RESEARCH REPORT

OCCASIONAL PAPER

Exploring Some Uses for Instrumental-Variable Calibration Weighting

Phillip S. Kott

January 2013

RTI Press

RTI INTERNATIONAL

About the Author

Phillip S. Kott, PhD, is a senior research statistician in RTI International's Survey Research Division.

RTI Press publication OP-0012-1302

Suggested Citation

Kott, P. S. (2013). *Exploring some uses for instrumental-variable calibration weighting*. RTI Press publication No. OP-0012-1302. Research Triangle Park, NC: RTI Press. Retrieved from http://www.rti.org/rtipress.

This publication is part of the RTI Research Report series. Occasional Papers are scholarly essays on policy, methods, or other topics relevant to RTI areas of research or technical focus.

RTI International
3040 E. Cornwallis Road
PO Box 12194
Research Triangle Park, NC
27709-2194 USA

Tel: +1.919.541.6000
Fax: +1.919.541.5985
E-mail: rtipress@rti.org
Web site: www.rti.org

http://dx.doi.org/10.3768/rtipress.2013.op.0012.1302

www.rti.org/rtipress

Exploring Some Uses for Instrumental-Variable Calibration Weighting

Phillip S. Kott

Abstract

The WTADJX procedure incorporated into the 2012 release of SUDAAN 11® does instrumental-variable calibration weighting using a flexible nonlinear weight-adjustment function. We review the theory behind this procedure and discuss two potential uses. The first tends to reduce mean squared errors in the absence of unit nonresponse or coverage errors. The second adjusts for unit nonresponse when the variables governing the response mechanism differ from the variables used to calibrate the weights. This occurs either because the survey variables of interest cannot be roughly modeled as a linear function of the response-model variables or because the values of the response-model variables are known only for the respondents.

Contents

RTI INTERNATIONAL

Introduction

Brewer (1995) proposed using instrumental variables in a calibration-weighted estimator for a finite-population total as a way of integrating the prediction form of model-based sampling theory with (design-based) randomization consistency. We explore instead more practical uses of instrumental-variable calibration, such as (1) adjusting for nonresponse when the variables governing the response/nonresponse mechanism are not always the same as the calibration variables and (2) creating nearly optimal weights under probability sampling theory that never fall below unity and (if desired) are bounded from above.

We first briefly review calibration weighting and the generalized exponential form of Folsom & Singh (2000). We then discuss the hows and whys of instrumental-variable calibration as implemented with the WTADJX procedure incorporated into SUDAAN 11® (RTI International, 2012). This is fleshed out with two numerical examples. Throughout the text, SUDAAN and its procedures appear in upper case.

Calibration Weighting

Linear Calibration Weighting

When there is no nonresponse, calibration is a weight-adjustment method that creates a set of weights, $\{w_k\}$, with two important properties. Given a p-vector $\mathbf{z}_k$ with known population totals and probabilities of selection $\{\pi_k\}$, the new weights:

1. Are asymptotically close to the original design weights $d_k = 1/\pi_k$ (i.e., as the sample size grows arbitrarily large, w_k converges to d_k) and therefore *nearly* unbiased under probability-sampling theory.
2. Satisfy a set of calibration equations (one for each component of $\mathbf{z}_k$):

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

where S denotes the set of units in the sample, and U is the set of units in the finite population.

When a total $T = \Sigma_U y_k$ is estimated with $t = \Sigma_S w_k y_k$ or a mean $\bar{y}_U = T/N$ with $\hat{\bar{y}}_U = \Sigma_S w_k y_k / \Sigma_S w_k$, *calibration weighting* will tend to reduce mean squared error when y_k is correlated with components of $\mathbf{z}_k$. Real surveys usually have more than a single y-value of interest. It is not uncommon, however, for an establishment survey to have a main survey value of interest. For example, in the Drug Abuse Warning Network survey (US Department of Health and Human Services, 2011), that variable is annual drug-related hospital visits.

One way to compute calibration weights is linearly with the following formula:

$$w_k = d_k\left[1+\left(\sum_U \mathbf{z}_j - \sum_S d_j \mathbf{z}_j\right)^T \left(\sum_S d_j \mathbf{z}_k \mathbf{z}_j^T\right)^{-1} \mathbf{z}_k\right]$$
$$= d_k\left[1+\mathbf{g}^T \mathbf{z}_k\right],$$

where $\mathbf{g} = (\Sigma_S d_j \mathbf{z}_j \mathbf{z}_j^T)^{-1} (\Sigma_U \mathbf{z}_j - \Sigma_S d_j \mathbf{z}_j)$. Observe that as the sample size grows arbitrarily large, $\mathbf{g}^T\mathbf{z}_k$ (which means $\mathbf{g}$) converges to $\mathbf{0}$.

This is the weighting scheme implied by the *generalized regression* estimator (GREG) since

$$\sum_S w_k y_k = \sum_S d_k y_k + \left(\sum_U \mathbf{z}_j - \sum_S d_j \mathbf{z}_j\right)^T \left(\sum_S d_j \mathbf{z}_j \mathbf{z}_j^T\right)^{-1} \sum_S d_k \mathbf{z}_k y_k$$
$$= \sum_S d_k y_k + \left(\sum_U \mathbf{z}_j - \sum_S d_j \mathbf{z}_j\right)^T \mathbf{b},$$

where $\mathbf{b} = (\Sigma_S d_j \mathbf{z}_j \mathbf{z}_j^T)^{-1} \Sigma_S d_k \mathbf{z}_k y_k$ is a survey-weighted estimated linear-regression coefficient.

Linear calibration weighting can be easily adapted to handle unit nonresponse by simply replacing the sample S with respondent sample R and redefining the GREG estimator and $\mathbf{g}$ as:

$$t_{GREG} = \sum_R w_k y_k + \sum_R d_k \left(1+\mathbf{g}^T \mathbf{z}_k\right) y_k.$$

In this context, $\mathbf{g}^T$ can be either

$$\mathbf{g}^T = \left(\sum_U \mathbf{z}_j - \sum_R d_j \mathbf{z}_j\right)^T \left(\sum_R d_j \mathbf{z}_j \mathbf{z}_j^T\right)^{-1},$$

which requires that $\Sigma_U \mathbf{z}_j$ be known, or

$$\mathbf{g}^T = \left(\sum_S d_j \mathbf{z}_j - \sum_R d_j \mathbf{z}_j\right)^T \left(\sum_R d_j \mathbf{z}_j \mathbf{z}_j^T\right)^{-1},$$

which requires that $\sum_S d_j \mathbf{z}_j$ be known. The first is called "calibrating to the population" and the latter "calibrating to the original sample."

Either way, the estimate is also nearly unbiased under a quasi-probability theory that treats response as a second phase of random sampling as long as each unit's probability of response has the form:

$$p_k = \frac{1}{1 + \boldsymbol{\gamma}^T \mathbf{z}_k}, \qquad (1)$$

and g is a consistent estimator for $\boldsymbol{\gamma}$. Put another way:

$$t_{GREG} = \sum_R w_k \mathbf{z}_k + \sum_R d_k \hat{p}_k^{-1} \mathbf{z}_k.$$

Notice that when calibration weighting is used to adjust for unit nonresponse, neither $(\sum_U \mathbf{z}_j = \sum_R d_j \mathbf{z}_j)^T$ nor $(\sum_S d_j \mathbf{z}_j - \sum_R d_j \mathbf{z}_j)^T$ converges to $\mathbf{0}^T$, and so neither does $\mathbf{g}^T$. This, at the time surprising, use of calibration weighting for nonresponse adjustment was proposed by Fuller et al. (1994).

Nonlinear Calibration Weighting

The problem with the probability-of-response function in equation (1) is that it can exceed unity or even be negative. A useful nonlinear form of calibration weighting suggested by Folsom & Singh (2000) finds a $\mathbf{g}$ (through repeated linearization, i.e., Newton's method) such that

$$\sum_R w_k \mathbf{z}_k \equiv \sum_R d_k \alpha\left(\mathbf{g}^T \mathbf{z}_k\right) \mathbf{z}_k = \sum_U \mathbf{z}_k$$

or (2)

$$\sum_R w_k \mathbf{z}_k \equiv \sum_R d_k \alpha\left(\mathbf{g}^T \mathbf{z}_k\right) \mathbf{z}_k = \sum_S d_k \mathbf{z}_k,$$

where $\alpha(\mathbf{g}^T \mathbf{z}_k)$ is a function of the form,

$$\alpha(\mathbf{g}^T \mathbf{z}_k) = \frac{\ell(u-c) + u(c-\ell)\exp(A\mathbf{g}^T \mathbf{z}_k)}{(u-c) + (c-\ell)\exp(A\mathbf{g}^T \mathbf{z}_k)}, \qquad (3)$$

and $A = (u-\ell)/[(u-c)(c-\ell)]$. The inclusion of A in equation (3) makes finding the derivative of $\alpha(\mathbf{g}^T \mathbf{z}_k)$ easier (which is needed for implementing Newton's method).

The *weight adjustment* $\alpha(\mathbf{g}^T \mathbf{z}_k)$ is centered at c in the sense that $\alpha(0) = 1$ with a lower bound $\ell \geq 0$ and an upper bound $u > c > \ell$, which can be infinite. The user sets these *centering* and *bounding* parameters. Equation (3) is a generalization of both raking, where $\ell = 0$, $c = 1$, $u = \infty$ (and the components of $\mathbf{z}_k$ are binary); and the implicit estimation of a logistic-regression response model, where $\ell = 1$, $c = 2$, $u = \infty$.

When $c = 1$, equation (3) is the generalized-raking adjustment introduced by Deville and Särndal (1992) so that the range of $\alpha(\mathbf{g}^T \mathbf{z}_k)$ could be bounded (and the components of $\mathbf{z}_k$ continuous). Centering at 1 was a requirement of calibration weighting in that landmark paper ($\alpha'(0) = 1$ was required as well), but setting $c > 1$ with $\ell = 1$ is more sensible when adjusting for unit nonresponse so that the implicitly estimated probability of response is never greater than 1.

Folsom & Singh (2000) proposed using the following generalized exponential form:

$$\alpha_k(\mathbf{g}^T \mathbf{z}_k) = \frac{\ell_k(u_k-c_k) + u_k(c_k-\ell_k)\exp(A_k \mathbf{g}^T \mathbf{z}_k)}{(u_k-c_k) + (c_k-\ell_k)\exp(A_k \mathbf{g}^T \mathbf{z}_k)},$$

which generalized equation (3) by allowing separate weight functions for each k but found a common $\mathbf{g}$ chosen to satisfy one of the two versions of the calibration equation (the population or original-sample version). This form of calibration weighting has been incorporated into the SUDAAN procedure WTADJUST (RTI International, 2012). See Kott and Liao (2012a) for a more rigorous treatment of this version of nonlinear calibration weighting. Kott (2009) provides a good background on calibration weighting in general.

Although WTADJUST allows $\alpha_k(\mathbf{g}^T \mathbf{z}_k)$ to be k-specific, when adjusting for nonresponse (or coverage), it is sensible to select a single value for the c_k parameter and a very limited number of ℓ_k and u_k values since different parameter values across the population elements mean different response functions are being fit. When each of the three parameters has a single value, it is not hard to see that the choice of c becomes irrelevant (again, see Kott & Liao, 2012a).

Instrumental Variables

Nonresponse

Now suppose unit response follows a model of the form:

$$p_k = [\alpha_k(\boldsymbol{\gamma}^T\mathbf{x}_k)]^{-1} = \frac{(u_k - c_k) + (c_k - \ell_k)\exp(A_k\boldsymbol{\gamma}^T\mathbf{x}_k)}{\ell_k(u_k - c_k) + u_k(c_k - \ell_k)\exp(A_k\boldsymbol{\gamma}^T\mathbf{x}_k)}, \quad (4)$$

where some components of the *response-model* vector $\mathbf{x}_k$ governing the unit response mechanism need not coincide with the components on the calibration $\mathbf{z}$-vector. In other words, replace equation (2) by

$$\sum_R w_k\mathbf{z}_k = \sum_R d_k\alpha_k\left(\mathbf{g}^T\mathbf{x}_k\right)\mathbf{z}_k = \sum_U \mathbf{z}_k$$

or (5)

$$\sum_R w_k\mathbf{z}_k = \sum_R d_k\alpha_k\left(\mathbf{g}^T\mathbf{x}_k\right)\mathbf{z}_k = \sum_S d_k\mathbf{z}_k,$$

such that $\mathbf{g}$ again estimates $\boldsymbol{\gamma}$.

Mathematically, finding a $\mathbf{g}$ that satisfies either the first or second line of equation (5) can often be done as long as the number of response-model variables in $\mathbf{x}_k$ is no greater than p, the number of calibration variables in $\mathbf{z}_k$. A routine to do that is available in SUDAAN 11: WTADJX. It applies most simply when the numbers of model and calibration variables coincide so that one of the two sets of calibration equations in (5) holds. Otherwise, there are more equations than unknowns, and the vector equations in (5) cannot hold exactly. See Chang and Kott (2008) for a discussion of minimizing the difference between, say, $\Sigma_R d_k\alpha_k(\mathbf{g}^T\mathbf{x}_k)\mathbf{z}_k$ and $\Sigma_U\mathbf{z}_k$ as a means for estimating $\boldsymbol{\gamma}$.

The components of $\mathbf{x}_k$ that are not components of $\mathbf{z}_k$ are called *instrumental variables*. The name derives from the linear-calibration form where

$$\sum_S w_k y_k = \sum_S d_k y_k + \left(\sum_U \mathbf{z}_j - \sum_S d_j\mathbf{z}_j\right)^T \left(\sum_S d_j\mathbf{x}_j\mathbf{z}_j^T\right)^{-1}\sum_S d_k\mathbf{x}_k y_k$$
$$= \sum_S d_k y_k + \left(\sum_U \mathbf{z}_j - \sum_S d_j\mathbf{z}_j\right)^T\mathbf{b}_{IV}$$

when $\mathbf{x}_k$, like $\mathbf{z}_k$, is a p-component vector. In the prediction-model framework where instrumental variables originated, $E(y_k|\mathbf{z}_k, \mathbf{x}_k) = \mathbf{z}_k\boldsymbol{\beta}$, and $\mathbf{b}_{IV}$ is an unbiased estimator for $\boldsymbol{\beta}$.

For an establishment survey with a main survey variable of interest, it often makes sense to calibrate to a size variable—call it q_k—known for all members of the population because the main survey variable is nearly linear in the size variable. Although the probabilities of response vary by size, all other things being equal, response may be better modeled as a logistic function of the *log* of the size variable, so that a one-percent increase in the size variable results in a c-percent change in the odds of response. Thus, $\log(q_k)$ is an instrument used in place of q_k.

Deville (2000) noted that it is possible for a response-model variable to be known only for respondents (i.e., a function of the main survey variable itself rather than the associated size variable). That is, it is possible for *nonresponse to not be missing at random*.

Nearly Pseudo-Optimal Calibration

Instrumental-variable calibration can be profitably used in the absence of nonresponse and coverage errors. A linear estimator often better (i.e., more efficient) than the usual GREG also calibrates on $\mathbf{z}_k$ but sets $\mathbf{x}_k = (d_k - 1)\mathbf{z}_k$. This produces the nearly unbiased linear estimator with the smallest asymptotic mean squared error under Poisson sampling and similarly under stratified simple random sampling with large stratum samples sizes. As a result, it has been called the *optimal estimator* under Poisson sampling (Rao, 1994) and the *pseudo-optimal estimator* more broadly (Bankier, 2002).

With WTADJX centered at 1, we can bound the weights and retain the asymptotic properties of the optimal estimator by setting $\mathbf{x}_k = (d_k - 1)\mathbf{z}_k$. In particular, when $d_k > 1$, we can set $\ell_k = 1/d_k$ to ensure that all w_k are at least unity. If some $d_k = 1$, we can set $w_k = 1$ and remove k from U and S before applying equation (2) (see Kott, 2011a). Alternatively, we can simply set ℓ_k at any value less than 1 for elements with $d_k = 1$ since $\mathbf{x}_k$ will be $\mathbf{0}$, forcing w_k to be 1 as well.

We have some freedom in setting the u_k as long as they are each greater than 1 and the calibration equation ($\sum_S w_k \mathbf{z}_k = \sum_U \mathbf{z}_k$)can be satisfied. Sometimes, rather than bounding the weight adjustment, a SUDAAN user may want to bound the weight itself by creating an upper bound of the form $u_k = U/d_k$. Often with establishment surveys, it is desirable to set an upper bound of the form $u_k = U/(d_k q_k)$ so that $(w_k q_k)$ is bounded.

Two Examples

As has been noted, the WTADJX procedure in SUDAAN 11 can perform instrumental-variable calibration. SUDAAN 11 is also able to compute (asymptotic) standard errors properly for means, totals, and ratios with weights adjusted by one round of WTADJX or WTADJUST calibration (Witt, 2010). When the adjustment is for nonresponse (or coverage error), this assumes that the underlying response (or coverage) model has been specified correctly—that is, the model in equation (4) holds and that response is independent across primary sampling units. When the logistic response model is correct, SUDAAN will also compute standard errors properly when the LOGISTIC procedure (RLOGIST in the SAS-callable version of SUDAAN used here) is used to estimate the probabilities of response and the inverses of those estimates employed to adjust the weights.

An Example With No Nonresponse or Coverage Error

For purposes of this exposition, we created a stratified simple random sample of 364 fictional hospital emergency departments using the public-use data set of the Drug Abuse Warning Network (US Department of Health and Human Services, 2011) as a starting point. The sample, stratified on size, location, urbanicity, and ownerships (public or private), with some collapsing and varying selection probabilities across strata, can be found in *SUDAAN 11 Examples* (WTADJX examples) on the SUDAAN website (http://www.rti.org/sudaan/). Much of the SAS-callable SUDAAN code discussed in this section can be found there as well.

Each hospital on the frame has attached to it a size variable: the number of emergency-department visits in a previous year, which we call *frame visits*. There are also indicators on the frame of each hospital's census region, whether it is publicly owned, and whether it is in a metropolitan area. Our goal is to estimate the total number of *drug-related* emergency-department visits in the survey year both across the United States and within each census region.

In addition to computing the estimates directly with their probability weights, we raked the weights—using WTADJUST with a center of 1, a lower bound of 0, and no upper bound—so that the following calibration-weighted totals equaled the corresponding frame counts: the number of hospitals in each region, the total number of publicly owned hospitals, and the number of hospitals in a metropolitan area. That is to say, the calibration vector $\mathbf{z}_k$ had six components, four regional indicator dummies (δ_{k1}, δ_{k2}, δ_{k3}, δ_{k4}), an indicator dummy for public ownership (δ_{k5}), and an indicator dummy for a metropolitan location (δ_{k6}).

As can be seen in Table 1, raking did not improve the coefficients of variation (CVs) in any of the regions (computed using SUDAAN 11, as were all the estimates in this section). If anything, the CVs were slightly higher than when using the direct estimator. That is because a hospital's annual number of drug-related emergency-department visits is not nearly a linear function of its region, ownership status, and urbanicity.

Table 1. Comparing direct estimation to raking and nearly quasi-optimal (NQO) calibration

Region	Direct	Raking	Size Raking	NQO	NQO Intercept	WTADJUST Only
Estimate (in 10,000s)						
All	538	537	553	552	553	552
East	73	73	79	79	79	78
South	175	175	184	183	183	183
Midwest	137	137	143	143	143	143
West	152	152	148	147	148	149
Coefficient of variation (standard error/estimate as a percentage)						
All	6.47	6.48	2.16	1.91	1.87	1.94
East	5.67	5.71	3.32	3.27	3.28	3.41
South	13.92	13.94	3.49	2.02	1.95	2.10
Midwest	7.55	7.55	3.23	3.22	3.26	3.22
West	14.58	14.58	5.77	5.69	5.61	5.70

A variant of raking for establishment surveys—introduced by Hidiroglou and Patak (2006)—is more applicable in this setting. *Size raking* calibrates the weights so that the weighted-total of the size variable (q_k) within each region equals the actual number on the frame, with analogous equalities holding for public and metropolitan hospitals. This variant on raking should decrease the standard errors of estimates for drug-related emergency-department visits at the US and regional levels if these survey variables are roughly linear functions of the calibration variables.

Size raking was done in WTADJX by letting the region, public, and metropolitan indicator dummies remain as the MODEL variables (with a "/NOINT" option since there was no intercept), while each of those indicator dummies times the number of frame visits made up the calibration variables, or CALVARS. Here, the "model" in MODEL refers to the *weight-adjustment model*, $\alpha(\mathbf{g}^T\mathbf{x}_k) = \exp(\mathbf{g}^T\mathbf{x}_k)$, used in WTADJX, where $\mathbf{x}_k = \boldsymbol{\delta}_k = (\delta_{k1}\ \delta_{k2}\ \delta_{k3}\ \delta_{k4}\ \delta_{k5}\ \delta_{k6})^T$ is the vector of the six indicator dummies, while the vector of calibration variables was $\mathbf{z}_k = q_k\boldsymbol{\delta}_k$. There was no response (or coverage) model.

Employing size raking decreased the CVs by region noticeably. Better still, as can be seen in Table 1, were the two variants of nearly quasi-optimal (NQO) calibration weighting. In one, the same CALVARS were used as in size raking ($\mathbf{z}_k = q_k\boldsymbol{\delta}_k$), but the MODEL variables included these calibration variables times $d_k - 1$ (i.e., $\mathbf{x}_k = (d_k - 1)q_k\boldsymbol{\delta}_k$). In the other, an intercept was added (NQO intercept). The vector of calibration variables was $\mathbf{z}_k = (1\ q_k\boldsymbol{\delta}_k{}^T)^T$, while the vector of model variables was $\mathbf{x}_k = (d_k - 1)(1\ q_k\boldsymbol{\delta}_k{}^T)^T$. (Note that when WTADJX is run, the "/NOINT" option must still be used, since $d_k - 1$ appears in the MODEL statement in place of an intercept.)

Finally, Table 1 shows what happens when the last vector of calibration variables, $\mathbf{z}_k = (1\ q_k\boldsymbol{\delta}_k{}^T)^T$, is used in WTADJUST without any instrumental variable. The CV results are similar to the quasi-optimal analogue, but mostly not as good—as expected. Surprisingly, they are mostly smaller than the CVs from size raking.

An Example With Nonresponse

We then used the same data set as in the previous example, but generated unit nonresponse as a logistic function of the *log* of drug-related emergency-department visits. Assuming first that response was a function of the log of the *frame* visits ($\mathbf{z}_k = (1\ \log(q_k))^T$), we employed SUDAAN to estimate survey-variable totals applying first RLOGIST and WTADJUST. We then applied WTADJX, again letting the log of frame visits be the MODEL variable ($\mathbf{x}_k = (1\ \log(q_k))^T$), but having frame visits become the calibration variable in CALVARS ($\mathbf{z}_k = (1\ q_k)^T$). The resulting CVs are shown in Table 2.

Table 2. CV for estimated number of drug-related emergency department visits by weight adjustment method

Weight Adjustment Method	CV
RLOGIST	7.33
RLOGIST + WTADJUST	8.30
RLOGIST + WTADJX	
Calibrating to the frame visits in the original sample	6.39
Calibrating to the frame visits in the population	3.40

It may come as a bit of a surprise that adjusting for nonresponse using RLOGIST was estimated to be more efficient than adjusting with WTADJUST. Given the nature of the data, however, it should be no surprise that using WTADJX and calibrating on frame visits rather than the log of those visits appeared more efficient than using either RLOGIST or WTADJUST even though the same variable (log of frame visits) was used to model response by all three. Moreover, calibrating to frame totals rather than full-sample totals increased the estimated efficiency even more.

The WTADJX procedure can also be used to test whether there is a significant difference between estimates derived under different assumed response models. In this case, the estimated relative bias (roughly 1.2 percent)—from incorrectly assuming response was a logistic function of the log of the *frame* variable rather than the log of the *survey* variable—was significant at the .08 level.

It may be tempting to conclude that bias was not an issue here because the statistical significance did not reach the magic .05 level. When testing for possible bias, however, we need to be more concerned with Type 2 error (failing to recognize a bias when it exists) than Type 1 error (finding a bias when none exists). As a result, statistical significance at the .08 level should be viewed as problematic.

In practical application, we rarely know the true response model. Even so, the test we used can be applied to determine whether different response models lead to significantly different estimates. To conduct the test, SUDAAN users should first duplicate each record, assigning the first version to a domain governed by one assumed response model and the second version to a domain governed by a different assumed model *while keeping both in the same primary sampling unit*; and then test the difference between domain estimates, treating the sample as if it were drawn with replacement. A test like this was proposed by Fuller (1984) for determining whether failing to incorporate sampling weights into a linear regression would produce biased coefficient estimates. Although not as powerful as a Hausman test (Hausman, 1978), which requires stronger assumptions, this test does benefit from using the same data twice.

Concluding Remarks

Although calibrating to the population is more efficient than calibrating to the full sample, it is better to calibrate in two steps, adjusting first to remove nonresponse bias (assuming one's response model is correct) and then to reduce variance (Kott & Liao, 2012b), using nearly pseudo-optimal calibration in the second step to make up for any inefficiency from instrumental-variable calibration in the first step.

Kott (2011b) points out that instrumental-variable calibration can aid in replication-based variance estimation when a bounded version of WTADJUST or WTADJX calibration is used. Empirical research on this use of WTADJX is under way.

References

Bankier, M. (2002, July). *Regression estimators for the 2001 Canadian Census.* Paper presented at the International Conference in Recent Advances in Survey Sampling, Carlton University, Ottawa, Ontario, Canada.

Brewer, K. R. W. (1995). Combining design-based and model-based inference. In B. G. Cox, D. A. Binder, B. N. Chinappa, A. Christianson, M. J. Colledge, and P. S. Kott (Eds.), *Business Survey Methods* (pp. 586–606). New York: Wiley.

Chang, T., & Kott, P. S. (2008). Using calibration weighting to adjust for nonresponse under a plausible model. *Biometrika, 95*, 557–571.

Deville, J. C. (2000). Generalized calibration and application to weighting for non-response. In J. G. Bethlehem & P. G. M. Van der Heijden (Eds.), *COMPSTAT: Proceedings in computational statistics; 14th symposium held in Utrecht, The Netherlands* (pp. 65–76). Heidelberg: Physica Verlag.

Deville, J. C., & Särndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association, 87*, 376–382.

Folsom, R. E., & Singh, A. C. (2000). The generalized exponential model for sampling weight calibration for extreme values, nonresponse, and poststratification. In *Proceedings of the American Statistical Association, Survey Research Methods Section* (pp. 598–603). Retrieved January 15, 2013, from http://www.amstat.org/sections/srms/Proceedings/

Fuller, W. A. (1984). Least squares and related analyses for complex survey designs. *Survey Methodology, 10*, 97–118.

Fuller, W. A., Loughin, M. M., & Baker, H. D. (1994). Regression weighting for the 1987–88 National Food Consumption Survey. *Survey Methodology, 20*, 75–85.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrika, 46*(6), 1251–1271.

Hidiroglou, M., & Patak, Z. (2006). An application to the Canadian Retail Trade Survey. *Journal of Official Statistics, 22*, 71–80.

Kott, P. S. (2009). Calibration weighting: Combining probability samples and linear prediction models. In D. Pfeffermann, & C. R. Rao (Eds.), *Handbook of statistics 29B: Sample surveys: Inference and analysis* (pp. 55–82). New York: Elsevier.

Kott, P. S. (2011a). A nearly pseudo-optimal method for keeping calibration weights from falling below unity in the absence of nonresponse or frame errors. *Pakistan Journal of Statistics, 27*(4), 391–396.

Kott, P. S. (2011b). WTADJX is coming: Calibration weighting in SUDAAN when unit nonrespondents are not missing at random and other applications. *Proceedings of the American Statistical Association, Survey Research Methods Section* (pp. 1746–1752). Retrieved January 15, 2013, from http://www.amstat.org/sections/srms/Proceedings/allyearsf.html

Kott, P. S., & Liao, D. (2012a). Providing double protection for unit nonresponse with a nonlinear calibration-weighting routine. *Survey Research Methods*, *6*(2), 105–111.

Kott, P. S., & Liao, D. (2012b, May–June). *One step or two? Calibration weighting for a complete frame with nonresponse*. Paper presented at the Fields Institute Symposium on the Analysis of Survey Data and Small Area Estimation in Honour of the 75th Birthday of Professor J. N. K. Rao, Carlton University, Ottawa, Ontario, Canada. Manuscript also submitted for publication.

Rao, J. N. K. (1994). Estimating totals and distribution functions using auxiliary information at the estimations stage. *Journal of Official Statistics, 25*, 1–21.

RTI International. (2012). *SUDAAN language manual, Release 11.0*. Research Triangle Park, NC: RTI International.

US Department of Health and Human Services. (2011). *Drug Abuse Warning Network (DAWN), 2008*. Computer file of survey conducted by the Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality. Ann Arbor, MI: Inter-University Consortium for Political and Social Research.

Witt, M. (2010). Estimating the R-indicator, its standard error, and other related statistics with SAS and SUDAAN. *Proceedings of the American Statistical Association, Section on Survey Research Methods* (pp. 5654–5668). Retrieved January 15, 2013, from http://www.amstat.org/sections/srms/Proceedings/

Acknowledgments

The author thanks two referees for their numerous suggestions which greatly improved the quality of the final manuscript.

www.rti.org/rtipress **RTI Press publication OP-0012-1302**