RESEARCH REPORT SERIES
(Statistics #2010-03)

Simultaneous Calibration and Nonresponse Adjustment

Eric V. Slud
Yves Thibaudeau

Statistical Research Division
U.S. Census Bureau
Washington, D.C. 20233

Report Issued: February 19, 2010

Simultaneous Calibration and Nonresponse Adjustment

by Eric V. Slud[1,2] & Yves Thibaudeau[1]

[1]Census Bureau SRD and [2]Statistics Program, Univ. of Maryland

December 10, 2009, corrected May 10, 2010

Abstract. Single and joint inclusion probabilities are generally available and known for complex survey designs up to the point where survey weights are modified due to nonresponse and population controls. Best practice by sophisticated survey practitioners generally includes weight modifications, first by calibration, ratio adjustment or raking to correct for nonresponse, next by further steps to impose population survey controls; and also, often, by final steps involving weight truncation or cell-collapsing to constrain the modified weights, usually so that the largest and smallest weights do not differ by more than a designated multiplicative factor. These adjustments are sometimes made in successive stages, the order of which may differ from one survey to another. In this article, generalized-raking calibration methodology is adapted to allow all of these adjustments, and possibly additional nonlinear constraints, to be accomplished in a single stage, after which linearization-based large-sample variance formulas are available.

Keywords. Consistency, inclusion probabilities, Lagrange multipliers, linearized variance, objective function, population controls, weight adjustment.

1 Introduction

Survey weights in large complex surveys are often modified from their designed values, for at least three reasons: to correct for nonresponse, to compensate for frame deficiencies by enforcing population controls (often by demographic categories), and to prevent the adjusted unit weights from being too different from one another. As a result of these modifications, there are usually no meaningful joint inclusion probabilities when complex survey results are analyzed.

Modifications to survey weights are generally applied in multiple stages, and although each stage may be well motivated, it is fair to say that in practice, the effect of propagating early-stage modifications through later adjustments is generally poorly understood. Moreover, the later adjustments (particularly the final population controls) are often repeated after trimming or compressing the most extreme weights until controls and moderate weights are achieved simultaneously. Usually, only the final population controls and weight-trimming criteria are imposed rigidly, with no assurance that the criteria used to adjust at earlier stages hold at the final stage.

This paper begins by summarizing the existing methods to correct survey data both for nonresponse and population controls, while retaining overall bounds on the weights. While some theoretically based methods do exist for enforcing two out of the three of these types of weight constraints, there do not seem to be methods which simultaneously incorporate all of them. A new framework is presented which handles all of these desirable weight adjustments *simultaneously* in a single stage, in a way which justifies linearized variance formulas for survey total estimates based on the adjusted weights, and which allows a tuning parameter to place more or less weight on initial nonresponse adjustment while strictly enforcing population controls.

1.1 Background Literature

There is a large body of literature on construction and modification of survey weights, much of which has been absorbed into standard survey methodology texts, like that of Särndal et al. (1992). A brief survey of the most important theoretical contributions includes:

- nonresponse adjustment by cell-based ratios or raking, as in Oh and Scheuren (1983), or by models fitted by model-assisted 'pseudo-likelihood' as in Kim and Kim (2007);
- weight modification via calibration leading up to generalized raking in Deville and Särndal (1992) and Deville, Särndal & Sautory (1993), papers which established linearized variance formulas for weighted survey totals;
- linear regression-based approaches to nonresponse weight adjustment surveyed in Fuller (2002), treated more fully in the monograph of Särndal and Lundström (2005) which discusses both simultaneous and two-stage calibration to benchmarks or controls along with nonresponse adjustment;
- (single-stage) methods combining weight-truncation with weight-adjusted calibration as in Singh and Mohl (1996) and Théberge (2000);
- Lu and Gelman (2003), a paper which estimates variances via a delta method for a survey in which weights are modified both by poststratification (an operation which includes cell-based ratio adjustment for nonresponse) and raking;
- methods like those of Elliott and Little (2000) or Beaumont (2008), which modify weights within a model-based framework, allowing use of auxiliary information either in the spirit of calibration or of weight smoothing;
- methods for handling 'informative' nonresponse, including those of Pfefferman and Sverchkov (2004) and Chang and Kott (2008), the latter treating weight adjustment for nonresponse using an estimating equation based on a generalized-linear response model;

- and many references which discuss adjustment for nonresponse followed by a separate calibration stage, e.g., Yung and Rao's (2000) discussion of jackknife variance estimation in such a setting.

1.2 Notation and Assumptions

Consider a sample survey with a frame $\mathcal{U}$ from which a probability sample $\mathcal{S}$ is drawn according to a plan with known single and double inclusion probabilities π_k, π_{kj}, for $k, j \in \mathcal{U}$. Assume that the total $Y = t_y = \sum_{k\in\mathcal{U}} y_k$ of a scalar attribute is of primary interest, and that $(y_k,\, x_k,\; k \in \mathcal{S})$ is (potentially) observable, i.e., the sample data include the auxiliary p-dimensional vector x_k. This setting corresponds to the *InfoS* sampling framework of Särndal and Lundström (2005), with auxiliary data available at sample but not frame level.

Assume that each sampled individual in the survey decides independently whether or not to respond. Without loss of generality, denote by r_k for all $k \in \mathcal{U}$ the indicator which is 1 or 0 respectively if the k'th individual *would* or *would not* have responded if sampled, and assume that these random variables are independent of each other and of the sample selection mechanism. (However, in some surveys this assumption could be applied only with 'individuals' replaced by households.) The observable data are now taken to be $(y_k \cdot r_k,\, r_k,\, x_k,\; k \in \mathcal{S})$. No restriction other than positivity is placed on the probabilities

$$P(r_k = 1) \;=\; Er_k \;\equiv\; \rho_k$$

with which individual units respond. However, these quantities ρ_k must be estimated in order to adjust weights for nonresponse, and this is typically done either by ratio-adjustment and raking (Oh and Scheuren 1983) or by using a working generalized-linear parametric model (Kim and Kim 2007)

$$1/\rho_k \;=\; \kappa(\lambda' x_k) \quad , \qquad k \in \mathcal{U}$$

where λ is a p-dimensional parameter vector which is estimated from sample data through the solution $\hat{\lambda}$ to an estimating equation. The most important example of such a working nonresponse model is the case treated in this paper,

$$1/\rho_k \;\equiv\; (Er_k)^{-1} \;=\; \kappa(\lambda' x_k) \;=\; 1 + \lambda'\, x_k \tag{1}$$

This model motivates the estimation of ρ_i through $\hat{\lambda}$ defined from a nonresponse-adjustment constraint equation (Särndal and Lundström 2005) which requires that modified weights $w_k \equiv r_k\, w_k^o / \hat{\rho}_k$ satisfy

$$\sum_{k\in\mathcal{S}} r_k\, w_k\, \mathbf{x}_k \;=\; \sum_{k\in\mathcal{S}} w_k\, \mathbf{x}_k \;=\; t^*_{\mathbf{x}} \tag{2}$$

The constrained totals $t^*_{\mathbf{x}}$, which may not always be close to the true frame totals, will generally arise in one of two ways. They may derive from a survey or census –

possibly by some form of projection or updating – which is believed to be larger or more accurate than the current survey. Alternatively, the $\mathbf{x}$ variables and totals may be known only for the sampled units in the current survey, in which case

$$t^*_{\mathbf{x}} \;\equiv\; \hat{t}_{\mathbf{x},\pi} \;\equiv\; \sum_{k\in\mathcal{S}} w^o_k\,\mathbf{x}_k \qquad \text{or} \qquad t^*_{\mathbf{x}} \;\equiv\; N\,\frac{\sum_{k\in\mathcal{S}} r_k\, w^o_k\,\mathbf{x}_k}{\sum_{k\in\mathcal{S}} r_k\, w^o_k} \tag{3}$$

where $w^o_k = 1/\pi_k$ denote the initial or design weights and N is the frame population size, assumed known. The census source for $\mathbf{x}_k$ totals corresponds to the *infoU* setting of Särndal and Lundström (2005), while the source (3) — which is far more common for nonresponse adjustment — corresponds to those authors' *infoS* setting.)

The nonresponse-adjusted weights $w_k \;=\; r_k\, w^o_k/\hat{\rho}_k \;=\; r_k\, w^o_k\,(1+\hat{\lambda}'\mathbf{x}_k)$ are often treated as a distinct weight-adjustment stage, and are used as input to further weight modification stages. A special case of the linear-calibration weight adjustment is the standard ratio adjustment corresponding to a set $C_1,\,\ldots,\,C_K$ of *adjustment cells* partitioning the frame $\mathcal{U}$, where the components of $\mathbf{x}_k$ are defined by $x_{k,j} = I_{[k\in C_j]}$.

The results of surveys designed to estimate totals and ratios of totals are often reported after controlling total numbers of units within designated population cells to be equal to the totals found in a more comprehensive survey or (updated) census, generally through constraints on final weights

$$\sum_{k\in\mathcal{S}} r_k\,\hat{w}_k\,\mathbf{z}_k \;=\; t^*_{\mathbf{z}} \tag{4}$$

Here $t^*_{\mathbf{z}}$ is a known vector approximating the frame total $t_{\mathbf{z}}$, for a vector $\mathbf{z}_k = (z_{1k},\ldots,z_{qk})$ of survey variables defined for each unit $k\in\mathcal{U}$. The constraint (4) is imposed on any system of survey weights $\{\hat{w}_k\}_{k\in\mathcal{S}}$ however obtained — by modifications for nonresponse, population controls, and weight compression or truncation — starting from a designed system $w^o_k \;=\; 1/\pi_k$ of inverse inclusion probabilities. The final weights $\hat{w}_k$ are ultimately used in estimating population totals of survey variables $y_k,\; k\in\mathcal{U}$, by weighted totals

$$\hat{t}_{y,adj} \;=\; \sum_{k\in\mathcal{S}} r_k\,\hat{w}_k\,y_k \tag{5}$$

2 A New Weight-Adjustment Framework

The objective of the present research is to accomplish nonresponse adjustment, population-control calibration, and weight-truncation in a single step, with a linearization-based formula for variance. The single step may require an iteratively calculated solution to estimating equations, but only a single objective function is being optimized. The framework is similar to that of Deville and Särndal (1992) and Deville, Särndal and Sautory (1993), where the magnitudes of weight modification are kept as small as possible through a loss function. As in those papers, the system of initial weights $w_k^o = 1/\pi_k$ arises from the design inclusion probabilities. The modification from design weights to final weights is viewed notionally as $\{w_k^o\} \mapsto \{w_k\} \mapsto \{\hat{w}_k\}$, with only the final weights appearing in the survey estimates (5), but now the two sets $\{w_k\}$, $\{\hat{w}_k\}$ of survey weights are created simultaneously to obey respective constraints (4) and (2), both contributing to an aggregated loss function. The auxiliary weight differences $w_k - w_k^o$ can be interpreted as the component of the overall weight-modification $\hat{w}_k - w_k^o$ that is due to nonresponse adjustment.

The three systems of weights $\{w_k^o\}_k$, $\{w_k\}_k$, $\{\hat{w}_k\}_k$ are related through the desire to minimize simultaneously the losses

$$\sum_{k \in \mathcal{S}} r_k \, w_k^o \, G_1(w_k / \, w_k^o \, - \, 1) \qquad \text{and} \qquad \sum_{k \in \mathcal{S}} r_k \, w_k^o \, G_2((\hat{w}_k - w_k)/w_k^o)$$

where each of $G_1(z)$, $G_2(z)$ is a convex loss-function which is locally of the form $z^2/2$ plus a term of smaller order (like z^3) near $z = 0$. The intermediate and final modified weights $\mathbf{w} = \{w_k\}_{k \in \mathcal{S}}$ and $\hat{\mathbf{w}} = \{\hat{w}_k\}_{k \in \mathcal{S}}$ are found together, subject to the constraints (2) and (4), by the objective-function minimization

$$\min_{\mathbf{w},\,\hat{\mathbf{w}}} \; \sum_{k \in \mathcal{S}} r_k \, w_k^o \Big\{ \, G_1(\frac{w_k}{w_k^o} \, - 1) \, + \, \alpha \, G_2(\frac{\hat{w}_k - w_k}{w_k^o}) \, + \, Q(\frac{\hat{w}_k}{w_k^o}) \, \Big\} \tag{6}$$

where $\alpha > 0$ is a constant chosen by the statistician and Q is a convex and piecewise smooth penalty term which is nonzero only for large or small weight ratios, and enters this single optimization step to enforce *weight-truncation* or *restricted weights* as in Singh and Mohl (1996) or Théberge (2000). The most important instances of (6) will have $G_1(z) = G_2(z) = z^2/2$, and Q a piecewise smooth function such that $Q(z) \equiv 0$ on an interval $[c_1, c_2]$, for fixed constants $0 < c_1 < 1 < c_2 < \infty$, and $Q(z)$ large when $\max(c_1 - z, \, z - c_2)$ is positive and not very small.

Unlike the design weights, both sets of modified weights are positive only for indices $k \in \mathcal{S}$ for which $r_k = 1$. Both sets of weight changes $w_k - w_k^o$, $\hat{w}_k - w_k$ might be meaningfully large, the first because of significant nonresponse and the second because of important differences between the coverage of the current survey and the (presumably more reliable) one to which it is being controlled. Since only the final weights $\hat{w}_k$ are ultimately used in survey estimation, these are the only weights that are truncated, by penalizing weights outside a certain range of multiples of the base weights w_k^o. However, both sets of modified weights are restricted by the calibration equation combining (2) and (4) into :

$$\sum_{k \in \mathcal{S}} r_k \begin{pmatrix} w_k \, \mathbf{x}_k \\ \hat{w}_k \, \mathbf{z}_k \end{pmatrix} = \begin{pmatrix} t^*_{\mathbf{x}} \\ t^*_{\mathbf{z}} \end{pmatrix} \tag{7}$$

The next two sub-sections compare limiting cases of the proposed adjustment (6) with previously studied two-stage and calibration methods.

2.1 Special and Limiting Cases

In (6), the (strict) convexity of the objective-function implies that the weights $\mathbf{w}$, $\hat{\mathbf{w}}$ (restricted to the finitely many indices in $\mathcal{S}$ for which $r_k = 1$) have a unique optimal solution. In several special and limiting cases, the solution relates simply to existing methods.

(Case 1. $r_k \equiv 1$ and $t^*_{\mathbf{x}} = \hat{t}_{\mathbf{x},\pi}$). In this full-response case, the $\mathbf{w}$ weights are unconstrained, and the minimization (6) becomes a pure 'generalized raking' problem with penalized weights, as in Singh and Mohl (1996) and Théberge (2000), following Deville and Särndal (1992). In the special sub-case where $G_1(z) \equiv G_2(z) \equiv z^2/2$, it is easy to check that $w_k - w^o_k = \alpha(\hat{w}_k - w^o_k)/(1+\alpha)$, and that $\hat{w}_k$ minimizes a penalized-linear-calibration objective function

$$\sum_{k\in\mathcal{S}} r_k \Big(\frac{\alpha}{2(1+\alpha)} \, \frac{(\hat{w}_k - w^o_k)^2}{w^o_k} \; + \; w^o_k \, Q(\frac{\hat{w}_k}{w^o_k}) \Big)$$

subject to the constraint (4). Within this sub-case, if $Q \equiv 0$, then the final weights $\hat{w}_k$ coincide with the calibrated 'g' weights arising in generalized regression (Särndal et al. 1992, Deville and Särndal 1992) subject to (4).

(Case 2. Omission of constraints (4)) If the population-control constraints (4) are omitted, or if it happens that $\hat{w}_k = w_k$, then when $G_1(z) \equiv z^2/2$ and $Q \equiv 0$, the weights $\hat{w}_k$ are the same as the calibrated nonresponse-adjusted weights found as in Särndal and Lundström (2005).

(Case 3. $\alpha \to \infty$) If the parameter α in (6) is taken to be very large, for fixed Q, then Appendix D shows for the sub-case $G_j(z) \equiv z^2/2$, $j = 1, 2$, and $Q \equiv 0$, that the limiting systems $\mathbf{w}$ and $\hat{\mathbf{w}}$ of weights are identical, i.e., in the large-α limit, $\hat{w}_k \equiv w_k$ for all $k \in \mathcal{S}$ such that $r_k = 1$ (and $\hat{w}_k = w_k = 0$ for all other indices $k \in \mathcal{U}$). A similar result can be expected to hold also in the case of general convex loss-functions G_j, since large α makes the G_1 loss term tiny compared to the G_2 term in (6) .

Thus, when α is large, the weight optimization problem (6) approximates the problem of finding $\{\hat{w}_k \,:\, k \in \mathcal{S},\ r_k = 1\}$, subject to

$$\sum_{k\in\mathcal{S}} r_k \, \hat{w}_k \begin{pmatrix} \mathbf{x}_k \\ \mathbf{z}_k \end{pmatrix} \; = \; \begin{pmatrix} t^*_{\mathbf{x}} \\ t^*_{\mathbf{z}} \end{pmatrix} \tag{8}$$

such that

$$\min_{\hat{\mathbf{w}}} \sum_{k\in\mathcal{S}} r_k \, w^0_k \, \{\alpha \, G_1(\frac{\hat{w}_k - w^o_k}{w^o_k}) \, + \, Q(\frac{\hat{w}_k}{w^o_k})\} \tag{9}$$

Within this Case, with $G_j(z) \equiv z^2/2$, $Q \equiv 0$, the problem is one of pure linear calibration as in Särndal and Lundström (2005). Although those authors did not explicitly treat the problem of simultaneous nonresponse adjustment and population controls in a calibration setting, that would have fallen easily within the framework of their book, since they formulate nonresponse adjustment as a calibration problem in their Chapter 6.

The simultaneous imposition of adjustment cell and population control constraints in (8) could lead to numerical instabilities, especially when these constraints share some variables for which the imposed totals do not agree closely. In that case, (8)–(9) with Q taking large values at extreme weight ratios provides a more stable, smoothed calibrated weight adjustment.

(Case 4. $\alpha \to 0$**)** In the limit as $\alpha \to 0$, the connection between w_k and $\hat{w}_k$ apparently grows weaker and weaker. However, the explicit limits for Lagrange multipliers and weights, found in Appendix E under the additional restriction $Q \equiv 0$, shows that the population controls as well as the nonresponse adjustments play an important role. This must be so, because the final weights identically satisfy the calibration constraint (4), while (2) generally does not hold precisely with w_k replaced by $\hat{w}_k$. See equation (34) in the Appendix for the small-α limit of $\sum_{k \in \mathcal{S}} \hat{w}_k \, \mathbf{x}_k$ when $Q \equiv 0$. This limit will be close in large superpopulation samples to $t^*_{\mathbf{x}}$ when both $t^*_{\mathbf{x}} \approx t_{\mathbf{x}}$ and $t^*_{\mathbf{z}} \approx t_{\mathbf{z}}$. In other cases, this small-α limiting form of nonresponse-adjustment constrained total may actually be preferable to $t^*_{\mathbf{x}}$ since the limit can be interpreted as an adjusted form of $t^*_{\mathbf{x}}$ to achieve greater compatibilty with the population-constrained $t^*_{\mathbf{z}}$ totals.

(Case 5. $Q \equiv 0$**)** When there is no penalty Q for extreme weight-ratios $\hat{w}_k / w^o_k$, the optimal weights (16) found below resemble a ridge-regression form of linearly calibrated weights, in which the population control condition (4) holds precisely but (2) does not. However, the particular form (16) together with (32)–(33) has not arisen before.

2.2 Combined Nonresponse Adjustment and Calibration

Many large-scale complex surveys, such as the Census Bureau's American Community Survey (ACS) and Survey of Income and Program Participation (SIPP), are analyzed by first adjusting for nonresponse by a cell-based ratio or raking method and then later by imposing population controls. (In Census Bureau surveys, those controls usually require that population totals in certain demographic and geographic categories match those of the demographically updated decennial census.) This two-step approach is implicit in much of the survey sampling journal literature, and explicit in some sources, such as Yung and Rao (2000) and Särndal and Lundström (2005, Ch. 8 & Sec. 11.4 on the 'Two-Step Method' $\hat{Y}_{W2A}$), which treat variance estimation at a realistic level of complexity. Lu and Gelman (2003) also provide variances of the two-step estimators based on Taylor linearization, in the setting where simple random sampling can be assumed within the intersections of strata and post-strata, essentially by using a numerical perturbation finite-difference method to obtain linearization coefficients.

The method of weight adjustment proposed here, in (6), seems not to have any special or limiting cases which exactly reproduce the known two-step method. In (6), the asymmetric role of the nonresponse adjustment contraints (2) and population controls is made explicit through choice of the parameter α. In accord with current practice, the population controls are required to hold exactly for the final weights actually used in survey estimation. If (2) is to hold at least approximately for w_k replaced by $\hat{w}_k$, then α should be chosen large. When that is done, and there is no penalty term Q in (6), **Case 3** above showed that the proposed method is a simultaneous calibration or generalized raking in the spirit of Särndal and Lundström (2005) or Deville and Särndal (1992). Yet calibration to satisfy (8), i.e., both constraints exactly, can exaggerate the dissimilarity among some design weights, resulting in large variances for some survey totals. For that reason, the method proposed here relaxes the $\mathbf{x}$ constraints in (8) and penalizes large and small weights through Q.

The remainder of this paper develops numerical algorithms and asymptotic properties of the adjusted weights, and of survey totals using them, for the case where $G_j(z) \equiv z^2/2$, which we refer to as the *linear* calibration case because, apart from penalty terms involving Q, the equations (11)–(12) for survey weights are linear in the weights and Lagrange multipliers, a helpful simplification. The superpopulation conditions **(A.0)-(A.4)** of Appendix A are assumed from now on.

3 Numerical Solution in Linear Case

In the important particular case where $G_1(z) \equiv G_2(z) = z^2/2$, the simultaneous adjustment and calibration step described above is the minimization subject to (7) over $(\lambda, \mu, \mathbf{w}, \hat{\mathbf{w}})$ of

$$\sum_{k \in \mathcal{S}} r_k \Big[\frac{(w_k - w_k^o)^2}{2\, w_k^o} \;+\; \alpha\, \frac{(\hat{w}_k - w_k)^2}{2\, w_k^o} \;+\; w_k^o\, Q(\frac{\hat{w}_k}{w_k^o}) \;-\; \hat{w}_k\, \mu' \mathbf{z}_k \;-\; w_k\, \lambda' \mathbf{x}_k \Big] \;+\; \mu' t^*_{\mathbf{z}} + \lambda' t^*_{\mathbf{x}} \qquad (10)$$

where $\lambda \in \mathbf{R}^p$ and $\mu \in \mathbf{R}^q$ are Lagrange multiplier vectors. The solution is determined by (7) together with the following equations obtained by equating to 0 the derivatives of (10) respectively with respect to w_k and $\hat{w}_k$:

$$\alpha\, \hat{w}_k \;=\; (1+\alpha)\, w_k \;-\; w_k^o \;-\; w_k^o\, \lambda' \mathbf{x}_k \qquad (11)$$

$$\alpha\, \hat{w}_k \;+\; w_k^o\, Q'(\frac{\hat{w}_k}{w_k^o}) \;=\; \alpha\, w_k \;+\; w_k^o\, \mu' \mathbf{z}_k \qquad (12)$$

for all $k \in \mathcal{S}$ such that $r_k = 1$.

Then the solution equations become, for $k \in \mathcal{S}$ with $r_k = 1$,

$$\begin{pmatrix} w_k \\ \hat{w}_k \end{pmatrix} = w_k^o \, (1 + \mu' \mathbf{z}_k + \lambda' \mathbf{x}_k) \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \frac{w_k^o}{\alpha} \begin{pmatrix} 0 \\ \mu' \mathbf{z}_k \end{pmatrix} - \frac{w_k^o}{\alpha} \, Q'(\frac{\hat{w}_k}{w_k^o}) \begin{pmatrix} \alpha \\ 1+\alpha \end{pmatrix}$$

first by subtracting (11) minus (12) and then by expressing $\hat{w}_k - w_k$ through (12). The second of these equations can be rewritten in the form

$$\hat{w}_k + \frac{1+\alpha}{\alpha} \, w_k^o \, Q'(\frac{\hat{w}_k}{w_k^o}) = w_k^o \Big\{ 1 + \frac{1+\alpha}{\alpha} \, \mu' \mathbf{z}_k + \lambda' \mathbf{x}_k \Big\} \tag{13}$$

Together with (7), the last equations imply

$$\begin{pmatrix} t^*_{\mathbf{x}} \\ t^*_{\mathbf{z}} \end{pmatrix} = \sum_{k \in \mathcal{S}} r_k \begin{pmatrix} w_k \, \mathbf{x}_k \\ \hat{w}_k \, \mathbf{z}_k \end{pmatrix} = \sum_{k \in \mathcal{S}} r_k \, w_k^o \begin{pmatrix} \mathbf{x}_k \\ \mathbf{z}_k \end{pmatrix} +$$

$$+ \sum_{k \in \mathcal{S}} r_k \, w_k^o \begin{pmatrix} \mathbf{x}_k^{\otimes 2} & \mathbf{x}_k \, \mathbf{z}_k' \\ \mathbf{z}_k \mathbf{x}_k' & (1 + \frac{1}{\alpha}) \, \mathbf{z}_k^{\otimes 2} \end{pmatrix} \begin{pmatrix} \lambda \\ \mu \end{pmatrix} - \sum_{k \in \mathcal{S}} r_k \, w_k^o \, Q'(\frac{\hat{w}_k}{w_k^o}) \begin{pmatrix} \mathbf{x}_k \\ (1 + \frac{1}{\alpha}) \, \mathbf{z}_k \end{pmatrix}$$

where we adopt from now on the simplifying notation, for any column vector v, that $v^{\otimes 2} = v \, v'$ and v' denotes the transpose of v. Thus, with the definition

$$M_\alpha = \sum_{k \in \mathcal{S}} r_k \, w_k^o \begin{pmatrix} \mathbf{x}_k^{\otimes 2} & \mathbf{x}_k \, \mathbf{z}_k' \\ \mathbf{z}_k \mathbf{x}_k' & (1 + \alpha^{-1}) \, \mathbf{z}_k^{\otimes 2} \end{pmatrix}$$

the Lagrange multipliers are determined by

$$\begin{pmatrix} \lambda \\ \mu \end{pmatrix} = (M_\alpha)^{-1} \Big\{ \begin{pmatrix} t^*_{\mathbf{x}} \\ t^*_{\mathbf{z}} \end{pmatrix} - \sum_{k \in \mathcal{S}} r_k \, w_k^o \begin{pmatrix} \mathbf{x}_k \\ \mathbf{z}_k \end{pmatrix} + \sum_{k \in \mathcal{S}} r_k \, w_k^o \, Q'(\frac{\hat{w}_k}{w_k^o}) \begin{pmatrix} \mathbf{x}_k \\ (1 + \frac{1}{\alpha}) \, \mathbf{z}_k \end{pmatrix} \Big\} \tag{14}$$

Equation (14) would immediately yield (λ, μ) if the penalty function Q were identically 0 on the range of weight ratios $\hat{w}_k / w_k^o$, through the equation

$$\begin{pmatrix} \lambda \\ \mu \end{pmatrix} = (M_\alpha)^{-1} \Big\{ \begin{pmatrix} t^*_{\mathbf{x}} \\ t^*_{\mathbf{z}} \end{pmatrix} - \sum_{k \in \mathcal{S}} r_k \, w_k^0 \begin{pmatrix} \mathbf{x}_k \\ \mathbf{z}_k \end{pmatrix} \Big\} \tag{15}$$

and in this case, by the equation preceding (13), the weights have closed-form solutions

$$\begin{pmatrix} w_k \\ \hat{w}_k \end{pmatrix} = r_k \, w_k^o \Big\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} \mathbf{x}_k & \mathbf{z}_k \\ \mathbf{x}_k & (1 + \alpha^{-1}) \mathbf{z}_k \end{pmatrix} M_\alpha^{-1} \begin{pmatrix} t^*_{\mathbf{x}} - \hat{t}_{r\mathbf{x},\pi} \\ t^*_{\mathbf{z}} - \hat{t}_{r\mathbf{z},\pi} \end{pmatrix} \Big\} \tag{16}$$

We now describe, in the case with nonzero Q satisfying $Q'(-\infty) = -\infty$ and $Q'(\infty) = \infty$, an iterative algorithm for determining simultaneously the multipliers (λ, μ) and the weights $\mathbf{w}, \hat{\mathbf{w}}$. The initial values $(\lambda^{(0)}, \mu^{(0)})$ are defined by the right-hand side of (15). Next, since Q' is increasing and has range $(-\infty, \infty)$, the function $x + (\alpha^{-1} + 1)\, Q'(x)$ is strictly increasing, with range the whole real line, and its inverse function h is uniquely specified on the whole real line by the properties

$$h(u) + \frac{1+\alpha}{\alpha}\, Q'(h(u)) \equiv u \qquad \text{and} \quad h(1) = 1 \tag{17}$$

In terms of the inverse function h, equation (13) is rewritten more handily as

$$\hat{w}_k = w_k^o \cdot h\Big(1 + \frac{1+\alpha}{\alpha}\, \mu' \mathbf{z}_k + \lambda' \mathbf{x}_k\Big) \tag{13$'$}$$

while (14) becomes

$$\frac{1}{N} M_\alpha \begin{pmatrix} \lambda \\ \mu \end{pmatrix} - \frac{1}{N} \sum_{k \in \mathcal{S}} r_k\, w_k^o\, Q' \circ h(1 + \frac{1+\alpha}{\alpha}\, \mu' \mathbf{z}_k + \lambda' \mathbf{x}_k) \begin{pmatrix} \mathbf{x}_k \\ (1+\alpha^{-1}) \mathbf{z}_k \end{pmatrix}$$

$$= \frac{1}{N} \left\{ \begin{pmatrix} t_{\mathbf{x}}^* \\ t_z^* \end{pmatrix} - \sum_{k \in \mathcal{S}} r_k\, w_k^o \begin{pmatrix} \mathbf{x}_k \\ \mathbf{z}_k \end{pmatrix} \right\} \tag{14$'$}$$

Thus, the iterative calculation for $(\lambda^{(j)}, \mu^{(j)})$ and $\{w_k^{(j)}, \hat{w}_k^{(j)} : k \in \mathcal{S}\}$ starting at $j = 0$ with (15), is given for $j \geq 1$ by

$$\hat{w}_k^{(j)} = w_k^o \cdot h(1 + \frac{1+\alpha}{\alpha}\, \mathbf{z}_k'\, \mu^{(j-1)} + \mathbf{x}_k'\, \lambda^{(j-1)}) \tag{18}$$

$$\begin{pmatrix} \lambda^{(j)} \\ \mu^{(j)} \end{pmatrix} = (M_\alpha)^{-1} \left\{ \begin{pmatrix} t_{\mathbf{x}}^* \\ t_{\mathbf{z}}^* \end{pmatrix} - \sum_{k \in \mathcal{S}} r_k\, w_k^0 \begin{pmatrix} \mathbf{x}_k \\ \mathbf{z}_k \end{pmatrix} + \sum_{k \in \mathcal{S}} r_k\, w_k^o\, Q'(\frac{\hat{w}_k^{(j)}}{w_k^o}) \begin{pmatrix} \mathbf{x}_k \\ (1 + \frac{1}{\alpha})\, \mathbf{z}_k \end{pmatrix} \right\} \tag{19}$$

The iteration (18)-(19) continues until $\|\lambda^{(j)} - \lambda^{(j-1)}\| + \|\mu^{(j)} - \mu^{(j-1)}\|$ falls below a pre-specified tolerance. In fact, the numerical properties of the iteration are very good as long as the totals $t_{\mathbf{x}}^*$ and $t_{\mathbf{z}}^*$ are close enough to their respective frame totals to produce moderate Lagrange multipliers (i.e., values λ, μ such that $Q'(1 + (1 + \alpha^{-1})\mu' \mathbf{z}_k + \lambda' \mathbf{x}_k)$ remain bounded). Theory supporting this assertion is contained in Appendix F, and numerics in Section 5.

4 Asymptotic Theory for Solutions

Following the approach of Deville and Särndal (1992), we sketch the theory of large-sample behavior of the solutions of the weight-equations just developed. Assume that Q is a piecewise smooth and convex function on an interval $(0, L)$, $L \leq \infty$, with $Q(u) \geq 0$ and $Q(u) \equiv 0$ on an interval (c_1, c_2) containing 1, and that its derivative Q' ranges from $-\infty$ to $+\infty$ on $(0, L)$. Assume in addition the regularity conditions **(A.0)-(A.4)** of Appendix A.

When the functions G_1, G_2, and Q are all convex, so is the objective function to be minimized subject to a linear constraint, and there will be a unique set of minimizing weights $\mathbf{w}$, $\hat{\mathbf{w}}$ equating the gradient of the objective function to $\mathbf{0}$. For this reason, the unique limit for any convergent subsequence along the iterative scheme (18)-(19) yields a unique solution to (11)-(12) along with (7) under the regularity conditions of Appendix A.

The left-hand side of equation (14′) is proved in Appendix B to be a nonsingular function of (λ, μ), with Jacobian bounded above and below by positive-definite matrices, in the positive-definite matrix ordering. The proof given by Deville and Särndal (1992) to establish large-sample (design) consistency of the calibrated weights, works in almost the same way in the present setting, as shown in Appendix C. However, the reasoning here reflects the possible misspecification of the constrained totals $t^*_{\mathbf{x}}$, $t^*_{\mathbf{z}}$ with respect to the initial survey weights $w^o_k = 1/\pi_k$, as expressed in the possibly non-zero limits in **(A.4)**. The solutions $(\hat{\lambda}, \hat{\mu})$ have finite in-probability limits, as do the calibrated weights w_k, $\hat{w}_k$, the latter being given by (13′) with $(\hat{\lambda}, \hat{\mu})$ substituted for (λ, μ).

The parameters (λ_*, μ_*) consistently estimated by $(\hat{\lambda}, \hat{\mu})$ in large samples are

$$\begin{pmatrix} \lambda_* \\ \mu_* \end{pmatrix} = \phi^{-1}\Big(\lim_N N^{-1} \sum_{k \in \mathcal{U}} (1 - \rho_k) \begin{pmatrix} \mathbf{x}_k \\ \mathbf{z}_k \end{pmatrix} \Big) \tag{20}$$

where, as proved in Appendix C, the invertible mapping ϕ on $\mathbf{R}^{p+q}$ is the limit of the mappings $\phi_{s,N}(\lambda, \mu)$ defined in Appendix B and C as the left-hand side of (14′) :

$$\phi(\lambda, \mu) \equiv \lim_N N^{-1} \sum_{k \in \mathcal{U}} \rho_k \Big\{ \begin{pmatrix} \mathbf{x}_k^{\otimes 2} & \mathbf{x}_k \, \mathbf{z}_k' \\ \mathbf{z}_k \, \mathbf{x}_k' & (1 + \alpha^{-1}) \, \mathbf{z}_k^{\otimes 2} \end{pmatrix} \begin{pmatrix} \lambda \\ \mu \end{pmatrix}$$

$$- Q' \circ h(1 + \frac{1+\alpha}{\alpha} \mu' \mathbf{z}_k + \lambda' \mathbf{x}_k) \begin{pmatrix} \mathbf{x}_k \\ (1 + \alpha^{-1}) \mathbf{z}_k \end{pmatrix} \Big\}$$

The estimator and limit definitions in (14′) and (20) imply

$$\sqrt{n} \left(\phi_{s,N}(\hat{\lambda}, \hat{\mu}) - \phi(\lambda_*, \mu_*) \right) = \tag{21}$$

$$\sqrt{n} \Big\{ \frac{1}{N} \begin{pmatrix} t^*_{\mathbf{x}} \\ t^*_{\mathbf{z}} \end{pmatrix} - \frac{1}{N} \sum_{k \in \mathcal{S}} r_k \, w^o_k \begin{pmatrix} \mathbf{x}_k \\ \mathbf{z}_k \end{pmatrix} - \lim_N \frac{1}{N} \sum_{k \in \mathcal{U}} (1 - \rho_k) \begin{pmatrix} \mathbf{x}_k \\ \mathbf{z}_k \end{pmatrix} \Big\}$$

which by **(A.4)** is equal to

$$\begin{pmatrix} k_{\mathbf{x}} \\ k_{\mathbf{z}} \end{pmatrix} - \frac{\sqrt{n}}{N} \sum_{k \in \mathcal{U}} (r_k \, w^o_k \, I_{[k \in \mathcal{S}]} - \rho_k) \begin{pmatrix} \mathbf{x}_k \\ \mathbf{z}_k \end{pmatrix} + o_P(1) \equiv \begin{pmatrix} k_{\mathbf{x}} \\ k_{\mathbf{z}} \end{pmatrix} + A_1 \tag{22}$$

The Mean Value Theorem implies that for $(\tilde{\lambda}, \tilde{\mu})$ lying on the ray between (λ_*, μ_*) and $(\hat{\lambda}, \hat{\mu})$, (depending on the sample and on N), expression (21) is equal to

$$D\phi_{s,N}(\tilde{\lambda}, \tilde{\mu}) \sqrt{n} \begin{pmatrix} \hat{\lambda} - \lambda_* \\ \hat{\mu} - \mu_* \end{pmatrix} + \sqrt{n} \left(\phi_{s,N}(\lambda_*, \mu_*) - \phi(\lambda_*, \mu_*) \right) \tag{23}$$

In the first of these two terms, the uniform convergence of $D\phi_{s,N}$ to the uniformly continuous function $D\phi$ as established in Appendix C implies that $D\phi_{s,N}(\tilde{\lambda},\, \tilde{\mu}) = D\phi(\lambda_*,\, \mu_*) + o_P(1)$; while the second term differs by $o_P(1)$ from the asymptotically normal variable

$$A_2 \;\equiv\; \frac{\sqrt{n}}{N} \sum_{k \in \mathcal{U}} (r_k\, w_k^o\, I_{[k \in \mathcal{S}]} \;-\; \rho_k) \Big[\begin{pmatrix} \mathbf{x}_k^{\otimes 2} & \mathbf{x}_k\, \mathbf{z}_k' \\ \mathbf{z}_k\, \mathbf{x}_k' & (1+\alpha^{-1})\, \mathbf{z}_k^{\otimes 2} \end{pmatrix} \begin{pmatrix} \lambda_* \\ \mu_* \end{pmatrix}$$

$$-\; Q' \circ h(1 + \frac{1+\alpha}{\alpha}\, \mu_*' \mathbf{z}_k \;+\; \lambda_*' \mathbf{x}_k) \begin{pmatrix} \mathbf{x}_k \\ (1+\alpha^{-1}) \mathbf{z}_k \end{pmatrix} \Big]$$

Assembling terms (21), (22), and (23), and using Lemma 1 in Appendix B to re-express $D\phi(\lambda_*, \mu_*)$, we have proved that

$$\sqrt{n} \begin{pmatrix} \hat{\lambda} - \lambda_* \\ \hat{\mu} - \mu_* \end{pmatrix} \;=\; D_*^{-1} \Big(A_1 \;-\; A_2 \;+\; \begin{pmatrix} k_{\mathbf{x}} \\ k_{\mathbf{z}} \end{pmatrix} \Big) \;+\; o_P(1) \tag{24}$$

where (by Lemma 1 in Appendix B)

$$D_* \;\equiv\; D\phi(\lambda_*, \mu_*) \;=\; \lim_N \frac{1}{N} \sum_{k \in \mathcal{U}} \rho_k \Big[\frac{1}{1+\alpha} \begin{pmatrix} \mathbf{x}_k \\ \mathbf{0} \end{pmatrix}^{\otimes 2} +$$

$$+\; \frac{\alpha}{1+\alpha}\, h'(1 + \;\lambda_*' \mathbf{x}_k + \frac{1+\alpha}{\alpha}\, \mu_*' \mathbf{z}_k) \begin{pmatrix} \mathbf{x}_k \\ (1+\alpha^{-1}) \mathbf{z}_k \end{pmatrix}^{\otimes 2} \Big]$$

is a nonsingular matrix. After some algebraic simplifications resembling the steps in Appendix B, it turns out that

$$\begin{aligned} A_2 - A_1 \;\;&=\;\; \frac{\sqrt{n}}{N} \sum_{k \in \mathcal{U}} (r_k\, w_k^o\, I_{[k \in \mathcal{S}]} \;-\rho_k) \Big[\frac{1}{1+\alpha}\, (1 + \lambda_*' \mathbf{x}_k) \begin{pmatrix} \mathbf{x}_k \\ \mathbf{0} \end{pmatrix} \\ &+\;\; \frac{\alpha}{1+\alpha}\, h(1 + \frac{1+\alpha}{\alpha}\, \mu_*' \mathbf{z}_k + \lambda_*' \mathbf{x}_k) \begin{pmatrix} \mathbf{x}_k \\ (1+\alpha^{-1}) \mathbf{z}_k \end{pmatrix} \Big] \end{aligned} \tag{25}$$

an asymptotically normally distributed random $(p+q)$-vector with mean $\mathbf{0}$.

4.1 Variance of Survey Estimators

The purpose of these derivations is to find and develop estimators for the survey-weighted totals (5), when the weights $\hat{w}_k$ incorporate the estimated Lagrange multipliers $\hat{\lambda}, \hat{\mu}$ through (13′). Note that the Assumptions **(A.0)-(A.4)** of Appendix A do not require the response indicators $\{r_k\}_{k \in \mathcal{S}}$ to be *noninformative* in the sense of having expectations ρ_k without functional pattern or correlation with respect to the attributes $\{y_k\}_{k \in \mathcal{S}}$.

By (24) and (13′),

$$\hat{w}_k \;=\; r_k\, w_k^o\, h(1 \;+\; \frac{1+\alpha}{\alpha}\, \hat{\mu}' \mathbf{z}_k \;+\; \hat{\lambda}' \mathbf{x}_k) \tag{26}$$

These expressions are linearized by Taylor-expansion of h around the argument obtained by replacing $(\hat{\lambda}, \hat{\mu})$ by (λ_*, μ_*). To this end, define for all $k \in \mathcal{U}$,

$$u_k^* \equiv 1 + \frac{1+\alpha}{\alpha} \mu_*' \mathbf{z}_k + \lambda_*' \mathbf{x}_k \quad , \qquad f_k^* \equiv h(u_k^*) \tag{27}$$

where recall from Appendix F that $h'(u) = 1/(1 + (1+\alpha^{-1})Q''(h(u))) < 1$. Then (26) implies

$$\sqrt{n}\,(\hat{w}_k - r_k\, w_k^o\, f_k^*) = r_k\, w_k^o\, h'(u_k^*)\,\sqrt{n} \begin{pmatrix} \hat{\lambda} - \lambda_* \\ \hat{\mu} - \mu_* \end{pmatrix}' \begin{pmatrix} \mathbf{x}_k \\ (1+\alpha^{-1})\mathbf{z}_k \end{pmatrix} + o_P(1)$$

and the discussion of Section 4 implies that this o_P term, like the others appearing below, is uniform in $k \in \mathcal{S}$. Substituting from (24) in the last expression yields

$$r_k\, w_k^o \left(A_1 - A_2 + \begin{pmatrix} k_{\mathbf{x}} \\ k_{\mathbf{z}} \end{pmatrix} \right)' D_*^{-1} \begin{pmatrix} \mathbf{x}_k \\ (1+\alpha^{-1})\mathbf{z}_k \end{pmatrix} + o_P(1)$$

Therefore,

$$\frac{\sqrt{n}}{N} \Big(\sum_{k\in\mathcal{S}} \hat{w}_k\, y_k - \sum_{k\in\mathcal{U}} \rho_k\, f_k^*\, y_k \Big) = \frac{\sqrt{n}}{N} \sum_{k\in\mathcal{U}} (r_k\, w_k^0\, I_{[k\in\mathcal{S}]} - \rho_k)\, f_k^*\, y_k +$$

$$+ \left(A_1 - A_2 + \begin{pmatrix} k_{\mathbf{x}} \\ k_{\mathbf{z}} \end{pmatrix} \right)' (N\, D_*)^{-1} \sum_{k\in\mathcal{S}} r_k\, w_k^o \begin{pmatrix} \mathbf{x}_k \\ (1+\alpha^{-1})\mathbf{z}_k \end{pmatrix} y_k + o_P(1)$$

As a result, we deduce via (25)

Proposition 1 *Under the Assumptions* **(A.0)-(A.4)**, *with notations (27),*

$$\frac{\sqrt{n}}{N} \Big[\sum_{k\in\mathcal{S}} \hat{w}_k\, y_k - \sum_{k\in\mathcal{U}} \rho_k\, f_k^*\, y_k \Big] = \begin{pmatrix} k_{\mathbf{x}} \\ k_{\mathbf{z}} \end{pmatrix}' D_*^{-1}\, b_* + \frac{\sqrt{n}}{N} \sum_{k\in\mathcal{U}} (r_k\, w_k^0\, I_{[k\in\mathcal{S}]} - \rho_k) \cdot$$

$$\cdot \Big[f_k^*\, y_k - b_*' D_*^{-1} \Big(\frac{1+\lambda_*'\mathbf{x}_k}{1+\alpha} \begin{pmatrix} \mathbf{x}_k \\ \mathbf{0} \end{pmatrix} + \frac{\alpha\, f_k^*}{1+\alpha} \begin{pmatrix} \mathbf{x}_k \\ (1+\alpha^{-1})\mathbf{z}_k \end{pmatrix} \Big) \Big] + o_P(1) \tag{28}$$

where D_* *is defined below (24) and* b_* *is defined by*

$$b_* = \lim_N N^{-1} \sum_{k\in\mathcal{U}} \rho_k\, y_k \begin{pmatrix} \mathbf{x}_k \\ (1+\alpha^{-1})\,\mathbf{z}_k \end{pmatrix} h'(u_k^*)$$

4.2 Consistency of Weighted Survey Totals

The consistency of survey-weighted totals in the present nonresponse and calibration framework must be justified by one or both of two essentially model-based assumptions, saying either that the weight terms $\rho_k\, f_k^*$ in the centering constant of Prop. 1 are almost uniformly close to 1 or that their differences from 1 are asymptotically orthogonal to the vector(s) $\{y_k\}_{k \in \mathcal{U}}$ of survey attributes in the frame population. This twofold path to consistency is a known instance of the concept of *double robustness* (Kang and Schafer 2007). The underlying assumptions and result are specified in the following Proposition.

Proposition 2 *Assume* **(A.0)-(A.4)***, and in addition one of the following:*

(i). There is a subset $\mathcal{U}_1 \subset \mathcal{U}$ *and* $\lambda \in \mathbf{R}^p$ *such that for all* $k \in \mathcal{U}_1$, $\rho_k \equiv (1 + \lambda' \mathbf{x}_k)^{-1}$*, and the interval* (c_1, c_2) *on which* $Q \equiv 0$ *contains* $[\min_{k \in \mathcal{U}_1} (1 + \lambda' \mathbf{x}_k), \max_{k \in \mathcal{U}_1} (1 + \lambda' \mathbf{x}_k)]$*, and also*

$$\sum_{k \in \mathcal{U} \setminus \mathcal{U}_1} (|y_k| + \|\mathbf{z}_k\|^2 + \|\mathbf{x}_k\|^2) = o(N/\sqrt{n})$$

(ii). Assume that $Q \equiv 0$ *and for some* $\beta \in \mathbf{R}^q$*,*

$$\lim_{N \to \infty} (\sqrt{n}/N) \sum_{k \in \mathcal{U}} \rho_k \, \mathbf{z}_k \, (y_k - \beta' \mathbf{z}_k) = 0$$

Then the left-hand side of the equality in Proposition 1 is equal to

$$(\sqrt{n}/N) \, (\sum_{k \in \mathcal{S}} \hat{w}_k \, y_k - t_y) + o_P(1)$$

Assumption (i) says that the 'working' quasi-randomization model is correct, while (ii) is a slight generalization (as long as n is much smaller than N) of the requirement that for some $\beta \in \mathbf{R}^q$, the vector of weighted residuals $\rho_k \, (y_k - \beta' \mathbf{z}_k)$ is asymptotically orthogonal to the columns of entries of the predictor variables $\mathbf{z}_k$. The second of these assumptions is the technical sense in which response should be noninformative for y_k. Other authors, such as Fuller (2002), prove consistency in superpopulation models by assuming that the residuals are independent mean-0 errors uncorrelated with the calibration variables $\mathbf{z}_k$. See Appendix G for a proof of Proposition 2, and for discussion of a generalization of Prop. 2.(i) in which the calibration terms involving $\mu_*' \mathbf{z}_k$ could possibly correct for a frame deficiency.

4.3 Consequences of Propositions 1 & 2

The two Propositions of the previous subsection have several important statistical implications. The first, already mentioned in connection with the assumptions (i), (ii) of Proposition 2, is that the consistency of the survey-weighted total estimators depends strongly on unverifiable model assumptions about the 'stochastic' mechanisms of unit nonresponse and omission from survey frames. A second implication is that even relatively slight (order of $1/\sqrt{n}$) discrepancies between the assumed calibration totals $t_{\mathbf{x}}^*$, $t_{\mathbf{z}}^*$ and the respective totals $t_{\mathbf{x}}$, $t_{\mathbf{z}}$ they are intended to approximate, *unavoidably* and *irreparably* have the effect in Proposition 1 (the first term on the right-hand side of the equality) of biasing the centering of confidence intervals based on survey estimates (5). Finally, the clear positive consequence of the new single-stage approach is a readily computable estimator for the linearized variance formula for $\hat{t}_{y,adj}$ under each of a number of possible origins for the totals $t^*_{\mathbf{x}}$.

There are three distinct scenarios under which totals $t^*_{\mathbf{x}}$ are known: first, $t^*_{\mathbf{x}}$ might be nonrandom (e.g., from an updated census) even if incorrect, subject to (A.4); second, these totals could have been generated as estimates from a previous or parallel survey with identical frame and comparable size; but the third and most common way to obtain these totals is by estimating them as Horvitz-Thompson totals from the current survey, as in (3), *using information available on sampled units whether or not they respond*. In both the second and third cases, the term $t^*_{\mathbf{x}}$ has its own $O(N/\sqrt{n})$ standard error. In all three scenarios, $t^*_{\mathbf{z}}$ is still viewed as a nonrandom constant.

First case. In terms of the natural design-consistent estimators $\hat{f}_k$, $\hat{b}$, and $\hat{D}$ for the respective quantities f^*_k, b_*, D_*, and of the (ratio-adjustment or calibration based) estimator $\hat{\rho}_k$ for ρ_k, the variance estimator $\hat{V}(\sqrt{n}\,\hat{t}_{y,adj}/N)$ in this case is

$$\frac{n}{N^2}\Big[\sum_{k,l\in\mathcal{S}}(w^o_k\,w^o_l\;-\;\frac{1}{\pi_{kl}})\,\hat{R}_k\,\hat{R}_l\,r_k\,r_l\;+\;\sum_{k\in\mathcal{S}}r_k\,w^o_k\,\hat{R}^2_k\,(1-\hat{\rho}_k)\Big] \tag{29}$$

where

$$\hat{R}_k\;\equiv\;\hat{f}_k\,y_k\;-\;\hat{b}'\,\hat{D}^{-1}\Big(\frac{1+\hat{\lambda}'\mathbf{x}_k}{1+\alpha}\begin{pmatrix}\mathbf{x}_k\\ \mathbf{0}\end{pmatrix}\;+\;\frac{\hat{f}_k\,\alpha}{1+\alpha}\begin{pmatrix}\mathbf{x}_k\\ (1+\alpha^{-1})\mathbf{z}_k\end{pmatrix}\Big)$$

The formula (29) for variance follows immediately by calculating

$$\mathrm{Var}(\cdot)\;=\;E(\,\mathrm{Var}(\,\cdot\,|\,\{r_j\}_{\mathcal{U}})\,)\;+\;\mathrm{Var}(E(\,\cdot\,|\,\{r_j\}_{\mathcal{U}})\,)$$

applied to the sum statistic $\sum_{k\in\mathcal{U}}(r_k w^o_k I_{[k\in\mathcal{S}]}\;-\;\rho_k)\,R_k$, where R_k is the final square-bracketed summand on the right-hand side of (28) in Prop. 1.

Second case. Here the totals $t^*_{\mathbf{x}}$ are themselves survey-based estimators, for which we must assume that we have a design-consistent survey estimator $\hat{V}_{\mathbf{x}}$ of $\mathrm{Var}(\sqrt{n}\,t^*_{\mathbf{x}}/N)$ which is approximately equal in distribution to a random variable (usually normally distributed) not depending on n or N. In that case, $k_{\mathbf{x}}$ should be regarded as a random variable independent of the current survey, with variance-covariance matrix approximately $\hat{V}_{\mathbf{x}}$. Then a design-consistent estimator for $\mathrm{Var}(\sqrt{n}\,\hat{t}_{y,adj}/N)$ is given by formula (29) plus the additional term $(\hat{D}^{-1}\,\hat{b})'_x\,\hat{V}_{\mathbf{x}}\,(\hat{D}^{-1}\,\hat{b})_x$, where $(\hat{D}^{-1}\,\hat{b})_x$ denotes the sub-vector of the first p components of the $(p+q)$-vector $\hat{D}^{-1}\,\hat{b}$.

Third case. Finally, we have a setting where $t^*_{\mathbf{x}}$ is given by (3), and for definiteness, we assume the second equality in (3). Moreover, in order that the top-order behavior of $k_{\mathbf{x}}$ be entirely due to sampling rather than bias, we assume that

$$\sum_{k\in\mathcal{U}}\rho_k\,\mathbf{x}_k\Big/\sum_{k\in\mathcal{U}}\rho_k\;-\;\bar{\mathbf{x}}_{\mathcal{U}}\;=\;o(1/\sqrt{n})$$

where $\bar{\mathbf{x}}_{\mathcal{U}} \equiv t_{\mathbf{x}}/N$. Then by Proposition 1,

$$k_{\mathbf{x}} = (\sqrt{n}/N)(t^*_{\mathbf{x}} - t_{\mathbf{x}}) = (\sqrt{n}/\sum_{\mathcal{U}} \rho_k) \sum_{k \in \mathcal{U}} (r_k w_k^o I_{[k \in \mathcal{S}]} - \rho_k)(x_k - \bar{x}_{\mathcal{U}}) + o_P(1)$$

and $\mathrm{Var}(\sqrt{n}\, \hat{t}_{y,adj}/N)$ is approximately the same as the variance of

$$\frac{\sqrt{n}}{N} \sum_{k \in \mathcal{U}} \Big\{ (r_k\, w_k^0\, I_{[k \in \mathcal{S}]} - \rho_k) \Big(R_k + \frac{N}{\sum_{k \in \mathcal{U}} \rho_k}\, b_*' D_*^{-1} \begin{pmatrix} \mathbf{x}_k - \bar{\mathbf{x}}_{\mathcal{U}} \\ \mathbf{0} \end{pmatrix} \Big) \Big\}$$

A variance formula is derived as in in the first case, following (29), and the design-consistent asymptotic variance estimation formula extending (29) takes exactly the same form as in (29) with $\hat{R}_k$ replaced by $\hat{T}_k$, where

$$\hat{T}_k = \hat{R}_k + \frac{N}{\sum_{j \in \mathcal{S}} r_j\, w_j^o}\, \hat{b}'\, \hat{D}^{-1} \begin{pmatrix} \mathbf{x}_k - \bar{x}_{\mathcal{S}} \\ \mathbf{0} \end{pmatrix} \quad , \qquad \bar{x}_{\mathcal{S}} \equiv \frac{\sum_{j \in \mathcal{S}} r_j\, w_j^o\, \mathbf{x}_j}{\sum_{j \in \mathcal{S}} r_j\, w_j^o}$$

5 Implementation on Simulated and Real Data

In this Section, the performance of the new single-stage weight adjustment algorithms is illustrated for $G_1(z) \equiv G_2(z) \equiv z^2/2$, first for a simulated sampled-superpopulation dataset, and then for the Survey of Income and Program Participation (SIPP) 1996 Panel Wave 1 data. The objective is to compare properties of adjusted weights and estimates for different choices of α and parameters used to specify Q. In both computed examples, the weight-penalty functions Q were defined on the interval $(0, 10)$, identically 0 on an interior interval (c_1, c_2) containing 1, and elsewhere

$$Q(u) = a_1 \frac{(c_1 - u)_+^{b_1}}{u^{\kappa_1}} + a_2 \frac{(u - c_2)_+^{b_2}}{(10 - u)^{\kappa_2}} \tag{30}$$

where $(x)_+ \equiv \max(x, 0)$. Computations in this Section were done in **R**.

First, we simulated a superpopulation sample ($n = 1000$, $N = 500,000$) from the following design. Four independent nonconstant latent variables $V_{i,j}$, $1 \le j \le 4$, were simulated (two $Binom(1, \frac{1}{2})$ and two standard normal) independently, and *iid* across 'sampled' individuals i. The weights for the sample were taken from a stratified design, with 5 strata of size 200 each corresponding to true frame stratum sizes of 85000, 125000, 100000, 90000, and 100000. The response-indicators $\{r_i\}_{i=1}^{1000}$ were simulated independently given $\mathbf{V} = \{V_{ij}\}$ from the logistic regression model

$$P(r_i = 1 \,|\, \mathbf{V}) = (1 + \exp(-1.50 - 0.24\, V_{i,1} + 0.36\, V_{i,2} - 0.18\, V_{i,3} - 0.27\, V_{i,4}))^{-1}$$

The calibration variables for nonresponse adjustment were taken to be $p = 12$ dummy variables $x_{i,k}$, $1 \le k \le 12$, constructed by the $2 \times 2 \times 3$ levels of $(V_{i,1}, V_{i,2}, I_{[V_{i,3} > -1]} + I_{[V_{i,3} > 1]})$. The calibration vectors $\mathbf{z}_i$ were taken to be of dimension 5, consisting of 4 independent $Expon(1)$ variables (independent of $\mathbf{V}$) together with the variable $V_{i,4}$. Finally, the control totals $t^*_{\mathbf{x}}$ were taken to be exactly the values $500 \sum_{i=1}^{1000} \mathbf{x}_i$ which would have been obtained as stratified sample estimates with complete observations of $\mathbf{x}_i$, while the totals $t^*_{\mathbf{z}}/10^5$ which should have been percentagewise close to $(5, 5, 5, 5, 0)$ were fixed at $(5, 5, 5.3, 4.7, 0)$. Overall, this example conforms well to the assumptions **(A.0)-(A.4)** except possibly for **(A.4)**, since the third and fourth components of $\sqrt{n}\,(t^*_{\mathbf{z}} - t_{\mathbf{z}})/N$ have magnitude $0.3\sqrt{1000} = 9.5$, while the components of $\sqrt{n}\,(t^*_{\mathbf{x}} - t_{\mathbf{x}})/N$ range from -4 to 4.

In this Example, with $\alpha = 1$ (and with $c_1 = .6$, $c_2 = 1.6$, $a_1 = a_2 = 10$, $b_1 = b_2 = 4$, $\kappa_1 = 4$, $\kappa_2 = 2$ in (30)), the iterative algorithm (18)-(19) converged quickly because the fitted weight-factors $1+2\hat{\mu}'\mathbf{z}_i+\hat{\lambda}'\mathbf{x}_i$ ranged only on the interval $(.65, 1.86)$, barely larger than the interval on which $Q \equiv 0$. However, the new single-stage calibrated weights for the responding sample units are correlated only 0.558 with the two-stage weights obtained by first performing ratio adjustment on the $\mathbf{x}_i$ defined cells and then calibrating these weights to the $t^*_{\mathbf{z}}$ totals. Both sets of calibrated weights solve the $t^*_{\mathbf{z}}$ constraints (4) almost exactly, but not (2): the relative discrepancies between the left- and right-hand sides of (2 are quite similar for the ($\alpha = 1$) single- and two-stage calibrated weights. However, a scatterplot and a 99% correlation shows that the $\alpha = 100$ single-stage calibrated weights are nearly identical to those with $\alpha = 1$, yet the former solve (2) with relative errors less than .001.

Further numerical experiments, with other simulated superpopulation samples and varying degrees of conformity to **(A.4)**, indicate that the new single-stage calibrated weights with large α (e.g., $\alpha = 100$) and Q functions rising sharply outside $(.6, 2.5)$, successfully compress the modified weights into ranges usually no wider than $(.4, 4)$ while accurately solving both the constraints (8).

As a second example, single-stage calibrated weights were fitted to the 1996 Wave 1 data from the SIPP survey, a large stratified complex longitudinal survey conducted by the U.S. Census Bureau. As described in Slud and Bailey (2010), SIPP nonresponse adjustment was based on poststratified ratio-adjustment using a system of $p = 149$ cells (involving demographics and some indicators of assets and of income compared with the poverty level), and SIPP population controls involved raking to census-based totals of race by family structure, race by age interval, and of Hispanic-origin persons by (coarser) age groups. The SIPP population-control constraints can be expressed in $q = 126$ linearly independent equations of the type (4). The weights w^o_k used in the SIPP file of 94444 individuals sampled and responding in 1996 Wave 1 are the base weights `GBASEWT` before nonresponse adjustment. The $t^*_{\mathbf{x}}$ constrained totals in (2) are the values $N \sum_{\mathcal{S}} \mathbf{x}_k \, w^o_k / \sum_{\mathcal{S}} w^o_k$ estimated with the base weights, and the population N and control totals $t^*_{\mathbf{z}}$ were, as in SIPP production estimates, derived from the 1990 census demographically updated to 1996.

Table 1: Estimated Totals and Std. Dev.'s in Thousands. 'Two-Stage' method and EHG standard deviation described in text; totals and standard deviations for $\hat{t}_{y,adj}$ based on (9) with $\alpha = 1$, 100, with variables $\mathbf{x}_k$, $\mathbf{z}_k$ and function Q described in text.

Item	2stage	$\hat{t}_{\alpha=1}$	$\hat{t}_{\alpha=100}$	EHG.sd	SD$_{\alpha=1}$	SD$_{\alpha=100}$
FOODST	27541	27454	26930	687.0	317.7	300.8
AFDC	14123	14089	13800	450.5	298.3	287.6
MDCD	28468	28399	27895	573.8	403.7	351.1
SOCSEC	36994	37071	37240	469.6	156.9	156.8
HEINS	194216	194475	195030	1625.1	438.8	422.6
POV	41951	41978	41475	747.3	360.1	357.1
EMP	190871	190733	190097	1477.3	254.8	239.8
UNEMP	6403	6379	6295	163.1	144.5	143.2
NILF	66979	67354	67864	626.7	231.4	216.7
MAR	111440	114457	114347	1088.1	159.1	157.7
DIV	18534	18542	18591	253.4	194.9	194.9

In the SIPP example, the single-stage calibrated weights were fitted with $\alpha = 1$ and 100, using Q as above except for the changes $c_1 = .6$, $c_2 = 3$, $a_1 = a_2 = .2$, $b_1 = b_2 = 3$ (but $c_1 = .8$, $c_2 = 2.5$ and $a_1 = a_2 = .5$ when $\alpha = 100$). The fitted weight ratios $\hat{w}_k/w_k^o$ fell in the range $(.55, 3.86)$ for $\alpha = 1$, and in $(.62, 4.08)$ for $\alpha = 100$. The 126 fitted Lagrange multiplier components of $\hat{\mu}$ had range $(-1.13, 0.72)$ for $\alpha = 1$, and $(-0.86, 1.77)$ for $\alpha = 100$. Both sets of single-stage modified weights satisfied (4) accurately; the modified weights with $\alpha = 1$ already satisfied (2) to within several percent, while those with $\alpha = 100$ satisfied (2) to within a few tenths of a percent. In addition, we experimented with values of α as small as 0.1. However, the weights were much more difficult to make converge with small α, and the estimated totals and standard deviations were very similar to and did not offer any improvement over those with $\alpha = 1$, so we do not present results for them here.

As an indication of the similarity of estimated results $\hat{t}_{y,adj}$ with the $\alpha = 100$ estimated weights from (9), with $\alpha = 1$ or 100, to those obtained in SIPP 1996 by what amounted to two-stage weight adjustment (with raking rather than linear calibration in the second population-control stage), we give in Table 1 the estimated totals (in thousands) for eleven of the important SIPP surveyed attributes. The labelling for the survey attributes is as in Slud and Bailey (2010) where further details can be found. Differences among the estimators based on the three sets of weights are quite small compared with the standard deviations of the two-stage estimated totals, given in the Table as the EHG (Ernst, Huggins and Grill 1986) estimated standard deviation, a slightly upwardly biased estimator of standard deviation which the VPLX Fay-method standard deviation estimator approximates).

Table 2: Estimated standard deviations in thousands for several different calibrated and nonresponse adjusted estimators in SIPP 96. EHG or VPLX standard deviation as in Table 1; SD_α^{fix} indicates standard deviation with $t^*_{\mathbf{x}}$ regarded as fixed, as in (29); SD_α^{est} indicates standard deviation with $t^*_{\mathbf{x}}$ regarded as estimated in (3) and Case 3.

Item	EHG.sd	SD_1^{fix}	SD_{100}^{fix}	SD_1^{est}	SD_{100}^{est}
FOODST	687.0	317.7	300.8	429.3	427.6
AFDC	450.5	298.3	287.6	317.7	308.3
MDCD	573.8	403.7	351.1	496.3	468.8
SOCSEC	469.6	156.9	156.8	169.5	176.2
HEINS	1625.1	438.8	422.6	580.7	586.2
POV	747.3	360.1	357.1	462.9	474.2
EMP	1477.3	254.8	239.8	343.5	390.0
UNEMP	163.1	144.5	143.2	146.0	145.0
NILF	626.7	231.4	216.7	320.0	369.1
MAR	1088.1	159.1	157.7	160.0	158.6
DIV	253.4	194.9	194.9	197.7	200.4

The VPLX variances calculated for Table 1 account for nonresponse adjustment only by scaling up the `GBASEWT` base weights so that their total is the known population size N. The Table also displays variances as calculated from formula (29), using $\alpha = 1$ or 100 and Q as specified above. Strikingly ,these estimated variances are much smaller than the VPLX variances, highlighting the fact that the variance formula (29) treats all calibration totals $t^*_{\mathbf{x}}$, $t^*_{\mathbf{z}}$ as though they were known from the outset and nonrandom. In fact, the values $t^*_{\mathbf{x}}$ were, unavoidably for SIPP 1996, in fact derived from sample estimates from the same SIPP 1996 survey as in the rightmost equality in (3). The last two columns of Table 2 give the standard deviations for the eleven survey attributes as estimated in Case 3, respectively with $\alpha = 1$ and $\alpha = 100$. The variances obtained as in Case 3 are notably larger than those with totals fixed as in Case 1, only for those survey items (`FOODST, MDCD, HEINS, POV`) which relate to programs applicable to entire households.

6 Discussion and Future Research

This paper has developed asymptotic theory and computational tools for a new, single-stage approach to the adjustment of weights in large complex surveys for nonresponse, population-controls, and weight-trimming. Numerical experiments using functions created in the **R** statistical programming language (2008) show that the single-stage modified weights are readily computable iteratively, and are generally similar, although not always extremely highly correlated among responders, with the two-stage weights obtained first by nonresponse-adjustment and then by population controls and trimming. The single-stage calibrated weights with large α (of order 100) penalized via function Q to lie in intervals such as $(.4, 3)$ seem particularly satisfactory.

Linearized variance estimators for survey estimates based on the new weights have been developed here, both for the ideal case where nonresponse adjustments are based on externally derived adjustment totals $t^*_{\mathbf{x}}$ and for the more usual case where these totals are instead based on estimates derived from the same survey to which the new weights will be applied. These latter types of weights are roughly similar across a broad range of α smoothing-parameter values, at least in the SIPP 1996 data example studied here. The comparison begun here between bias and variance properties of estimators based on the new weights will be elaborated in future research, using data from other surveys and external checks on those surveys' estimates. In addition, future research will extend the theory and algorithms to the case of nonlinear $G'_j(z) = z \log z - z + 1$, the so-called 'multiplicative' form of the adjustment-discrepancy loss-function which leads to classical raking in the absence of weight-penalization.

7 References

Beaumont, J.-F. (2008), A new approach to weighting and inference in sample surveys. *Biometrika* **95**, 539-553.

Chang, T. and Kott, P. (2008), Using calibration weighting to adjust for nonresponse under a plausible model. *Biometrika* **95**, 555-571.

Deville, J.-C. and Särndal, C.-E. (1992), Calibration estimators in survey sampling, *Jour. Amer. Statist. Assoc.* **87**, 376-382.

Deville, J.-C., Särndal, C.-E. and Sautory, O. (1993), Generalized raking procedures in survey sampling. *Jour. Amer. Statist. Assoc.* **88**, 1013-1020.

Elliott, M. and Little, R. (2000), Model-based alternatives to trimming survey weights. *Jour. Official Statist.* **16**, 191-209.

Ernst, L., Huggins, V., and Grill, D. (1986), Two new variance estimation techniques. *ASA Proc. Survey Res. Meth.*, Washington DC, pp. 400-405.

Fuller, W. (2002), Regression estimation for survey samples. *Survey Metho-dology* **28**, 5-23.

Kang, J. and Schafer, J. (2007), Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science* **22**, 523-539.

Kim, Jae-Kwang and Kim, Jay J. (2007), Nonresponse weighting adjustment using estimated response probability. *Canad. Jour. Statist.* **35**, 501-514.

Lu, H. and Gelman, A. (2003), A method for estimating design-based sampling variances for surveys with weighting, poststratification, and raking. *Jour. Official Statist.* **19**, 133-151.

Oh, H. and Scheuren, F. (1983) Weighting adjustment for unit nonresponse. In: *Incomplete Data in Sample Surveys*, vol. 2, Eds. Madow, W., Olkin, I. and Rubin, D. New York: Academic Press, 143-184.

R Development Core Team (2008), *R: A Language and Environment for Statistical Computing*, ISBN 3-900051-07-0, `http://www.R-project.org`.

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

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

Singh, A. and Mohl, C. (1996), Understanding calibration estimation in survey sampling. *Survey Methoology* **22**, 107-115.

Slud, E. and L. Bailey (2010), Evaluation and selection of models for attrition nonresponse adjustment, *Jour. Official Statist.*, **26** 127-143.

Théberge, A. (2000) Calibration and restricted weights. *Survey Methodology* **26**, 99-107.

Yung and Rao, J.N.K. (2000) Jackknife variance estimation under imputation for estimators using poststratification information. *Jour. Amer. Statist. Assoc.* **95**, 903-915.

A Superpopulation Regularity Conditions

Although less stringent assumptions may be sufficient for the validity of the asymptotic theory we present, we follow the pattern of assumptions of Deville and Särndal (1992, p. 379) concerning the superpopulation weights and variables, and these assumptions must be augmented in the present context by assumptions involving the response probabilities $E(r_i) = \rho_i$.

(A.0) The frame population $\mathcal{U}$ indexed by its population size N is one of a larger and larger system of ('super'-) populations, with $N \to \infty$, and for large N the design (joint) inclusion probabilities $\{\pi_i\}_{i \in \mathcal{U}}$, $\{\pi_{i,j}\}_{i,j \in \mathcal{U}}$ are such that the sample size $n = |\mathcal{S}|$ also gets large, with $n/N \le c < 1$ for some constant c not depending upon N. Recall that the initial weights w_i^o of the paper are assumed defined as $w_i^o = 1/\pi_i$.

(A.1) The response indicators $r_i,\ i \in \mathcal{U}$, are independent $\{0,1\}$-valued random variables with $P(r_i = 1) = \rho_i$ such that $\inf_{N,i} \rho_i > 0$.

(A.2) Let the attributes u_i be given by any linear combinations of the components of $\mathbf{x}_i, \mathbf{z}_i, \mathbf{z}_i^{\otimes 2}, \mathbf{x}_i^{\otimes 2}, y_i\,(1, \mathbf{x}_i', \mathbf{z}_i')$, or $f(1 + \mu_*'\mathbf{z}_i + \lambda_*'\mathbf{x}_i)$, where (λ, μ) is fixed but arbitrary and f is one of $h, h',$, or h'' and where h is defined in (17) and (λ_*, μ_*) in (20). Then for v_i defined either by u_i or $u_i\,\rho_i$, the following conditions hold:

(i) $\lim_N N^{-1}\, t_v \equiv \lim_N N^{-1} \sum_{i\in\mathcal{U}} v_i$ exists $= t_v/N + o(n^{-1/2})$, and

(ii) with $t_{v,\pi} \equiv \sum_{i\in\mathcal{S}} v_i/\pi_i$, the scaled and centered weighted survey totals $n^{1/2}\, N^{-1}\,(\hat{t}_{v,\pi} - t_v)$ converges in distribution as $N, n \to \infty$ to a normal random variable with mean 0 and finite variance σ_v^2.

In several places where we make use of auxiliary calibration variables $\mathbf{x}_i$ and $\mathbf{z}_i$, we must also assume one of the following two conditions:

(A.3) The limits of $N^{-1} \sum_{i\in\mathcal{U}} \mathbf{x}_i^{\otimes 2}$ and $N^{-1} \sum_{i\in\mathcal{U}} \mathbf{z}_i^{\otimes 2}$ are positive definite.

(A.3′) The limit of $N^{-1} \sum_{i\in\mathcal{U}} \binom{\mathbf{x}_i}{\mathbf{z}_i}^{\otimes 2}$ is positive definite.

Finally, to accomodate the possible incorrectness of the population total $t^*_{\mathbf{z}}$ to which we control our weighted total estimates of $\mathbf{z}_k$ attributes, we assume

(A.4) The population-total constraints $t^*_{\mathbf{z}},\ t^*_{\mathbf{x}}$ satisfy as $N \to \infty$:

$$\lim_N \frac{n^{1/2}}{N} \begin{pmatrix} t^*_{\mathbf{x}} - t_{\mathbf{x}} \\ t^*_{\mathbf{z}} - t_{\mathbf{z}} \end{pmatrix} = \begin{pmatrix} k_{\mathbf{x}} \\ k_{\mathbf{z}} \end{pmatrix} \qquad \text{exists} \;, \qquad k_{\mathbf{z}} \in \mathbf{R}^q \quad , \quad k_{\mathbf{x}} \in \mathbf{R}^p$$

B Nonsingular Dependence on (λ, μ) in (14′)

The left-hand side of (14′) is a function of (λ, μ) given by

$$\phi_{s,N}(\lambda,\mu) \;=\; \frac{1}{N} \sum_{k\in\mathcal{S}} r_k\, w_k^o \Big[\begin{pmatrix} \mathbf{x}_k^{\otimes 2} & \mathbf{x}_k\, \mathbf{z}_k' \\ \mathbf{z}_k \mathbf{x}_k' & (1+\alpha^{-1})\, \mathbf{z}_k^{\otimes 2} \end{pmatrix} \begin{pmatrix} \lambda \\ \mu \end{pmatrix}$$

$$- Q' \circ h(1 + \frac{1+\alpha}{\alpha}\, \mu'\mathbf{z}_k \,+\, \lambda'\mathbf{x}_k) \begin{pmatrix} \mathbf{x}_k \\ (1+\alpha^{-1})\mathbf{z}_k \end{pmatrix} \Big]$$

The bracketed summand in this expression is equal, by definition of h and the identity $h(u) + (1+\alpha^{-1})Q'(h(u)) \equiv u$, to

$$\Big[\begin{pmatrix} \mathbf{x}_k^{\otimes 2} & \mathbf{x}_k\, \mathbf{z}_k' \\ \mathbf{z}_k \mathbf{x}_k' & (1+\alpha^{-1})\, \mathbf{z}_k^{\otimes 2} \end{pmatrix} - \frac{\alpha}{1+\alpha} \begin{pmatrix} \mathbf{x}_k \\ (1+\alpha^{-1})\mathbf{z}_k \end{pmatrix}^{\otimes 2} \Big] \begin{pmatrix} \lambda \\ \mu \end{pmatrix}$$

$$+\frac{\alpha}{1+\alpha}\Big(h(1+\frac{1+\alpha}{\alpha}\,\mu'\mathbf{z}_k\,+\,\lambda'\mathbf{x}_k)\,-\,1\Big)\Big(\begin{array}{c}\mathbf{x}_k\\(1+\alpha^{-1})\mathbf{z}_k\end{array}\Big)$$

By algebraic reduction of the first term, we find that the differential in (λ,μ) of the last displayed expression is

$$\frac{1}{1+\alpha}\left(\begin{array}{cc}\mathbf{x}_k^{\otimes 2} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{array}\right)+\frac{\alpha}{1+\alpha}\,h'(1+\lambda'\mathbf{x}_k+\frac{1+\alpha}{\alpha}\,\mu'\mathbf{z}_k)\Big(\begin{array}{c}\mathbf{x}_k\\(1+\alpha^{-1})\mathbf{z}_k\end{array}\Big)^{\otimes 2}$$

Our assumption on asymptotic positive definiteness of $N^{-1}\sum_{l\in\mathcal{S}} r_k w_k^o\,((\mathbf{x}_k',\mathbf{z}_k')')^{\otimes 2}$ now implies asymptotic positive definiteness of the sum over $k\in\mathcal{S}$ of $N^{-1}\,r_k\,w_k^o$ times the last expression. This completes the proof that differential in (λ,μ) of the left-hand side of (14′) is everywhere a positive definite matrix. In passing, we have proved

Lemma 1 *The mapping $\phi_{s,N}(\cdot)$ of (λ,μ) defined by the left-hand side of (14′) has the equivalent expression*

$$\phi_{s,N}(\lambda,\mu)\;=\;\frac{1}{N}\sum_{k\in\mathcal{S}} r_k\,w_k^o\Big[\frac{1}{1+\alpha}\,(\lambda'\mathbf{x}_k)\begin{pmatrix}\mathbf{x}_k\\ \mathbf{0}\end{pmatrix}\quad+$$

$$+\frac{\alpha}{1+\alpha}\Big(h(1+\frac{1+\alpha}{\alpha}\,\mu'\mathbf{z}_k\,+\,\lambda'\mathbf{x}_k)\,-\,1\Big)\Big(\begin{array}{c}\mathbf{x}_k\\(1+\alpha^{-1})\mathbf{z}_k\end{array}\Big)\Big]$$

with differential

$$D\phi_{s,N}(\lambda,\mu)\;=\;\frac{1}{N}\sum_{k\in\mathcal{S}} r_k\,w_k^o\Big[\frac{1}{1+\alpha}\begin{pmatrix}\mathbf{x}_k\\ \mathbf{0}\end{pmatrix}^{\otimes 2}+$$

$$+\frac{\alpha}{1+\alpha}\,h'(1+\lambda'\mathbf{x}_k+\frac{1+\alpha}{\alpha}\,\mu'\mathbf{z}_k)\Big(\begin{array}{c}\mathbf{x}_k\\(1+\alpha^{-1})\mathbf{z}_k\end{array}\Big)^{\otimes 2}\Big]$$

C Random Inverse Functions defining $(\hat{\lambda},\,\hat{\mu})$

As in Deville and Särndal (1992) [from now on cited as **DS92**], we determine the Lagrange multipliers (λ,μ) implicitly by appealing to a theorem extending the Inverse Function Theorem to a series of convergent superpopulation-based random functions. Following **DS92**, define $\phi_{s,N}$ as the the function of $(\lambda,\mu)\in\mathbf{R}^{p+q}$ given by the left-hand side of (14′), and we rely on our superpopulation assumptions to ensure that

$$\phi_{s,N}(\lambda,\mu)\;\xrightarrow{P}\;\phi(\lambda,\mu)\quad,\qquad D\phi_{s,N}(\lambda,\mu)\;\xrightarrow{P}\;D\phi(\lambda,\mu)\tag{31}$$

pointwise and uniformly on compact sets in $\mathbf{R}^{p+q}$, where ϕ is continuously differentiable with everywhere symmetric positive-definite differential $D\phi(\lambda,\mu)\;=\;B(\lambda,\mu)\;=\;B$.

We now elaborate the extension of the essential argument of (**DS92**, App. 1) in a series of steps.

1^o. *The continuously differentiable limiting function ϕ is everywhere defined and one-to-one onto its range.*

Since the domain of ϕ is convex, the positive-definiteness of its differential together with the Mean Value Theorem immediately implies that it is one-to-one. (If $\phi(u_2) = \phi(u_1)$ then for some u_* on the ray from u_1 to u_2, $0 = \phi(u_2) - \phi(u_1) = B(u_*)(u_2 - u_1)$, from which it follows that $u_2 = u_1$.)

2^o. *The range of ϕ is open.*

For each point $\phi(u_0) \in \text{range}(\phi)$, the Inverse Function Theorem implies by local nonsingularity of ϕ near u_0 that there exists an open neighborhood A_{u_0} of $\phi(u_0)$ on which the inverse map ϕ^{-1} is well-defined and continuously differentiable with uniformly bounded $\|D\phi^{-1}\|$ where $\|\cdot\|$ is any matrix norm. It follows that the range of ϕ contains a neighborhood around each of its elements, and therefore is open.

3^o. *On each compact subset $C \subset range(\phi)$, with probability approaching* 1 *as the superpopulation size N gets large, $\phi_{s,N}$ is invertible on C.*

For all $v \in C \subset \text{range}(\phi)$ there is a convex open ball $B = B_\epsilon(v)$ about v which also lies within range(ϕ). By the mean value theorem, for all $w \in B_\epsilon(u)$ (with probability 1) there exists $\tilde{v} \in B_\epsilon(u)$ on the ray from v to w for which

$$\|\phi_{s,N}(v) - \phi_{s,N}(w)\| \;=\; \|D\phi(\tilde{v})(v-w) \,+\, D(\phi_{s,N} - \phi)(\tilde{v})(v-w)\|$$

$$\geq\; \inf_{v\in C} \|D\phi(v)\| \,\cdot\, \|v-w\| \,-\, \sup_{v\in C} \|D(\phi_{s,N} - \phi)(v)\| \,\cdot\, \|v-w\|$$

which with probability converging to 1 as N gets large, is bounded below by a positive constant times $\|v-w\|$. After covering C with finitely many open sets B found in this way, on each of which $\|D\phi_{s,N}\|$ is bounded away from 0 with probability approaching 1 as $N \to \infty$, we conclude that $P(\inf_{v\in C} \|D\phi_{s,N}(v)\| \,>\, \delta) \to 1$ for all sufficiently small $\delta > 0$. As a consequence, $\phi_{s,N}$ is invertible on all of C with probability converging to 1.

D Large α Limit in (6)

In this and the next Section, assume only **(A.3)**, not **(A.3′)**, since there can often be redundant variables among the combined coordinates of $\mathbf{x}_k$, $\mathbf{z}_k$. Throughout, let

$$M \;=\; M_\infty \;=\; \sum_{k\in\mathcal{S}} r_k\, w_k^o \begin{pmatrix} \mathbf{x}_k \\ \mathbf{z}_k \end{pmatrix}^{\otimes 2} \;\equiv\; \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}$$

where M_{11} is $p\times p$ and $M_{21} = M_{12}'$, and we will replace all matrix inverses by notations A^- to cover the case where they are generalized inverses.

Since the calculations in these Sections are done with $Q \equiv 0$, (15) can be re-expressed using standard partitioned matrix identities to say

$$\lambda \;=\; \Big(M_{11} - \frac{\alpha}{1+\alpha}\, M_{12} M_{22}^{-} M_{21}\Big)^{-} \Big(t^*_{\mathbf{x}} - \hat{t}_{r\mathbf{x},\pi} \;-\; \frac{\alpha}{1+\alpha}\, M_{12} M_{22}^{-} (t^*_{\mathbf{z}} - \hat{t}_{r\mathbf{z},\pi}\Big) \qquad (32)$$

$$\mu \;=\; \Big((1+\alpha^{-1})\, M_{22} - M_{21} M_{11}^{-} M_{12}\Big)^{-} \Big(t^*_{\mathbf{z}} - \hat{t}_{r\mathbf{z},\pi} \;-\; M_{21} M_{11}^{-} (t^*_{\mathbf{x}} - \hat{t}_{r\mathbf{x},\pi}\Big) \qquad (33)$$

In the setting of the present Section, with α large, these equations lead to

$$\begin{pmatrix} \lambda \\ \mu \end{pmatrix} \;=\; M_\alpha^{-} \begin{pmatrix} t^*_{\mathbf{x}} - \hat{t}_{r\mathbf{x},\pi} \\ t^*_{\mathbf{z}} - \hat{t}_{r\mathbf{z},\pi} \end{pmatrix} \;=\; M_\infty^{-} \begin{pmatrix} t^*_{\mathbf{x}} - \hat{t}_{r\mathbf{x},\pi} \\ t^*_{\mathbf{z}} - \hat{t}_{r\mathbf{z},\pi} \end{pmatrix} \;+\; O(\frac{1}{\alpha})$$

and, since $\mathbf{x}_k$, w_k^o are uniformly bounded, (11) implies

$$w_k \;=\; \frac{\alpha}{1+\alpha}\, \hat{w}_k \;+\; \frac{1}{\alpha}\, w_k^o\, (1 + \lambda' \mathbf{x}_k) \;=\; \hat{w}_k \;+\; O(\frac{1}{\alpha})$$

and $\hat{w}_k$ differs by $O(1/\alpha)$ (uniformly in k) from the calibrated weights

$$r_k\, w_k^o \,\Big(1 \;+\; \begin{pmatrix} \mathbf{x}_k \\ \mathbf{z}_k \end{pmatrix}' M_\infty^{-} \begin{pmatrix} t^*_{\mathbf{x}} - \hat{t}_{r\mathbf{x},\pi} \\ t^*_{\mathbf{z}} - \hat{t}_{r\mathbf{z},\pi} \end{pmatrix}\Big)$$

E Small α Limit in (6)

Now start from equations (32) and (33) in the asymptotic setting with $\alpha \to 0$, finding (again under the restriction $Q \equiv 0$)

$$\lambda \;=\; M_{11}^{-}\,(t^*_{\mathbf{x}} - \hat{t}_{r\mathbf{x},\pi}) + O(\alpha) \quad , \quad \mu \;=\; \alpha\, M_{22}^{-} \Big(t^*_{\mathbf{z}} - \hat{t}_{r\mathbf{z},\pi} - M_{21} M_{11}^{-}\,(t^*_{\mathbf{x}} - \hat{t}_{r\mathbf{x},\pi})\Big) + O(\alpha^2)$$

Then (13) gives (uniformly in k)

$$\hat{w}_k \;=\; r_k\, w_k^o \Big(1 \;+\; \mathbf{x}_k' M_{11}^{-}\,(t^*_{\mathbf{x}} - \hat{t}_{r\mathbf{x},\pi}) \;+\; \mathbf{z}_k' M_{22}^{-}\,((t^*_{\mathbf{z}} - \hat{t}_{r\mathbf{z},\pi} - M_{21} M_{11}^{-} (t^*_{\mathbf{x}} - \hat{t}_{r\mathbf{x},\pi}))\Big)$$

and the definitions of the blocks in matrix M, together with some algebra, give

$$\lim_{\alpha \to 0} \sum_{k \in \mathcal{S}} \hat{w}_k\, \mathbf{z}_k \;=\; \hat{t}_{r\mathbf{z},\pi} \;+\; M_{21} M_{11}^{-} (t^*_{\mathbf{x}} - \hat{t}_{r\mathbf{x},\pi}) \;+\; \Big(t^*_{\mathbf{z}} - \hat{t}_{r\mathbf{z},\pi} - M_{21} M_{11}^{-} (t^*_{\mathbf{x}} - \hat{t}_{r\mathbf{x},\pi}))\Big)$$

which is precisely equal to $t^*_{\mathbf{z}}$, and this must be so because the conditions of minimization which led to (13) required that the calibration constraint (4) be satisfied exactly for each α. However, similar algebraic simplification shows that the nonresponse adjustment condition (2), which holds precisely only for the auxiliary weights w_k, leads to the final-weight totals

$$\lim_{\alpha \to 0} \sum_{k \in \mathcal{S}} \hat{w}_k\, \mathbf{x}_k \;=\; t^*_{\mathbf{x}} \;+\; M_{12}\, M_{22}^{-} \Big(t^*_{\mathbf{z}} - \hat{t}_{r\mathbf{z},\pi} - M_{21} M_{11}^{-} (t^*_{\mathbf{x}} - \hat{t}_{r\mathbf{x},\pi})\Big) \qquad (34)$$

F Contraction Property of (18)-(19) in (λ, μ)

Substitute the identity $Q'(h(u)) = (\alpha/(1+\alpha))\,(u - h(u))$ into equation (19) and apply the Mean Value Theorem to the function h to find that for some values $u_k^{(j)}$ between $\hat{w}_k^{(j+1)}/w_k^o$ and $\hat{w}_k^{(j)}/w_k^o$,

$$\begin{pmatrix} \lambda^{(j+1)} - \lambda^{(j)} \\ \mu^{(j+1)} - \mu^{(j)} \end{pmatrix} \;=\; \frac{\alpha}{1+\alpha}\, M_\alpha^{-1} \sum_{k \in \mathcal{S}} r_k\, w_k^o \,\cdot \tag{35}$$

$$(1 - h'(u_k^{(j)}))\, \begin{pmatrix} \mathbf{x}_k \\ (1+\alpha^{-1})\mathbf{z}_k \end{pmatrix}^{\otimes 2} \begin{pmatrix} \lambda^{(j)} - \lambda^{(j-1)} \\ \mu^{(j)} - \mu^{(j-1)} \end{pmatrix}$$

Next observe by (17) that for all u,

$$0 \;<\; h'(u) \;=\; \Big(1 + \frac{1+\alpha}{\alpha}\, Q''(h(u))\Big)^{-1} \;<\; 1$$

Then the definition of M_α immediately shows that the norm of $((\lambda^{(j)} - \lambda^{(j-1)})', (\mu^{(j)} - \mu^{(j-1)})')'$ is a decreasing function of $j \geq 1$, and in fact decreases by at least the constant factor $\inf_{u \in B}\,(1 - h'(u))$ as long as the quantities $h(1 + (1+\alpha^{-1})\mu^{(j)\prime}\mathbf{z}_k + \lambda^{(j)\prime}\mathbf{x}_k)$ and $h(1 + (1+\alpha^{-1})\mu^{(j-1)\prime}\mathbf{z}_k + \lambda^{(j-1)\prime}\mathbf{x}_k)$ both lie within a bounded interval B.

G Proof & Discussion of Proposition 2

Assume the hypotheses of Proposition 2, first with condition (i) but not (ii). Note that $(\lambda_*,\, \mu_*) \in \mathbf{R}^{p+q}$ has been shown to be uniquely determined satisfying (20). On the other hand, (i) implies the same limiting equation holds with $(\lambda_*,\, \mu_*)$ replaced by $(\lambda, 0)$, as we now show. First, by the assumption that (i) imposes on indices $k \in \mathcal{U}_1$,

$$\sum_{k \in \mathcal{U}_1} \rho_k \begin{pmatrix} \mathbf{x}_k^{\otimes 2} & \mathbf{x}_k\, \mathbf{z}_k' \\ \mathbf{z}_k\, \mathbf{x}_k' & (1+\alpha^{-1})\, \mathbf{z}_k^{\otimes 2} \end{pmatrix} \begin{pmatrix} \lambda \\ \mathbf{0} \end{pmatrix} \;=\; \sum_{k \in \mathcal{U}_1} (1 - \rho_k) \begin{pmatrix} \mathbf{x}_k \\ \mathbf{z}_k \end{pmatrix}$$

and $Q' \circ h(1 + \lambda'\mathbf{x}) \equiv 0$, while (i) together with the boundedness of $Q' \circ h$ implies by (17) that

$$\sum_{k \in \mathcal{U} \setminus \mathcal{U}_1} \Big(\rho_k \begin{pmatrix} \mathbf{x}_k^{\otimes 2} & \mathbf{x}_k\, \mathbf{z}_k' \\ \mathbf{z}_k\, \mathbf{x}_k' & (1+\alpha^{-1})\, \mathbf{z}_k^{\otimes 2} \end{pmatrix} \begin{pmatrix} \lambda \\ \mathbf{0} \end{pmatrix} \;-\; \rho_k\, Q'(h(1 + \lambda'\mathbf{x}_k)) \begin{pmatrix} \mathbf{x}_k \\ (1+\alpha^{-1})\mathbf{z}_k \end{pmatrix} \Big)$$

$$=\; \sum_{k \in \mathcal{U} \setminus \mathcal{U}_1} (1 - \rho_k) \begin{pmatrix} \mathbf{x}_k \\ \mathbf{z}_k \end{pmatrix} \;+\; \mathcal{O}(\frac{N}{\sqrt{n}})$$

Putting these displayed lines together shows that (20) holds with $\lambda_* = \lambda$ and $\mu_* = \mathbf{0}$, so that $(\lambda_*,\, \mu_*) = (\lambda,\, \mathbf{0})$. Again using the properties assumed for indices in $\mathcal{U}_1$, it follows that

$$\sum_{k \in \mathcal{U}_1} (\rho_k\, f_k^* - 1)\, y_k \;=\; 0 \quad , \qquad \sum_{k \in \mathcal{U} \setminus \mathcal{U}_1} (\rho_k\, f_k^* + 1)\, |y_k| \;=\; o(\frac{N}{\sqrt{n}})$$

from which the assertion $\sum_{\mathcal{U}} \rho_k\, f_k^*\, y_k \;= t_y \;+\; o(N/\sqrt{n})$ follows immediately.

Next assume (ii) but not (i). The displayed limit in (ii) implies that β is uniquely determined as the limiting least squares coefficient vector for a regression over $\mathcal{U}$ of y_k on $\mathbf{z}_k$ with weights ρ_k, i.e.,

$$\beta \;=\; \lim_N \Big(\sum_{k \in \mathcal{U}} \rho_k\, \mathbf{z}_k^{\otimes 2} \Big)^{-1} \sum_{k \in \mathcal{U}} \rho_k\, y_k\, \mathbf{z}_k$$

and then, since $Q' \equiv 0$ implies $f_k^* \;=\; 1 + (1+\alpha^{-1})\, \mu_*' \mathbf{z}_k \;+\; \lambda_*' \mathbf{x}_k$

$$\sum_{k \in \mathcal{U}} \rho_k\, f_k^*\, y_k \;=\; \sum_{k \in \mathcal{U}} \rho_k\, y_k \;+ \sum_{k \in \mathcal{U}} \rho_k \begin{pmatrix} \mathbf{x}_k \\ (1+\alpha^{-1}) \mathbf{z}_k \end{pmatrix}' \begin{pmatrix} \lambda_* \\ \mu_* \end{pmatrix} (y_k - \beta' \mathbf{z}_k \;+\; \beta' \mathbf{z}_k)$$

$$=\; \sum_{k \in \mathcal{U}} \rho_k\, y_k \;+\; \beta' \sum_{k \in \mathcal{U}} \rho_k \begin{pmatrix} \mathbf{z}_k\, \mathbf{x}_k' \\ (1+\alpha^{-1}) \mathbf{z}_k^{\otimes 2} \end{pmatrix}' \begin{pmatrix} \lambda_* \\ \mu_* \end{pmatrix} \;+\; o(\frac{N}{\sqrt{n}})$$

Now the (last q components in) identity (20), together with **(A.2)** saying that the limits in (20) and the definition of ϕ are attained with order of convergence $o(1/\sqrt{n})$, imply that the last expression is

$$\sum_{k \in \mathcal{U}} \rho_k\, y_k \;+ \sum_{k \in \mathcal{U}} (1-\rho_k)\, \beta'\, \mathbf{z}_k \;+\; o(\frac{N}{\sqrt{n}})$$

Appealing again to the displayed assumption (ii) shows the last expression is

$$\sum_{k \in \mathcal{U}} \rho_k\, y_k \;+ \sum_{k \in \mathcal{U}} (1-\rho_k)\, y_k \;+\; o(\frac{N}{\sqrt{n}}) \;=\; t_y \;+\; o(\frac{N}{\sqrt{n}})$$

as asserted. □

Remark 1 *In the preceding proof, exactly the same steps could be followed but with regression of* y_k *on* $\mathbf{z}_k$ *replaced by regression on* $(\mathbf{x}_k', \mathbf{z}_k')'$, *if* $\alpha = \infty$. *The point here is that the full equation (20) can be used, and not simply the last q components, when the matrix arising in the summation term of* $\phi(\lambda, \mu)$ *is* $(\mathbf{x}_k', \mathbf{z}_k')^{\otimes 2}$, *which is true only if* $\alpha = \infty$.

G.1 Discussion of Population Controls with $\mu_* \neq 0$

Population controls are often imposed in order to remove frame deficiencies, e.g., when a survey collects useful information but would estimate population counts smaller than those available from a larger survey or census. How might superpopulation consistency theory be reformulated to cover this situation ?

The most straightforward reformulation is to regard the sample as being drawn with inclusion probabilities π_k from a frame $\mathcal{U}_0 \subset \mathcal{U}$ about which one can assume a working model something like

$$\sum_{k \in \mathcal{U}} v_k \left(I_{[k \in \mathcal{U}_0]} \left(1 + \mu' \mathbf{z}_k\right) - 1\right) = o(\frac{N}{\sqrt{n}}) \tag{36}$$

for v_k equal to any components of $y_k, \mathbf{x}_k, \mathbf{z}_k,\, y_k\,(\mathbf{x}_k', \mathbf{z}_k')$, or $(\mathbf{x}_k', \mathbf{z}_k')^{\otimes 2}$. Although it would usually make little sense to view frame-inclusion indicators $I_{[k \in \mathcal{U}_0]}$ as independent binary random variables like the response indicators in the Oh and Scheuren (1983) quasi-randomization model, the assumption (36) has much the same effect. With full response, it is not hard to see that this model assumption would ensure that the population-control calibration on $\mathbf{z}_k$ consistently corrects for the difference between frames $\mathcal{U}_0$ and $\mathcal{U}$, although the limiting Lagrange multipliers μ_* would generally be different from 0. In the presence both of nonresponse and deficient frame $\mathcal{U}_0$, assuming (36) together with **(A.0)-(A.4)**, it is also possible to check that the two-stage method (first calibrating weights w_k via $\mathbf{x}_k$ to correct for nonresponse within $\mathcal{U}_0$ and then starting from w_k, calibrating to population totals via $\mathbf{z}_k$) would be consistent. It is not clear whether the new single-step generalized calibration methods of this paper would also be consistent, or whether that is particularly worrisome. The path to consistency via assumption (ii) in Proposition 2, including the more general version mentioned in Remark 1, would still be available.