since each target provides an additional "equation" in (3), one would suspect that addition of targets will improve the calibration as long as the additional targets do not introduce correlation problems in the $\widehat{t}_z(\beta_*)$. More precise results in this direction are a topic for future research.

The parameter β can be estimated by minimizing an objective function of the form

$$\rho(\beta) = -\text{logdet}(W) + (T_z - \widehat{t}_z(\beta))^T W (T_z - \widehat{t}_z(\beta)) \tag{4}$$

for some appropriately chosen $P \times P$ positive definite matrix W (which may depend upon β), where logdet(W) denotes the log of the determinant of W. Note that (4) is the log likelihood of $\widehat{t}_z(\beta)$ under an asymptotic normal distribution with covariance matrix W^{-1}. In maximum likelihood estimation, where W is estimated from the sample, the $-$logdet(W) term prevents the minimization from attempting to send $W \to 0$. It is important to realize that when $P > Q$, no choice of β forces $\widehat{t}_z(\beta)$ to equal T_z as in conventional $P = Q$ calibration.

The nonlinear regression formulation of (3) suggests setting $W = V^{-1}$ for some suitably defined variance matrix V of ϵ. When this choice of W depends on β, we propose an iterative procedure analogous to what would be used with a fixed W. Given a guess $\widetilde{\beta}_0$ of β_*, we can linearize the regression (3) at $\widetilde{\beta}_0$. The solution to the linearized regression is the next guess $\widetilde{\beta}_1$. This procedure is described more thoroughly in Section 3.

An obvious candidate for V is $\widehat{\text{var}}_{qr}(\widehat{t}_z(\beta)|\beta = \beta_*)$, an estimator for the quasi-randomization variance of $\widehat{t}_z(\beta)$ assuming $\beta = \beta_*$. If the sampling scheme is without replacement so that $\widehat{t}_z(\beta)$ of (3) is the Horvitz Thompson estimator, then one such quasi-randomization variance estimator is

$$\widehat{\text{var}}_{qr}(\widehat{t}_z(\beta)|\beta = \beta_*) = \sum_{i,j \in \mathcal{R}} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij} \pi_i \pi_j p_i p_j} z_i z_j^T + \sum_{i \in \mathcal{R}} \frac{1 - p_i}{p_i^2 \pi_i} z_i z_i^T, \tag{5}$$

where $p_i = p(x_i^T \beta)$. Equation (5) is derived in Appendix 1.

In the quasi-randomization framework supporting (5) one assumes the set $\mathcal{R}$ of respondents results from a two-phase sample of the target population $\mathcal{U}$. In the first phase, the sample $\mathcal{S}$ is drawn without replacement from the population with inclusion probabilities $\pi_i = Pr[\ i \in \mathcal{S}]$ and $\pi_{ij} = Pr[\ i, j \in \mathcal{S}]$. Note that $\pi_{ii} = \pi_i$. In this case, $d_i = \pi_i^{-1}$. In the second phase, $\mathcal{R}$ is a Poisson subsample of $\mathcal{S}$ with unit selection probabilities of the form $p_i = p(x_i^T\beta)$.

If the targets T_z are themselves previously estimated population totals, then it is reasonable to let

$$V = \widehat{\text{var}}_{qr}(\widehat{t}_z(\beta)|\beta = \beta_*) + \widehat{\text{var}}(T_z)$$

where $\widehat{\text{var}}(T_z)$ is a good externally determined estimate of the variance of T_z. In some applications $\widehat{\text{var}}(T_z)$ may be much greater than $\widehat{\text{var}}_{qr}(\widehat{t}_z(\beta)|\beta = \beta_*)$, in the sense that the eigenvalues of $\widehat{\text{var}}(T_z) - \widehat{\text{var}}_{qr}(\widehat{t}_z(\beta)|\beta = \beta_*)$ are large. In this case, it would be reasonable to set W to $(\widehat{\text{var}}(T_z))^{-1}$, which is not a function of β at all.

3. Partial minimization

Given a guess $\widetilde{\beta}_0$ of β and matrix $W(\widetilde{\beta}_0)$ we linearize (3) at $\widetilde{\beta}_0$ and obtain

$$T_z - \widehat{t}_z(\widetilde{\beta}_0) = \widehat{t}_z(\widetilde{\beta}) - \widehat{t}_z(\widetilde{\beta}_0) + \epsilon \approx \widehat{H}(\widetilde{\beta}_0)(\beta - \widetilde{\beta}_0) + \epsilon, \tag{6}$$

where $\widehat{H}(\widetilde{\beta}_0)$ is the $Q \times P$ matrix

$$\widehat{H}(\widetilde{\beta}_0) = \frac{\partial \widehat{t}_z(\widetilde{\beta})}{\partial \beta}\Big|_{\beta=\beta_0} = \sum_{i\in\mathcal{R}} d_i g_1(x_i^T\widetilde{\beta}_0) z_i x_i^T, \tag{7}$$

and $g_1(x_i^T\widetilde{\beta}_0)$ is the first derivative of $g(\eta) = 1/p(\eta)$ evaluated at $x_i^T\widetilde{\beta}_0$.

The (weighted) linear regression estimate $\widetilde{\beta}_1$ corresponding to (6) minimizes the objective function $U^T W(\widetilde{\beta}_0) U$ where $U = T_z - \widehat{t}_z(\widetilde{\beta}_0) - \widehat{H}(\widetilde{\beta}_0)(\beta - \widetilde{\beta}_0)$. It is given by the "update equation":

$$\widetilde{\beta}_1 = \widetilde{\beta}_0 + \left\{\widehat{H}(\widetilde{\beta}_0)^T W(\widetilde{\beta}_0)\widehat{H}(\widetilde{\beta}_0)\right\}^{-1} \widehat{H}(\widetilde{\beta}_0)^T W(\widetilde{\beta}_0)(T_z - \widehat{t}_z(\widetilde{\beta}_0)) \tag{8}$$

For simplicity, we assume $\widehat{H}(\beta)$ and $W(\beta)$ are of full rank everywhere. This will allow us to always be able to invert matrices when the need arises.

Iteration continues with $\widetilde{\beta}_1$ serving the role of $\widetilde{\beta}_0$, and $\widehat{H}$ and W updated to $\widetilde{\beta}_1$, in (8), and so on until we reach a step K, if such a step can be reached, where (at least to some small pre-specified tolerance)

$$\widehat{H}^T W(T_z - \widehat{t}_z(\widetilde{\beta})) = 0 \tag{9}$$

with the matrices $\widehat{H}$ and W evaluated at $\widetilde{\beta} = \widetilde{\beta}_K$. As the iteration starting with (8) is a Newton-Raphson type method, it is sometimes helpful to limit the size of the step from, say, $\widetilde{\beta}_{k-1}$ to $\widetilde{\beta}_k$ for some $k = 1, ..., K$. What this means is that when $\left\{\widehat{H}(\widetilde{\beta}_{k-1})^T W(\widetilde{\beta}_{k-1})\widehat{H}(\widetilde{\beta}_{k-1})\right\}^{-1}\left\{\widehat{H}(\widetilde{\beta}_{k-1})^T W(\widetilde{\beta}_{k-1})(T_z - \widehat{t}_z(\widetilde{\beta}_{k-1}))\right\}$ is deemed too large, it can be replaced with some fraction of itself in the update equation.

If W is the inverse of an estimate for the variance matrix of $\widehat{t}_z(\widetilde{\beta})$, an alternative derivation and justification of the update equation (8) follows from Thompson (1997), section 6.3. Consider the equations $\widehat{t}_z(\beta) - T_z = 0$ as P estimating equations for the Q coefficients β. If A is a $Q \times P$ matrix of constants, let $\widetilde{\beta}_A$ denote the solution to the estimating equations

$$A\widehat{t}_z(\beta) = AT_z. \tag{10}$$

A choice for A such that $\widetilde{\beta}_A$ has a minimum asymptotic variance is

$$A^* = \widehat{H}(\widetilde{\beta}_{A^*})^T \mathrm{var}(\widehat{t}_z(\widetilde{\beta}_{A^*}))^{-1},$$

where $\widetilde{\beta}_{A^*}$ results from the convergence of update equation (8).

Observe that when $W = W(\beta)$ depends upon β, our suggested procedure for estimating β_* does not minimize the objective function (4). This is because if (4) were differentiated with respect to β, then there would be a term for the derivative of W which is not accounted for in the linearization (6). In other words, letting

$$\ddot{\rho}(\beta, \gamma) = -\mathrm{logdet}(W(\gamma)) + (T_z - \widehat{t}_z(\beta))^T W(\gamma)(T_z - \widehat{t}_z(\beta)),$$

full minimization would solve the equation

$$0 = \frac{\partial \ddot{\rho}}{\partial \beta}(\widehat{\beta}, \widehat{\beta}) + \frac{\partial \ddot{\rho}}{\partial \gamma}(\widehat{\beta}, \widehat{\beta}) \tag{11}$$

whereas the partial minimization, obtained through iterated use of (8), sets only the first term of the right hand side of (11) to zero. In what follows we will distinguish partial minima from full minima by using the notation $\widetilde{\beta}$ for the former and $\widehat{\beta}$ for the latter.

In the appendices we first discuss an example of conditions sufficient to establish the consistency of $\widehat{\beta}$ under full minimization and then the asymptotic equivalence of the partial minimum $\widetilde{\beta}$. Full and partial minimization usually yielded very similar results in our own empirical investigations, with the former taking longer to compute. In fact, finding a full minimum was extremely difficult when there was no partial-minimum solution to use as an initial guess in the iterative process to full minimization. Given that, the asymptotic equality of the two minima, and the reliance on an asymptotic framework in our analysis, most of the remainder of text concerns partial-minimization results.

4. Some choices for W

When $P = Q$, equation (8) reduces to

$$\widetilde{\beta}_1 = \widetilde{\beta}_0 + \widehat{H}^{-1}(T_z - \widehat{t}_z(\widetilde{\beta}_0)). \tag{12}$$

Thus, in this case, the form of W is irrelevant for the update equation. Indeed (12) is the Newton-Raphson update equation for solving the equation $T_z = \widehat{t}_z(\widetilde{\beta})$, and the solution, if it exists, will also minimize (4), at least for any W which does not depend upon β.

When $P > Q$, we propose setting $W(\beta)$ to the inverse of $V = \widehat{\mathrm{var}}_{qr}(\widehat{t}_z(\beta)|\beta = \beta_*) + \widehat{\mathrm{var}}(T_z)$, where the latter term is provided to us from external sources (and may be 0). If the sample $\mathcal{S}$ is selected without replacement, then (5) is an unbiased estimate of $\mathrm{var}_{qr}(\widehat{t}_z(\beta)|\beta = \beta_*)$.

For a stratified multistage design with primary sampling units (PSU's) chosen with replacement, let $h = 1, \ldots, H$ index the strata. Suppose n_h PSU's are chosen with replacement in stratum h. Let $n_h q_{hi}$ be the expected number of times PSU i from stratum h will appear in the first stage sample

Changing the meaning of i slightly, let $\mathcal{R}_{hi}$ be the set of subsampled and then responding elements in the i-th PSU selected from stratum h, and let d_{hij} be the sampling weight (before nonresponse adjustment) for element $j \in \mathcal{R}_{hi}$. Define

$$\widehat{t}_{zhi}(\beta) = n_h q_{hi} \sum_{j \in \mathcal{R}_{hi}} \frac{d_{hij}}{p(x_{hij}^T \beta)} z_{hij}, \text{ and}$$

$$\widehat{t}_{zh}(\beta) = n_h^{-1} \sum_{i=1}^{n_h} q_{hi}^{-1} \widehat{t}_{zhi}(\beta).$$

Then $\widehat{t}_z(\beta) = \sum_h \widehat{t}_{zh}(\beta)$ is an unbiased estimate of $t_z(\beta) = \sum_{\mathcal{U}} p(x^T\beta)^{-1} p(x^T\beta_*) z$. When $\beta = \beta_*$ its quasi-randomization variance can be unbiasedly estimated by

$$\widehat{\text{var}}_{wr}(\widehat{t}_z(\beta)|\beta = \beta_*) = \tag{13}$$
$$\sum_h \{n_h(n_h - 1)\}^{-1} \sum_{i=1}^{n_h} (q_{hi}^{-1}\widehat{t}_{zhi}(\beta) - \widehat{t}_{zh}(\beta))(q_{hi}^{-1}\widehat{t}_{zhi}(\beta) - \widehat{t}_{zh}(\beta))^T.$$

For simplicity we are assuming in (13) that no element is selected more than once into the sample before nonresponse (see the remark following Proposition 2 in Appendix 2).

Finally, notice that when $\widehat{\text{var}}(T_z) = 0$, the estimate $\widetilde{\beta}$ is unchanged when $\text{var}_{qr}(\widehat{t}_z(\beta))$ is only estimated up to a multiplicative constant within W^{-1}. This suggests invoking the spirit of design effects (Kish 1965), that is assuming that the true variance is proportional to the variance that would have resulted from a with-replacement simple random sample of size r drawn from a population of size N, where r is the respondent size. The latter variance can be estimated, taking into account the actual unequal selection and response probabilities of the respondent sample, by:

$$\widehat{\text{var}}_{srs}(\widehat{t}_z(\beta)|\beta = \beta_*) = \frac{N^2}{r(N-1)} \sum_{i \in \mathcal{R}} d_i p_i^{-1} (z_i - \overline{z})(z_i - \overline{z})^T,$$

where

$$\overline{z} = (\sum_{i \in \mathcal{R}} d_i p_i^{-1})^{-1} \sum_{i \in \mathcal{R}} d_i p_i^{-1} z_i,$$

and, as before, $p_i = p(x_i^T \beta)$. In practice, the scalar multiple $N^2 r^{-1}(N-1)^{-1}$ can be dropped from $\widehat{V}$ within $W = \widehat{V}^{-1}$.

5. The calibration-weighted estimator for a population total and its quasi-randomization variance

Let y_i be a vector of variables of interest. The purpose of calibration weighting is to estimate population totals for the components of y_i or functions of those totals, like population ratios. Typically, the components of y_i will not include the benchmark variables in z_i as the latter's totals are already known or have been alternatively estimated. They may include model variables in x_i. In principle, however, they could include both types of variables.

Our calibration-weighted estimator for the total t_y is

$$\widehat{t}_y(\widetilde{\beta}) = \sum_{i \in \mathcal{R}} \frac{d_i}{p(x_i^T \widetilde{\beta})} y_i = \sum_{i \in \mathcal{R}} d_i g(x_i^T \widetilde{\beta}) y_i = \sum_{i \in \mathcal{R}} c_i y_i, \tag{14}$$

where $c_i = d_i g(x_i^T \widetilde{\beta})$ is the calibration weight for unit i. Define

$$\begin{aligned} \widehat{H}_y &= \frac{\partial \widehat{t}_y}{\partial \beta}(\widetilde{\beta}) = \sum_{i \in \mathcal{R}} d_i g_1(x_i^T \widetilde{\beta}) y_i x_i^T, \text{ and} \\ B &= \widehat{H}_y (\widehat{H}^T W \widehat{H})^{-1} \widehat{H}^T W, \end{aligned}$$

where $\widehat{H} = \widehat{H}_z$ is the matrix of partial derivatives (7) which, like W, is evaluated at $\widetilde{\beta}$. Note that the components of $\widehat{H}_y / N$ are $O_P(1)$.

Writing y_i as $Bz_i + (y_i - Bz_i)$, we have

$$\begin{aligned} \widehat{t}_y(\widetilde{\beta}) &= B\widehat{t}_z(\widetilde{\beta}) + \sum_{i \in \mathcal{R}} \frac{d_i}{p(x_i^T \widetilde{\beta})}(y_i - Bz_i) \\ &= BT_z + \sum_{i \in \mathcal{R}} d_i g(x_i^T \widetilde{\beta})(y_i - Bz_i). \end{aligned}$$

The latter equality follows from (9). Note that

$$g(x_i^T\beta_*) - g(x_i^T\widetilde{\beta}) = g_1(x_i^T\widetilde{\beta})x_i^T(\widetilde{\beta} - \beta_*) + \frac{1}{2}g_2(\xi_i)\{x_i^T(\widetilde{\beta} - \beta_*)\}^2, \qquad (15)$$

for some ξ_i between $x_i^T\widetilde{\beta}$ and $x_i^T\beta_*$, where $g_2(.)$ is the second derivative of g. Thus

$$\begin{aligned}
\widehat{t}_y(\widetilde{\beta}) &= BT_z + \sum_{i\in\mathcal{R}} d_i g(x_i^T\beta_*)(y_i - Bz_i) \\
&- (\widehat{H}_y - B\widehat{H})(\widetilde{\beta} - \beta_*) - \frac{1}{2}\sum_{i\in\mathcal{R}} d_i g_2(\xi_i)\{x_i^T(\widetilde{\beta} - \beta_*)\}^2(y_i - Bz_i) \qquad (16) \\
&= BT_z + \sum_{i\in\mathcal{R}} d_i g(x_i^T\beta_*)(y_i - Bz_i) + O_P(N/n),
\end{aligned}$$

assuming all the components of $\sum_{i\in\mathcal{R}} d_i|g_2(\xi_i)|(x_ix_i^T)|y_i - Bz_i|$ are $O_P(N)$. This leads to

$$(\widehat{t}_y(\widetilde{\beta}) - t_y)/N = (\widehat{t}_u(\beta_*) - t_u)/N + O_P(1/n), \qquad (17)$$

where $u_i = y_i - Bz_i$.

Unfortunately, u_i depends on values for responding units other than i through B, so t_u is not really a vector of population totals. Under suitable additional assumptions on the y_i, $B - B_* = O_P(1/n^{1/2})$, where B_* is the probability limit of B, we can infer that $(B - B_*)(\widehat{t}_z(\beta_*) - T_z)/N = O_P(1/n)$. Thus, we can replace the u_i in (17) by $u_{i*} = y_i - B_*z_i$ and still have

$$(\widehat{t}_y(\widetilde{\beta}) - t_y)/N = (\widehat{t}_{u*}(\beta_*) - t_{u*})/N + O_P(1/n). \qquad (18)$$

This suggests estimating $\mathrm{var}_{qr}(N^{-1}\widehat{t}_y(\widetilde{\beta}))$ with an estimate of $\mathrm{var}_{qr}(N^{-1}\widehat{t}_{u*}(\beta_*))$.

Recall that $\widehat{t}_y(\widetilde{\beta})$ has made use of a quasi-random model to adjust for nonresponse. The variance of the respondent indicator function for sampled unit i (which is 1 if i responds and 0 otherwise) is $p(x_i^T\beta_*)(1 - p(x_i^T\beta_*))$. We will make the standard assumption that these indicator functions are independent conditionally on the x_i.

Assume now that the T_z is a known vector of population totals; that is the benchmark totals are known without error.

The component of the randomization variance of $\widehat{t}_y(\widetilde{\beta})$ due to nonresponse can be estimated by

$$\widehat{\text{var}}_{add} = \sum_{i \in \mathcal{R}} \frac{d_i^2}{p(x_i^T \widetilde{\beta})^2}(1 - p(x_i^T \widetilde{\beta}))(y_i - Bz_i)(y_i - Bz_i)^T. \tag{19}$$

Note that the right hand side of (19) does not have a component that accounts for the uncertainty in estimating β_* because the terms involving $\widetilde{\beta} - \beta_*$ in equation (16) are asymptotically ignorable.

When the original sample is drawn randomly without replacement, a reasonable estimator for the (quasi-randomization) variance of $\widehat{t}_y(\widetilde{\beta})$ can be computed using the right hand side of (5) with p_i set to $p(x_i^T \widetilde{\beta})$ and z_i replaced by u_i. That is to say,

$$\widehat{\text{var}}(\widehat{t}_y(\widetilde{\beta})) = \sum_{i,j \in \mathcal{R}} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij} \pi_i \pi_j p_i p_j} u_i u_j^T + \sum_{i \in \mathcal{R}} \frac{1 - p_i}{p_i^2 \pi_i} u_i u_i^T, \tag{20}$$

where $p_i = p(x_i^T \widetilde{\beta})$. Proving $\widehat{\text{var}}(\widehat{t}_y(\widetilde{\beta}))$ is asymptotically unbiased is not trivial when the right hand side of of (20) has $O(n^2)$ terms. See Kim et al. (2006) for sufficient additional restrictions on the coefficients in that situation.

Similarly when the original sample is drawn using a stratified multi-stage routine and the first stage drawn with replacement, one can compute the variance estimator for $\widehat{t}_y$ under a first-stage-with-replacement design as described in Section 4, again with p_i set to $p(x_i^T \widetilde{\beta})$ and u_i replacing z_i.

There is an additional component in the asymptotic variance of $\widehat{t}_y(\widetilde{\beta})$ when T_z is not known with certainty but is estimated from an independent external source or sources. An obvious measure for this component is

$$\widehat{\text{var}}_{ext} = B \widehat{\text{var}}(T_z) B^T, \tag{21}$$

where $\widehat{\text{var}}(T_z)$ is an externally-provided estimate for the variance of T_z.

Finally we note that when a full minimum $\widehat{\beta}$ is used, equation (9) can replaced by equation (35) of Appendix 3. This implies that, to the order of approximation used here, the results of this section still apply if the partial minimum $\widetilde{\beta}$ is everywhere replaced by $\widehat{\beta}$.

6. Example: the 2002 Census of Agriculture

We focus our attention in this section on adjusting for unit nonresponse in the 2002 Census of Agriculture conduction by the U.S. National Agricultural Statistics Service (NASS). The agency used a weighting-class/poststratification approach to adjust for whole-unit nonresponse in each state (the two are the same for a census). Simplifying slightly, sampling units receiving Census forms were divided in mutually exclusive response groups (poststrata) based on size class as measured by expected sales and on whether the unit had responded to an agency survey since 1997.

We first parallel the routine actually used by NASS and reweighted responding units in a state using these five mutually exclusive groups:

- Z-Group 1: Expected 2002 sales less than \$2,500;
- Z-Group 2: Expected 2002 sales between \$2,500 and \$9,999;
- Z-Group 3: Expected 2002 sales between \$10,000 and \$49,999 and previously reported survey data from 1997 or later;
- Z-Group 4: Expected 2002 sales greater than or equal to \$50,000 and reported survey data from 1997 or later;
- Z-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 each state data set belonged to one of these five groups. Mathematically, $z_i = (z_{i1}, z_{i2}, \cdots, z_{i5})^T$, where $z_{ih} = 1$ when i is in Z-Group h and $z_{ih} = 0$ otherwise.

All $d_i = 1$ because units in the list frame were sent Census forms. The poststatification weight for responding unit i is $c_i = a_i = N_h/r_h$ when the unit is in Z-Group h, where N_h is the number of sampling units in the agency list frame in Z-group h and r_h is the number of responding entities in Z-group h.

We contrast this poststratification with an alternative calibration using these same five Z-group indicator variables as benchmarks variables but different model variables. We present here results from five state, California (CA), Delaware (DE), Illinois (IL), Louisiana (LA), and South Dakota (SD).

In the alternative calibration, we model response as a logistic function of three calibration x-variables: an intercept, *logsales* the logarithm of the actual annual sales in 2002, truncated to the range \$1,000 to \$100,000, and $s97$ an indicator variable for whether or not the farm responded to a survey since the 1997 Census of Agriculture. For these runs, the fitted benchmark totals $\widehat{t}_z(\widehat{\beta})$ differ from the benchmark target totals in each state. Both sets are displayed in Table 1.

The table shows how well calibration does at estimating the benchmark totals (recall that exact equality is not expected because there are less model variables than benchmark variables). This is a simple check on the appropriateness of the model. Except for Illinois, most of the estimates are within two standard errors the benchmark totals.

In the next section we perform a simulation study using this data. Our conclusion is that if the response model (that is, that the probability of response is the logistic back link of a linear combination of an intercept, *logsales*, and $s97$) is correct then calibration works well. In contrast to that Table 1 appears to show that this response model is, in fact, unlikely to be correct in Illinois.

We now use calibration to estimate the total number of active farms. In this context our y variable is a 0-1 variable for being an active farm. We compare the existing NASS approach of poststratification within Z-groups with calibration using the three-variable model (intercept, *logsales* and *s97*). The results are given in Table 2. Standard errors are calculated using (20) with all the $\pi_i = 1$, which is equivalent to (19). For poststratification, variance estimation assuming the respondent sample results from simple random sampling within poststrata produces identical answers asymptotically and is within roundoff error in this application.

In interpreting Table 2, it is important to realize that the standard errors are

computed assuming the underlying model for that fit is correct. It is well known, and simulations in the next section document, that a misspecified response model can introduce biases that dwarf the estimated standard errors calculated in Table 2.

The poststratification standard error assumes that response probability takes on one of five possible values. The choice of one of these five possible values depends upon NASS assigned expected sales and participation in surveys since 1997.

The three-variable-calibration model assumes that response probability varies continuously with actual sales given survey participation. Although this seems reasonable and results in estimated standard error only slightly higher than those for poststratification, we saw in Table 1 that this model is not supported by the data. Clearly, more research on model-fitting techniques for response modelling through calibration is needed.

7. Simulations

As discussed in the previous section, an incorrect model is a possible explanation for the poor fit provided by the three-variable models evidenced in Table 1. To explore this question further, we conducted simulations. Our conclusion is that, if the form of the response model is correct, then calibration performs just fine.

Our approach was as follows. For each of the five states, assuming the fitted three-variable response model is correct, response probabilities were calculated for each of the respondents to the Census. Then a synthetic state was created using only the respondents to the Census, together with their response probabilities. In particular, new target benchmarks were calculated as sizes of the Z-groups within the synthetic state populations, that is the respondents within the original states.

Each Monte Carlo replication consisted of creating respondents within the synthetic states using the assigned response probabilities and fitting a calibration

model to the synthetic state using these respondents. One thousand Monte Carlo simulations were done for each state. Table 3 gives for each of the synthetic state simulations, the mean of the fitted targets and their sample standard deviation, together with square root of the mean estimated variance given by equation (20).

Examining Table 3, all mean fitted targets are within two standard errors of their benchmarks; in CA, DE, and LA they are well within one standard error. It should be noted that in CA, DE, and LA the maximum of the assigned response probabilities is 0.925, 0.859, and 0.917 respectively. By contrast, 21.5% of the synthetic IL population has an assigned probability in excess of 0.95. In SD the corresponding percentage is 23.2%. Nonlinearities in the response probability link function are most severe for probabilities close to 1 or to 0. Thus the biases in the fitted targets that possibly exist for these two populations are likely due to the large proportion of these synthetic populations with high response probabilities.

When $W(\beta) = \widehat{\text{var}}(\widehat{t}_z(\beta))^{-1}$, equations (36) and (37) from Appendix 3 suggest estimating $\text{var}(\widehat{\beta})$ and $\text{var}(\widetilde{\beta})$ by $(\widehat{H}^T W \widehat{H})^{-1}$ (where $\widehat{H}$ and W are evaluated at $\widehat{\beta}$ and $\widetilde{\beta}$ respectively). Table 4 gives, for each synthetic state, the true β_*, the mean of the fitted $\widetilde{\beta}$, their sample standard deviations, and the square root of the mean of the variance estimates.

Table 4 also gives the results for each state, the total number of farms t_y and its estimate $\widehat{t}_y(\widetilde{\beta})$ given by equation (14). The sample estimate of the variance of $\widehat{t}_y(\widetilde{\beta})$ is given by equation (20).

Table 4 in general shows good fit.

We also fit two misspecified models to the synthetic states. In the first, post-stratification/reweighting using the five Z-groups was employed. In the second, five mutually-exclusive X-groups were formed among the respondents paralleling the five Z-groups but using census-reported sales in place of expected sales. A five-component vector of X-group-membership indicators provided the model variables for calibration while the five-component vector of Z-group indicators provided the benchmark variables.

The calibration equations in the five X-group variable calibrations can be expressed as

$$\sum_{g=1}^{5} \widetilde{a}_g r_{hg} = N_h \tag{22}$$

$$\widetilde{a}_g = (p(\widehat{\beta}_g))^{-1} \tag{23}$$

where r_{hg} is the number of respondents in X-group g and Z-group h. Unless the 5×5 matrix $(r_{hg})_{g,h=1,\cdots,5}$ is singular, (22) has a unique solution $\widetilde{a}_g$, and $c_i = a_i = \widetilde{a}_g$ is the calibration weight for respondent i in X-Group g.

Sometimes the unique solution to (22) corresponds to response probabilities outside the range $(0, 1)$. This anomaly can be handled by setting the response probabilities of the units in the offending X-group to one and removing them from the calibration. We now have four model and five benchmark variables for the remaining units, so the fitted targets can differ from preset targets. This anomaly occurred for 0%, 26.3%, 89.3%, 24.5%, and 48.8% of the runs in CA, DE, IL, LA, and SD respectively. An analogous anomaly is mathematically impossible with poststratification.

Recall that in the synthetic states, the response probability was constructed to depend upon actual sales and participation in a previous survey. The three variable calibrations, whose results are given in Tables 3 and 4, hypothesize the correct functional form for the dependence of the response probability on these two variables (but fit the coefficients). The five X-group calibrations use the correct variables to model the response probability but implicitly hypothesize an incorrect functional form for its dependence upon these variables. By contrast, poststratification models response probability using the wrong variables (NASS expected sales instead of actual sales).

For these two response models, Table 5 gives the total number of farms, and the mean of the sample estimates, their empirical standard deviation, and the square root of the mean of the sample estimates of their variance as calculated using equation (20). Examining these two misspecified models, it is clear that total

empirical error is almost always dominated by bias. Although both models produced downward biased estimates, the bias is in most cases substantially reduced when the appropriate variables are used for modelling the response mechanism. This provides a strong argument for separating the calibration variables from the response model variables.

8. Some Concluding Remarks

8.1 Calibration

Quasi-randomization modelling of nonresponse (also known as "response-propensity modelling") assumes that each element in a sample has an independent, conditionally on the model variables, probability of survey response. Conventionally, the functional form of the model is assumed known even if the parameters of the model are not. Moreover, the values of the model variables need to be known for all sampled elements, respondents and nonrespondents alike. This allows the fitting of the response model directly to the full-sample data treating response/nonresponse as a binary dependent variable. Alternatively, if a benchmark population (or full-sample) total is known for each of the model variables, the model can be fit implicitly through calibration, as shown by Folsom and Singh (2000) among others.

We have shown how to extend the theory and practice of calibration to fitting a response function with Q model variables given $P \geq Q$ benchmark (calibration) variables. Although population totals, which can be estimates from external sources, need to be known for the benchmark variables, they do not need to known for model variables. In fact, it is not even necessary to know the model-variable values for sampled nonrespondents.

When $P > Q$, calibration-weighted estimates of benchmark-variable aggregates will not generally equal externally-provided benchmark totals (as they will when $P = Q$). This is paradoxically advantageous because it allows statistically testing of the difference between these two quantities (as was done in Tables 3 and 5). This provides one means for assessing the validity of the response model.

8.2 Partial minimization

The appendices establish conditions for the existence of both a full minimum to the objective function (equation (4)) or a partial minimum solution to (9), at least given a large-enough sample. These solutions both serve as consistent estimators for the response-model parameter. In our own empirical investigations, we found it much simpler to reach an iterative "partial-minimization" solution to equation (9). The two solutions were generally close. We recommend statisticians use the partial-minimization approach in practice.

The two approaches differ only when the matrix W is itself a function of β, for example, when it is be the inverse of the variance of $\widehat{t}_z(\beta)$. We found in simulations (not shown) that using the identity matrix in place of W in (9) decreased the efficiency of the resulting estimates when derived under the correct response model. Nevertheless, these estimates were consistent, as our theory anticipated.

By setting W equal to the inverse of a quasi-randomization with-replacement variance estimator for $\widehat{t}_z(\beta)$ in Appendix 2, we established sufficient conditions for the existence of a minimum of the objective function given a large-enough sample. In practice, we recommend using the without-replacement variance estimator in this context when practical (e.g., for single-stage original samples with known joint selection probabilities). Recomputing W assuming a with-replacement original sample and $n = N$ in our simulations did not appreciably change the results for the estimators or their variances.

8.3 Coverage error

It is a simple matter to adapt most of the results in this paper to the situation where calibration is used to adjust for coverage errors, whether from frame under-coverage or unit duplication. In this context, p_i is the expected number of times population unit i is in the frame. Under the assumed quasi-randomization model

this value is independent of the actual sample drawn and of how many times other elements of the population are in the frame.

The possibility a unit appearing more than once on a frame does cause some small differences in variance estimation. When there is no possibility of unit duplication, the variance of the function indicating whether population unit i is in the frame is $p(x_i^T\beta_*)(1-p(x_i^T\beta_*))$, paralleling the situation with survey response. When the frame contains potential duplication, however, the variance of the expected number of times a population unit is no longer $p(x_i^T\beta_*)(1-p(x_i^T\beta_*))$. Consequently, its value will need to be assumed for the analogue of (19), a measure of the added variance due to frame errors, to be derived.

When using calibration to adjust for coverage errors, it is likely that the components of T_z will be estimates coming from external sources. If the (internal) sample size is fairly large, it may be reasonable to set $W = (\widehat{\text{var}}(T_z))^{-1}$, where $\widehat{\text{var}}(T_z)$ is externally provided.

8.4 Final remarks

Returning to response modelling, although the following applies equally well to coverage modelling, much work is needed in determining how to select model and benchmark variables in practice and assessing the usefulness of an asymptotic theory on finite samples. Additional complication arise when the targets of the benchmark variables are themselves potentially subject to sampling and measurement errors. Nevertheless, by allowing a separation between the model and benchmark variables, the approach to calibration developed here may open the door to more plausible modelling of the response mechanism.

Avoiding response and coverage modelling as much as possible remains a prudent policy. Even the cleverest model assumptions are difficult to test and prone to failure. Unfortunately, eschewing models will not be viable option as surveys based on incomplete frames or suffering from small response rates become increasingly common.

Acknowledgements

The authors wish to thank the referees for their close reading of the manuscript and many thoughtful suggestions.

Appendix 1

Proof of variance formula (5)

Let $\widehat{t}_z$ denote $\widehat{t}_z(\beta_*)$. Then

$$\mathrm{var}(\widehat{t}_z) = \mathrm{var}(E(\widehat{t}_z|\mathcal{S})) + E(\mathrm{var}(\widehat{t}_z|\mathcal{S}))$$

Using Särndal et al. (1992) Result 9.3.1

$$\begin{aligned} \mathrm{var}(E(\widehat{t}_z|\mathcal{S})) &= \sum_{i,j\in\mathcal{U}} \frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j} z_i z_j^T \\ E(\mathrm{var}(\widehat{t}_z|\mathcal{S})) &= \sum_{i\in\mathcal{U}} \frac{1-p_i}{p_i\pi_i} z_i z_i^T. \end{aligned}$$

and the sample estimates of these variance components are

$$\begin{aligned} \widehat{\mathrm{var}}(E(\widehat{t}_z|\mathcal{S})) &= \sum_{i\neq j\in\mathcal{R}} \frac{\pi_{ij}-\pi_i\pi_j}{\pi_{ij}\pi_i\pi_j p_i p_j} z_i z_j^T + \sum_{i\in\mathcal{R}} \frac{1-\pi_i}{\pi_i^2 p_i} z_i z_i^T \\ \widehat{E}(\mathrm{var}(\widehat{t}_z|\mathcal{S})) &= \sum_{i\in\mathcal{R}} \frac{1-p_i}{p_i^2\pi_i^2} z_i z_i^T. \end{aligned}$$

Combining these two yields equation (5). □

Appendix 2

Quasi-randomization consistency of the full minimum $\widehat{\beta}$

The estimate $\widehat{\beta}$ minimizes the slightly redefined objective function

$$\begin{aligned} n^{-1}\rho(\beta) &= -n^{-1}\mathrm{logdet}\left(\frac{N^2}{n}W(\beta)\right) \\ &+ \left(N^{-1}T_z - N^{-1}\widehat{t}_z(\beta)\right)^T \left(\frac{N^2}{n}W(\beta)\right) \left(N^{-1}T_z - N^{-1}\widehat{t}_z(\beta)\right) \end{aligned} \tag{24}$$

where n denotes the sample size before nonresponse, N the population size, and T_z the vector of calibration targets. Now

$$\widehat{t}_z(\beta) = \sum_{i \in \mathcal{R}} \frac{d_i}{p(x_i^T \beta)} z_i \tag{25}$$

has expected value under the quasi-randomization model

$$t_z(\beta) = \sum_{i \in \mathcal{U}} \frac{p(x_i^T \beta_*)}{p(x_i^T \beta)} z_i. \tag{26}$$

Conditionally on $\mathcal{S}$, the expected value of $\widehat{t}_z(\beta)$ is

$$\widehat{t}_{z\mathcal{S}}(\beta) = E(\widehat{t}_z(\beta)|\mathcal{S}) = \sum_{i \in \mathcal{S}} \pi_i^{-1} \frac{p(x_i^T \beta_*)}{p(x_i^T \beta)} z_i. \tag{27}$$

Assumption 1 *We assume that x_i and β are constrained to lie in compact sets and that these compact sets are such that $p(x^T\beta)$ is bounded away from 0.*

Assumption 2 *The limits $\lim N^{-1} t_z(\beta)$ and $\lim N^{-1} T_z$ converge, with the former limit converging uniformly in β. These limits satisfy*

$$\lim N^{-1} t_z(\beta_*) = \lim N^{-1} T_z$$
$$\lim N^{-1} t_z(\beta) \neq \lim N^{-1} T_z$$

for $\beta \neq \beta_*$.

Let

$$\rho_0(\beta) = (\lim N^{-1} T_z - \lim N^{-1} t_z(\beta))^T W_0(\beta) (\lim N^{-1} T_z - \lim N^{-1} t_z(\beta)),$$

where $W_0(\beta)$ is a positive definite symmetric matrix to be specified shortly.

If Assumption 2 is true, then $\rho_0(\beta)$ is uniquely minimized at $\beta = \beta_*$. Wald's proof of the consistency of the maximum likelihood estimate (see, for example, Silvey (1975)) can be used to show the consistency of $\widehat{\beta}$ if it can be established that

$$n^{-1} \rho(\beta) \rightarrow_P \rho_0(\beta)$$

uniformly in β. In other words, we need to establish that, uniformly in β,

$$\begin{aligned} N^{-1}\widehat{t}_z(\beta) - N^{-1}t_z(\beta) &\to_P 0 \\ \frac{N^2}{n}W(\beta) &\to_P W_0(\beta), \end{aligned} \tag{28}$$

where $\to_P$ refers to convergence in probability with respect to both the sampling design and the response model.

Wald's proof establishes that for large enough samples, the minimum $\widehat{\beta}$ to equation (24) exists.

A formal asymptotic structure must postulate a sequence of populations and sampling designs, indexed in what follows by r, so that the population size N_r can grow along with the sample size n_r. Several examples of such asymptotic structure can be found in Fuller and Isaki (1981). This paper can be applied to many useful fixed sample size without replacement designs. Lemma 1 of Fuller and Isaki (1981) would apply, for example, if the sampling design is a stratified two stage cluster sample in which the number of strata is fixed (in r), the number of PSUs in each stratum increases as $r \to \infty$, the first stage sampling fraction within each stratum of PSU's is bounded away from 0 and 1, and the second stage sampling fractions are bounded away from 0. We will establish consistency of $\widehat{\beta}$ within this asymptotic framework, using Fuller and Isaki type assumptions on the original sampling scheme. Although not all designs of interest can be accommodated within this asymptotic framework, we offer these proofs to illustrate the reasoning that might apply elsewhere.

Let the inclusion probabilities for the r-th universe and design be denoted by $\pi_{i(r)}$ and $\pi_{ij(r)}$, so that $d_i = \pi_{i(r)}^{-1}$. We require that the original element sample size n_r be fixed, but the respondent sample size is random. Assumption 3 is a special case of the assumptions in Fuller and Isaki's Lemma 1.

Assumption 3 *Assume that for all r and all $i \neq j$*

$$\pi_{i(r)}\pi_{j(r)} - \pi_{ij(r)} \leq \alpha n_r^{-1}\pi_{i(r)}\pi_{j(r)}$$

$$N_r^{-2} n_r \sum_{i \in \mathcal{U}_r} \pi_{i(r)} \left[\pi_{i(r)}^{-1} z_i - n_r^{-1} t_{z(r)}(\beta_*) \right] \left[\pi_{i(r)}^{-1} z_i - n_r^{-1} t_{z(r)}(\beta_*) \right]^T \ll M_2$$

where α is a fixed constant, M_2 a fixed positive definite symmetric matrix, $t_{z(r)}(\beta)$ is defined using equation (26) from the r-th universe $\mathcal{U}_r$, and, for symmetric matrices A and B, $A \ll B$ means $B - A$ is positive semi-definite.

Proposition 1 *Under assumptions 1, 2, and 3,*

$$N^{-1}\widehat{t}_z(\beta) - N^{-1} t_z(\beta) = O_P(n^{-.5})$$

uniformly in β.

Proof: For notational simplicity, we drop the index r in this proof. Let $u_i = p(x_i^T\beta)^{-1} p(x_i^T\beta_*) z_i$. Using Assumption 1, $p(x_i^T\beta)^{-1} p(x_i^T\beta_*) < C$ for all β and all x_i. Thus since $n = \sum_{i \in \mathcal{U}} \pi_i$,

$$\begin{aligned}
& N^{-2} n \sum_{i \in \mathcal{U}} \pi_i \left\{ \pi_i^{-1} u_i - n^{-1} t_z(\beta) \right\} \left\{ \pi_i^{-1} u_i - n^{-1} t_z(\beta) \right\}^T \\
= \;& N^{-2} n \sum_{i \in \mathcal{U}} \pi_i^{-1} u_i u_i^T - N^{-2} t_z(\beta) t_z(\beta)^T \\
\ll \;& C^2 N^{-2} n \sum_{i \in \mathcal{U}} \pi_i^{-1} z_i z_i^T \\
\ll \;& C^2 \left\{ M_2 + N^{-2} t_z(\beta_*) t_z(\beta_*)^T \right\}.
\end{aligned}$$

Examining the proof of Lemma 1 in Fuller and Isaki (1981), it follows that $N^{-1}\widehat{t}_{z\mathcal{S}}(\beta) - N^{-1} t_z(\beta)$ is $O_P(n^{-.5})$, uniformly in β.

Notice also that the above establishes that $N^{-2} \sum_{i \in \mathcal{U}} \pi_i^{-1} z_i z_i^T$ is $O(n^{-1})$. Let $T = N^{-1}\widehat{t}_z(\beta) - N^{-1}\widehat{t}_{z\mathcal{S}}(\beta)$. Then $E(T|\mathcal{S}) = 0$ and

$$\begin{aligned}
\mathrm{var}(T|\mathcal{S}) &= N^{-2} \sum_{i \in \mathcal{S}} \pi_i^{-2} \frac{p(x_i^T\beta_*) - p(x_i^T\beta_*)^2}{p(x_i^T\beta)^2} z_i z_i^T \\
E[\mathrm{var}(T|\mathcal{S})] &= N^{-2} \sum_{i \in \mathcal{U}} \pi_i^{-1} \frac{p(x_i^T\beta_*) - p(x_i^T\beta_*)^2}{p(x_i^T\beta)^2} z_i z_i^T,
\end{aligned}$$

which is $O(n^{-1})$, uniformly in β. Thus $T = O_P(n^{-.5})$, uniformly in β which establishes the Proposition. □

We now turn to establishing (28), that is $N^2 n^{-1} W(\beta) \to_P W_0(\beta)$, where $W_0(\beta)$ is positive definite symmetric. First of all we note that $W(\beta)$ can be scaled by a constant without changing $\widehat{\beta}$. Thus if $c_N W(\beta)$ converges uniformly to a positive definite symmetric matrix, we can redefine $W(\beta)$ so that (28) holds. Thus, for example, if $W(\beta) = \widehat{\text{var}}(T_z)^{-1}$ and $\widehat{\text{var}}(T_z)$ is externally defined, one has to assume the existence of a c_N which makes (28) true.

Here we will show that if $W(\beta)$ is the inverse of a with replacement estimated covariance matrix, then (28) will often hold (under the original design).

Proposition 2 *If an element sample $\mathcal{S}$ is drawn with replacement and expected counts π_i, then the variance $\text{var}_{wr}(\widehat{t}_z)$ of $\widehat{t}_z = \sum_{i \in \mathcal{R}} \pi_i^{-1} p_i^{-1} z_i$ is*

$$\text{var}_{wr}(\widehat{t}_z) = \sum_{i \in \mathcal{U}} \pi_i^{-1} p_i^{-1} z_i z_i^T - n^{-1} t_z t_z^T. \tag{29}$$

It can be unbiasedly estimated by

$$\widehat{\text{var}}_{wr}(\widehat{t}_z) = \frac{n}{n-1} \sum_{i \in \mathcal{R}} \left(\pi_i^{-1} p_i^{-1} z_i - n^{-1} \widehat{t}_z\right) \left(\pi_i^{-1} p_i^{-1} z_i - n^{-1} \widehat{t}_z\right)^T \tag{30}$$

Proof: Since $E(\widehat{t}_z | \mathcal{S}) = \sum_{i \in \mathcal{S}} \pi_i^{-1} z_i$, Särndal et al. (1992) formulas (3.6.14) and (3.5.5) yield

$$\begin{aligned} \text{var}(E(\widehat{t}_z|\mathcal{S})) &= \sum_{i \in \mathcal{U}} \pi_i \left(\pi_i^{-1} z_i - n^{-1} t_z\right) \left(\pi_i^{-1} z_i - n^{-1} t_z\right)^T \\ E(\text{var}(\widehat{t}_z|\mathcal{S})) &= \sum_{i \in \mathcal{U}} \frac{1 - p_i}{\pi_i p_i} z_i z_i^T. \end{aligned}$$

These two sum to the expression in (29). Now

$$\begin{aligned} E(\widehat{\text{var}}_{wr}(\widehat{t}_z)) &= \frac{n}{n-1} E\left(\sum_{i \in \mathcal{R}} \frac{z_i z_i^T}{\pi_i^2 p_i^2} - n^{-1} \widehat{t}_z \widehat{t}_z^T\right) \\ &= \frac{n}{n-1} \left[\sum_{i \in \mathcal{U}} \frac{z_i z_i^T}{\pi_i p_i} - n^{-1} \left\{\text{var}_{wr}(\widehat{t}_z) + t_z t_z^T\right\}\right] \end{aligned}$$

which yields (29) with some algebra. □

Remark: Recall our intention is to use equation (30) to define a suitable $W(\beta)$ for estimating β with a nonreplacement sample. Consequently we felt free to

take certain liberties in the statement of Proposition 2. Formally, we require that $\mathcal{S} = (j_1, ..., j_n)$ is an ordered sample with $Pr[j_k = i] = n^{-1}\pi_i$ for all $k = 1, \ldots, n$ and that $\mathcal{R}$ is an ordered Poisson subsample with $Pr[j_k \in \mathcal{R}] = p_{j_k}$. In particular it is assumed that the events $j_k \in \mathcal{R}$ and $j_l \in \mathcal{R}$ are independent, conditionally on $\mathcal{S}$. Although this independence assumption is dubious when $j_k = j_l$ in the interpretation of $\mathcal{R}$ as respondent set, this contingency does not arise in a nonreplacement sample.

Assumption 4 *Write* $p_i = p(x_i^T\beta_*)$ *and let*

$$V_{z(r)}(\beta) = \sum_{i\in\mathcal{U}_r} \pi_{i(r)} p_i \left(\frac{z_i}{\pi_{i(r)} p(x_i^T\beta)} - n^{-1} t_{z(r)}(\beta)\right)\left(\frac{z_i}{\pi_{i(r)} p(x_i^T\beta)} - n^{-1} t_{z(r)}(\beta)\right)^T$$

Assume that for some positive definite symmetric matrix $V_0(\beta)$,

$$\frac{n_r}{N_r^2} V_{z(r)}(\beta) \to V_0(\beta),$$

uniformly in β.

Assumption 5 *Write* $z_i = [z_{i1} \cdots z_{ik} \cdots]^T$ *and* $t_z(\beta) = [t_z(\beta)_1 \cdots t_z(\beta)_k \cdots]^T$. *Assume that for all* $k_1, \ldots, k_4$

$$\begin{aligned} N_r^{-3} n_r^2 \sum_{i\in\mathcal{U}_r} \left\{ \pi_{i(r)} \prod_{s=1}^{3} \left(\pi_{i(r)}^{-1} z_{ik_s} - n_r^{-1} t_{z(r)}(\beta_*)_{k_s}\right)\right\} &\leq M_3 \\ N_r^{-4} n_r^3 \sum_{i\in\mathcal{U}_r} \left\{ \pi_{i(r)} \prod_{s=1}^{4} \left(\pi_{i(r)}^{-1} z_{ik_s} - n_r^{-1} t_{z(r)}(\beta_*)_{k_s}\right)\right\} &\leq M_4 \end{aligned}$$

for some constants M_3, M_4.

Proposition 3 *Let the sample* $\mathcal{S}$ *be chosen without replacement and with inclusion probabilities* π_i *and* π_{ij}. *Let*

$$\widehat{V}_z(\beta) = \frac{n}{n-1} \sum_{i\in\mathcal{R}} \left(\pi_i^{-1} p(x_i^T\beta)^{-1} z_i - n^{-1}\widehat{t}_z(\beta)\right)\left(\pi_i^{-1} p(x_i^T\beta)^{-1} z_i - n^{-1}\widehat{t}_z(\beta)\right)^T.$$

That is $\widehat{V}_z(\beta)$ *is the matrix in equation (30).*

Under Assumptions 1, 2, 3, and 5

$$\widehat{V}_z(\beta) = V_z(\beta) + O_P(\frac{N^2}{n^{1.5}}) \tag{31}$$

uniformly in β. Hence, assuming in addition Assumption 4,

$$\frac{n}{N^2}\widehat{V}_z(\beta) \to_P V_0(\beta)$$

uniformly in β.

Proof: For simplicity of notation, we assume that z is univariate. In addition, as in Proposition 1, Assumption 1 yields uniformity in β as long as we can establish the proposition for $\beta = \beta_*$. Thus we will suppress the β in what follows.

Furthermore, it is easily checked that Assumptions 1, 2, 3, and 5 imply

$$N^{-4}n^3 \sum_{i\in\mathcal{U}} \pi_i p_i \left(\frac{z_i}{\pi_i p_i} - n^{-1}t_z\right)^4 \;\leq\; M_0$$

for some constant M_0.

With a little algebra and using Proposition 1

$$\begin{aligned}
\widehat{V}_z &= \big(1 + O(n^{-1})\big)\left\{\sum_{i\in\mathcal{R}}(\frac{z_i}{\pi_i p_i} - n^{-1}t_z)^2 - n^{-1}(\widehat{t}_z - t_z)^2\right\} \\
&= \big(1 + O(n^{-1})\big)\left\{\sum_{i\in\mathcal{R}}(\frac{z_i}{\pi_i p_i} - n^{-1}t_z)^2 + O_P(N^2/n^2)\right\}.
\end{aligned}$$

Let

$$\widetilde{V}_z = \sum_{i\in\mathcal{R}}(\frac{z_i}{\pi_i p_i} - n^{-1}t_z)^2. \tag{32}$$

Now

$$E(\widetilde{V}_z) = \sum_{i\in\mathcal{U}} \pi_i p_i (\frac{z_i}{\pi_i p_i} - n^{-1}t_z)^2 = V_z$$

which can be shown to be $O(N^2/n)$ using Assumptions 1 and 3.

Furthermore

$$\begin{aligned}
E(\widetilde{V}_z|\mathcal{S}) &= \sum_{i\in\mathcal{S}} p_i (\frac{z_i}{\pi_i p_i} - n^{-1}t_z)^2 \\
\mathrm{var}(E(\widetilde{V}_z|\mathcal{S})) &= \frac{-1}{2}\sum_{i,j\in\mathcal{U}}(\pi_{ij} - \pi_i\pi_j)\left\{p_i(\frac{z_i}{\pi_i p_i} - n^{-1}t_z)^2 - p_j(\frac{z_j}{\pi_j p_j} - n^{-1}t_z)^2\right\}^2 \\
&\leq \frac{\alpha}{n}\sum_{i,j\in\mathcal{U}}\pi_i\pi_j\left\{p_i^2(\frac{z_i}{\pi_i p_i} - n^{-1}t_z)^4 - p_i p_j(\frac{z_i}{\pi_i p_i} - n^{-1}t_z)^2(\frac{z_j}{\pi_j p_j} - n^{-1}t_z)^2\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \alpha \sum_{i \in \mathcal{U}} \pi_i p_i^2 (\frac{z_i}{\pi_i p_i} - n^{-1} t_z)^4 - \frac{\alpha}{n} V_z^2 = O(N^4/n^3) \\
\text{var}(\widetilde{V}_z|\mathcal{S}) &= \sum_{i \in \mathcal{S}} (p_i - p_i^2)(\frac{z_i}{\pi_i p_i} - n^{-1} t_z)^4 \\
E(\text{var}(\widetilde{V}_z|\mathcal{S})) &= \sum_{i \in \mathcal{U}} \pi_i (p_i - p_i^2)(\frac{z_i}{\pi_i p_i} - n^{-1} t_z)^4 = O(N^4/n^3),
\end{aligned}
$$

where we have applied Assumption 5 and noted that $p_i^2 < p_i$. Equation (31) follows. □

Appendix 3

Quasi-randomization consistency of the partial minimum $\widetilde{\beta}$

In Section 3, we defined

$$\ddot{\rho}(\beta, \gamma) = -\text{logdet}(W(\gamma)) + (T_z - \widehat{t}_z(\beta))^T W(\gamma)(T_z - \widehat{t}_z(\beta)).$$

The partial minimum $\widetilde{\beta}$ solves the equation

$$\frac{\partial \ddot{\rho}}{\partial \beta}(\widetilde{\beta}, \widetilde{\beta}) = 0 \tag{33}$$

whereas the full minimum $\widehat{\beta}$ satisfies

$$\frac{\partial \ddot{\rho}}{\partial \beta}(\widehat{\beta}, \widehat{\beta}) + \frac{\partial \ddot{\rho}}{\partial \gamma}(\widehat{\beta}, \widehat{\beta}) = 0.$$

Lemma 4 *Let f and f_r be C^1 functions from $R^Q \to R^Q$. Suppose $f_r(\beta) \to f(\beta)$ and $f_r^{'}(\beta) \to f^{'}(\beta)$ uniformly in β and that for some β_*, $f(\beta_*) = 0$, and $f^{'}(\beta_*)$ is nonsingular. Then if r is large enough, there is a sequence $\beta_r \to \beta_*$ with $f_r(\beta_r) = 0$.*

Proof: For a linear transformation $A : R^Q \to R^Q$, let $||A|| = sup_{\{x \in R^Q \ | \ |x|=1\}} |Ax|$. The following is often proven as part of the proof of the inverse function theorem (see, for example, Rudin (1964), 193-194):

Suppose f is C^1 with $f^{'}(\beta_*)$ nonsingular. Let $\lambda = (4||f^{'}(\beta_*)^{-1}||)^{-1}$. Let $B(\beta_*, \delta)$ be the ball of radius δ centered at β_* and let $\overline{B}(\beta_*, \delta)$ be its closure. Then if δ

is sufficiently small so that $||f^{'}(\beta) - f^{'}(\beta_*)|| < 2\lambda$ *for all* $\beta \in \overline{B}(\beta_*, \delta)$*, then* $f(B(\beta_*, \delta))$ *contains* $B(f(\beta_*), \lambda\delta)$.

Pick a compact neighborhood $\mathcal{C}$ of β_* that is in the domain of all the f_r and f. Pick N so that $||f_r^{'}(\beta_*)^{-1} - f^{'}(\beta_*)^{-1}|| \leq \frac{1}{2}||f^{'}(\beta_*)^{-1}||$ for all $r > N$. Then $\frac{2}{3}\lambda \leq \lambda_r = (4||f_r^{'}(\beta_*)^{-1}||)^{-1}$.

Since the convergence is uniform, the family is equicontinuous on $\mathcal{C}$. Hence there exists a δ_0 such that if $|\beta - \beta_*| \leq \delta_0$, then and $|f_r^{'}(\beta) - f_r^{'}(\beta_*)| < \frac{4}{3}\lambda \leq 2\lambda_r$ for all $r > N$. Thus if $\delta < \delta_0$, $f_r(B(\beta_*, \delta)) \supseteq B(f_r(\beta_*), \lambda_r\delta)$. So given δ, we can increase N so that if $r > N$, $||f_r(\beta_*)|| < \frac{2}{3}\lambda\delta \leq \delta\lambda_r$. It follows that $0 \in f_r(B(\beta_*, \delta))$ and this proves the lemma. □

Proposition 5 *Let* $\widetilde{\beta}_r$ *be the partial minimum closest to the full minimum if both minima exist and* 0 *otherwise. Then, under suitable regularity conditions,* $\widetilde{\beta}_r \rightarrow_P \beta_*$.

Proof: Let

$$\begin{aligned} f_r(\beta) &= n^{-1}\widehat{H}_r(\beta)W_r(\beta)[T_z - \widehat{t}_z(\beta)] \\ f(\beta) &= H_\infty(\beta)W_0(\beta)[\lim N^{-1}T_z - \lim N^{-1}t_z(\beta)], \end{aligned}$$

where $H_\infty(\beta)$ is the Jacobian matrix of $\lim N^{-1}t_z(\beta)$.

The previous section established that under the assumptions of that section $f_r(\beta) \rightarrow_P f(\beta)$ uniformly in β. It is clear that similar additional assumptions can be made to insure that $f_r^{'}(\beta) \rightarrow_P f^{'}(\beta)$.

Using Skorobod's theorem (see Billingsley (1995), p. 333), we can replace the convergence in probability to convergence almost surely. Then Lemma 4 applies.□

Assume the correctness of the targets, that is $T_z = t_z(\beta_*)$. Having established the consistency of $\widehat{\beta}$, the usual series argument implies that $\widehat{\beta} - \beta_*$ is $O_P(n^{-0.5})$.

Thus

$$\begin{aligned} n^{-1}\rho'(\beta) &= -n^{-1}\frac{\mathrm{d}\ \mathrm{logdet}\frac{N^2}{n}W(\beta)}{\mathrm{d}\beta} \\ &- 2N^{-1}\left(\frac{\mathrm{d}\widehat{t}_z(\beta)}{\mathrm{d}\beta}\right)^T \frac{N^2}{n}W(\beta)\left(N^{-1}T_z - N^{-1}\widehat{t}_z(\beta)\right) \qquad (34) \\ &+ \left(N^{-1}T_z - N^{-1}\widehat{t}_z(\beta)\right)^T \frac{\mathrm{d}\frac{N^2}{n}W(\beta)}{\mathrm{d}\beta}\left(N^{-1}T_z - N^{-1}\widehat{t}_z(\beta)\right). \end{aligned}$$

Since

$$\begin{aligned} N^{-1}\left(T_z - \widehat{t}_z(\beta)\right) &= N^{-1}\left(t_z(\beta_*) - \widehat{t}_z(\beta_*) + \widehat{H}(\widehat{\beta})(\beta_* - \widehat{\beta})\right) + O_P(n^{-1}) \\ &= O_P(n^{-0.5}), \end{aligned}$$

when evaluated at $\widehat{\beta}$, the first and third terms of the right hand side of (34) are $O_P(n^{-1})$. Hence

$$\widehat{H}(\widehat{\beta})^T W(\widehat{\beta})\left(T_z - \widehat{t}_z(\widehat{\beta})\right) = O_P(1) \qquad (35)$$

On the other hand, the partial minimum satisfies (see equation (9))

$$\widehat{H}(\widetilde{\beta})^T W(\widetilde{\beta})\left(T_z - \widehat{t}_z(\widetilde{\beta})\right) = O_P(1)$$

Finally we have

$$\widehat{\beta} = \beta_* + \left\{\widehat{H}(\widehat{\beta})^T W(\widehat{\beta})\widehat{H}(\widehat{\beta})\right\}^{-1}\widehat{H}(\widehat{\beta})^T W(\widehat{\beta})(T_z - \widehat{t}_z(\beta_*)) + O_P(n^{-1}) \quad (36)$$

$$\widetilde{\beta} = \beta_* + \left\{\widehat{H}(\widetilde{\beta})^T W(\widetilde{\beta})\widehat{H}(\widetilde{\beta})\right\}^{-1}\widehat{H}(\widetilde{\beta})^T W(\widetilde{\beta})(T_z - \widehat{t}_z(\beta_*)) + O_P(n^{-1}). \quad (37)$$

References

Billingsley, P. (1995) *Probability and Measure, 3rd ed.*, Wiley, New York.

Crouse, C. and Kott, P.S. (2004). "Evaluating Alternative Calibration Schemes for an Economic Survey with Large Nonresponse," *ASA Proceedings of the Survey Reseach Methods Section.*

Folsom., R.E. (1991). Exponential and Logistic Weight Adjustment for Sampling and Nonresponse Error Reduction, *ASA Proceedings of the Social Statistics Section*, 197-202.

Folsom., R.E. and Singh, A.C. (2000). The Generalized Exponential Model for Sampling Weight Calibration for Extreme Values, Nonresponse, and Post-stratification, *ASA Proceedings of the Section on Survey Research Methods*, 598-603.

Fuller, W.A. and Isaki, C.T. (1981). *Survey Design Under Superpopulation Models* in *Current Topics in Survey Sampling*, eds. D. Krewshi, J.N.K. Rao, and R. Platek, New York: Academic Press.

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.

Kim, J.K., Navarro, A., and Fuller, W.A. (2006). Replicate Variance Estimation for Two-Phase Stratified Sampling, *J. Am. Statist. Assoc.*, forthcoming.

Kish, L. (1965). *Survey Sampling*, Wiley, New York.

Kott, P.S. (2006). Using Calibration Weighting to Adjust for Nonresponse and Coverage Errors, *Survey Methodology* **32**, 133-42.

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

Lundström, S. and Särndal, C.-E. (1999). Calibration as a Standard Method for the Treatment of Nonresponse, *J. Official Statist.* **15**, 305-327.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*, Second edition, London: Chapman & Hall.

Rudin, W. (1964). *Principles of Mathematical Analysis, 2nd ed.*, McGraw-Hill, New York.

Särndal, C.-E. and Lundström, S. (2005). *Estimation in Surveys with Nonresponse*, New York: Wiley.

Särndal, C.-E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*, New York: Springer-Verlag.

Silvey, J. D. (1975). *Statistical Inference*, London: Chapman & Hall.

Thompson, M. E. (1997). *Theory of Sample Surveys*, London: Chapman & Hall.

Table 1: Benchmark targets and fitted totals for three variable calibrations

	N_1	N_2	N_3	N_4	N_5
CA targets	21804	14622	14309	14777	4769
CA fit	21861.5 (30.2)	14578.0 (34.3)	14273.8 (36.5)	14816.0 (25.7)	4751.6 (9.4)
DE targets	628	369	334	517	216
DE fit	638.9 (5.9)	370.3 (7.0)	311.5 (6.5)	535.4 (5.2)	207.9 (2.5)
IL targets	20220	15959	14241	24505	6513
IL fit	21044.1 (35.0)	15597.6 (36.2)	13560.2 (32.3)	25291.4 (30.4)	5874.4 (23.2)
LA targets	10390	7850	4275	2638	1040
LA 3 fit	10394.3 (20.9)	7880.7 (24.8)	4225.2 (19.8)	2664.8 (12.8)	1027.9 (9.3)
SD targets	5847	5304	7278	11134	2416
SD fit	5864.2 (15.9)	5322.9 (18.8)	7198.1 (22.2)	11209.4 (21.4)	2380.7 (10.1)

Standard errors, calculated using (20) with $\pi_i = \pi_{ij} = 1$ and $y_i = z_i$, are given in parentheses.

Table 2: Estimated number of farms

	poststratification using 5 Z-groups	calibration using intercept, *logsales*, *s97*
CA	45312.5 (45.8)	46178.8 (56.3)
DE	1390.9 (9.2)	1400.6 (11.0)
IL	57332.4 (52.1)	58925.7 (53.7)
LA	16139.6 (30.3)	16425.3 (37.6)
SD	23260.7 (27.7)	23821.1 (25.5)

Standard errors in parentheses.

Table 3: Benchmark targets and fitted totals for three model variable calibrations, simulated populations

	N_1	N_2	N_3	N_4	N_5
CA targets	19603	13034	12502	12463	3967
CA fitted	19603.28	13033.60	12502.34	12462.83	3966.96
	28.11	32.42	34.01	23.26	8.64
	28.42	32.16	33.75	23.37	8.75
DE targets	537	311	259	435	167
DE fitted	537.02	310.99	259.04	434.96	166.99
	5.39	6.40	6.04	4.75	2.31
	5.42	6.42	5.96	4.73	2.30
L targets	18542	14000	12048	20268	4388
IL fitted	18540.11	14001.55	12048.90	20266.67	4388.91
	31.11	31.97	27.91	25.51	20.44
	30.88	32.39	28.72	25.91	20.55
LA targets	9054	6938	3673	2182	806
LA fitted	9054.15	6938.15	3672.75	2182.09	805.87
	19.64	22.65	18.39	11.37	8.64
	19.30	22.61	18.24	11.14	8.45
SD targets	5358	4876	6499	9378	1837
SD fitted	5357.63	4876.97	6498.43	9378.30	1836.66
	14.57	17.52	20.22	19.24	9.34
	14.37	17.23	20.01	19.06	9.23

Based upon 1000 simulations. For each fitted model, the first line gives the mean of the fitted targets, the second their empirical standard deviations, and the third the square root of the mean of the sample estimated variances calculated using (20).

Table 4: Coefficients β_* and total number of farms t_y, together with estimates $\widehat{\beta}$ and $\widehat{t}_y(\widehat{\beta})$, three model variable calibrations on simulated populations

	Coefficients			
	int	*logsales*	*s97*	Number of Farms
CA β_* and t_y	3.7478	-0.2341	0.3841	39568
CA fitted $\widehat{\beta}$ and $\widehat{t}_y(\widehat{\beta})$	3.7433	-0.2335	0.3835	39565.53
	0.1051	0.0132	0.0510	52.81
	0.1062	0.0132	0.0507	53.57
DE β_* and t_y	2.3161	-0.0964	0.1543	1147
DE fitted $\widehat{\beta}$ and $\widehat{t}_y(\widehat{\beta})$	2.3292	-0.0957	0.1379	1146.61
	0.4934	0.0638	0.2227	10.63
	0.4908	0.0628	0.2230	10.16
IL β_* and t_y	5.2025	-0.4548	1.1894	48608
IL fitted $\widehat{\beta}$ and $\widehat{t}_y(\widehat{\beta})$	5.2009	-0.4545	1.1886	48605.32
	0.1193	0.0152	0.0525	48.59
	0.1207	0.0154	0.0540	48.30
LA β_* and t_y	3.5327	-0.2617	0.6822	13944
LA fitted $\widehat{\beta}$ and $\widehat{t}_y(\widehat{\beta})$	3.5277	-0.2609	0.6813	13944.63
	0.1899	0.0269	0.0958	35.56
	0.1898	0.0269	0.0952	35.21
SD β_* and t_y	6.0846	-0.5188	1.2285	20231
SD fitted $\widehat{\beta}$ and $\widehat{t}_y(\widehat{\beta})$	6.0915	-0.5193	1.2281	20230.13
	0.2287	0.0272	0.0846	24.20
	0.2316	0.0276	0.0856	23.95

Based upon 1000 simulations. These calibrations fit a correct model. For each state, the first line contains the true values, the second contains the mean of the fitted values, the third their empirical standard deviations, and the fourth the square root of the mean of the estimated variances.

Table 5: Estimated number of farms, simulated populations

	Poststratification: 5 Z-groups			Calibration: 4/5 X-variables		
		standard error			standard error	
No. farms	mean	empirical	eqn (20)	mean	empirical	eqn (20)
CA 39568	38815.03	39.62	42.86	39365.00	154.60	156.68
DE 1147	1138.96	8.01	8.41	1145.66	15.97	16.91
IL 48608	47369.47	37.58	45.38	48133.22	73.04	79.66
LA 13944	13708.36	26.03	27.92	13713.93	81.04	91.46
SD 20231	19766.06	19.51	25.17	20081.41	43.86	51.50

Based upon 1000 simulations. Both of these calibrations fit incorrect models. Poststratification fits the wrong model variables; the calibration model fits the correct model variables with the wrong functional form. Notice that although both models have estimates that appear to be biased downward, the bias is worse for the model with the incorrect model variables and that this bias is much more important than the standard error of the estimates.