BIAS CORRECTED ESTIMATING FUNCTION APPROACH FOR VARIANCE ESTIMATION ADJUSTED FOR POSTSTRATIFICATION

A.C. Singh and R.E. Folsom, Jr., Research Triangle Institute
A.C. Singh, Statistics Research Division, RTI, Research Triangle Park, NC 27709 asingh@rti.org

Key Words: Generalized Raking, g-weights, Sandwich Variance Formula

1. Introduction

We consider the problem of finding a Taylor linearization variance estimator of the poststratified or the general calibration estimator such that four goals, somewhat analogous to those of Särndal, Swensson, and Wretman (1989), are met. These are : (i) the variance estimator should be consistent under a joint design-model distribution, i.e., in a quasi-design based framework, (ii) should have a simple form with general applicability, (iii) the model postulated for the quasi-design based framework should be driven by the real need for unbiased point estimation, and (iv) should be expected to have sensible conditional properties under a conditional inference outlook whenever it can be suitably defined. By way of notation define a finite population U from which a sample of size n is selected using the design p(s). Denote the data by (y_k, x_k, d_k), $k \in s$, where for the k^{th} unit in the sample, y_k is the study variable (considered scalar here for simplicity), x_k is a p-vector of covariates or predictor variables; and d_k is the design weight. Now consider a simple example of the often recommended Taylor linearization variance estimator of the ratio estimator (a simple type of poststratification (ps)) for domain or area a. The population total estimator is given by (the symbol $k \in a$ is used to denote the membership of the unit k in the population domain a)

$$\hat{T}_{ya,ratio} = (\hat{T}_{ya} / \hat{N}_a) N_a =: \Sigma_s y_k I_{k \in a} d_k g_k \qquad (1.1)$$

where $\hat{T}_{ya} = \Sigma_s y_k I_{k\in a} d_k$, $\hat{N}_a = \Sigma_s d_k I_{k\in a}$, $N_a = \Sigma_U I_{k\in a}$, $g_k = N_a / \hat{N}_a$, and the variance estimate $\hat{V}_{ratio,g}$ is

$$(N_a/\hat{N}_a)^2 \, \hat{V}(\hat{T}_y - R_{ya}\hat{N}_a)|_{R=\hat{R}} = \hat{V}(\Sigma_s e_k I_{k\in a} d_k g_k) \qquad (1.2)$$

where $\quad e_k = y_k - \hat{R}_{ya} x_k, \quad \hat{R}_{ya} = \hat{T}_{ya}/\hat{N}_a$.

In the above, $\hat{V}$ denotes the design-based variance estimator. The standard Taylor linearization variance estimator (Cochran, 1977, p.135), on the other hand, is given by

$$\hat{V}_{ratio} = \hat{V}(\Sigma_s e_k I_{k\in a} d_k) \qquad (1.3)$$

i.e., here the g-weight (= $N_a / \hat{N}_a$) or the ratio adjustment factor is not used to adjust the design weight d_k in the linearized estimator. In the sequel, we will refer to (1.2) as the g-weighted variance estimator while (1.3) as the standard variance estimator.

If there is no coverage bias, then under regularity conditions (see e.g., Isaki and Fuller, 1982) the estimator $\hat{V}_{ratio,g}$ is design-consistent. It has a simple but general form of variance applicable to the class of generalized raking estimators, and is expected to have reasonable conditional properties (Rao, 1985) in that unlike $\hat{V}_{ratio}$, it is sensitive to whether or not the realized $\hat{N}_a$ is near its expectation N_a, i.e., the variance estimator gets inflated if it is below the expectation and deflated otherwise; this is clearly a desirable feature. To justify the use of the factor g_k in the g-weighted variance estimator, a model expressing the relation between the study variable (y) and the auxiliary variable (x) is postulated to get guidance in choosing among a multitude of design-consistent variance estimators. Särndal et al. (1989) use a regression superpopulation model (for the present example it is simply a domain-specific constant mean model) to show that the g-weighted variance estimator unlike the standard estimator has the important property of being approximately model unbiased. This is often termed as a model-assisted approach. On the other hand, Royall and Cumberland (1981) using a pure model-based approach in the context of simple random sampling also obtained a variance estimator only slightly different from the g-weighted estimator. An alternative justification of the use of the g-weighted variance estimator was given by Yung and Rao (1996) by showing that it is identical (under the commonly made assumption of with replacement selection of clusters or PSUs) to the linerarized jackknife variance estimator, and using the fact that jackknife variance estimator is known to have good conditional properties as demonstrated by Royall and Cumberland (1981). Note that the term linearization of the jackknife signifies that the nonlinear random part in the estimator arising from the pseudo-replicated subsample is linearized about its full sample counterpart. Incidentally, the well known SUDAAN software also calculates the Taylor variance of the ratio estimator using the g-weighted variance estimator.

Despite the popularity of the g-weighted variance estimator, there is still uncertainty among the practitioners regarding its use because of the speculative nature of the superpopulation model used as a guide to choose among the variance estimators. Deville (1999) describes the use of g-weighted variance estimator for superior performance as a mystery. The main purpose of this paper is to provide a simple realistic justification of the preference of the g-weighted variance estimator by arguing that there is indeed a need of a suitable model for unbiased point estimation under the joint design-model distribution when ps is viewed as adjusting for the coverage bias reduction. Note that ps is often used in practice in the dual role of both variance and coverage bias reductions. In the sequel, we use the term ps for coverage bias adjustment. However, this definition is quite general in that the case of no coverage bias can be obtained as a special case by letting the bias tend to zero.

In this paper we approach the problem of variance estimation adjusted for coverage bias (via ps) by using the conceptual similarity to the variance estimation adjusted for nonresponse(nr) bias, results for which are well known. To this end, it is observed that use of calibration equations for nr and ps adjustments leads quite naturally to the nonoptimal estimating function (EF) framework introduced by Binder (1983) for variance estimation for finite population parameters under a quasi-design based framework, i.e., under the joint superpopulation model and design-based distribution, and the optimal EF framework introduced by Godambe (1960) and Godambe and Thompson (1989) for finite or infinite population parameters. In view of the superpopulation model required for correcting coverage bias or nr or both, we therefore propose a bias corrected estimating function (BCEF) approach under a quasi-design based framework for estimating variance adjusted for ps or nr or both. The BCEF are functions of data including design weights, finite population parameters, and parameters in the modeling of adjustment factors (g-weights) for bias due to nr and coverage errors. The proposed approach is based on a simple idea, and has been used implicitly by Folsom (1991) in the context of nonresponse bias adjusted variance estimation, but the authors are not aware of its use in the context of coverage bias adjustment. Section 2 first provides a review of variance estimation adjusted for nr, and then a motivation for the proposed method of variance estimation adjusted for ps using similarity with the problem of adjustment for nr. The BCEF method is described in Section 3. It is shown that (i) ps has the dual property of variance reduction as well as bias reduction because both objectives give rise to the same set of estimating equations, (ii) by linearizing the EF about the estimated bias-model parameters, one obtains a simple direct justification of why g-weights should be used in variance estimates adjusted for calibration; earlier justifications come from model-assisted and empirical considerations, and (iii) the variance estimator advocated by Deville-Särndal (1992) for the generalized raking estimators can be justified for a subclass having exponential-type adjustment factors. Note that the above observations can also be made by considering directly the calibration estimators (rather than EF) and linearizing them about the true bias-model parameters. However, for multivariate or complex nonlinear estimators, BCEF is expected to offer a simpler framework. Moreover, using properties inherent to EF, it is further shown that (iv) although, asymptotically the calibration estimator is unique only up to a constant multiple tending to unity, and so is the variance estimator, the finite sample Godambe optimality of EF (analogous to score functions) does provide a unique estimator (analogous to mle) whenever it exists, and (v) for the general multivariate case, the calibration-adjusted variance estimate can be obtained by a simple sandwich-type formula (analogous to the inverse of the information matrix) which is relatively easy for computer automation. Illustrative examples are presented in Section 4. Finally, Section 5 contains a brief summary.

2. Review and Motivation

It would be useful to review variance estimation adjusted for nr because of the conceptual similarity between the problems of variance estimation adjusted for the nr bias, and adjusted for the coverage bias, and the fact that the g-weight (in the form of nr adjustment factor) does show up in the variance estimation adjusted for nr under the joint design-model distribution. Based on this observation, it follows that handling coverage bias via ps would indeed require a model to achieve design-model unbiasedness of the bias-adjusted point estimator (recall objective (iv) mentioned in the introduction), and that the g-weights would appear naturally in the variance estimator. Thus we are borrowing ideas from methods to deal with the nr problem to the problem of coverage bias. It may be of interest to note that ever since the publication of the seminal paper on calibration estimation (viewed as a general form of ps) by Deville and Särndal (1992), research efforts have been among others to take the reverse route, i.e., to try to use calibration methods for the purpose of nr adjustment. However, in the process the coverage bias reduction aspect of ps has apparently been overlooked. Selected papers on the use of calibration methods for dealing with nr are due to Folsom (1991) who used the raking (traditionally used for ps) idea to fit the inverse logistic model for nr adjustment, thus ensuring that the adjustment factor is at least 1; Fuller, Loughin, and Baker (1994) used the usual regression model with known totals from external sources to adjust for both nr and ps; Singh, Wu, and Boyer (1995), similar to Folsom (1991), proposed a raking-type calibration method for the inverse logistic nr model except that the unit-level information for the nonrespondents was not deemed to be available, and instead the calibration feature of the estimating equations was exploited to use external controls obtained from alternative sources such as census or administrative data; and Lundstrom and Sarndal (1999) proposed use of a single common regression model with full sample level controls used for nr and census -level controls for ps. The recent paper by Folsom and Singh (2000) also describes a new calibration method which provides a unified approach of weight adjustment for extreme values, nr, and ps.

In this paper, as mentioned above, we turn to ideas used for point and variance estimation in the presence of nr to deal with the problem of coverage errors. To this end, we first review the calibration approach to nonreponse bias adjustment. It is assumed that specific to the survey objectives and conditions, we can assign a random variable δ_k taking values of 1 or 0 for the response indicator to each unit in the finite population. Thus we suppose a superpopulation model ξ_1 for the response indicator for eack k in U, given by

$$P_{\xi 1}\text{ (kth unit in U responds)} = g_{1k}^{-1}(\lambda_1), \qquad (2.1)$$

which implies that the adjustment factor is inverse of the response propensity, and the adjusted estimator $\Sigma_s y_k d_k g_{1k}(\lambda_1)$, based on the respondent subsample s, becomes unbiased for T_y under the joint $p\xi_1$- distribution. For the inverse logistic model $g_{1k}(\lambda) = 1 + e^{-x'_k\lambda_1}$ (which ensures that the adjustment factor is at least 1). Folsom (1991) proposed to use the estimating equation for estimating λ,

$$\Sigma_s y_k d_k (1 + e^{-x'_k\lambda_1}) = \tilde{T}_x \qquad (2.2)$$

where $\tilde{T}_x$ denotes random controls obtained from the full sample. The above equation can be solved by the usual raking method used for ps. A similar adjustment was also proposed by Singh, Wu, and Boyer (1995) in the context of nr with longitudinal surveys where for some covariates the unit-specific information for the nonrespondents and hence the control total $\tilde{T}_x$ was not available. They suggested to replace $\tilde{T}_x$ by a reasonable alternative set of nonrandom values T_x obtained from some other source. Note that the above approach for nr is clearly similar to the calibration for ps except that the adjustment factor is restricted to be at least 1. Now using the idea of sandwich-type variance estimation for estimators of finite population parameters via estimating functions as pioneered by Binder (1983), Folsom proposed an equivalent version in terms of residuals as in the case of generalized regression estimators. This variance estimate can also be derived directly by linearizing the estimators $\hat{T}_y(\hat{\lambda}_1)$, and $\hat{T}_x(\hat{\lambda}_1)$ about λ_1 as follows. Since

$$\hat{T}_x(\hat{\lambda}_1) - \tilde{T}_x \approx \hat{T}_x(\lambda_1) - \tilde{T}_x - H_{22}(\lambda_1)(\hat{\lambda}_1 - \lambda_1)$$

we have, $\hat{T}_y(\hat{\lambda}_1)$ given by

$$\Sigma_s y_k d_k g_{1k}(\hat{\lambda}_1) \approx \hat{T}_y(\lambda_1) - H_{12}(\lambda_1) H_{22}^{-1}(\lambda_1)(\hat{T}_x(\lambda_1) - \tilde{T}_x) \qquad (2.3)$$

where $H_{12}(\lambda_1) = -\Sigma_s y_k d_k (\partial g_{1k}(\lambda_1)/\partial\lambda_1)'$

$$H_{22}(\lambda_1) = -\Sigma_s x_k d_k (\partial g_{1k}(\lambda_1)/\partial\lambda_1)'.$$

Therefore,

$$\hat{T}_y(\hat{\lambda}_1) \approx \Sigma_s e_k d_k g_{1k}(\lambda_1) + B(\lambda_1)\tilde{T}_x, \qquad (2.4)$$

where e_k are the residuals $y_k - B(\lambda_1)x_k$, and $B(\lambda_1) = H_{12}(\lambda_1) H_{22}^{-1}(\lambda_1)$.

Now the Taylor linearized variance estimate of the nr adjusted point estimator can be obtained from standard methods in survey sampling. In fact, the design variance is conditional about $\Sigma_U y_k \delta_k g_{1k}(\lambda_1)$ given ξ_1, and the second term in the unconditional variance about $T_y = \Sigma_U y_k$ is negligible by comparison under regularity conditions. Note that λ_1 is replaced by its consistent estimate $\hat{\lambda}_1$ in the variance expression obtained from the right hand side of (2.4). The above linearization (2.4) is useful for interpretation in that the residuals e_k are expected to have less variability than y_k due to correlation between y and x, and so a net variance reduction is expected to be realized, in general, despite additional variance due to the presence of the g-weights and the random controls $\tilde{T}_x$. Here, we wish to emphasize the presence of the g-weights, i.e., $g_k(\lambda)$ in the expression (2.4) which gives rise to the g-weighted variance estimator. The alternative but equivalent variance estimate obtained as a sandwich (see next section) using EF as mentioned in Section 1 has a simple but general form, and is more amenable for automation.

Similarly, for ps, if we postulate a model (ξ_2) for coverage bias in the sense that for each k in U,

$$E_{\xi 2}\text{ (\# times the kth unit in U is enumerated)} = g_{2k}^{-1}(\lambda_2), \qquad (2.5)$$

then as in the case of nr we will naturally get the desired g-weighted Taylor variance estimator. In the model (2.5), it is assumed that each unit k in U corresponds to a realization of a random variable η_k (taking nonnegative integer values) defining the number of times the unit is listed or enumerated, and the model gives its expected value. The problem of undercoverage is quite common due to outdated frames for which η_k takes the value of 0 while the problem of overcoverage might arise due to multiple listings, e.g., children may get listed both at their homes by parents as well as in the dorms in which case η_k is at least 1. Thus the expectation of η_k may be more or less than or equal to 1. For adjusting for both nr and coverage biases, we can postulate two independent superpopulation models (ξ_1, ξ_2) such that under the joint $p\xi_1\xi_2$-distribution, the adjusted estimator is unbiased, and then the standard linearization gives rise to the appropriate g-weighted variance estimate. This is the motivation for the method proposed in the next section.

3. The Proposed Method

The calibration equations used in estimating λ-parameters of the superpopulation models defined in the previous section give rise to a set of bias corrected estimating functions (these are functions of data including design weights, finite population parameters, and parameters in the modeling of bias adjustment factors) with mean zero and appropriate covariance that can be consistently estimated under the $p\xi_1\xi_2$-distribution. In fact, along the lines of the argument mentioned in the previous section, only the design-based variance estimation is sufficient for consistent estimation, and this is usually given by a conservative approximation under the assumption of with replacement selection of first stage units. The above defines a semiparametric (in the sense of first two moments) finite population model in a quasi-design based framework from which estimators and their variances can be obtained analogous to the framework of score functions derived from the log-likelihood. However, unlike score functions (whose dimension matches the number of unknown parameters), there may be more (elementary) estimating functions than the number of parameters. This could happen in the case of panel data where several panel specific estimating functions may contain the same set of parameters. Godambe and Thompson (1989) provide an optimal method for combining elementary estimating functions. The estimating function approach is quite appealing in practice because like maximum likelihood estimation and information matrix obtained from score functions, it provides a simple but general formulation of point and variance estimation . This is the approach we take for the proposed method of BCEF to obtain Taylor linearization variance estimates. Note that the resampling methods provide an alternative method for variance estimation, but they may be computationally tedious for large data sets involving a large number of parameters in nr and coverage bias models which need to be estimated for each resample.

For the problem under consideration, the number of estimating functions is equal to the number of parameters, and so we don't need to consider the problem of combining elementary estimating functions. All we need is to define the sandwich matrix (analogue of the information matrix) for Taylor variance estimation by linearizing the estimating functions. We consider BCEF for coverage bias. The case for for both nr and coverage biases is analogous.

3.1 BCEF for Coverage Bias Adjusted via Poststratification

Under the $p\xi_2$-distribution, we have unbiased estimating functions $h_y(T_y, \lambda_2)$, $h_x(\lambda_2)$ defined by

$$\begin{pmatrix} h_y(T_y, \lambda_2) \\ h_x(\lambda_2) \end{pmatrix} := \begin{pmatrix} \Sigma_s\, y_k d_k g_{2k}(\lambda_2) - T_y \\ \Sigma_s\, x_k d_k g_{2k}(\lambda_2) - T_x \end{pmatrix} \tag{3.1}$$

The semiparametric finite-population model for estimating the population total T_y where λ_2 are the nuisance parameters, is now approximately given by

$$\begin{pmatrix} h_y(T_y, \lambda_2) \\ h_x(\lambda_2) \end{pmatrix} \sim \left(0,\ V_h(\lambda_2)\right), \tag{3.2}$$

where $V_h(\lambda)$ is a design-consistent estimate of the $p\xi_2$-covariance matrix of the vector of estimating functions. Now linearizing the h-functions about $\hat{T}_y$ and $\hat{\lambda}_2$, we get (note that unlike the estimators, the EFs are linearized about the model parameter estimates and not the parameters),

$$\begin{pmatrix} \Sigma_s\, y_k d_k g_{2k}(\lambda_2) - T_y \\ \Sigma_s\, x_k d_k g_{2k}(\lambda_2) - T_x \end{pmatrix} \approx \begin{pmatrix} H_{11}(\hat{\lambda}_2) & H_{12}(\hat{\lambda}_2) \\ H_{21}(\hat{\lambda}_2) & H_{22}(\hat{\lambda}_2) \end{pmatrix} \begin{pmatrix} \hat{T}_y(\hat{\lambda}_2) - T_y \\ \hat{\lambda}_2 - \lambda_2 \end{pmatrix} \tag{3.3}$$

where H_{11}, H_{12}, H_{21}, H_{22} are the partitioning submatrices of the negative Hessian H. In this case, H_{11}=1, H_{21}=0, and others are defined as in (2.3). We have

$$H^{-1} = \begin{pmatrix} 1 & -H_{12}H_{22}^{-1} \\ 0 & H_{22}^{-1} \end{pmatrix} \tag{3.4}$$

Therefore, $\hat{T}_y(\hat{\lambda}_2) - T_y$ is approximately

$$\left(1 \quad -H_{12}(\hat{\lambda}_2)H_{22}(\hat{\lambda}_2)^{-1}\right) \begin{pmatrix} \Sigma_s\, y_k d_k g_{2k}(\lambda) - T_y \\ \Sigma_s\, x_k d_k g_{2k}(\lambda) - T_x \end{pmatrix} \tag{3.5}$$

It follows that the Taylor variance estimate of $\hat{T}_y$ can be approximated as a sandwich-type variance $V\left(\hat{T}_y(\hat{\lambda}_2)\right)$ and is given by

$$\left(1 \ -H_{12}(\hat{\lambda}_2)H_{22}^{-1}(\hat{\lambda}_2)\right) V_h(\hat{\lambda}_2) \left(1 \ -H_{12}(\hat{\lambda}_2)H_{22}^{-1}(\hat{\lambda}_2)\right)' \tag{3.6}$$

Clearly the above variance estimate is g-weighted because $V_h(\lambda)$ is. In view of the correlation between y and x, we expect a net variance reduction in the g-weighted variance estimator despite variance inflationary effect due to the presence of the g-weights; see also the comment below (2.4). Also observe that If y is replaced by x, then the poststratified estimator reproduces T_x perfectly, and its variance should be zero. This is indeed the case with the sandwich variance (3.6). To check this, note that in the conformally partitioned matrices,

$$H = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix},\ V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix},$$

when y = x_1, the first x-variable (say), H_{11} is simply the (1,1) element of H_{22}, H_{12} is the first row of H_{22}, and H_{21} is the 1st column of H_{22}. Similar relationship exists in the submatrices of V. It follows that $H_{12}H_{22}^{-1}$ reduces to the row vector $(1, 0, ..., 0)$, and the sandwich variance (3.6) degenerates to zero.

If there is no coverage bias, then $\lambda_2 = 0$, and $g_{2k}(\lambda)=1$, and the Taylor expansion (3.3) can still be justified in the limiting sense as $\lambda_2 \to 0$ using continuity of the EFs. Note that we need this limiting argument because if we set $\lambda_2=0$ in the left hand side, the g-weights disappear from the EFs. However, theoretically it is valid to leave $g_{2k}(\hat{\lambda}_2)$ in the variance estimate $V_h(\hat{\lambda}_2)$ because $g_{2k}(\hat{\lambda}_2) \to 1$ (in p-probability) uniformly in k under regularity conditions. This would be beneficial if $g_{2k}(\hat{\lambda}_2)$ is not close to 1 or if there is some doubt about the coverage bias. In addition g-weighted variance estimator is known to have good conditional properties. The above gives a heuristic justification why use of g-weights in the variance estimate should lead to good empirical results in practice even when the objective is not coverage bias reduction.

Moreover in the absence of coverage bias for the class of generalized raking estimators, Deville and Särndal (1992, equation 3.4) advocate not only use of the g-weighted variance estimator involving residuals e_k (equivalent to the sandwich variance), but also use of somewhat different residuals defined as

$$e_k^{DS} = y_k - B^{DS}(\hat{\lambda}_2) \text{where } B^{DS}(\lambda_2) = H_{12}^{DS}(\lambda_2)\left(H_{22}^{DS}(\lambda_2)\right)^{-1},$$
$$H_{12}^{DS}(\lambda_2) = \Sigma_s y_k x'_k d_k g_{2k}(\lambda_2) \quad ,$$
$$H_{22}^{DS}(\lambda_2) = \Sigma_s x_k x'_k d_k g_{2k}(\lambda_2).$$

The residual version of the sandwich variance (3.6) does not give rise to the residuals e_k^{DS} because of the difference in H-matrices defined as in (2.3). It is easily seen that the two versions coincide if $\partial g_{2k}(\lambda_2)/\partial\lambda_2 = g_{2k}(\lambda_2)x_k$, i.e., the adjustment factor is of the exponential type. So for the subclass of generalized raking methods involving exponential type adjustment factors, (e.g., the usual raking), the proposed BCEF approach provides a direct theoretical justification of the Deville-Särndal's suggestion based on model-assisted considerations. However, for nonexponential-type adjustment factors such as the generalized regression, we are not able to reconcile the two versions.

Furthermore, it may be noted that in the absence of coverage bias, the estimator after ps does not change because the adjustment factor $g_{2k}(\lambda_2)$ (the functional form of which is motivated from some distance minimization criterion, see e.g., Deville and Särndal, 1992) is obtained such that for auxiliary variables x , the true value of T_x is perfectly estimated, i.e., λ_2 is estimated to satisfy

$$\Sigma_s x_k d_k g_{2k}(\hat{\lambda}_2) = T_x \tag{3.7}$$

The reason for using the above estimating equation is that in view of the anticipated high correlation between y and x, the estimator of T_y given by (3.5) is expected to have higher precision than the unadjusted estimator $\Sigma_s x_k d_k$, because when y is replaced by x, the error becomes zero. Thus with or without the presence of coverage bias, the expression for $\hat{T}_y(\hat{\lambda}_2)$ is identical, and therefore ps is expected to have the dual property of variance reduction as well as bias reduction.

Lastly we note that the above properties of the estimator $\hat{T}_y(\hat{\lambda}_2)$ could have been derived without using EFs although EFs do make it convenient. However, we mention two other properties which inherently need the framework of EFs. One is, of course, the computational simplicity realized via sandwich-type variance estimation already mentioned in the previous section. Other is the uniqueness of the point and variance estimators. Clearly, one cannot distinguish asymptotically between various estimators (and the corresponding variance estimators) differing by constant multiples tending to 1. However, using the finite sample optimality criterion of regular EFs in the sense of Godambe (1960) and Godambe and Thompson (1989) which does give rise to a unique EF up to a constant multiple, it is known that the resulting estimator is unique (provided a solution to the EF exists) and the corresponding variance estimator is given uniquely by the inverse of the Godambe information matrix. Notice that it is analogous to the properties of the score functions and maximum likelihood estimators. It follows that since the number of equations in BCEF is equal to the number of unknown parameters, BCEF are optimal in the Godambe sense, and hence the corresponding point and variance estimators are unique. $g_{1k}(\hat{\lambda}_1)$ and $g_{2k}(\hat{\lambda}_2)$. Vaish, Gordek, and Singh (2000) provide computational details for calibration-adjusted variance estimation for a general exponential model.

4. Examples of BCEF for Coverage Bias Adjustments via Poststratification

The following examples are presented assuming that there is coverage bias. Identical results also hold for the case of no coverage bias by considering the limit as the model parameter vector λ tends to zero. Here we suppress the subscript 2 used earlier for denoting the model parameters under ps.

4.1 Ratio Estimation

For the estimator defined in Section 1, we have $g_k(\lambda) = 1 + \lambda, \hat{\lambda} = N_a/\hat{N}_a - 1, \forall\, k \in U_a$

$$\begin{pmatrix} \Sigma_s\, y_k I_{k\in a} d_k (1+\lambda) - T_{ya} \\ \Sigma_s\, I_{k\in a} d_k (1+\lambda) - N_a \end{pmatrix} \approx \begin{pmatrix} 1 & -\hat{T}_{ya} \\ 0 & -\hat{N}_a \end{pmatrix} \begin{pmatrix} \hat{T}_{ya,\,ratio} - T_y \\ \hat{\lambda} - \lambda \end{pmatrix}$$

$$V\left(\hat{T}_{y,ratio}\right) \approx \left(1 \quad -\hat{T}_{ya}\hat{N}_a^{-1}\right) V_h(\hat{\lambda}) \left(1 \quad -\hat{T}_{ya}\hat{N}_a^{-1}\right)'$$
$$= \hat{V}\left(\Sigma_s (y_k I_{k\in a} - (\hat{T}_{ya}/\hat{N}_a) I_{k\in a}) d_k (1+\hat{\lambda})\right)$$
$$= (1+\hat{\lambda})^2\, \hat{V}(\Sigma_s e_k d_k)$$

which is exactly the g-weighted variance estimator $\hat{V}_{ratio,g}$ given in Section 1.

4.2 Generalized Regression (greg) Estimation

The greg estimator is given by

$$\hat{T}_{y,greg} = \Sigma_s\, y_k\, d_k (\,1 + x'_k\, \lambda\,),\; g_k(\lambda) = 1 + x'_k\, \lambda$$

Note that the ratio estimator is a special case of greg when the covariate x is a scalar indicating the domain a. Now, linearization of the BCEF gives

$$\begin{pmatrix} \Sigma_s y_k d_k (1 + x_k \lambda) - T_y \\ \Sigma_s x_k d_k (1 + x_k \lambda) - T_x \end{pmatrix} \approx \begin{pmatrix} 1 & -\Sigma_s y_k x_k' d_k \\ 0 & -\Sigma_s x_k x_k' d_k \end{pmatrix} \begin{pmatrix} \hat{T}_{y,greg} - T_y \\ \hat{\lambda} - \lambda \end{pmatrix}$$

and $V\left(\hat{T}_{y,greg}\right)$ is approximately

$$\hat{V}\left(\Sigma_s (y_k - (\Sigma_s y_k x_k' d_k)\,(\,\Sigma_s x_k x_k' d_k)^{-1} x_k\,)\, d_k\,(\,1 + x_k' \lambda)\right)$$

which is the well known g-weighted variance estimator $\hat{V}(\Sigma_s e_k d_k (\,1 + x_k' \lambda))$, see Särndal et al. (1989).

4.3 Raking Ratio Estimation

We have the estimator given by

$$\hat{T}_{y,\,raking} = \Sigma_k\, y_k\, d_k\, e^{x_k' \lambda}$$

and the corresponding BCEF is

$$\begin{pmatrix} \Sigma_s y_k d_k e^{x_k' \lambda} - T_y \\ \Sigma_s x_k d_k e^{x_k' \lambda} - T_x \end{pmatrix} \approx \begin{pmatrix} 1 & -\Sigma_s y_k x_k' d_k e^{x_k' \lambda} \\ 0 & -\Sigma_s x_k x_k' d_k e^{x_k' \lambda} \end{pmatrix} \begin{pmatrix} \hat{T}_{y,raking} - T_y \\ \hat{\lambda} - \lambda \end{pmatrix}$$

So the sandwich variance is obtained as

$$\hat{V}\left(\Sigma_s (y_k - (\Sigma_s y_k x_k' d_k e^{x_k' \lambda})(\Sigma_s x_k x_k' d_k e^{x_k' \lambda})^{-1} x_k) d_k e^{x_k' \lambda}\right)$$

which is the g-weighted variance estimator of the type recommended by Deville and Särndal.

5. Summary

The proposed bias corrected estimating function method was motivated by observing the similarity between ps and nr when one takes the perspective of coverage bias reduction in ps. The BCEF method is based on a simple semiparametric model built on estimating functions that are commonly used for estimating parameters for modeling nr and ps adjustment factor. It uses the joint pξ-distribution for specifying the model in a wide sense (i.e., up to first two moments). It provides a simple justification of why g-weights should be used in the Taylor linearized variance estimate for calibrated estimators. Using the property inherent to estimating functions, the BCEF method provides a sandwich variance estimate adjusted for ps or nr, or both (this is simply the inverse of the Godambe information matrix), which has a simple yet general form useful for computer automation. Also, using the Godambe finite sample optimality criterion of estimating functions, it is shown that the point and variance estimators (whenever the solution of the estimating function exists) are unique analogous to maximum likelihood estimation for parametric models.

References

Binder, D.A. (1983). On the variances of asymptotically normal estimates for complex surveys. *International Statistical Review*, 51, 279-92.

Deville, J.C. (1999). Variance estimation for complex statistics and estimators: linearization and residual techniques. *Survey methodology*, 25, 193-203.

Deville, J.-C., and Särndal, C.E. (1992). Calibration estimation in survey sampling. *JASA*, 87, 376-382.

Folsom, R.E., Jr. (1991). Exponential and logistic weight adjustments for sampling and nonresponse error reduction. *ASA Proc. Soc. Statist. Sec.*, 197-202.

Folsom, R.E., Jr., and Singh, A.C. (2000). A generalized exponential model for sampling weight calibration for extreme values, nonresponse, and poststratification. *ASA Proc. Surv. Res. Meth. Sec.*

Fuller, W. A., Loughin, M.M., and Baker, H.D. (1994). Regression weighting in the presence of nonresponse with application to the 1987-88 Nationwide Food Consumption Survey. *Survey Meth.*, 20, 75-85.

Godambe, V.P. (1960). An optimum property of regular maximum likelihood estimation. *Ann. Math. Statist.*, 31, 1208-1212.

Godambe, V.P. and Thompson, M.E. (1989). An extension of quasilikelihood estimation (with discussion). *J. Statist. Plan. Inf.*, 22, 137-172.

Isaki, C, T. and. Fuller, W.A.(1982). Survey design under the regression superpopulation model. *JASA.*, 77, 89-96.

Rao, J.N.K. (1985). Conditional inference in survey sampling. *Survey Methodology*, 11, 15-31.

Royall, R.M. and Cumberland, W.G. (1981). An empirical study of the ratio estimator and estimators of its variance. *J. Am. Statist. Assoc.*, 76, 66-88.

Lundström, S. and Särndal, C.E. (1999). Calibration as a standard method for treatment of nonresponse. *J. Off. Statist.*, 15, 305-327.

Särndal,C. E., Swensson, B., and Wretman, J.H. (1989). The weighted residual technique for estimating the variance of the general regression estimator of the finite population total. *Biometrika*, 76, 527-37.

Singh, A.C., Wu, S., and Boyer, R. (1995). Longitudinal survey nonresponse adjustment by weight calibration for estimation of gross flows. *ASA Proc. Surv. Res. Meth Sec.*, 396-401.

Yung, W. and Rao, J. N. K. (1996). Jackknife linearization variance estimators under stratified multistage sampling. *Survey Methodology*, 22, 23-31.