

Methods for Adjusting Survey Weights When Estimating a Total

Kimberly Henry¹ and Richard Valliant²

¹Statistics of Income, Internal Revenue Service

²Joint Program in Survey Methodology, University of Michigan.

¹77 K Street NW, Washington DC, 20001, Kimberly.A.Henry@irs.gov.

Abstract: Several design-based, model-based, and model-assisted methods have been developed to adjust survey weights for nonresponse or coverage errors, to reduce variances through the use of auxiliary data or by restricting the range of the weights themselves. Some methods directly change the weights, like nonresponse adjustments, calibration weighting, and design-based ad hoc weight trimming methods. Other methods, like the robust superpopulation modeling approaches, implicitly adjust the associated case weights. A newly developed method called generalized design-based inference involves modeling the weights as a function of the survey response variables and using the smoothed weights predicted from the model to estimate finite population totals. This paper first provides a review of the alternative weighting adjustment methods proposed in the literature. We also discuss the methods' strengths and weaknesses, as related to implications for sample-based inference.

Introduction

Weighting samples is important to reflect not only sample design decisions made at the planning stage and use of auxiliary data to improve the efficiency of estimators, but also practical issues that arise during data collection and cleaning that necessitate weighting adjustments. Planned and unplanned adjustments to survey weights are used to account for these practical issues. The standard, theoretically-based methods for weighting are summarized by Kalton and Flores-Cervantes (2003). Often these adjustments lead to variations in the survey weights. When large variation in survey weights exists, it can potentially produce unreasonable point estimates of population means and totals and decrease the precision of these point estimates (e.g., Valliant 2004). Survey practitioners commonly face this problem when producing weights for analysis datasets (Kish 1990; Liu *et al.* 2004; Chowdhury *et al.* 2007). Even when the estimators are unbiased, extreme weights can produce inefficient estimates. Trimming or truncating large weights can reduce unreasonably large estimated totals and substantially reduce the variability due to the weights. This reduces variance of the finite population estimates at the expense of introducing bias; if the variance reduction is larger than the squared bias increase, then the net result is an overall decrease in mean square error (MSE) of the estimate. The various existing trimming methods use either design-based or model-based approaches to meet this MSE-reduction goal.

Variation in survey weights can arise at the sample design, data collection, and post-data collection stages of sampling. First, intentional differential base weights, the inverse of the probability of selection, are created under different sampling designs. For example, multiple survey analysis objectives may lead to disproportionate sampling of population subgroups. Issues that occur during data collection can also impact probabilities of selection, e.g., in area probability samples, new construction developments with a large number of housing units that were not originally listed may be discovered. These are usually subsampled to reduce interviewer workloads, but this subsampling may create a subset of units with extremely large base weights. Another example is subsampling cases for nonresponse follow up; subsequent weighted analysis incorporates subsampling adjustments. Last, post-data collection adjustments to base survey weights are also commonly used to account for multistage sampling. Examples include subsampling persons within households (Liu *et al.* 2004), adjusting for nonresponse to the survey (Oh and Scheuren 1983), calibrating to external population totals to control for nonresponse and coverage error (Holt and Smith 1979; Särndal *et al.* 1992; Bethlehem 2002; Särndal and Lundström 2005), and combining information across multiple frames (such as telephone surveys collected from landline telephone and cell phone frames, e.g., Cochran 1967; Hartley 1962).

Often multiple adjustments are performed at each stage of sample design, selection, and data editing. For example, the Bureau of Transportation Statistics (2002) used the following steps to produce weights in their National Transportation Availability and Use Survey, a complex sample using in-person household interviews to assess

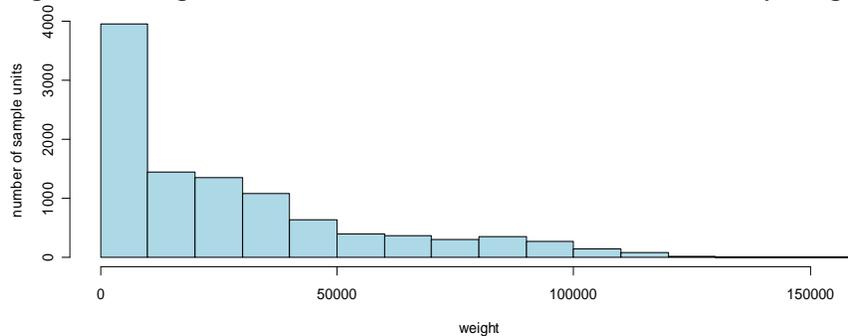
people’s access to public and private transportation in the U.S.:

- Household-level weights: base weights for stratified, multistage cluster sampling; unknown residential status adjustment; screener nonresponse adjustment; subsampling households for persons with and without disabilities; multiple telephone adjustment; poststratification.
- Person-level weights: the initial weight is the product of the household-level weight from above and a subsampling adjustment for persons within the household; an extended interview nonresponse adjustment; trimming for disability status; raking; and a non-telephone adjustment.

This example also illustrates that there can be a need for multiple weights based on the level of analysis (here household-level vs. person-level) that is desired from the sample data.

The effect of such weighting adjustments, when applied across the sample’s design strata and PSUs, can be that the variability of the weights is increased (Kish 1965, 1992; Kalton and Kasprzyk 1986). This can increase the variances of the sample-based estimators, thereby decreasing precision. A few very large weights can also be created such that the product $w_i y_i$ creates an unusually large estimate of the population total. To see this, consider the following empirical example from the 2005-2006 National Health and Nutrition Examination (NHANES) public-use file dataset (NHANESa). The interview-level weights for 10,348 people have post-stratification adjustments to control totals estimated from the Current Population Survey (NHANESb). These weights range from 1,225 to 152,162 and are quite skewed:

Figure 1. Histogram of NHANES 2005-2006 Adult Interview Survey Weights



The weights in the tails of the distribution, such as the one shown in Figure 1, can lead to overly large totals with associated large variances, particularly when the weight is combined with a large survey response value. This problem can increase for domain-level estimates, particularly in establishment data, when variables of interest can also be highly skewed toward zero. On the other hand, large weights for some subgroups of units may be necessary to produce estimators that are, in some sense, unbiased. Beaumont and Rivest (2008) estimate establishment sample units that were assigned large design weights based on incorrect small measures of size accounted for 20-30% of an estimated domain total. These outlying weights also drive variance estimates.

There is limited literature and theory on design-based weight trimming methods, most of which are not peer-reviewed publications and focus on issues specific to a single survey or estimator. The most cited work is by Potter (1988; 1990), who presents an overview of alternative procedures and applies them in simulations. Other studies involve a particular survey (e.g., Battaglia *et al.* 2004; Chowdhury *et al.* 2007; Griffin 1995; Liu *et al.* 2004; Pedlow *et al.* 2003; Reynolds and Curtin 2009). All design-based methods involve establishing an upper cutoff point for large weights, reducing weights larger than the cutoff to its value, then “redistributing” the weight above the cutoff to the non-trimmed cases. This ensures that the weights before and after trimming sum to the same totals (Kalton and Flores-Cervantes 2003). The methods vary by how the cutoff is chosen. There are three general approaches: (1) ad hoc methods that do not use the survey response variables or an explicit model for the weight to determine the cutoff (e.g., trimming weights that exceed five times the median weight to this value); (2) Cox and McGrath’s (1981) method that uses the empirical MSE of a particular estimator and variable of interest; and (3) methods assuming that the right-tail of the weights follow some skewed parametric distribution, then use the cutoff associated with an arbitrarily small probability from the empirical distribution (Chowdhury *et al.* 2007; Potter 1988).

Examples of Post-Data Collection Weight Adjustment Methods

Here we provide some examples of various weighting adjustment methods that are performed in sample surveys. This is not an all-encompassing list; refer to the post-data collection adjustment references and Chapters 4, 8, 16, and 25 in Pfeffermann and Rao (2009) for more detail.

Example 1. Cell-based Nonresponse Adjustments. These adjustments involve categorizing the sample dataset into cells using covariates available for both respondents and nonrespondents that are believed to be highly correlated with response propensity and key survey variables. Assuming that nonresponse is constant within each cell (“missing at random;” age/race/sex are often used in household surveys), the reciprocal of the cell-based response rate is used to increase the weights of all units within the cell. Propensity models across all cells using the cell-based covariates can also be used to predict the response rate. Other nonresponse weighting adjustments are discussed in Brick and Montaquila (2009). ■

Example 2. Dual Frame Adjustments. For surveys that use multiple frames, e.g., telephone and area probability samples or landline and cell phone surveys, additional weighting adjustments may be used to account for units that are contained in more than one frame. Often composite estimators of totals are used, which are a weighted average of the separate frame estimates (Hartley 1962; Cochran 1967) incorporates the known or estimated overlap of units on both frames. These can be explicitly expressed as adjustments to each sample unit’s weight. ■

Example 3. Poststratification. Here survey weights are adjusted such that they add up to external population counts by available domains. This widely-used approach allows us to correct the imbalance that can occur between the sample design and sample completion, i.e., if the sample respondent distribution within the external categories differs from the population (which can occur if, e.g., subgroups respond or are covered by the frame at different rates), as well as reduce potential bias in the sample-based estimates. Denoting the poststrata by $d = 1, \dots, D$, the poststratification estimator for a total involves adjusting the base-weighted domain totals (\hat{T}_d) by the ratio of known (N_d) to estimated (\hat{N}_d) domain sizes: $\hat{T}_{PS} = \sum_{d=1}^D N_d \hat{T}_d / \hat{N}_d$. ■

Example 4. Calibration. Case weights resulting from calibration on benchmark auxiliary variables can be defined with a global regression model for the survey variables (Huang and Fuller 1978; Bardsley and Chambers 1984; Bethlehem and Keller 1987; Särndal *et al.* 1992; Sverchkov and Pfeffermann 2004; Beaumont 2008; Kott 2009; Särndal and Lundström 2005; Slud and Thibaudeau 2010). Deville and Särndal (1992) proposed a model-assisted calibration approach that involves minimizing a distance function between the base weights and final weights to obtain an optimal set of survey weights. Here “optimal” means that the final weights produce totals that match external population totals for the auxiliary variables \mathbf{X} within a margin of error. Specifying alternative distance functions produces alternative estimators. ■

In addition to the design-based methods, model-based estimation methods incorporate a model that is assumed between the survey variable of interest and some auxiliary variables whose individual values are assumed known for the entire finite population or for which totals are known. When the relationship between the survey and auxiliary variables is strong, the model-based estimators can have superior performance over design-based ones. There are two approaches: superpopulation modeling and its Bayesian extension. In both methods, the model fit must be justified and the potential for variable-specific weights exists. We also include a relatively new form of estimation called generalized design-based estimation under the category of model-based estimation. While the form of inference is different, the incorporation of a model for weight trimming leads us to also include this method under the general category of model-based.

Another approach, the model-assisted, appeals to the use of models, but particular design-based properties hold when the model does not. By optimizing the weights such that benchmark auxiliary totals are met, this method can produce estimators with desirable design-based properties. Recently, penalized (p -) spline estimators have been developed to produce estimators of a total that are minimally influenced by extreme data values. Zheng and Little (2003, 2005) used p -spline estimators to improve estimates of totals from probability-proportional-to-size (pps) samples. Breidt and Opsomer (2000) and Breidt *et al.* (2005) proposed and developed a model-assisted p -spline estimator that was more robust to misspecification of the linear model, resulting in minimum loss in efficiency

compared to other calibration estimators. The p -spline estimators are a specific case of a robust model-prediction estimator. The model-assisted approaches involve unit-level adjustments to weights, potentially creating more variability in the weights.

Existing Design-Based Methods

The most common sample-based inference for a finite population total involves the Horvitz-Thompson (HT, 1952) estimator. This section briefly introduces the HT estimator and examples of methods that trim the HT weights. For s denoting a probability sample of size n drawn from a population of N units, the *Horvitz-Thompson estimator* (Horvitz and Thompson 1952) for a finite population total of the variable of interest y is

$$\hat{T}_{HT} = \sum_{i \in s} y_i / \pi_i = \sum_{i \in s} w_i y_i. \quad (1)$$

Here the inverse of the probability of selection, $\pi_i = P(i \in s)$, is used as the weight, $w_i = \pi_i^{-1}$. This estimator is unbiased for the finite population total in repeated π p s sampling, but can be quite inefficient due to variation in the selection probabilities if π_i and y_i are not closely related. The design-based variance of (1) is

$$\text{Var}(\hat{T}_{HT}) = \sum_{i \in U} \sum_{j \in U} (\pi_{ij} - \pi_i \pi_j) y_i y_j / \pi_i \pi_j, \quad (2)$$

where π_{ij} is the joint selection probability of units i and j in the population set U . Influential observations in estimating a population total using (1) and the variance estimator associated with the variance in (2) arise simply due to the combination of probabilities of selection and survey variable values. Alternative sample designs, such as probability proportional to a measure of size, introduce variable probabilities of selection in (1). The variability in selection probabilities can increase under complex multistage sampling and multiple weighting adjustments. Thus, the HT-based estimates from one particular sample may be far from the true total value, particularly if the probabilities of selection are negatively correlated with the characteristic of interest (see discussion in Little 2004).

Examples of existing design-based trimming methods are presented in the remainder of this section. In all methods, outlier weights are flagged in the survey dataset, usually through data inspection, editing, and/or computation of domain-level estimates, then trimmed to some arbitrary value. The remaining portion of the weight, called the “excess weight,” is then “redistributed” to other survey units (Kalton and Cervantes 2003). This increases the weights on non-trimmed cases; if the increase is slight, then the associated bias is small, assuming that the initial set of weights produced unbiased estimates. Redistributing the weight is done to ensure that the weights after trimming still add up to target population sizes. The underlying assumption is that decreasing the variability caused by the outlying weights offsets the increase in bias incurred by units that absorb the excess weight. The most extreme windsorized value for outlying weights is one; other possibilities include using weights within another stratum, adjacent weighting class, or some percentile value from an assumed weight distribution.

The alternative design-based weight trimming methods differ in how the cutoff boundary to identify outlying weights is chosen, but they can be grouped into ad hoc methods (Ex. 5 and 6), methods that use the empirical MSE of the estimator of interest (Ex. 7), and methods that assume a specified parametric (skewed) distribution for the weight (Ex. 8 and 9).

Example 5. *The NAEP method.* To reduce extremely large HT-estimator weights in (1.1), Potter (1988) proposed trimming all weights $w_i > \sqrt{c \sum_{i \in s} w_i^2 / n}$ to this cutoff value. This method was used to trim weights in the 1986 National Association of Educational Progress sample (Johnson *et al.* 1987). The other sample units’ weights are adjusted to reproduce the original weighted sum in (1). The value of c is “arbitrary and is chosen empirically by looking at values of $nw_i^2 / \sum_{i \in s} w_i^2$ ” (p. 457 in Potter 1988). The sum of squared adjusted weights is computed iteratively until no weights exceed the cutoff value, then the windsorized weights replace w_i in (1.1) to estimate the total. Potter (1990) claims this method outperformed other MSE-minimizing alternatives, despite the fact it does not incorporate the survey response variables of interest. ■

Example 6. *The NIS method.* Chowdhury *et al.* (2007) describe the weight trimming method used to estimate proportions in the U.S. National Immunization Survey (NIS). The “current” (at the time of the article) cutoff value

was $\text{median}(w_i) + 6IQR(w_i)$, where $IQR(w_i)$ denotes the inter-quartile range of the weights. Versions of this cutoff (e.g., a constant times the median weight or other percentiles of the weights) have been used by other survey organizations (Battaglia *et al.* 2004; Pedlow *et al.* 2003; NCES 2003; Appendix A in Reynolds and Curtin 2009). ■

Example 7. Cox and McGrath’s MSE Trimming Method. Cox and McGrath (1981) proposed using the empirical MSE for a sample mean estimated at various trimming levels:

$\widehat{MSE}(\bar{y}_t) = (\bar{y}_t - \bar{y}_w)^2 + \widehat{Var}(\bar{y}_t) + 2\sqrt{\widehat{Var}(\bar{y}_t)\widehat{Var}(\bar{y}_w)}$, $t = 1, \dots, T$, where t denotes the trimming level ranging from $t=1$ for the unweighted sample mean estimator to $t=T$ denoting the fully-weighted sample estimator \bar{y}_w (the sample-based estimate of the mean with no weights trimmed). Assuming that \bar{y}_w is the true population mean, the MSE expression is calculated for possible values of t , which correspond to different weight trimming cutoffs, and the cutoff associated with the minimum MSE value is chosen as “optimal.” Determining a trimming level by assuming that \bar{y}_w is the true population mean is dubious because \bar{y}_w itself may have a large variance and be far from the population mean in a particular sample. Potter (1988) also used this approach, estimating the MSE for a few survey variables at twenty trimming levels. He determined the “optimal” trimming by ranking the MSE (from 1 to 20) for each variable/trimming level combination, calculating the average rank across variables, and identifying the trimming level with the lowest average rank. ■

Example 8. Inverse Beta Distribution Method. Potter (1988) also considered a method that assumes the survey weights (w) follow an inverted Beta, $IB(\alpha, \beta)$, distribution:

$$f(w) = n \left(\frac{1}{nw} \right)^{\alpha+1} \left(1 - \frac{1}{nw} \right)^{\beta-1} \Gamma(\alpha + \beta) / \Gamma(\alpha)\Gamma(\beta), \quad n^{-1} \leq w \leq \infty, \quad \text{where } \Gamma(\bullet) \text{ denotes the gamma function.}$$

The IB distribution was proposed since it is a right-tailed skewed distribution. The mean and variance of the empirical IB distribution generated from the sample weights are used to estimate the IB model parameters. The trimming level is then set according to some pre-specified quantile of the cumulative IB distribution and weights in the tail of the distribution are trimmed. ■

Example 9. Exponential Distribution Method. Chowdhury *et al.* (2007) propose an alternative weight trimming method to the ad hoc method in Ex. 1.2. They assume that the weights in the right-tail of the weight distribution follow an exponential distribution, $Exp(\lambda)$, $f(w) = \lambda e^{-\lambda w} = \mu^{-1} e^{-w\mu^{-1}}$, $0 \leq w \leq \infty$, where $\lambda = \mu^{-1}$ and μ is the mean of the weights within the right-tail of the weight distribution. Using the $Exp(\lambda)$ cumulative distribution function, for p denoting an arbitrarily small probability, they obtain the trimming level $-\mu \log(p)$. In application to NIS data, they assume $p = 0.01$, using the cutoff 4.6μ . They also account for “influential weights” (above the median) in estimating μ by adjusting the trimming level to $\text{median}(w_i) + \log(p) \sum_{i \in S} Z_i / n$, where $Z_i = w_i - \text{median}(w_i)$ for weights exceeding the median weight and zero otherwise. They also use Fuller’s (1991) minimum MSE estimator to estimate μ to avoid extreme values influencing $\sum_{i \in S} Z_i / n$ and derive the bias/MSE for children’s vaccination rates (proportions). While they found the proposed method produced estimates with lower variance than the Ex. 1.2 method, the offset in the empirical MSE estimates was negligible. ■

Additional methods intended to bound the survey weights exist. Two general approaches to bounding weights have been proposed in the calibration literature: bounding the range of the weights or bounding their change before and after calibration. Deville and Särndal (1992) and Folsom and Singh (2000) describe a form of calibration that involves simply bounding the weighting adjustment factors. Isaki *et al.* (1992) show that quadratic programming can easily accomplish this bounding of the calibration weights themselves, rather than bounding the adjustment factors. Kott and Chang (2010) give another method of bounding the weights while calibrating and adjusting for nonresponse.

Example 10. Bounding the Range of Weights. One method proposed to bound the range of weights uses quadratic programming (Isaki *et al.* 2004). Quadratic programming seeks the vector \mathbf{k} to minimize the function

$\Phi = 0.5\mathbf{k}^T \Sigma \mathbf{k} - \mathbf{z}^T \mathbf{k}$ subject to the constraint $\mathbf{c}^T \mathbf{k} \geq c_0$, where Σ is a symmetric matrix of constants and \mathbf{z} a vector of constants. For $\mathbf{d} = (d_1, \dots, d_n)^T$ and $\mathbf{w} = (w_1, \dots, w_n)^T$ denoting the set of input and final weights, respectively, calibration weights are produced when minimizing a distance function specified for Φ subject to the constraints that the final weights reproduce population control totals but fall within specified bounds: $\sum_{i \in S} w_i \mathbf{X}_i = \mathbf{T}_x$ and $L \leq w_i \leq U$. The extent to which the constraints affect the input weights depends on which units are randomly sampled. In addition, there is no developed theory that a consistent and asymptotically unbiased variance estimator is produced when the weights are constrained using quadratic programming. ■

Example 11. Bounding the Relative Change in Weights. Another weight bounding method is to constrain the adjustment factors by which weights are changed (see Singh and Mohl 1996 for a summary). Folsom and Singh (2000) propose minimizing a constrained distance function using the generalized exponential model for poststratification. For the unit-specified upper and lower bounds L_i, U_i and centering constant C_i such that

$$L_i < C_i < U_i, \text{ the bounded adjustment factor for the weights is } a_i(\lambda) = \frac{L_i(U_i - C_i) + U_i(C_i - L_i) \exp(A_i \mathbf{X}_i^T \lambda)}{(U_i - C_i) + (C_i - L_i) \exp(A_i \mathbf{X}_i^T \lambda)},$$

where $A_i = \frac{U_i - L_i}{(U_i - C_i)(C_i - L_i)}$ can control the behavior of $a_i(\lambda)$. For example, as $L_i \rightarrow 1, C_i \rightarrow 2, U_i \rightarrow \infty$,

$a_i(\lambda) \rightarrow 1 + \exp(\mathbf{X}_i^T \lambda)$. It can be shown that the resulting estimator with $C_i = 1$ is asymptotically equivalent to the GREG estimator. This method is incorporated in SUDAAN's `proc wtadjust` (RTI 2010). ■

There are also a number of somewhat simpler methods of dealing with unusual y values by either trimming the weights and/or the y 's that we have not covered here. Hidiroglou and Srinath (1981) considered several that are appropriate for simple random samples. Lee (1995) also reviews many of the older proposals for dealing with outliers.

Model-Based Weighting Methods

This section mixes summaries of existing methods of estimation and examples of the approaches proposed in the Bayesian and the superpopulation model-based approach, including both the Best Linear Unbiased Prediction (BLUP) and robust estimators. While both approaches incorporate a model for the survey variable, each method has an associated set of case-level weights, even if only implicitly defined. After introducing the two approaches, we provide some simple BLUP-based examples and existing robust methods proposed to estimate a total.

Bayesian Methods

While Bayesian inference for finite populations is not new (e.g., Basu 1971; Ericson 1969, 1988; Ghosh and Meeden 1997; Rubin 1983, 1987; Scott 1977), Bayesian model-based approaches related to weight trimming have been recently developed. The general Bayesian inference approach first specifies a model for the population values \mathbf{Y} as a function of some unknown parameter $\boldsymbol{\theta}$, denoted $p(\mathbf{Y}|\boldsymbol{\theta})$. We denote \mathbf{X} as the matrix of covariates and \mathbf{I} as the vector of sample inclusion indicators. To make all inferences for the finite population quantities, we use the posterior predictive distribution $p(\mathbf{y}_r|\mathbf{y}_s, \mathbf{I})$, where \mathbf{y}_r are the $N-n$ non-sampled units of \mathbf{Y} and \mathbf{y}_s the n sampled values (Little 2004). The distribution $p(\mathbf{Y}|\boldsymbol{\theta})$ is combined with a prior for $\boldsymbol{\theta}$, denoted by $p(\boldsymbol{\theta})$, to produce the posterior distribution. From Bayes' theorem, the posterior predictive distribution of \mathbf{y}_r is

$$p(\mathbf{y}_r|\mathbf{y}_s) \propto \int p(\mathbf{y}_r|\mathbf{y}_s, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{y}_s) d\boldsymbol{\theta}. \quad (3)$$

where $p(\boldsymbol{\theta}|\mathbf{y}_s) = p(\boldsymbol{\theta}, \mathbf{y}_s)/p(\mathbf{y}_s) = p(\boldsymbol{\theta}) p(\mathbf{y}_s|\boldsymbol{\theta})/p(\mathbf{y}_s)$ is the posterior distribution of the model parameters, $p(\mathbf{y}_s|\boldsymbol{\theta})$ the likelihood (as a function of $\boldsymbol{\theta}$), and $p(\mathbf{y}_s)$ a normalizing constant. The distribution in (3) is used to

make all inferences about the non-sampled population values \mathbf{y}_r . To make inference to the population total $T = \sum_{i=1}^N y_i$, we use the posterior distribution $p(T|\mathbf{y}_s)$.

Elliott and Little (2000) and Lazzeroni and Little (1993; 1998) propose using the Bayesian model-based framework to pool or collapse strata when estimating finite population means under post-stratification adjustments applied within strata. They first establish that a model is assumed under various methods that pool data at either the weight trimming (related to weight pooling) or estimation (weight smoothing) stages. In both the weight pooling and weight smoothing approaches, strata are first created using the size of the weights. These strata may either be formal strata from a disproportional stratified sample design (“inclusion strata”) or “pseudo-strata” based on collapsed/pooled weights created from the selection probabilities, poststratification, and/or nonresponse adjustments. These inclusion strata are ordered by the inverse of the probability of selection, the strata above a predetermined boundary (the “cutoff” or “cutpoint stratum”) are identified, and data above the cutpoint are smoothed.

To obtain the final estimate of the finite population mean in both approaches, estimates of means are calculated for each possible smoothing scenario; the key distinction is that weights are smoothed in weight pooling while the means are smoothed in weight smoothing. The final estimate is a weighted average across the means for all possible pooling scenarios, where each mean estimate produced under a trimming scenario is “weighted” by the probability that the associated trimming scenario is “correct.” Since the probability that the trimming scenario is correct is calculated using the posterior probability of each smoothing cut point, conditional on the observed data and proposed Bayesian model, this method becomes variable-dependent. Although the Bayesian models look similar, the two approaches are different and their technical details are discussed separately.

Weight Pooling Details

After dividing the sample into “strata,” as defined above, and sorting the strata by size of weights w_h , the untrimmed (or “fully weighted”) sample-based estimate for a mean under stratified simple random sampling is

$$\bar{y}_w = \sum_{h=1}^H \sum_{i \in h} w_h y_{hi} / \sum_{h=1}^H \sum_{i \in h} w_h = \sum_{h=1}^H N_h \bar{y}_h / N, \quad (4)$$

where $w_h = N_h/n_h$. Elliott and Little (2000) show that when the weights for all units within a set of strata (separated by a “cut point,” denoted by l) are trimmed to the predetermined cutoff w_0 , estimate (4) is written as

$$\bar{y}_t = \sum_{h=1}^{l-1} \gamma N_h \bar{y}_h / N + \sum_{h=l}^H w_0 n_h \bar{y}^* / N, \quad (5)$$

where $\gamma = (N - w_0 \sum_{h=l}^H n_h) / \sum_{h=1}^{l-1} N_h$ is the amount of “excess weight” (the weight above the cutpoint) absorbed into the non-trimmed cases and $\bar{y}^* = \sum_{h=l}^H n_h \bar{y}_h / \sum_{h=l}^H n_h$. They also show that choosing $w_0 = \sum_{h=l}^H N_h / \sum_{h=l}^H n_h$, which gives $\gamma = 1$, makes the trimmed estimator (5) correspond to a model-based estimator from a model that assumes distinct stratum means (μ_h) for smaller weight strata and a common mean (μ_l) for larger weight strata:

$$y_{hi} | \mu_i \sim \overset{\text{ind}}{N}(\mu_h, \sigma^2), h < l; \quad y_{hi} | \mu_h \sim \overset{\text{ind}}{N}(\mu_l, \sigma^2), h \geq l. \quad (6)$$

$\mu_h, \mu_l \propto \text{constant}$

They extend model (6) to include a noninformative prior for the weight pooling stratum, denoted $P(L=l) = H^{-1}$, and recognize that this model is a special case of a Bayesian variable selection problem (see their references):

$\mathbf{y} | \boldsymbol{\beta}_l, l, \sigma^2 \sim \overset{\text{ind}}{N}(\mathbf{Z}_l^T \boldsymbol{\beta}_l, \Sigma)$, where \mathbf{Z}_l^T is a $n \times l$ matrix with an intercept and dummy variables for each of the first $l-1$ strata, the parameters $\mu_1 = \beta_0, \dots, \mu_l = \beta_0 + \beta_{l-1}$ in $\boldsymbol{\beta}_l$ are the model parameters associated with each smoothing scenario, and Σ is σ^2 times an identity matrix. That is, each smoothing scenario corresponds to a dummy variable parameter (1 for smoothing the weights within all strata above the cut point l under the pooling scenario, 0 for not smoothing) in \mathbf{Z}_l^T . Elliott and Little (2000) incorporate additional priors for the unknown

parameters σ^2 and β_l from the Bayesian variable selection literature and propose

$$\begin{aligned} y_{hi} | \mu_i &\stackrel{\text{ind}}{\sim} N(\mu_h, \sigma^2), h < l; \quad y_{hi} | \mu_h \stackrel{\text{ind}}{\sim} N(\mu_l, \sigma^2), h \geq l \\ P(L=l) &= H^{-1}; \quad P(\sigma^2 | L=l) \propto \sigma^{-(l+1/2)}; \quad P(\beta_l | \sigma^2, L=l) \propto (2\pi)^{-l}, \end{aligned} \quad (7)$$

where $\mu_1 = \beta_0, \dots, \mu_l = \beta_0 + \beta_{l-1}$. Since the probability that the trimming scenario is correct is calculated using the posterior probability of each cut point l , conditional on the observed data and the hierarchical Bayesian model (7), this method becomes variable-dependent. That is, pooling scenarios that are identified as the “most correct” for one variable may not be for others. In addition, the explicit form of the case weights was not derived. Assuming the prior (7) produces a posterior distribution that does not have a closed-form estimate like (6); while the model is more flexible, Elliott and Little found it can be susceptible to “over-pooling.” Elliott (2008) used Bayesian analytic methods, such as data-based priors (Bayes Factors) and pooling conterminous strata, which improved robustness of the model and reduced the over-pooling.

Weight Smoothing Details

For their weight smoothing model, Lazzeroni and Little (1993, 1998) assume that both the survey response variables and their strata means follow Normal distributions:

$$Y_{hi} | \mu_h \stackrel{\text{ind}}{\sim} N(\mu_h, \sigma^2); \quad \boldsymbol{\mu} \stackrel{\text{ind}}{\sim} N_H(\mathbf{X}^T \boldsymbol{\beta}, \mathbf{D}), \quad (8)$$

where $\boldsymbol{\mu}$ is the vector of stratum means, $\boldsymbol{\beta}$ is a vector of unknown parameters, \mathbf{D} a covariance matrix, and H the total number of strata. Under model (8), each data value is normally distributed around the true stratum mean μ_h with constant variance σ^2 . Since each μ_h is unknown, the model assumes each stratum mean follows a Normal distribution with a mean that is a linear combination of a vector of known covariates \mathbf{X} and jointly follow an H -multivariate Normal distribution. They use this model to predict post-strata means. For y_r , observations not in the sample, Y_{hi} is estimated by $\hat{\mu}_h$, the expected value of Y_{hi} given the data. The estimated finite population mean is the mean of the posterior distribution obtained assuming prior (8). It can be written as:

$$\bar{y}_{wt} = \sum_{h=1}^H [n_h \bar{y}_h + (N_h - n_h) \hat{\mu}_h] / N. \quad (9)$$

The $\hat{\mu}_h$ term in (9) is an estimate of \bar{y}_h that is smoothed toward $\mathbf{X}^T \boldsymbol{\beta}$. The (9) mean has a lower variance than the fully weighted mean (in (4), with no trimming). Also, the weights have less influence, since we borrow information from $\hat{\mu}_h$, the means that are predicted using $\mathbf{X}^T \boldsymbol{\beta}$. In large samples, estimator (9) behaves like estimator (4) (as the $\hat{\mu}_h$ term tends to \bar{y}_h), but it smoothes stratum means toward $\mathbf{X}^T \boldsymbol{\beta}$ for small sample sizes.

Lazzeroni and Little (1998) use linear and exchangeable random effects models to estimate the parameter $\boldsymbol{\beta}$. They also consider the groups (“strata”) used for establishing trimming levels (denoted by h) as being fixed. Elliott and Little (2000) extend model (9) and relax the assumption that h is fixed. They create the strata under particular fixed pooling patterns and use non-informative priors (in model (8)) for the unknown model (9) parameters. Estimates of smoothed means are calculated for each possible smoothing scenario. Like the weight pooling method, the final estimate is a weighted average across means for all possible pooling scenarios, where each mean estimate is “weighted” by the probability that the smoothing scenario is “correct.” Using their proposed prior produces a posterior distribution that does not have a closed-form estimate like (9), but the model is more flexible. Elliott (2007) extends weight smoothing models to estimate the parameters in linear and generalized linear models. Elliott (2008) extends model (7) for linear regression, allows pooling all conterminous strata (which extends model-robustness and prevents over-pooling), and uses a fractional Bayes factor prior to compare two weight smoothing models (which increases efficiency). Elliott (2009) extends this method, using Laplace approximations to draw from the posterior distribution and estimate generalized linear model parameters. Elliott and Little (2000) also show that a semi-parametric penalized spline produces estimators of means that are more robust under model misspecification related to the necessity of pooling.

Superpopulation Model Prediction Approaches

Here we describe the other model-based approach in survey inference, other than the Bayesian approach. This approach involves assuming that the population survey response variables \mathbf{Y} are a random sample from a larger (“super”) population and assigned a probability distribution $P(\mathbf{Y}|\boldsymbol{\theta})$ with parameters $\boldsymbol{\theta}$. Typically the Best Linear Unbiased Prediction (BLUP, e.g., Royall 1976) method is used to estimate the model parameters. The theoretical justification for this is described next. Here, for observation i , we assume that the population values of \mathbf{Y} follow the model

$$E_M(y_i|\mathbf{X}_i) = \mathbf{X}_i^T \boldsymbol{\beta}, \quad \text{Var}_M(y_i|\mathbf{X}_i) = \sigma^2 D_i, \quad (10)$$

where \mathbf{X}_i denotes a p -vector of benchmark auxiliary variables for unit i , which is known for all population units. A full model-based approach uses the BLUP method to estimate the parameter $\boldsymbol{\beta}$ (Royall 1976). The BLUP-based estimator of a finite population total is the sum of the observed sample units’ total plus the sum of predicted values for the non-sample units, denoted by r :

$$\hat{T}_{BLUP} = \sum_{i \in s} y_i + \sum_{i \in r} \mathbf{X}_i^T \hat{\boldsymbol{\beta}}. \quad (11)$$

Valliant *et al.* (2000) demonstrate that, when the corresponding model holds, estimator (11) is the best estimator of the total. However, when the model does not hold, model-misspecification related bias is introduced. Note that estimator (11) is also variable-specific; a separate model must be formulated for each y -variable of interest. In addition, the resulting BLUP-based weights for a particular sample unit can be negative or less than one, which is undesirable from a design-based perspective.

When x_i is a scalar, every design- and model-based estimator can be written in a form resembling (11) using the following expression (p. 26 in Valliant *et al.* 2000):

$$\hat{T} = \sum_{i \in s} y_i + \left[\frac{\hat{T} - \sum_{i \in s} y_i}{\sum_{i \in r} x_i} \right] \sum_{i \in r} x_i. \quad (12)$$

Alternatively, we can write (12) as

$$\hat{T} = \sum_{i \in s} y_i + \sum_{i \in r} \hat{y}_i = \sum_{i \in s} y_i + \sum_{i \in s} (w_i - 1) y_i, \quad (13)$$

where the component $\sum_{i \in s} (w_i - 1) y_i$ is an estimator of the term $\sum_{i \in r} y_i$. All subsequent estimators can be written in these forms. Both (12) and (13) can be used to explicitly define the associated case weights. In general, for $\mathbf{Y} = (y_1, \dots, y_N)^T$ denoting the vector of population y -values, since the total is a linear combination of \mathbf{Y} , we can write $T = \boldsymbol{\gamma}^T \mathbf{Y}$, where $\gamma_i = 1$. We partition both components into the sample and non-sample values, $\mathbf{Y} = (\mathbf{y}_s^T, \mathbf{y}_r^T)^T$ and $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_s^T, \boldsymbol{\gamma}_r^T)^T$. The population total is then $T = \boldsymbol{\gamma}_s^T \mathbf{y}_s + \boldsymbol{\gamma}_r^T \mathbf{y}_r$. If we denote the linear estimator as $\hat{T} = \mathbf{w}_s^T \mathbf{y}_s$, where $\mathbf{w}_s = (w_1, \dots, w_n)^T$ is a n -vector of coefficients, the estimator error in \hat{T} is

$$\hat{T} - T = \mathbf{w}_s^T \mathbf{y}_s - (\boldsymbol{\gamma}_s^T \mathbf{y}_s + \boldsymbol{\gamma}_r^T \mathbf{y}_r) = \mathbf{a}^T \mathbf{y}_s - \boldsymbol{\gamma}_r^T \mathbf{y}_r, \quad (14)$$

where $\mathbf{a} = \mathbf{w}_s \boldsymbol{\gamma}_s$. The term $\mathbf{a}^T \mathbf{y}_s$ in (14) is known from the sample, but $\boldsymbol{\gamma}_r^T \mathbf{y}_r$ must be predicted using the model parameters estimated from the sample and the \mathbf{x} -values in the population that are not in the sample. Thus, we

similarly partition the population covariates $\mathbf{x} = \begin{bmatrix} \mathbf{X}_s \\ \mathbf{X}_r \end{bmatrix}$, where \mathbf{X}_s is the $n \times p$ matrix and \mathbf{X}_r is $(N-n) \times p$, and

variances $\mathbf{V} = \begin{bmatrix} \mathbf{V}_{ss} & \mathbf{V}_{sr} \\ \mathbf{V}_{rs} & \mathbf{V}_{rr} \end{bmatrix}$, where \mathbf{V}_{ss} is $n \times n$, \mathbf{V}_{rr} is $(N-n) \times (N-n)$, \mathbf{V}_{sr} is $n \times (N-n)$, and $\mathbf{V}_{rs} = \mathbf{V}_{sr}^T$.

Then, under the general prediction theorem (Thm. 2.2.1 in Valliant *et al.* 2000), the optimal estimator of a total is

$$\hat{T}_{opt} = \boldsymbol{\gamma}_s^T \mathbf{y}_s + \boldsymbol{\gamma}_r^T \left(\mathbf{X}_r^T \hat{\boldsymbol{\beta}} + \mathbf{V}_{rs} \mathbf{V}_{ss}^{-1} (\mathbf{y}_s - \mathbf{X}_s^T \hat{\boldsymbol{\beta}}) \right), \quad (15)$$

where $\hat{\boldsymbol{\beta}} = (\mathbf{X}_s^T \mathbf{V}_{ss}^{-1} \mathbf{X}_s)^{-1} \mathbf{X}_s^T \mathbf{V}_{ss}^{-1} \mathbf{y}_s$. Using Lagrange multipliers, Valliant *et al.* (2000) obtain the optimal value of

\mathbf{a} as $\mathbf{a}_{opt} = \mathbf{V}_{ss}^{-1} \left[\mathbf{V}_{sr} + \mathbf{X}_s \mathbf{A}_s^{-1} \left(\mathbf{X}_r^T - \mathbf{X}_s^T \mathbf{V}_{ss}^{-1} \mathbf{V}_{sr} \right) \right] \boldsymbol{\gamma}_r$, where $\mathbf{A}_s = \mathbf{X}_s^T \mathbf{V}_{ss}^{-1} \mathbf{X}_s$. This leads to the optimal vector of BLUP coefficients $\mathbf{w}_s = \mathbf{a}_{opt} + \boldsymbol{\gamma}_s$, where the i th component is the “weight” on sample unit i . This weight depends on the regression component of the model $E_M(\mathbf{Y})$, the variance $Var_M(\mathbf{Y})$, and how the sample and non-sample units are designated. In general, the case weights for a total are $\mathbf{w}_s = \mathbf{V}_{ss}^{-1} \left[\mathbf{V}_{sr} + \mathbf{X}_s \mathbf{A}_s^{-1} \left(\mathbf{X}_r^T - \mathbf{X}_s^T \mathbf{V}_{ss}^{-1} \mathbf{V}_{sr} \right) \right] \mathbf{1}_r + \mathbf{1}_s$, where $\mathbf{1}_s, \mathbf{1}_r$ are $n \times 1$ and $(N-n) \times 1$ vectors of units with elements that are all 1’s, respectively. For the total under the general linear model with constant variance, this expression reduces to $\mathbf{w}_s = \mathbf{X}_s \left(\mathbf{X}_s^T \mathbf{X}_s \right)^{-1} \mathbf{X}_r^T \mathbf{1}_r + \mathbf{1}_s$.

In the remainder of this section, we provide examples of simple BLUP-based estimators of totals and the associated case weights, as well as more robust alternatives that have been proposed in the related literature.

Simple Model-based Weight Examples

Example 12. HT Estimator, simple random sampling. In simple random sampling, where $w_i = N/n$ and the model is $y_i | X_i = \mu + e_i$, $e_i \stackrel{\text{ind}}{\sim} (0, \sigma^2)$, $\hat{T}_{HT} = \sum_{i \in s} y_i + (N-n) \bar{y}_s$. In this case, every unit in the population but not in the sample is predicted with the same value, the sample mean $\bar{y}_s = n^{-1} \sum_{i \in s} y_i$. We also see how the HT estimator does not incorporate any auxiliary information when formulated this way, which corresponds to a very simple model. ■

Example 13. Ratio Estimator, simple random sampling. For a single auxiliary variable X_i , suppose the true model is the ratio model, $y_i | X_i = X_i \beta + e_i$, $e_i \sim (0, X_i \sigma^2)$, or a regression through the origin with a variance proportional to X_i . The optimal estimator associated with this model is the ratio estimator $\hat{T}_R = N \bar{y} \bar{X} / \bar{x}$ for \bar{y}, \bar{x} denoting the sample means and \bar{X} the population mean. The ratio estimator has the equivalent form to (13) as $\hat{T}_R = \sum_{i \in s} y_i + \sum_{i \in r} \hat{\beta} X_i = \sum_{i \in U} X_i \sum_{i \in s} y_i / \sum_{i \in s} X_i = N \bar{X}_U \bar{y}_s / \bar{X}_s$, where $\bar{X}_s = \sum_{i \in s} X_i / n$, $\bar{X}_U = \sum_{i \in U} X_i / N$, and the weights are the same for all units, i.e., $w_i = N \bar{X}_U / n \bar{X}_s$. ■

Example 14. Simple Linear Regression Estimator, simple random sampling. Here the model for the regression estimator is $y_i | X_i = \beta_0 + \beta_1 X_i + e_i$, $e_i \stackrel{\text{ind}}{\sim} (0, \sigma^2)$, $\hat{T}_{REG} = \sum_{i \in s} y_i + \hat{\beta}_1 (N \bar{X}_U - n \bar{X}_s)$, where $\hat{\beta}_1 = \sum_{i \in s} (y_i - \bar{y}_s)(X_i - \bar{X}_s) / \sum_{i \in s} (X_i - \bar{X}_s)^2$. The weights here are $w_i = \frac{N}{n} \left[1 + n (\bar{X}_U - \bar{X}_s)(X_i - \bar{X}_s) / \sum_{i \in s} (X_i - \bar{X}_s)^2 \right]$. ■

Robust Model-based Weight Examples

Since the efficiencies of the simple methods in Ex. 13-14 depend on how well the associated model holds, these methods can be susceptible to model misspecification. When comparing a set of candidate weights to a preferable set of weights, the difference in the estimated totals under the “preferable” model attributed to model misspecification is a measure of design-based inefficiency or model bias. To overcome the bias, the superpopulation literature has developed a few robust alternatives, with examples given here. Generally, each approach involves using the preferable alternative model to produce an adjustment factor that is added to the BLUP.

Example 15. Dorfman’s Kernel Regression Method. Dorfman (2000) proposed outlier-robust estimation of $\boldsymbol{\beta}$ in the BLUP estimator in (15) using kernel regression smoothing. Suppose that the true model is

$y_i|X_i = m(X_i) + v_i e_i$, where $m(X_i)$ is a smooth and (at least) twice-differentiable function. For $j \in r$, he proposed estimating $m(X_j)$ with $\hat{m}(X_j) = \sum_{i \in s} w_{ij} y_i$, where $\sum_{i \in s} w_{ij} = 1$ and larger w_{ij} 's imply that $X_i, i \in s$ is closer to $X_j, j \in r$. He proposed to use kernel regression smoothing to produce the weights $w_{ij} = K_b(X_i - X_j) / \sum_{i \in s} K_b(X_i - X_j)$, where $K(u)$ denotes a density function that is symmetric around zero, from which a family of densities is produced from using the scale transformation $K_b(u) = b^{-1}K(u/b)$, and the scale b is referred to as a "bandwidth." His nonparametric estimator for the finite population total is thus given by

$$\hat{T}_D = \sum_{i \in s} y_i + \sum_{i \in r} \hat{m}(X_i) = \sum_{i \in s} y_i + \sum_{i \in s} \sum_{j \in r} w_{ij} y_i = \sum_{i \in s} (1 + w_{i+}) y_i, \quad (16)$$

where $w_{i+} = \sum_{j \in r} w_{ij}$. Here the case weights are $w_i = 1 + w_{i+}$. While Dorfman found this estimator can produce totals with lower MSE's, it was sensitive to the choice of bandwidth. When X_i is categorical, this method is not appropriate since there is not a range of X_i over which to smooth. ■

Example 16. Chambers et al.'s NP Calibration Method. Chambers et al. (1993) proposed an alternative to Dorfman's kernel regression approach (Ex. 15) that applies a model-bias correction factor to linear regression case weights. This bias correction factor is produced using a nonparametric smoothing of the linear model residuals against frame variables known for all population units is applied to the BLUP estimator (15). Suppose that the true model is $y_i|X_i = m(X_i) + v_i e_i$, with working model variance $Var(y_i|X_i) = \sigma^2 D_i$, where D_i is a measure of size for population unit i . If the BLUP estimator was used to estimate the finite population total, then the model bias in the total is $E_M(\hat{T}_{BLUP} - T) = \sum_{i \in r} \delta(X_i)$, where $\delta(X_i) = \mathbf{x}_i^T E_M(\hat{\boldsymbol{\beta}}) - m(X_i)$. Since the residual $\hat{e}_i = y_i - \mathbf{X}_i^T \hat{\boldsymbol{\beta}}$ is an unbiased estimator of $-\delta(X_i)$, they used sample-based residuals to estimate the nonsample $\delta(X_i)$ values. This produced the *nonparametric calibration estimator* for the finite population, given by

$$\hat{T}_{C1} = \sum_{i \in s} y_i + \sum_{i \in r} \mathbf{X}_i^T \hat{\boldsymbol{\beta}} - \sum_{i \in s} \hat{e}_i = \hat{T}_{BLUP} + \sum_{i \in s} \hat{\delta}(X_i), \quad (17)$$

Here, the associated case weights are $\mathbf{w}_s = \mathbf{X}_s (\mathbf{X}_s^T \mathbf{X}_s)^{-1} \mathbf{X}_r^T \mathbf{1}_r + \mathbf{1}_s + \mathbf{m}_s$, where \mathbf{m}_s contains the residual-based estimates of $\delta(X_i)$. ■

Example 17. Chambers' Ridge Regression Method. Chambers (1996) proposed an alternative outlier-robust estimation of $\boldsymbol{\beta}$ in the BLUP estimator in (15) using a GREG-type approach. He proposes to find the sets of weights \mathbf{w} that minimize a λ -scaled, cost-ridged loss function

$$Q_\lambda(\boldsymbol{\Omega}, \mathbf{g}, \mathbf{c}) = \sum_{i \in s} \Omega_i \left((w_i - g_i)^2 / g_i \right) + \lambda^{-1} \sum_{j=1}^p c_j \left(\hat{\mathbf{T}}_{x_j \mathbf{w}} - \mathbf{T}_{x_j} \right)^2, \quad (18)$$

where Ω_i and g_i are both pre-specified terms (e.g., $\Omega_i = D_i$ and $g_i = 1$ for the BLUP), w_i is the original weight, and $\mathbf{c} = \text{diag}(c_j), j = 1, \dots, p$ is a vector of prespecified non-negative constants representing the cost of the case weighted estimator not satisfying the calibration constraint $\hat{\mathbf{T}}_{x_j \mathbf{w}} - \mathbf{T}_{x_j}$, where $\hat{\mathbf{T}}_{x_j \mathbf{w}} = \sum_{i \in s} w_i x_{ij}$, $\hat{\mathbf{T}}_{x_j} = \sum_{i \in U} x_{ij}$ is the population total of variable j , and λ is a user-specified scale function. Minimizing (18) produces the ridge-regression weights:

$$\mathbf{w}_\lambda = \mathbf{g}_s + \mathbf{A}_s^{-1} \mathbf{X}_s \left(\lambda \mathbf{c}^{-1} + \mathbf{X}_s^T \mathbf{A}_s^{-1} \mathbf{X}_s \right)^{-1} \left(\hat{\mathbf{T}}_{x_j \mathbf{w}} - \mathbf{T}_{x_j} \right), \quad (19)$$

where \mathbf{T}_x is the vector of population totals and \mathbf{A}_s is the diagonal variance matrix with i th diagonal element $\Omega_i g_i^{-1}$ (e.g., $\text{diag}(\mathbf{A}_s) = D_i$ for Chambers' BLUP). Chambers showed how $\lambda = 0$ reduces expression (19) to the calibration weights and $\lambda > 0$ produces weights that produce estimators that are biased, but have lower variance. The population total is estimated by

$$\hat{T}_{C2} = \sum_{i \in s} y_i + \sum_{i \in r} \mathbf{X}_i^T \hat{\boldsymbol{\beta}}_\lambda, \quad (20)$$

where $\hat{\boldsymbol{\beta}}_\lambda$ is the ridge-weighted estimator of $\boldsymbol{\beta}$ using the weights (19) and the linear model $E_M(\mathbf{Y}) = \mathbf{X}^T \boldsymbol{\beta}$. ■

Example 18. Chambers' NP Bias Correction Ridge Regression Method. Chambers (1996) also proposed a nonparametric approach to obtaining the ridge regression weights. His NP version of the weights (21) is

$$\mathbf{w}_{\lambda,m} = \mathbf{1}_s + \mathbf{m}_s + \mathbf{A}_s^{-1} \mathbf{X}_s \left(\lambda \mathbf{c}^{-1} + \mathbf{X}_s^T \mathbf{A}_s^{-1} \mathbf{X}_s \right)^{-1} \left(\mathbf{T}_x - \mathbf{X}_s^T \mathbf{1}_s - \mathbf{X}_s^T \mathbf{m}_s \right), \quad (22)$$

where $\mathbf{1}_s$ is a vector of 1's of length n and \mathbf{m}_s the NP-corrected weights (e.g., the kernel-smoothing weights in Ex. 2.4). The (22) weights depend on the choice of λ and choices related to how the NP weights are constructed. For example, using the kernel smoothing-based weights (Ex. 15), the weights in (22) depend on the bandwidths of the kernel smoother, which is a separate choice from determining λ . He also recommends choosing λ such that all weights $\mathbf{w}_{\lambda,m} \geq 1$. Assuming that the ridge estimator model is correct, but the BLUP model was used to estimate the total, he incorporates a bias correction factor into his ridge estimator:

$$\hat{T}_{C3} = \sum_{i \in s} y_i + \sum_{i \in r} \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_\lambda + \sum_{i \in s} m_i \left(y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_\lambda \right) = \hat{T}_{C2} + \sum_{i \in s} m_i \left(y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_\lambda \right), \quad (23)$$

where $\hat{\boldsymbol{\beta}}_\lambda$ is the ridge-weighted estimator $\boldsymbol{\beta}$ (estimated using the weights (22) and linear model $\mathbf{Y} = \mathbf{X}^T \boldsymbol{\beta}$) and $m_i \left(y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_\lambda \right)$ are the nonparametric predicted estimates obtained by summing the contributions of unit i to the NP prediction of the linear model residual, calculated for all $N - n$ units in r . The case weights applied to each y_i are given in expression (22). ■

Example 19. Firth and Bennett's Method. Firth and Bennett (1998) produce a similar bias-correction factor to Chambers *et al.* (1993, see Ex. 15) for a difference estimator (Särndal *et al.* 1992) as follows:

$$\hat{T}_D = \hat{T}_{BLUP} + \sum_{i \in s} (w_i - 1) \left(y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}} \right). \quad (24)$$

The associated weights are $\mathbf{w}_s = \mathbf{V}_{ss}^{-1} \left[\mathbf{V}_{sr} + \mathbf{X}_s \mathbf{A}_s^{-1} \left(\mathbf{X}_r^T - \mathbf{X}_s^T \mathbf{V}_{ss}^{-1} \mathbf{V}_{sr} \right) \right] \mathbf{1}_r + \mathbf{1}_s + \mathbf{m}_s$, where \mathbf{m}_s contains the residual-based estimates of $(w_i - 1) \left(y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}} \right)$, and w_i are the original BLUP weights. Firth and Bennett also define the *internal bias calibration* property to hold when $\sum_{i \in s} w_i \left(y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}} \right) = 0$, for all s under the given sample design. They also provide examples of when this property holds, e.g., using generalized linear models with a canonical link function to predict $\mathbf{x}_i^T \hat{\boldsymbol{\beta}}$ and incorporating the survey weights in the estimating equations for the model parameters or the regression model. ■

As shown in the preceding examples, the superpopulation inference approach can indirectly induce weight trimming by producing estimators that take advantage of an underlying model relationship, but they have not been directly developed specifically for this purpose. The main advantage to the model-based approach is that when the underlying superpopulation model holds, estimators of totals have a lower MSE due to a decrease in variance and no bias. Examples 15 through 19 illustrate solutions to problems of model misspecification and robustness to extreme values. However, this has not been developed for weight trimming models. When the assumed underlying model does not hold, the bias of the estimates increases and can offset the MSE gains achieved by having lower variances. It is also necessary to postulate and validate a model for each variable of interest, which leads to variable-specific estimators. This can be practically inconvenient when analyzing many variables.

Generalized Design-Based Method

A recently developed weight trimming approach uses a model to trim large weights on highly influential or outlier observations. This method was first formulated in a Bayesian framework by Sverchkov and Pfeffermann (2004); independently Beaumont and Alavi (2004) propose a similar method by extending bounded calibration (e.g., Singh and Mohl 1996) to improve the efficiency and MSE of the general regression estimator. These articles separately examine specific examples of models; the general framework and theory for estimating finite population totals was developed later by Beaumont (2008). For applications, Beaumont and Rivest (2009) use an analysis-of-variance

model for “stratum jumpers,” units that received incorrect base weights due to incorrect information at the time of sample selection. Beaumont and Rivest (2009) describe this method as a general approach for handling outliers in survey data.

Before introducing this method, we provide a general discussion on the approach and introduce the notation. Generally, within a given observed sample, we fit a model between the weights and the survey response variables. The weights predicted from the model then replace the weights and estimate the total. The hope is that using regression predictions of the weights will eliminate extreme weights. The underlying theory uses the properties of the model with respect to the weights; this is very different from conventional “model-based” approaches (i.e., the Bayesian modeling and superpopulation modeling approaches), where the properties are with respect to a model fit to the survey variable.

For the notation, denote M as the model proposed for the weights, and π the design used to select the sample. The model M trims weights by removing variability in them. This is different from the other model-based approaches, where the model describes the relationships between a survey response variable and a set of auxiliary variables. In the generalized design-based approach, only one model is fit and one set of smoothed weights is produced for all variables. Denote $\mathbf{I} = (I_1, \dots, I_N)^T$ as the vector of sample inclusion indicators, i.e., I_i is 1 if unit i is in the sample, 0 otherwise, and $\mathbf{Y} = (y_1, \dots, y_N)^T$ the values of the survey response variable y . Generalized design-based inference is defined as “any inference that is conditional on \mathbf{Y} but not \mathbf{I} ” (p. 540 in Beaumont 2008). Noninformative probability sampling is assumed, such that $p(\mathbf{I}|\mathbf{Z}, \mathbf{Y}) = p(\mathbf{I}|\mathbf{Z})$. For inferential purposes, we also consider $\mathbf{Z} = (z_1, \dots, z_N)^T$, the vector of design-variables, and $\mathbf{H}_i = \mathbf{H}_i(y_i)$, a vector of specified functions of different y -values for unit i . Beaumont (2008) makes specific inferences (i.e., taking expectations) with respect to the joint distribution of \mathbf{Z} and \mathbf{I} , conditional on \mathbf{Y} , denoted by $F_{\mathbf{Z}, \mathbf{I}|\mathbf{Y}}$. Despite confusing notation, \mathbf{I} is thus the only random quantity (not \mathbf{Z}). In order to evaluate the estimators with respect to both the sample design and the model for the weights, denote such expectations by $E_F = E_M[E_\pi(\bullet)]$ or $E_F = E_\pi[E_M(\bullet)]$, where the E_M denotes expectation with respect to the model for the weights. In the simple case of one design variable z_i and one response variable y_i , we denote the smoothed weight by $\tilde{w}_i = E_M(w_i|I_i, z_i, y_i)$.

By definition, $\hat{T}_{HT} = \sum_{i \in s} w_i y_i$, where $w_i = \pi_i^{-1}$, is the HT estimator in (1). For particular single-stage sample designs, such as probability proportional to size sampling, this weight can vary considerably due to varying selection probabilities and result in a few extreme outliers. The Beaumont (2008) estimator, proposed to reduce the variability in the w_i 's, replaces them with their expected value under the weights model: $\tilde{T}_B = E_M(\hat{T}_{HT}|\mathbf{I}, \mathbf{Y}) = \sum_{i \in s} E_M(w_i|\mathbf{I}, \mathbf{Y}) y_i = \sum_{i \in s} \tilde{w}_i y_i$. Beaumont (2008) gave two examples for the model M , the linear and exponential model. Examples for our simple one-survey variable model (Beaumont provides equivalent expressions for multiple y -variables) are given next.

Example 20. Linear model. $E_M(w_i|\mathbf{I}, \mathbf{Y}) = \mathbf{H}_i^T \boldsymbol{\beta} + v_i^{1/2} \varepsilon_i$, where \mathbf{H}_i and $v_i > 0$ are known functions of \mathbf{y}_i , the errors are $\varepsilon_i \stackrel{\text{i.i.d.}}{\sim} (0, \sigma^2)$, and $\boldsymbol{\beta}, \sigma^2$ are unknown model parameters. This model produces the smoothed weight $\hat{w}_i = \mathbf{H}_i^T \hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is the generalized LS estimate of $\boldsymbol{\beta}$. ■

Example 21. Exponential model. $E_M(w_i|\mathbf{I}, \mathbf{Y}) = 1 + \exp(\mathbf{H}_i^T \boldsymbol{\beta} + v_i^{1/2} \varepsilon_i)$, where \mathbf{H}_i , v_i , ε_i , $\boldsymbol{\beta}$, and σ^2 are given in Ex. 21. The exponential model produces the smoothed weight $\hat{w}_i = 1 + \exp(\mathbf{H}_i^T \hat{\boldsymbol{\beta}}) \sum_{i \in s} \exp(v_i^{1/2} \hat{\varepsilon}_i) / n$, where $\hat{\varepsilon}_i = [\log(w_i - 1) - \mathbf{H}_i^T \hat{\boldsymbol{\beta}}] v_i^{-1/2}$. ■

Since $\tilde{w}_i = E_M(w_i | \mathbf{I}, \mathbf{Y})$ is unknown, it is estimated with \hat{w}_i , found by fitting a model to the sample data. The estimator for the finite population total is then $\hat{T}_B = \sum_{i \in S} \hat{w}_i y_i$. If the weights model is correct, then $E_M(\hat{T}_B | \mathbf{Y}) = E_M(\sum_{i \in S} \hat{w}_i y_i | \mathbf{Y}) = \sum_{i \in S} \tilde{w}_i y_i = \tilde{T}_B$.

Beaumont (2008) demonstrates that several properties hold under the generalized design-based approach. First, the HT estimator is always unbiased across the weights model and sample designs. Also, if the model for the weights is correct, then the Beaumont estimator is also unbiased. However, the Beaumont estimator is biased when the weights model does not hold. Third, under relaxed assumptions, Beaumont (2008) also showed that his estimator is also consistent.

Model-Assisted Weighting Methods

Here two model-assisted approaches are discussed: the generalized regression estimator and a penalized spline estimator. Both methods are forms of calibration estimators. In calibration (see Ex. 4), we incorporate an underlying model for the survey and auxiliary variables and evaluate estimators with respect to their design-based properties.

Example 22. GREG estimator. As described in Example 4, specifying alternative distance functions in the calibration equations produces alternative estimators. A linear distance function produces the *general regression estimator (GREG)* $\hat{T}_{GREG} = \hat{T}_{HT} + \hat{\mathbf{B}}^T (\mathbf{T}_X - \hat{\mathbf{T}}_{XHT}) = \sum_{i \in S} g_i y_i / \pi_i$, where $\hat{\mathbf{T}}_{XHT} = \sum_{i \in S} w_i \mathbf{x}_i = \sum_{i \in S} \mathbf{x}_i / \pi_i$ is the vector of Horvitz-Thompson totals for the auxiliary variables, $\mathbf{T}_X = \sum_{i=1}^N \mathbf{x}_i$ is the corresponding vector of known totals, $\hat{\mathbf{B}}^T = \mathbf{A}_s^{-1} \mathbf{X}_s^T \mathbf{V}_{ss}^{-1} \mathbf{\Pi}_s^{-1} \mathbf{y}_s$, with $\mathbf{A}_s = \mathbf{X}_s^T \mathbf{V}_{ss}^{-1} \mathbf{\Pi}_s^{-1} \mathbf{X}_s$, \mathbf{X}_s^T is the matrix of \mathbf{x}_i values in the sample, $\mathbf{V}_{ss} = \text{diag}(v_i)$ is the diagonal of the variance matrix specified under the model, and $\mathbf{\Pi}_s = \text{diag}(\pi_i)$ is the diagonal matrix of the probabilities of selection for the sample units. In the second expression for the GREG estimator, $g_i = 1 + (\mathbf{T}_X - \hat{\mathbf{T}}_{XHT})^T \mathbf{A}_s^{-1} \mathbf{x}_i v_i^{-1}$ is called the ‘‘g-weight.’’

The GREG estimator for a total is model-unbiased under the associated working model and is approximately design-unbiased when the sample size is large (Deville and Särndal 1992). When the model is correct, the GREG estimator achieves efficiency gains; if the model is incorrect, then the efficiency gains will be dampened (or nonexistent) but the approximate design-unbiased property still holds. One disadvantage to the GREG approach is that the resulting weights can be negative or less than one. Calibration can also introduce considerable variation in the survey weights. To overcome the first problem, extensions to limit the range of calibration weights have been developed that involve either using a bounded distance function (Rao and Singh 1999; Singh and Mohl 1996; Théberge 1999) or bounding the range of the weights using an optimization method (such as quadratic programming, Isaki *et al.* 1992). Chambers (1996) proposed penalized calibration optimization function to produce non-negative weights and methods that impose additional constraints on the calibration equations. ■

Recent survey methodology research has focused on a class of estimators based on penalized (p -) spline regression to estimate finite population parameters (Zheng and Little 2003, 2005; Chen *et al.* 2009; Breidt *et al.* 2005; Krivobokova *et al.* 2008; Claeskens *et al.* 2009). Separately, Eilers and Marx (1996) introduced penalized spline estimators; Ruppert *et al.* (2003), Ruppert and Carroll (2000), and Wand (2003) developed them further theoretically. Breidt *et al.* (2005) develop a model-assisted p -spline estimator similar to the GREG estimator. In application, they showed their p -spline estimator is more efficient than parametric GREG estimators when the parametric model is misspecified, but the p -spline estimator is approximately as efficient when the parametric specification is correct. However, this method applies only for quantitative covariates.

Breidt *et al.* (2005) convert the Ruppert *et al.* (2003) model into finite population sampling by assuming that quantitative auxiliary variables x_i are available and known for all population units. The details related to Ex. 23 are

described here. They propose the following superpopulation regression model: $y_i = m(x_i) + \varepsilon_i$, $\varepsilon_i \stackrel{\text{ind}}{\sim} N(0, v(x_i))$. Treating $\{(x_i, y_i) : i \in U\}$ as one realization from this model, the p -spline function using a linear combination of truncated polynomials is $m(x, \boldsymbol{\beta}) = \beta_0 + \beta_1 x + \cdots + \beta_p x^p + \sum_{q=1}^Q \beta_{q+p} (x - \kappa_q)_+^p$, $i = 1, \dots, N$, where the constants $\kappa_1 < \dots < \kappa_L$ are fixed “knots,” and the term $(u)_+^p = u^p$ if $u > 0$ and zero, otherwise, p is the degree of the spline, and $\boldsymbol{\beta} = (\beta_0, \dots, \beta_{p+Q})^T$ is the coefficient vector. The splines here are piecewise polynomial functions that are smoothed to a certain degree, and can be expressed as a linear combination of a set of basis functions defined with respect to the number of knots. The truncated polynomial version shown is often chosen for its simplicity over other alternatives (e.g., B -splines, as used in Eilers and Marx 1996). Zheng and Little (2003) adjusted the superpopulation model to produce a p -spline estimator that accounts for the effect of non-ignorable design weights: $y_i = m(\pi_i, \boldsymbol{\beta}) + \varepsilon_i$, $\varepsilon_i \stackrel{\text{ind}}{\sim} N(0, \pi_i^{2k} \sigma^2)$, where the constant $k \geq 0$ reflects knowledge of the error variance heteroskedasticity and $m(\pi_i, \boldsymbol{\beta}) = \beta_0 + \sum_{j=1}^p \beta_j \pi_i^j + \sum_{q=1}^Q \beta_{q+p} (\pi_i - \kappa_q)_+^p$, $i = 1, \dots, N$ is the spline function.

Ruppert (2002) and Ruppert *et al.* (2003) recommend using a relatively large number of knots (15 to 30) at pre-specified locations, such that the smoothing is achieved by treating the parameters $\beta_{p+1}, \dots, \beta_{p+Q}$ as random effects centered at zero. Otherwise, using a least-squares approach to estimate $\beta_{p+1}, \dots, \beta_{p+Q}$ can result in over-fitting the model. While knot selection methods exist (Friedman and Silverman 1989; Friedman 1991; Green 1995; Stone *et al.* 1997; Denison *et al.* 1998), in *penalized (p)-spline regression* the number of knots is large, but their influence is bounded using a constraint on the Q spline coefficients. One such constraint with the truncated polynomial model is to bound $\sum_{q=1}^Q \beta_{q+p}^2$ by some constant, while leaving the polynomial coefficients β_0, \dots, β_p unconstrained. This smoothes the $\beta_{p+1}, \dots, \beta_{p+Q}$ estimates toward zero. Adding the constraint as a Lagrange multiplier, denoted by α , in the least squares equation gives $\hat{\boldsymbol{\beta}} = \arg_{\boldsymbol{\beta}} \min \sum_{i \in U} (y_i - m(\mathbf{x}_i, \boldsymbol{\beta}))^2 + \alpha \sum_{q=1}^Q \beta_{q+p}^2$ for a fixed constant $\alpha \geq 0$. The smoothing of the resulting fit depends on α ; larger values produce smoother fits.

We can also see how the proposed p -spline estimators are special forms of the model-based difference estimator (Ex. 19) by summarizing them in the following two examples.

Example 23. Zheng and Little’s Bayesian P -splines. Zheng and Little (2003) recognized that treating the $\beta_{p+1}, \dots, \beta_{p+Q}$ as random effects, given the penalty $\alpha \sum_{q=1}^Q \beta_{q+p}^2$, is equivalent to using a multivariate normal prior $\beta_{p+q}, \dots, \beta_{p+Q} \sim N_L(0, \tau^2 \mathbf{I}_Q)$, where $\tau^2 = \sigma^2 / \alpha$ is an additional parameter estimated from the data and \mathbf{I}_Q is a $Q \times Q$ identity matrix. Their estimator for the total is

$$\hat{T}_{psp} = \sum_{i \in S} y_i + \sum_{i \in r} \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{psp} = \sum_{i \in S} \left[\frac{1}{\pi_i} + \sum_{j \in U} \left(1 - \frac{I_j}{\pi_j} \right) \boldsymbol{\Pi}_j^T \mathbf{g}_s e_i \right] y_i = \sum_{i \in S} w_i^* y_i, \quad (25)$$

where the covariate matrix \mathbf{X} involved powers and knots of the probabilities of selection, w_i^* is defined by the last equality in (25), $\hat{\boldsymbol{\beta}}_{psp} = \left[\boldsymbol{\Pi}^T \mathbf{W}_s \boldsymbol{\Pi} + \hat{\mathbf{D}}_\alpha \right]^{-1} \boldsymbol{\Pi}^T \mathbf{W}_s \mathbf{y}_s = \mathbf{g}_s \mathbf{y}_s$, $I_j = 1$ if $j \in s$ and zero otherwise, $\hat{m}_i = m(\mathbf{x}_i, \hat{\boldsymbol{\beta}}_{psp})$, $i \in s$, $e_i = y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{psp}$, and fixed α . The case weights are w_i^* . ■

Example 24. Model-Assisted P -splines. For $m_i = m(\mathbf{x}_i, \boldsymbol{\beta}_U)$, $i \in U$ denoting the p -spline fit obtained from the

hypothetical population fit at \mathbf{x}_i , Breidt *et al.* (2005) incorporate m_i into survey estimation by using a difference estimator $\sum_{i \in U} m_i + \sum_{i \in S} (y_i - m_i) / \pi_i$. Given a sample, m_i here can be estimated using a sample-based estimator \hat{m}_i . For $\mathbf{W} = \text{diag}(1/\pi_i), i \in U$ and $\mathbf{W}_s = \text{diag}(1/\pi_i), i \in S$ as the matrices of the HT weights in the population and sample, for fixed α , the π -weighted estimator for the p -spline model coefficients is $\hat{\boldsymbol{\beta}}_\pi = (\mathbf{X}_s^T \mathbf{W}_s \mathbf{X}_s + \mathbf{D}_\alpha)^{-1} \mathbf{X}_s^T \mathbf{W}_s \mathbf{y}_s = \mathbf{G}_\alpha \mathbf{y}_s$, such that $\hat{m}_i = m(\mathbf{x}_i, \hat{\boldsymbol{\beta}}_\pi)$. Their model-assisted estimator is then

$$\hat{T}_{mpsp} = \sum_{i \in U} \hat{m}_i + \sum_{i \in S} \frac{y_i - \hat{m}_i}{\pi_i} \doteq \sum_{i \in S} \left[\frac{1}{\pi_i} + \sum_{j \in U} \left(1 - \frac{I_j}{\pi_j} \right) \mathbf{x}_j^T \mathbf{G}_\alpha e_i \right] y_i = \sum_{i \in S} w_i^* y_i, \quad (26)$$

where $I_j = 1$ if $j \in S$ and zero otherwise and $e_i = y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_\pi$. From (26), their p -spline estimator is a linear estimator. The case weights here are w_i^* in (26). Chambers' ridge regression estimator in (20) has a similar form, with ridge matrix $\text{diag}(\alpha_1, \dots, \alpha_p)$, where $\alpha_i = 0$ for covariates corresponding to the calibration constraints that must be met. ■

Breidt *et al.* (2005) also showed that \hat{T}_{mpsp} shares many of the desirable properties of the GREG estimator. However, since it uses a more flexible model, the p -spline estimator (26) had improved efficiency over the GREG when the linear model did not hold. However, these methods also generally work well when there are quantitative auxiliary variables, as opposed to categorical, which have to be available for all of the population units. This may not hold well for some household surveys, where there is limited information aside from population counts by various categories.

Discussion: Implications for Inference and Practice

The goal behind all of the methods that we have summarized is to somehow make inferences robust to anomalous values of weights or y 's or both. There are many variations on how this can be attempted. Among them are:

- Smooth or trim the weights;
- Smooth or trim the y 's;
- Use nonparametric estimators that are minimally affected by outlying weights, y 's, or combinations of the two.

In some cases, explicit formulas for weights are obtained; in others the smoothing must be done using iterative methods that give only implicit sets of weights. Practitioners fighting deadlines gravitate toward methods where weights are trimmed or smoothed without consideration of the analysis variables with which the weights will be used. This is pragmatic because the process of weight computation often proceeds on a parallel track from the editing of the analytic variables. However, the weight-trimming approach can be inefficient for some variables. If an outlying y_i or $w_i y_i$ product causes an estimator to have an unnecessarily large variance, weight trimming alone may not correct the problem. Plus, values of weights or y 's that are innocuous for full population estimates may be quite influential for some domain estimates. The pros and cons of the different approaches are summarized below.

As Examples 1-9 illustrate, design-based weight adjustment methods vary widely. Most are simple to understand relative to the model-based approaches and implement in practice. All methods aim to change the most extreme weight values to make the largest reduction in the variance such that the overall MSE of an estimator is reduced. However, these methods are ad hoc, data-driven, and estimator-driven, so one method that works well in a particular sample may not work in other samples. Practitioners do not accept any one method as being the standard. Redistribution also requires careful judgment by the weight trimmer (Kish 1990). The empirical MSE method is the most theoretical method – from a design-based perspective – but it is also variable-dependent. To produce one set of weights for multiple variables in practice, some ad hoc compromise – like Potter's average MSE across variables (Ex. 7) – must be used. In addition, the sample at hand may not produce very accurate estimates of the MSE or the weights' distribution function.

Alternatively, methods that incorporate realistic models will improve the estimates of totals. By incorporating the relationship between the survey variable and some known auxiliary information, estimates of totals can have lower mean square errors. When the model is correctly specified, the associated estimators are optimal (e.g., the BLUP in Valliant *et. al* 2000). However, when the model does not hold or the sample contains outliers, several robust alternative estimators have been developed.

Bayesian methods that pool or group data together have been recently proposed for weight trimming. There are two complementary approaches: “weight pooling” and “weight smoothing.” While both use models and appear similar, weight pooling is the Bayesian extension of design-based trimming, but weight smoothing is the Bayesian extension of classical random effect smoothing. In weight pooling models, cases are grouped into strata, some of which are collapsed into groups, and the group weight replaces the original weights. In weight smoothing, a model that treats the group means as random effects smoothes the survey response values. In this approach, the influence of large weights on the estimated mean and its variance is reduced under the smoothing model. In both methods, Bayesian models are used to average the means across all possible trimming points, which are obtained by varying the smoothing cut point. Both methods can produce variable-dependent weights. Neither weight pooling nor weight smoothing incorporates auxiliary variables other than indicators for stratum membership.

The Bayesian trimming procedures are a theoretical breakthrough for weight trimming. They lay the foundation for forms of estimation that adapt to the sample configuration of the weights and the analysis variables. They account for realized survey response values by making posterior distribution estimates that are conditional on the observed sample data. Their main advantage here is that, when the underlying model holds, the resulting trimmed weights produce point estimates with lower MSE due to a decrease in variance that is larger than the increase in squared bias (Little 2004). Elliott and Little (2000) and Elliott (2007, 2008, 2009) demonstrate empirically with simulations and case studies that their methods can potentially increase efficiency under repeated sampling and decrease the MSE. Other work by Zheng and Little (2003, 2005) showed that p -spline estimation can produce results that are more robust when the Bayesian model does not hold, without much loss of efficiency. That research applies only to the case of having quantitative auxiliary variables for use in estimation.

From the traditional point-of-view, a drawback to the Bayesian method is that the smoothing occurs for a particular set of circumstances: weighting adjustments performed within strata, under noninformative and equal probability sampling, and for one estimation purpose (e.g., means, regression coefficients, etc.) of a small number of (often one) variables of interest. For model-based methods, in general, it is necessary to propose and validate a model for each variable of interest, which may then lead to variable-specific sets of weights. One might view this as a strength because it makes the statistician think more deeply about what needs to be done. On the other hand, although model-based approaches may be statistically efficient, they may be practically inconvenient when there are many variables of interest. The presence of variable-dependent weights on a public use file is potentially confusing to data users, particularly when they conduct multivariate analyses on the data. However, the MSE-minimization benefit of the Bayesian method may outweigh these practical limitations for particular estimates that are of extreme importance. In some applications using quantitative auxiliaries, the simplicity and flexibility of penalized splines can improve the model robustness and may reduce the need for variable-dependent weights. This method also requires the availability of quantitative auxiliary information; the p -spline methods do not apply to categorical and binary covariates.

While the superpopulation model-based approaches (e.g., Chambers 1986, 1996; Chambers *et al.* 1993; Firth and Bennett 1998) introduce implicitly defined weights, their impact on estimation varies based on the method used. For example, the “robust” alternative methods incorporate a residual-based adjustment to improve estimates of the finite population total by reducing the bias. These methods are able to handle both categorical and quantitative auxiliaries. As with the Bayesian approach, here the limitation is the existence of auxiliary information for all population units, as well as identifying an adequate model. However, the superpopulation model-based approach has a means of expressing case weights in a standardized form, which is the first step towards developing diagnostic measures to gauge the impact of alternative weighting adjustments.

The generalized design-based method (Beaumont 2008) smoothes weights by modeling them as functions of the y 's. The weight for each unit is then replaced by its regression prediction. Although the method may be an improved weight trimming method in some applications, much of the associated theory and its effectiveness in practice need to be further studied. While the variability in estimated totals may be reduced through a reduction of variance in the

weights, this method seems easy to misapply. More extensive comparisons of the bias and mean square error of the Beaumont estimator with other alternatives are needed. In addition, this method modifies all survey weights (perhaps substantially), while the typical design-based approaches aim to make sizeable changes to only a small number of cases. The wholesale changing of all weights by the generalized design-based approach may damage some estimates for domains even if overall population estimates are improved.

Generally, all weight trimming or modification methods have the potential to “undo” the effects of previous steps in weight calculation, like base weighting, nonresponse adjustment, and calibration to external controls. Nonresponse and calibration adjustments are designed to reduce biases and/or variances. In some cases, variable weights can be more efficient and their beneficial bias/variance reductions could be needlessly removed through arbitrary trimming of large weights. Thus, there is a need for diagnostic measures of the impact of weight trimming or modification on survey inference that extend past the existing “design effect” type of summary measures, most of which do not incorporate the survey variable of interest. The current methods do not quantify such “loss of information;” i.e., there is no indication of how various methods’ distortion of the original weight distribution potentially impacts inference about full population or domain estimates.

References

- Bardsley, P., and Chambers, R.L. (1984), “Multipurpose estimation from unbalanced samples,” *Applied Statistics*, **33**, 290-299.
- Basu, D. (1971), “An essay on the logical foundations of survey sampling, Part 1 (with discussion),” in V. P. Godambe and D. A. Sprott (Eds.), *Foundations of Statistical Inference*, Toronto: Holt, Reinhart, and Winston, 203-242.
- Battaglia M.P., Izrael, D., Hoaglin D.C., and Frankel, M.R. (2004), “Tips and Tricks for Raking Survey Data (a.k.a. Sample Balancing),” *American Association for Public Opinion Research*. Available at: <http://www.amstat.org/sections/srms/proceedings/y2004/files/Jsm2004-000074.pdf>.
- Beaumont, J.P. (2008), “A new approach to weighting and inference in sample surveys,” *Biometrika*, **95** (3), 539-553.
- Beaumont, J.P., and Alavi, A. A. (2004), “Robust Generalized Regression Estimation,” *Survey Methodology*, **30** (2), 195-208.
- Beaumont, J.P., and Rivest, L. P. (2009), “Dealing with outliers in survey data,” in D. Pfeffermann and C. R. Rao (Eds.), *Handbook of Statistics, Sample Surveys: Design, Methods and Application*, **29A**, Amsterdam: Elsevier BV.
- Bethlehem, J. G. (2002), “Weighting nonresponse adjustments based on auxiliary information,” in R. M. Groves,, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse*. New York: Wiley, 275-287.
- Bethlehem, J.G., and Keller, W. J. (1987), "Linear Weighting of Sample Survey Data", *Journal of Official Statistics*, **3**, 141-153.
- Breidt, F.J., Claeskens, G., and Opsomer, J.D. (2005), “Model-assisted estimation for complex surveys using penalized splines,” *Biometrika*, **92**, 831-846.
- Breidt, F.J., and Opsomer, J.D. (2000), “Local polynomial regression estimators in survey sampling,” *The Annals of Statistics*, **28**, 1026–1053.
- Brick, M., and Montaquila, J. (2009), “Nonresponse,” in D. Pfeffermann and C. R. Rao (Eds.), *Handbook of Statistics, Sample Surveys: Design, Methods and Application*, **29A**, Amsterdam: Elsevier BV.

- Bureau of Transportation Statistics (2002), National Transportation Availability and Use Survey Public-Use File Weight Adjustments, available at:
http://www.bts.gov/programs/omnibus_surveys/targeted_survey/2002_national_transportation_availability_and_use_survey/public_use_data_file/html/weighting_and_calculating_variance_estimates.html.
- Chambers, R. L., Dorfman, A.H., and Wehrly, T.E. (1993), "Bias Robust Estimation in Finite Populations Using Nonparametric Calibration," *Journal of the American Statistical Association*, **88**, 260-269.
- Chambers, R. L. (1996), "Robust Case-Weighting for Multipurpose Establishment Surveys," *Journal of Official Statistics*, **12** (1), 3-32.
- Chen, Q., Elliott, M. R., and Little, R. J. A., (2010), "Bayesian penalized spline model-based inference for finite population proportion in unequal probability sampling," *Survey Methodology*, **36**, 23-34.
- Chowdhury, S., Khare, M., and Wolter, K. (2007), "Weight Trimming in the National Immunization Survey," *Proceedings of the Joint Statistical Meetings, Section on Survey Research Methods, American Statistical Association*.
- Claeskens, G., Krivobokova, T., and Opsomer, J.D. (2009), "Asymptotic properties of penalized spline estimators," *Biometrika*, **96**, 529-544
- Cochran, R. S. (1967), "The Estimation of Domain Sizes When Frames are Interlocking," *Proceedings of the Joint Statistical Meetings, Section on Survey Research Methods, American Statistical Association*, 332-335
- Cox, B. G., and McGrath, D. S. (1981), "An examination of the effect of sample weight truncation on the mean square error of estimates," presented at Biometrics Society ENAR meeting, Richmond VA.
- Denison, D., Mallick, B., and Smith, A. (1998), "Automatic Bayesian curve fitting," *Journal of the Royal Statistical Society, Series B*, **60**, 333-350.
- Dorfman, A.H. (2000), "Non-parametric Regression for Estimation of Totals" *Section on Survey Research Methods, American Statistical Association*, 47-54.
- Eilers, P.H.C., and Marx, B.D. (1996), Flexible smoothing using B-splines and penalized likelihood (with Comments and Rejoinder). *Statistical Science* **11** (2), 89-121.
- Elliott, M.R., and Little, R. J. A. (2000), "Model-based Alternatives to Trimming Survey Weights," *Journal of Official Statistics*, **16**, 191-209.
- Elliott, M.R. (2007), "Bayesian Weight Trimming for Generalized Linear Regression Models," *Survey Methodology*, **33**, 23-34.
- Elliott, M.R. (2008), "Model Averaging Methods for Weight Trimming," *Journal of Official Statistics*, **24**, 517-540.
- Elliott, M.R. (2009), "Model Averaging Methods for Weight Trimming in Generalized Linear Regression Models," *Journal of Official Statistics*, **25**, 1-20.
- Ericson, W.A. (1969), "Subjective Bayesian Models in Sampling Finite Populations," *Journal of the Royal Statistical Society, Series B*, **31**, 195-234.
- Ericson, W.A. (1988), "Bayesian Inference in Finite Populations," in *Handbook of Statistics*, **6**, P.R. Krishnaiah and C. R. Rao (Eds.), Amsterdam: North-Holland, 213-246.
- Firth, D. and Bennett, K.E. (1998), "Robust models in probability sampling," *Journal of the Royal Statistical Society, Series B*, **60**, 3-21.

- Friedman, J.H. (1991), "Multivariate adaptive regression splines, with discussion," *The Annals of Statistics*, **19**, 1–67.
- Friedman, J. H., and Silverman, B. W. (1989), "Flexible parsimonious smoothing and additive modeling (with discussion)," *Technometrics*, **31**, 3–21.
- Folsom, R. E. and Singh, A. C. (2000), "The Generalized Exponential Model for Sampling Weight Calibration for Extreme Values, Nonresponse, and Poststratification," *Proceedings of the American Statistical Association, Survey Research Methods Section*, 598-603.
- Fuller, W.A. (1991), "Sample Estimators for the Mean of Skewed Populations," *Statistica Sinica*, **1**, 137-158.
- Gelman, A. (2006), "Prior distributions for variance parameters in hierarchical models" *Bayesian Analysis*, **3**, 515-533.
- Ghosh, M., and Meeden, G. (1997), *Bayesian Methods for Finite Population Sampling*, London: Chapman & Hall.
- Green, P.J. (1995), "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, **82**, 711-732.
- Griffin, R. A. (1995), "Dealing with Wide Weight Variation in Polls," *Proceedings of the Social Statistics Section, American Statistical Association*, 908-911.
- Hartley, H. O. (1962), "Multiple Frame Surveys," *Proceedings of the Social Statistics Section, American Statistical Association*, 203-205.
- Hidiroglou and Srinath (1981), "Some estimators of a population total containing large units," *Journal of the American Statistical Association*, **78**, 690-695.
- Horvitz, D., and Thompson, D. (1952), "A Generalisation of Sampling without Replacement from a Finite Universe," *Journal of the American Statistical Association*, **47**, 663-685.
- Holt, D. and Smith, T.M.F. (1979), "Post Stratification," *Journal of the Royal Statistical Society A*, **142**, 33-46.
- Huang, Elizabeth T., and Fuller, Wayne A. (1978), "Nonnegative Regression Estimation For Sample Survey Data," *Proceedings of the Social Statistics Section, American Statistical Association*, 300-305.
- Isaki, C. T., and Fuller, W. A. (1982), "Survey Design Under the Regression Superpopulation," *Journal of the American Statistical Association*, **77**, 89-96.
- Isaki, C. T., Tsay, J. H., and Fuller, W. A. (2004), "Weighting sample data subject to independent controls," *Survey Methodology*, **30** (1), 35-44.
- Johnson, E.G., *et al.* (1987), "Weighting procedures, in Implementing the new design: The NAEP 1983-1984 technical report," by A. E. Beaton. Princeton, N.J.: *National Assessment of Educational Progress*, 493-504.
- Kalton, G., and Flores-Cervantes, A. (2003), "Weighting Methods," *Journal of Official Statistics*, **19** (2), 81-97.
- Kalton, G., and Kasprzyk, D. (1986), "The Treatment of Missing Survey Data," *Survey Methodology*, **12**, 1-16.
- Kish, L. (1965), *Survey Sampling*, New York: John Wiley & Sons.
- Kish, L. (1990), "Weighting: Why, When, and How?" *Proceedings of the Joint Statistical Meetings, Section on Survey Research Methods, American Statistical Association*, 121-129.
- Kish, L. (1992), "Weighting for unequal Pi," *Journal of Official Statistics*, **8**, 183-200.

- Kott, P. (2009), "Calibration weighting: combining probability samples and linear prediction models," in D. Pfeffermann and C. R. Rao (Eds.), *Handbook of Statistics, Sample Surveys: Design, Methods and Application*, **29B**, Amsterdam: Elsevier BV.
- Kott, P.S., and Chang, T. (2010), "Using calibration weighting to adjust for nonignorable unit nonresponse," *Journal of the American Statistical Association*, **105** (491):1265-1275.
- Krivobokova, T., Crainiceanu, C.M., and Kauermann, G. (2008), "Fast Adaptive Penalized Splines," *Journal of Computational and Graphical Statistics*, **17** (1), 1-20.
- Lazzeroni, L.C., and Little, R. J. A. (1993), "Models for Smoothing Post-stratification Weights," *Proceedings of the Joint Statistical Meetings, Section on Survey Research Methods*, American Statistical Association, 761-769.
- Lazzeroni, L.C., and Little, R. J. A., (1998), "Random-Effects Models for Smoothing Post-stratification Weights," *Journal of Official Statistics*, **14**, 61-78.
- Lee, H. (1995), "Outliers in Business Surveys," in *Business Survey Methods*, eds., Cox, B., Binder, D. Chinappa, B., Christianson, A., Colledge, M., and Kott, P., New York: John Wiley, 503-526.
- Little, R.J.A. (2004), "To Model or Not to Model? Competing Modes of Inference for Finite Population Sampling," *Journal of the American Statistical Association*, **39**, 546-556.
- Liu, B., Ferraro, D., Wilson, E., and Brick, M.J. (2004), "Trimming extreme weights in household surveys," *Proceedings of the Joint Statistical Meetings, Section on Survey Research Methods*, American Statistical Association, 3905-3912.
- National Center for Education Statistics, (2008), "NAPE Weighting Procedures: 2003 Weighting Procedures and Variance Estimation," available at http://nces.ed.gov/nationsreportcard/tdw/weighting/2002_2003/weighting_2003_studtrim.asp
- NHANES 2009a. Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Data. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2009, <http://www.cdc.gov/nchs/nhanes.htm>.
- NHANES 2009b. Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey 2005-2006 CPS Control Totals, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2008, http://www.cdc.gov/nchs/data/nhanes/cps_totals_0506.pdf.
- Oh, H.L., and Scheuren, F.J. (1983), "Weighting adjustments for unit nonresponse," in *Incomplete Data in Sample Surveys* Volume 2, W.G. Madow, I. Olkin, and D.B. Rubin (Eds.), Academic Press, 143-184.
- Pedlow, S., Porras, J., O.Muircheartaigh, C., and Shin, H. (2003), "Outlier Weight Adjustment in Reach 2010," *Proceedings of the Joint Statistical Meetings, Section on Survey Research Methods*, American Statistical Association, 3228-3233.
- Pfeffermann, D., and Rao, C. R. (Eds.), *Handbook of Statistics, Sample Surveys: Design, Methods and Application*, **29A-B**, Amsterdam: Elsevier BV.
- Potter, F.A. (1988), "Survey of Procedures to Control Extreme Sampling Weights," *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 453-458.
- Potter, F. A. (1990), "Study of Procedures to Identify and Trim Extreme Sample Weights," *Proceedings of the Survey Research Methods Section, American Statistical Association*, 225-230.

- Rao, J.N.K. and Singh, A.C. (1997), "A ridge-shrinkage method for range-restricted weight calibration in survey sampling," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, Washington, D.C., 57-65.
- Research Triangle Institute (2010), *SUDAAN Language Manual, Release 10.0*. Research Triangle Park, NC: Research Triangle Institute.
- Reynolds, P.D., and Curtin, R.T. (2009), *Business Creation in the United States: Initial Explorations with the PSED II Data Set*. New York: Springer.
- Royall, R. M. (1976), "The Linear Least-squares Prediction Approach to Two-stage Sampling," *Journal of the American Statistical Association*, **71**, 657–664.
- Rubin, D. B. (1983), "Inference and Missing Data," *Biometrika*, **53**, 581-592.
- Rubin, D. B. (1987), *Multiple Imputation for Nonresponse in Surveys*, New York: Wiley.
- Ruppert D., Wand, M.P. and Carroll, R., (2003), *Semiparametric Regression*, Cambridge University Press, Cambridge, New York.
- Ruppert, D. (2002), "Selecting the number of knots for penalized splines," *Journal of Computational and Graphical Statistics*, **11**, 735–737.
- Särndal, C. E., and Lundström, S. (2005), *Estimation in Surveys with Nonresponse*, New York: John Wiley and Sons.
- Särndal, C.E., Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*, Springer: Berlin, New York.
- Scott, A. J. (1977), "Large-Sample Posterior Distributions for Finite Populations," *The Annals of Mathematics*, **42**, 1113-1117.
- Singh, A.C., and Mohl, C.A. (1996), "Understanding Calibration Estimators in Survey Sampling." *Survey Methodology*, **22**, 107-115.
- Slud, E, and Thibaudeau, Y. (2010), "Simultaneous Calibration and Nonresponse Adjustment," *Research Report Series #2010-03, Statistical Research Division, U.S. Census Bureau*, available at <http://www.census.gov/srd/papers/pdf/rrs2010-03.pdf>.
- Stone, C.J., Hansen, M., Kooperberg, C., Truong, Y.K. (1997), "Polynomial splines and their tensor products in extended linear modeling" (with discussion). *Annals of Statistics*, **25**, 1371–1470.
- Sverchkov, M., and Pfeffermann, D. (2004), "Prediction of finite population totals based on the sample distribution," *Survey Methodology*, **30**, 79–92.
- Théberge, A. (1999), "Extensions of Calibration Estimators in Survey Sampling," *Journal of the American Statistical Association*, **94**, 635-644.
- Valliant, R. (2004), "The effect of multiple weight adjustments on variance estimation," *Journal of Official Statistics*, **20** (1), 1-18.
- Valliant, R. (2009), "Model-Based Prediction of Finite Population Totals," in D. Pfeffermann and C. R. Rao (Eds.), *Handbook of Statistics, Sample Surveys: Design, Methods and Application*, **29B**, Amsterdam: Elsevier BV.
- Valliant, R., Dorfman, A., and Royall, R. M. (2000), *Finite Population Sampling and Inference*, New York: Wiley & Sons.

Wand, M. P. (2003), "Smoothing and mixed models," *Computational Statistics*, **18**, 223-249.

Zheng, H. and Little, R.J.A. (2003), "Penalized spline model-based estimation of finite population total from probability-proportional-to-size samples," *Journal of Official Statistics*, **19**, 99-117.

Zheng, H., and Little, R.J.A. (2005), "Inference for the population total from probability-proportional-to-size samples based on predictions from a penalized spline nonparametric model," *Journal of Official Statistics*, **21**, 1-20.