

Nonidentifiability of Mixed-Model Parameters under Informative Sampling using only Single-Inclusion Weights

Eric V. Slud*[†]

Abstract

This paper studies the problem of consistent estimation of two-level superpopulation parameters from informatively cluster-sampled survey data, using possibly random weights. Conditions that random weights should satisfy are first formulated. Next an inclusive definition is given for informative sampling allowing for random weights. The notion of identifiability of superpopulation parameters is defined under the standard survey data structure in which (possibly random) single-inclusion weights are observed together with attribute data for sampled units. A review is given of previously proposed methods of estimation using the survey design only through the observed single-inclusion weights. Finally, a new class of informative within-cluster sampling designs is introduced and used to demonstrate exact nonidentifiability of parameters for the infinite-population-cluster biased sampling case, and asymptotic nonidentifiability for finite (large) population clusters.

Key Words: Biased Sampling, Consistency, Informative sampling, Nonidentifiability, Random weights, Superpopulation, Two-level model.

1. Introduction – Survey Sampling with Random Weights

In design-based sampling theory (Särndal et al. 1992, Lohr 2009), the finite population \mathcal{U} being sampled is viewed as an index set for a fixed but not directly observable set of multiple attributes $(Y_i, Z_i) \in \mathbb{R} \times \mathbb{R}^p$ associated with each population unit, and the probability-sampling design specifies probabilities with which the various potential samples \mathcal{S} are drawn as subsets from \mathcal{U} . The sampling mechanism is often summarized through the single-inclusion probabilities $\pi_i = P(i \in \mathcal{S})$ and weights $w_i = 1/\pi_i$. In practice, non-response usually intervenes, so that the only units i that are truly observable are those within the *respondent set* \mathcal{R} within the sample \mathcal{S} . The individual response indicators R_i are often modeled as conditionally independent given $\mathcal{D} \equiv \{(Y_l, Z_l) : l \in \mathcal{U}\}$, with probability or *propensity* of response by the i 'th unit taking the form

$$P(R_i = 1 | \mathcal{D}) = g(Y_i, Z_i) \quad (1)$$

(Pfeffermann and Sverchkov 2009, Kim & Shao 2014) depending on the underlying population data \mathcal{D} only through a function g of the i 'th unit's variables (Y_i, Z_i) . The weights applied to respondents are often manipulated into adjusted weights w_i^* through raking or calibration (Deville and Särndal 1992) in order to make estimates of explanatory-variable totals $\sum_{i \in \mathcal{S}} R_i w_i^* Z_{i,b}$ conform to known population totals $\sum_{i \in \mathcal{U}} Z_{i,b}$. Because these manipulations involve the random sample-inclusion indicators $I_{[i \in \mathcal{S}]}$, the response indicators R_i , and sometimes also values of missing explanatory-variable components that may be randomly imputed in model-based fashion, the adjusted weights w_i^* themselves become random variables, which are released together with the sampled data (generally only the respondent data) as part of the generic survey data structure.

Many authors allow random weights to enter explicitly into their survey models. In a series of contributions over many years, Pfeffermann and Sverchkov (2009) studied

*US Census Bureau, 4600 Silver Hill Road, Washington DC 20233

[†]University of Maryland, Mathematics Department

methods of estimating propensities (1) via regression-type models in survey sampling, in terms of regressors X_i (and sometimes of informative dependence on Y_i) and creating inverse-probability survey-weighted estimates from them. Use of random weights in inverse-probability-weighted estimators has a history also in econometrics (Magee 1998) and biostatistics (cf. Tsiatis 2006, Chapter 6). Other survey researchers directly incorporate random survey weights into models (Zheng and Little 1993, Beaumont 2008) as predictors in modeling, not paying too much attention to large-sample asymptotic frameworks and design- and model-consistency. The role of random survey weights in large-sample asymptotics is an understudied topic: there is much left to do beyond the theory of generalized raking initiated in Deville and Särndal (1992).

Theoretical treatments of large-sample survey inference (e.g., Krewski and Rao 1980, Rao and Scott 1983, Särndal et al. 1992, Deville and Särndal 1992, Fuller 2009) often view the finite-population survey data $(Y_i, Z_i, i \in \mathcal{U})$ as one stage of a triangular array of increasingly large super-population data consisting either of constants (as per Rao and Särndal) or of random variables (Fuller 2009) generated by a stochastic mechanism – often unknown to the statistical data analyst, but sometimes specified in the form of a parametric model — including sufficient independence structure to enable rigorous proofs of limit theorems for weighted sums of functions of unit-level data over survey respondents.

Thus, survey sampling theory even when design-based often incorporates a stochastic data-generating mechanism for both the population attributes and explanatory variables as well as the survey (single-inclusion) weights. A general formal structure simultaneously accounting for superpopulation data and random sampling that might depend stochastically on that data has been given by Rubin-Bleuer and Kratina (2006). Seen in this way, a distinguishing characteristic of the survey data-structure is the assumed observability of the (adjusted) weights w_i^* along with the data from sampled survey respondents (those $i \in \mathcal{U}$ for whom $i \in \mathcal{S}$ and $R_i = 1$).

The published weights, even if random, are generally assumed in sample survey theory to possess characteristics of design weights enabling consistent estimation of population totals, in the sense that for bounded functions $h(Y_i, Z_i)$ of respondent data, as superpopulation size $N = |\mathcal{U}|$ and sample size $n = |\mathcal{S}|$ increase to ∞ ,

$$N^{-1} \left[\sum_{i \in \mathcal{S}} R_i w_i^* h(Y_i, Z_i) - \sum_{i \in \mathcal{U}} h(Y_i, Z_i) \right] \rightarrow 0 \quad \text{in probability} \quad (2)$$

under the probabilistic mechanism or model (if there is one) for the superpopulation jointly with the probabilistic mechanism of sampling. If the Y_i, Z_i are random variables, then the natural and nearly universal way to guarantee (2) is to assume $E(w_i^* I_{[i \in \mathcal{S}]} R_i | Y_i, Z_i) = 1$ or $1/w_i^* = E(I_{[i \in \mathcal{S}]} R_i | Y_i, Z_i)$. Wherever the convergence (2) reflects consistency of estimation, that property would also be called *model and design-consistency*. In settings where the superpopulation models express independence (possibly across clusters) of identically distributed unit-level data, the quantities $N^{-1} \sum_{i \in \mathcal{U}} h(Y_i, Z_i)$ would themselves obey laws of large numbers and tend to a superpopulation expectation $h^* = E(h(Y_i, Z_i))$ not depending on i .

Under the conditional independence of R_i 's assumed in (1), with the function values (1) bounded away from 0 and some assumed independence (e.g., across clusters) of (Z_i, Y_i) , (2) would hold when $w_i^* = w_i/g(Y_i, Z_i)$. While property (2) cannot generally be proved for weights manipulated by common steps of data-imputation followed by raking or calibration, it seems universally if implicitly to be assumed in survey analysis, sometimes with adjusted weight-ratios w_i^* based on consistent estimates of the propensity function g . The assumption (2), which we impose throughout the present paper, is quite different from assumptions related to unobserved data elements being *missing at random* (MAR).

That issue is related to the dependence of the propensity $g(Y_i, Z_i)$ on attributes Y_i not observable for all $i \in \mathcal{U}$ when weights must be developed in terms of covariates Z_i whose empirical distribution over \mathcal{U} is known. We are instead concerned in this paper with general forms of biased sampling, not discernible from single-inclusion weights, that could impair identifiability of parameters even when the superpopulation model is parametric.

This paper is organized as follows. Section 2 formulates an identifiability problem for estimation of superpopulation parameters from data in a complex survey sampling design using only the observed single-inclusion weights and sampled data, under general informative sampling. The superpopulation and sampling design are restricted throughout to have a two-level cluster structure. Next, we give a brief review in Section 3 of methods that have been proposed, including methods for which model-consistent estimation has been proved to be possible based on the additional (seldom available) information of joint inclusion probabilities. Of the methods using only single-inclusion weights, the only known consistency results considerably restrict the generality of informative sampling allowed. Then, in Section 4, a class of (theoretical) informative-sampling designs is introduced for the purpose of proving exact and asymptotic nonidentifiability of superpopulation parameters from the sampled data together with single-inclusion weights. Exact nonidentifiability is proved when pairs are sampled from infinite population clusters, and an asymptotic non-identifiability result is sketched when pairs are sampled from large finite clusters whose size is at least of the same order as the overall number of clusters.

2. Non-identifiability in a Parametric Survey Sampling Context

In the previous Section, we summarized the characteristic features of a survey sampling data structure, consisting of superpopulation data tuples $(Y_i, Z_i) \in \mathbb{R} \times \mathbb{R}^p$, observable along with a system of (single-inclusion) random weights w_i^* for all population units in a sampled-respondent set $\mathcal{R} = \{i \in \mathcal{S} : R_i = 1\}$, for which the convergence (2) holds for all bounded functions h , as the size of \mathcal{R} increases to ∞ (possibly in a triangular-array framework). The data structure is that of biased sampling with observable random weights, but with otherwise unrestricted higher-order joint probabilities governing the sample selection. Our objective in this section is to make the framework and estimation problem much more concrete, restricted to a two-level population with independent and independently sampled clusters, in order to formulate a nonidentifiability result concerning the estimation of parameters in an underlying superpopulation.

2.1 Parameter Estimation in Complex Surveys

In superpopulations where outcome variables (Y_i, Z_i) are modeled as independent identically distributed (*iid*) across $i \in \mathcal{U}$, with parametric joint density $f(y_i, z_i, \theta)$, and then sampled by a complex probability design, the *census loglikelihood* $\sum_{i \in \mathcal{U}} \log f(Y_i, Z_i, \theta)$ is asymptotically approximated by the survey-weighted *pseudo-loglikelihood*

$$\sum_{i \in \mathcal{S}} w_i \log f(Y_i, Z_i, \theta) \quad (3)$$

introduced by Binder (1983), and this asymptotic approximation has been used to justify consistent and asymptotically normal inference based on the maximizer of (3) in θ . The same idea works under more general conditions on random or nonrandom (Y_i, Z_i) for (3) or for the conditional pseudo-loglikelihood $\sum_{i \in \mathcal{S}} w_i \log f(Y_i | Z_i, \theta)$. The applicability and justification of this estimation technique at population level when no model is actually assumed at unit level (and the target estimand is the maximizer of the census-loglikelihood

with respect to θ) has given rise to so-called *model-assisted* inferential procedures (Särndal et al. 1992). The same approach yields design- and model-consistent inferences for θ under general conditions including informative sampling. However, in superpopulation models with dependence across units, such as those in which only clusters are independent, this pseudo-loglikelihood approach to inference must be modified. We will consider how this may be done in two-level superpopulations with random effects shared within clusters, that are sampled from a complex cluster design.

2.2 Cluster Superpopulation, with Independent Cluster Sampling

This section establishes notation for cluster samples, allowing superpopulation data to be dependent within clusters, but independent across them, sharing a single random effect among all observations Y_i within cluster. Then we specialize to *two-level* models with a single random-effect variate shared within each superpopulation cluster, *iid* across clusters, and no explanatory variables. While it is not always true that design inclusion-probabilities are given with separate known factors for sampling of clusters and sampling within the clusters, this hierarchical sampling structure is assumed for single-inclusion probabilities throughout the present paper.

The underlying frame population index set is \mathcal{U} , with $|\mathcal{U}| = N$ elements, but its indices i are viewed as standing in one-to-one correspondence with double indices $i \leftrightarrow (j, k)$ where \mathcal{U} is partitioned into clusters \mathcal{U}_k , $k = 1, \dots, M$, with respective numbers of elements N_k , and where $k = k(i)$ is the cluster such that $i \in \mathcal{U}_k$, and j indexes units within cluster. In this way, $Y_{j,k}, Z_{j,k}, w_{j,k}$ can be written interchangeably for Y_i, Z_i, w_i . Assume further that the sampling design is a hierarchical cluster design, so that $w_{j,k} = \omega_k \cdot w_{j|k}$ for $j = 1, \dots, N_k$, $k = 1, \dots, M$, where ω_k is a *single-inclusion cluster weight* and $w_{j|k}$ is a *within-cluster single-inclusion conditional weight* with

$$\mathcal{S}_C = \{k(i) \in \{1, \dots, M\} : i \in \mathcal{S}\}, \quad \mathcal{S}_k = \{j = 1, \dots, N_k : (j, k) \in \mathcal{S}\} \quad (4)$$

$$\frac{1}{\omega_k} = P(k \in \mathcal{S}_k \mid \{Y_i, Z_i : i \in \mathcal{U}\}), \quad \frac{1}{w_{j|k}} = P((j, k) \in \mathcal{S} \mid k \in \mathcal{S}_k, \{Y_i, Z_i : i \in \mathcal{U}_k\}) \quad (5)$$

and **within-cluster sampling is assumed to be done independently across clusters**. Denote the numbers of sampled clusters, of sampled units within clusters, and of total sampled units respectively by

$$m = |\mathcal{S}_C|, \quad n_k = |\mathcal{S}_k|, \quad n = |\mathcal{S}| = \sum_{k \in \mathcal{S}_C} n_k \quad (6)$$

Then a parametric, clustered superpopulation model could take the form that for *iid* random-effect variables $a_k \sim f_C(a, \eta_2)$, $k = 1, \dots, M$, conditionally given $\underline{Z}_k \equiv (Z_{j,k} : j = 1, \dots, N_k)$,

$$\underline{Y}_k \equiv (Y_{j,k} : j = 1, \dots, N_k) \sim f_k(\underline{y} \mid \underline{Z}_k, a_k, \eta_1) \text{ independently across } k = 1, \dots, M \quad (7)$$

with $\theta = (\eta_1, \eta_2)$ as unknown parameter. The rest of the paper restricts attention to a more specific, two-level form of such a model and assumes the variables $Y_{j,k}$ are conditionally *iid* across j within cluster k , and that there are no explanatory variables $Z_{j,k}$. That is, the model takes the simplified two-level form

$$a_k \stackrel{iid}{\sim} f_C(a, \eta_2) \quad \text{and} \quad Y_{j,k} \stackrel{iid}{\sim} f(y \mid a_k, \eta_1) \quad (8)$$

where the functions f_C, f are assumed known and the parameters $\theta = (\eta_1, \eta_2)$ unknown.

2.3 Informative Sampling

Various authors have used the terminology of “informative sampling” to reflect aspects of biases in survey sampling, in ways that are not always the same. The common thread is that sampling of a population of random variables is *informative* if the probability law, or empirical distribution, of variables in the sample differs from the corresponding law or distribution in the underlying population. It is very common for probability sampling with unequal probabilities to have this property (Lohr 2009). A more helpful version of the term applies in a conditional setting, where at least some components (collectively X_i) of the variables Z_i are known or available for all members of a population before sampling, and where interesting outcomes Y_i do not have the same probability or empirical distribution given X_i , or within subpopulations defined by X_i . This is the sense in which *informative sampling* is discussed by many survey research statisticians (in lines of research summarized, e.g., by Pfeffermann and Sverchkov 2009, Kim and Shao 2014). Informative sampling is then closely related to the conditional dependence of $I_{i \in \mathcal{S}}$ or $I_{[i \in \mathcal{R}]}$ on Y_i given Z_i , and is recognizable as a form of nonignorable missing data. In the setting of a quasi-randomization model (1) with non-random base weights and $Z_i = X_i$, sampling is noninformative when the propensity function g does not depend on its Y argument.

However, in nearly all instances where noninformative sampling is discussed, it is described in terms of single-inclusion probabilities, because the great majority of inverse probability weighting strategies employ only single-inclusion probabilities. This is particularly true in treatments of biased sampling and missing data in biostatistics (following J. Robins and co-authors as cited and summarized in Tsiatis 2006), where finite-population aspects of sampling are less emphasized and individual experimental units are generally treated as independently sampled even if (through shared random effects) stochastically dependent. But in survey-sampling, also, almost all phenomena studied by survey methodologists under the heading of informative sampling could arise also if the sampling were done by independent unit-level (Poisson) sampling, or by Probability Proportional to Size (PPS) sampling with replacement. Exceptions to this are Korn and Graubard (2003), Rao et al. (2013), Yi et al. (2016), Savitsky and Williams (2019, 2020) and Slud (2020), who are all concerned with estimation of superpopulation model parameters in complex cluster-sampled surveys, and all explicitly allow sampling mechanisms that may depend on superpopulation cluster data so as to influence joint inclusion probabilities. We provide now an inclusive definition of informative sampling.

In the cluster setting of Section 2.2, sampling is called *informative* if for at least some sets $s_C = \{k_1, \dots, k_r\}$, $s_k \subset \mathcal{U}_k$, the joint densities of observed variables differ from the corresponding joint densities in the frame population given that these are the selected sets \mathcal{S}_C , \mathcal{S}_k , i.e.,

$$\begin{aligned} & \mathcal{L}\left(a_k, (Y_{j,k} : j \in s_k), k \in s_C \mid \mathcal{S}_C = s_C, \mathcal{S}_k = s_k \quad \forall k \in \mathcal{S}_C, (Z_{j,k} : j \in s_k, k \in s_C)\right) \\ & \neq \mathcal{L}\left(a_k, (Y_{j,k} : j \in s_k), k \in s_C \mid (Z_{j,k} : j \in s_k, k \in s_C)\right) \end{aligned}$$

where $\mathcal{L}(\cdot \mid \cdot)$ denotes (conditional) probability law. Here \mathcal{S}_C and \mathcal{S}_k respectively denote random sets of selected clusters and of selected units within cluster k , and s_C , s_k respectively denote particular sets of clusters and units that might have been selected. Under our assumptions of independent sampling and independent data across clusters, informativeness means that for some k , s_k ,

$$\begin{aligned} & \mathcal{L}\left(a_k, (Y_{j,k} : j \in s_k) \mid \mathcal{S}_k = s_k, (Z_{j,k} : j \in s_k), k \in S_C\right) \\ & \neq \mathcal{L}\left(a_k, (Y_{j,k} : j \in s_k) \mid Z_{j,k} : j \in s_k\right) \end{aligned}$$

In what follows, potential informativeness of sampling is considered both at the cluster and the within-cluster stages, where by definition *clusters are sampled noninformatively* if

$$\mathcal{L}\left(a_k, \underline{Y}_k \mid \underline{Z}_l, l \in S_C\right) = \mathcal{L}\left(a_k, \underline{Y}_k \mid \underline{Z}_k\right) \quad (9)$$

where recall that \underline{Y}_k denotes the vector of population attributes and \underline{Z}_k denotes the vector of observed explanatory variables, if any, for units within cluster k . Similarly, *within-cluster sampling is noninformative* if for all $k = 1, \dots, M$ and all within-cluster index sets $s_k \subset \mathcal{U}_k \equiv \{1, \dots, N_k\}$,

$$\begin{aligned} & \mathcal{L}\left((Y_{j,k} : j \in s_k) \mid a_k, (Z_{j,k} : j \in s_k), \mathcal{S}_k = s_k, k \in S_C\right) = \\ & \mathcal{L}\left((Y_{j,k} : j \in s_k) \mid a_k, (Z_{j,k} : j \in s_k), k \in S_C\right) \end{aligned} \quad (10)$$

As we will see below, in the construction of Section 4, informative cluster sampling in the sense just described may not manifest itself at all through single-inclusion probabilities or weights. This observation is important because of the prominence of methods published in the survey literature that estimate mixed-effect parameters in clustered superpopulations using only single-inclusion weights and no other aspects of informative sampling. In the next Section, we give an overview of such methods – for which consistency of estimation cannot generally be proved – by contrast with several recent papers that do justify consistent estimation in general two-level informatively sampled surveys by taking account of joint (second-order) inclusion probabilities.

2.4 Nonidentifiability

Identifiability is a basic concept of mathematical statistics, a minimal requirement for the statistician to know that parameters can in principle be recovered from large enough data collections. A parameter $\theta \in \Theta$ is called *identifiable* from data \mathbf{X} governed by a model, say one with joint density $f(\mathbf{x}, \theta)$, if the mapping $\theta \mapsto f(\mathbf{x}, \theta)$ from the parameter space to the space of possible density functions (of \mathbf{x}) is one-to-one. When this is not true – a condition we refer to below as *exact nonidentifiability* – at least two different parameter values θ_0, θ_1 would give rise under the model to data with exactly the same probabilistic behavior. In that case, the statistician’s desire to learn from data about the parameter of the data-generating model has no unique solution. A slightly different type of nonidentifiability arises when a family of statistical problems with data $\mathbf{X}_n \sim f_n(\mathbf{x}, \theta)$ has the same parameter regardless of sample size n , but there are pairs of parameters θ_0, θ_1 for which the likelihood ratios $f_n(\mathbf{X}_n, \theta_1)/f_n(\mathbf{X}_n, \theta_0)$ remain *stochastically bounded* or $O_P(1)$ as $n \rightarrow \infty$ for the θ_1 model, which means that

$$\text{for all } \epsilon > 0, \exists C_\epsilon < \infty : \text{for all } n, \int I_{[f_n(\mathbf{x}, \theta_0) > C_\epsilon f_n(\mathbf{x}, \theta_1)]} f_n(\mathbf{x}, \theta_1) d\mathbf{x} \leq \epsilon \quad (11)$$

When (11) holds, a standard argument using the Neyman-Pearson Lemma proves that no size- α hypothesis test with $\alpha < 1$ of $H_0 : \theta = \theta_0$ versus alternative $H_1 : \theta = \theta_1$ can have power converging to 1 as $n \rightarrow \infty$. We refer to this second type of nonidentifiability as *approximate or asymptotic nonidentifiability*. Under either exact or asymptotic nonidentifiability, there can be no consistent estimator $\hat{\theta}_n$ of $\theta = \theta_1$ as $n \rightarrow \infty$.

3. Proposed Methods of Inference in Two-Level Complex Surveys

3.1 Methods using only Single-inclusion Weights

By 1990, it was known that superpopulation random effects shared throughout clusters induce unit-level dependence making pseudolikelihood methods inapplicable. Various authors proposed new methods of inference for complex surveys with clustered designs like those introduced in Section 2.2. They began by searching for an appropriate design- and model-consistent methodology for parameter estimation in two-level cluster samples, and they proposed methods intended to apply also to samples that are informative within cluster. Pfeffermann et al. (1998) devised a complicated iterative weighted least squares procedure involving weight-rescaling, to apply to cluster-sampled linear-model data

$$Y_{j,k} = \beta' Z_{j,k} + a_k + \epsilon_{j,k}, \quad a_k \stackrel{iid}{\sim} \mathcal{N}(0, \eta_1), \quad \epsilon_{j,k} \stackrel{iid}{\sim} \mathcal{N}(0, \eta_2) \quad (12)$$

They provided algorithms and simulations supporting their method, but no proofs. Korn and Graubard (2003) restricted attention to the case without covariates, and produced simulations that showed the Pfeffermann et al. (1998) methods were not model- and design-consistent under several sorts of highly informative sampling designs, but might be in designs where sampling within clusters was noninformative.

To estimate $\theta = (\eta_1, \eta_2)$, Rabe-Hesketh and Skrondal (2006) and Asparouhov (2006) independently proposed to maximize an approximate log-likelihood

$$\log\text{Lik}_1 = \sum_{k \in \mathcal{S}_C} \omega_k \log \int \exp \left(\sum_{j \in \mathcal{S}_k} w_{j|k} \log f(Y_{jk} | \mathbf{Z}_{jk}, a_k, \eta_1) \right) f_C(a_k, \eta_2) da_k \quad (13)$$

But the integral expression within (13) is not a likelihood, and consistency of estimation can be justified generally only when the within-cluster sample-sizes go to ∞ . In similar vein, Savitsky and Williams (2019, 2020) developed a Bayesian ‘pseudo-posterior’ methodology applicable when clusters of bounded size are sampled approximately independently. Their posteriors are formed with the approximate loglikelihood

$$\log\text{Lik}_2 = \sum_{k \in \mathcal{S}_C} \log \int \exp \left(\sum_{j \in \mathcal{S}_k} w_{j,k} \log f(Y_{jk} | \mathbf{Z}_{jk}, a_k, \eta_1) + \omega_k \log f_C(a_k, \eta_2) \right) da_k \quad (14)$$

With some technical restrictions, they provide Bayesian large-sample theory, claiming that their method exhibits large-sample posterior concentration (a Bayesian analogue of consistency) under general informative sampling. Although their method uses only single-inclusion weights, their theoretical arguments require assumptions that considerably restrict the generality of informative sampling they consider. However, their simulation results show their methods perform well (in a way not justified by their theory) under some informative sampling designs extending those of Korn and Graubard (2003).

A simulation study in Slud (2020) compared frequentist pseudo-maximum likelihood survey estimates of mixed-model parameters η_1, η_2 in the setting of (12) without covariates, under various designs sampling clusters informatively, but with noninformative sampling within clusters. The methods studied were only those using single-inclusion and possibly random weights, including those of Rabe-Hesketh and Skrondal 2006, frequentist analogs of those of Savitsky and Williams 2020, as well as a new EM method Slud (2020) proposed based on (13) that explicitly used the design property that sampling is noninformative within clusters. Slud’s (2020) simulations and heuristic arguments showed that none of the available estimation methods were generally design- and model-consistent under noninformative-within-cluster designs except his own EM method.

3.2 Consistent Estimation in Two-Level Complex Surveys

A different line of research developed the idea that consistent estimation of two-level mixed-model parameters was possible from informatively cluster-sampled survey data, in models like (8) or (12) with cluster-level random effects, based on known joint inclusion probabilities. Korn and Graubard (2003) presented weighted method of moments estimates accomplishing this. Under general sampling with constant known single and joint inclusion probabilities, Rao et al. (2013) extended the idea of inverse joint-inclusion-weighted moment estimates, using different moment-based estimating equations in the linear-model setting, and to much more general two-level models by appeal to well-developed theory for Lindsay's (1988) Composite Likelihoods. Yi et al. (2016) provided still further extensions to prove model- and design consistent estimation of parameters in two-level survey data, using weighted composite likelihoods. However, between Korn and Graubard (2003) and these later papers based on composite likelihoods, something was lost: namely the applicability to random weights. Korn and Graubard explicitly allowed random weights, using them in criticizing Pfeffermann et al. 1998. But the Composite Likelihood literature, and the sampling papers of Rao et al. (2013) and Yi et al. (2016), did not. Rao et al. (2013) speak of weights obtained from inverse single- and joint-inclusion probabilities that are known from the design. However, their 'unified weighted log composite likelihood approach' expressed in equations (4.2) and (4.3) of section 4 of their paper could also directly use random weights. Under nonrestrictive assumptions, unbiased estimating equation theory would also lead to consistent weighted survey estimates, as long as the joint within-cluster inclusion weights satisfy $E(w_{jj'|k} I_{[j,j' \in S_k]} | Y_{jk}, Y_{j'k}) = 1$. Similar comments could be made about the weighted survey estimates of Yi et al. (2016), although both of these papers would have to modify their variance estimation techniques if the single and joint inclusion weights were random.

The consequence of this line of research for the present paper is fairly straightforward. Within the setting assumed above of possibly informative within-cluster sampling, consistent estimation of model parameters is generally feasible when random (conditional) joint inclusion probabilities are known and satisfy $E(I_{[j,j' \in S_k]} / \pi_{jj'|k} | Y_{jk}, Y_{j'k}) = 1$. Such an assumption covers all of the informative within-cluster designs contemplated in this paper. The issue addressed in the rest of the paper is that single-inclusion weights alone do not provide sufficient information about the design, in general, to ensure consistent estimation of two-level model parameters.

4. A New Construction of an Informative Sampling Design

Throughout the present section, we maintain the assumptions introduced in Section 2.2 of a two-level superpopulation model (8), with a sampling design such that the single-inclusion weights ω_k are independent across clusters and may depend on the cluster-level random effects a_k . The purpose of the Section is to present a family of within-cluster probability sampling designs for $P(S_k = s_k | a_k, \{Y_{j,k}\}_{j=1}^{N_k})$ within which the parameters η_1, η_2 may be non-identifiable, either exactly or asymptotically, as defined above.

For each cluster index $k = 1, \dots, M$, define within-cluster *ranks* $r_{j,k}$ (with ties broken lexicographically by increasing j -index) and *order-statistics* $Y_{(j),k}$ within $\{Y_{1,k}, \dots, Y_{N_k,k}\}$ uniquely by

$$\{r_{j,k}\}_{j=1}^{N_k} = \text{lexicographical ranks of } \{(Y_{j,k}, j)\}_{j=1}^{N_k} = \sum_{t=1}^{N_k} I_{[(Y_{t,k}, t) \leq (Y_{j,k}, j)]} \quad (15)$$

and

$$Y_{(1),k} \leq Y_{(2),k} \leq \cdots \leq Y_{(N_k),k} \quad \text{with} \quad Y_{j,k} \equiv Y_{(r_{j,k}),k} \quad (16)$$

Assume that $Q = Q^{(k)} = (Q_{b,b'}^{(k)})$ is an $N_k \times N_k$ doubly stochastic matrix (with entries ≥ 0 and row- and column-sums equal to 1), with 0's on the diagonal. The key idea of the sampling design introduced below is the definition for cluster sample-size $n_k = 2$, which is then generalized to $n_k > 2$. The construction formalizes the idea that the selected pair of order-statistic indices (b_1, b_2) , among the values $\{Y_{j,k}\}_j$, will have b_1 uniform in $\{1, \dots, N_k\}$ and b_2 given b_1 distributed with mass function equal to the b_1 row of $Q^{(k)}$.

Definition 1 A **Q-sampling design** of a two-level superpopulation with notations as in Section 2.2 is one where

(a) sampling of clusters may depend on cluster variates a_k , is independent across k , and random single-inclusion weights ω_k satisfy $\omega_k = 1/P(k \in \mathcal{S}_C | a_k)$, and

(b) in any cluster k where $n_k = 1$, $\mathcal{S}_k \equiv \{j_1\}$ where j_1 is drawn equiprobably from $\{1, \dots, N_k\}$,

(c) within cluster k , if $n_k = 2$, the sample $\mathcal{S}_k = \{j_1, j_2\}$ drawn from cluster k is defined in the following three steps:

(i) j_1 is drawn equiprobably from $\{1, \dots, N_k\}$, where $Y_{j_1,k} = Y_{(b_1),k}$ with $b_1 = r_{j_1,k}$ by (16),

(ii) $b_2 \in \{1, \dots, N_k\}$ is drawn conditionally given b_1 from the discrete probability distribution given by the b_1 row of $Q^{(k)}$, and

(iii) j_2 is defined using (15)-(16) by $Y_{j_2,k} \equiv Y_{(b_2),k}$, or equivalently $b_2 = r_{j_2,k}$.

(d) within cluster k , if $n_k > 2$, the sample $\mathcal{S}_k = \{j_1, j_2, \dots, n_k\}$ drawn from cluster k is defined in the following three steps:

(i) j_1 and then b_1 are defined as in step (i) of (c),

(ii) conditionally given j_1 and b_1 , order-statistic indices b_2, \dots, b_{n_k} are drawn from $\{1, \dots, N_k\}$ by any Probability Proportional to Size without Replacement (PPSWOR) method (such as the Hájek Rejective Method, Tillé 2003), using size vector equal to the b_1 row of $Q^{(k)}$; and

(iii) j_2, \dots, j_{n_k} are defined from b_2, \dots, b_{n_k} by the identities $r_{j_i,k} \equiv b_i$, $2 \leq i \leq n_k$.

Lemma 1 For doubly stochastic $N_k \times N_k$ matrices $Q = Q^{(k)}$ with 0's on the diagonal, the within-cluster sampling scheme in Definition 1 results in marginal single-inclusion probability $1/w_{j|k}$ of selecting $j \in \mathcal{U}_k$ equal to n_k/N_k .

Proof. For clusters k where $n_k = 1$, there is nothing to prove. In cases where $n_k \geq 2$, begin by noting that for each fixed superpopulation cluster $\{Y_{j,k} : j \in \mathcal{U}_k\}$, the definition (15) establishes a bijective mapping between within-cluster indices j and their corresponding within-cluster ranks $r_{j,k}$. When $n_k = 2$, the construction results in a selected pair of random order-statistic indices (b_1, b_2) such that b_1 is uniform and $b_2 \neq b_1$ because all diagonal elements of Q are 0. Moreover,

$$P(b_2 = b') = \sum_{b=1}^{N_k} P(b_1 = b) \cdot P(b_2 = b' | b_1 = b) = N_k^{-1} \sum_{b=1}^{N_k} Q_{b,b'} = 1/N_k$$

by the doubly-stochastic property of Q , and it follows immediately that (conditionally given the superpopulation, and therefore also unconditionally) $P(j \in \mathcal{S}_k) = P(b_1 = r_{j,k}) + P(b_2 = r_{j,k}) = 2/N_k$, so that $w_{j|k} \equiv N_k/n_k$ for all j . Finally, when $n_k > 2$, by assumption in Definition 1(d), conditionally given the \mathcal{U}_k data,

$$P(r_{j,k} \in \{b_2, \dots, b_{n_k}\}) = (n_k - 1)N_k^{-1} \sum_{b=1}^{N_k} Q_{b,r_{j,k}} = (n_k - 1)/N_k$$

and therefore $P(j \in \{j_1, \dots, j_{n_k}\}) = 1/N_k + (n_k - 1)/N_k = n_k/N_k$ and $w_{j|k} = N_k/n_k$. \square

Example 1 A Q -sampling design that might reflect the purposefully biased selection of clusters of size $n_k = 2$ from population clusters of size $N_k = 4$ (such as 4-person households) is the following. The matrix Q could be specified as

$$Q = Q(p) = \begin{pmatrix} 0 & p & 1-p & 0 \\ p & 0 & 0 & 1-p \\ 1-p & 0 & 0 & p \\ 0 & 1-p & p & 0 \end{pmatrix}$$

with any $p \in [0, 1]$. When p is close to 0, the sampled pair within a 4-unit population cluster is highly likely to consist of the 1st and 3rd or 2nd and 4th order-statistics among the $Y_{j,k}$ values, while p near 1 makes it highly that the sampled pair consists of the 1st and 2nd or 3rd and 4th order statistics. In the latter case ($p \approx 1$), the sampled pair will tend to be closer together, and share the same sign more often, than a SRS-sampled pair. \square

The Definition and Lemma so far are restricted to the case where $n_k < N_k < \infty$, but both have natural extensions to the case where $N_k = \infty$. In either case, samples of size n_k are drawn in potentially biased and dependent fashion. For simplicity, we describe the extension to $N_k = \infty$ only in the case where $n_k = 2$ and the conditional distribution of $Y_{j,k}$ given a_k in (8) is continuous. The key point of this extension is that instead of selecting from order-statistic indices b_1, b_2 , we select random pairs of quantiles u_1, u_2 . These quantiles are marginally uniform, so that their continuous joint density $c(u_1, u_2)$ is termed a *copula*.

Definition 2 In a cluster-sampling setting satisfying (8) and all the other assumptions of Definition 1(a) except that now $N_k = \infty$ and $n_k = 2$ in some or all clusters, assume that the conditional distribution function $F(y | a_k, \eta_2)$ for $Y_{j,k}$ given a_k is continuous. A Q -sampling design selecting a pair of marginally $f(\cdot | a_k, \eta_2)$ distributed variates $Y_{j,k}$ has the form:

(i) u_1, u_2 are marginally uniform variables with joint density given by a specified copula $c(u_1, u_2)$;

(ii) the selected pair of cluster k variates are defined as $F^{-1}(u_t | a_k, \eta_2)$, $t = 1, 2$.

In this case, the marginal uniformity of selected quantiles u_t implies that univariate selection from the infinite cluster is *not* biased, although the joint selection of pairs is biased. The copula property is analogous to the doubly stochastic property of matrices $Q^{(k)}$ used in Definition 1.

4.1 Exact and Near Nonidentifiability in a Two-Level Survey Model

The main goal of this Section and this paper is to show that cluster-level biased selection of $n_k \geq 2$ parametrically distributed superpopulation variates, restricted only by the marginal selection mechanism, may render the superpopulation parameters nonidentifiable. We recalled the general meaning of nonidentifiability, both in an exact and asymptotic sense, in Section 2.4 above. To obtain clear examples of nonidentifiability in the complex survey setting, we restrict from now on to two-level clustered superpopulations and sampling designs where the mechanism of sampling of clusters is Simple Random Sampling (SRS) of m clusters, so that the single-inclusion cluster weights are all $\omega_k \equiv M/m$. All of the within-cluster sampling designs we consider will also have the property that the single-inclusion within-cluster sampling weights are $w_{j|k} \equiv N_k/n_k$ in clusters k with $N_k < \infty$, and in clusters with $N_k = \infty$, the selected attributes $Y_{j_1,k}, \dots, Y_{j_{n_k},k}$ are all marginally distributed with density $f(\cdot | a_k, \eta_2)$ in (8). The immediate point of these restrictions is to ensure that the exactly observable single-inclusion weights of the sampled observations convey no statistical information at all. Thus the nonidentifiability definitions are applied to the probability laws of the observed sampled data $\{Y_{j,k} : (j, k) \in \mathcal{S}\}$.

Now specialize further to the one-way ANOVA case,

$$f_C(a_k, \eta_1) \sim \mathcal{N}(0, \eta_1) \quad , \quad f(y|a, \eta_2) \sim \mathcal{N}(a, \eta_2) \quad (17)$$

For fixed $\rho \in (-1, 1)$, let $c(u_1, u_2, \rho)$ denote the bivariate normal copula density (with 0 means and variances 1) with correlation ρ ,

$$\begin{aligned} c(u_1, u_2, \rho) &\equiv \frac{\{\phi(\Phi^{-1}(u_1)) \phi(\Phi^{-1}(u_2))\}^{-1}}{2\pi \sqrt{1-\rho^2}} \cdot \\ &\cdot \exp\left(-\frac{(\Phi^{-1}(u_1))^2 + (\Phi^{-1}(u_2))^2 - 2\rho \Phi^{-1}(u_1)\Phi^{-1}(u_2)}{2(1-\rho^2)}\right) \\ &= \frac{1}{\sqrt{1-\rho^2}} \exp\left(\frac{2\rho \Phi^{-1}(u_1)\Phi^{-1}(u_2) - \rho^2\{(\Phi^{-1}(u_1))^2 + (\Phi^{-1}(u_2))^2\}}{2(1-\rho^2)}\right) \end{aligned} \quad (18)$$

With this definition, the quantile $u \in (0, 1)$ of the population-cluster outcome variables $\{Y_{j,k}\}_{j=1}^\infty = \{a_k + \epsilon_{j,k}\}_{j=1}^\infty$ is equal to $a_k + \sqrt{\eta_2} \Phi^{-1}(u)$, since the $\epsilon_{j,k}$ variates are all $\mathcal{N}(0, \eta_2)$ distributed and independent of a_k . Now the random quantiles (u_1, u_2) generated (independently of a_k and $\{Y_{j,k}\}_{j=1}^\infty$) according to Definition 2 have the property that $(\Phi^{-1}(u_1), \Phi^{-1}(u_2))$ are jointly bivariate-normal distributed with means 0, variances 1, and correlation ρ , or equivalently, can be represented as

$$(\Phi^{-1}(u_1), \Phi^{-1}(u_2)) = \left(\rho Z_0 + \sqrt{1-\rho^2} \cdot Z_1, \rho Z_0 + \sqrt{1-\rho^2} \cdot Z_2\right)$$

where Z_0, Z_1, Z_2 are $\mathcal{N}(0, 1)$ random variables independent of each other and of $(a_k, \{Y_{j,k}\}_{j=1}^\infty)$. Thus the sampled quantiles u_1, u_2 among $\{Y_{j,k}\}_{j=1}^\infty$ are

$$\begin{aligned} (Y_{j_1,k}, Y_{j_2,k}) &\equiv (a_k + \sqrt{\eta_2} \Phi^{-1}(u_1), a_k + \sqrt{\eta_2} \Phi^{-1}(u_2)) \\ &= \left(a_k + \sqrt{\eta_2} \{Z_0 \rho + Z_1 \sqrt{1-\rho^2}\}, a_k + \sqrt{\eta_2} \{Z_0 \rho + Z_2 \sqrt{1-\rho^2}\}\right) \end{aligned}$$

and it follows immediately that $a_k + \rho \sqrt{\eta_2} Z_0 \sim \mathcal{N}(0, \eta_1 + \rho^2 \eta_2)$, and the joint distribution of $(Y_{j_1,k}, Y_{j_2,k})$ is bivariate-normal with means 0, variances $\eta_1 + \rho^2 \eta_2 + (1-\rho^2)\eta_2 = \eta_1 + \eta_2$, and correlation $\rho^* \equiv (\eta_1 + \rho^2 \eta_2)/(\eta_1 + \eta_2)$. With these steps, we have proved

Proposition 1 *Under superpopulation assumptions (8) and (17), with $n_k = 2$ units sampled within infinite population cluster k according to Definition 2 using the bivariate-normal copula density $c(u_1, u_2) = c(u_1, u_2, \rho)$ defined in (18), the joint distribution of the sampled cluster elements $(Y_{j_1,k}, Y_{j_2,k})$ is bivariate-normal with means 0, variances $\eta_1 + \eta_2$, and covariance $\eta_1 + \rho^2\eta_2$. This implies that the distribution of sampled $(Y_{j_1,k}, Y_{j_2,k})$ is exactly the same when the underlying parameters η_1, η_2 are σ_1^2, σ_2^2 under a sampling design with iid sampling within cluster ($\rho = 0$) as when $(\eta_1, \eta_2) = (\sigma_1^2 + \rho^2\sigma_2^2, (1 - \rho^2)\sigma_2^2)$ under the sampling scheme of Definition 2 using copula (18) with fixed $\rho \in (-1, 1)$.*

Proposition 1 is an exact nonidentifiability result, in the context of biased sampling from an infinite population rather than a finite cluster such as would arise in sample surveys. We provide next an example motivated by the infinite-cluster construction of Definition 2 under (17)-(18) but applicable to the case of large finite clusters using the design of Definition 1.

Fix a population cluster k of finite size N_k , and let (u_1, u_2) as above be a pair of random variables with joint density given by the copula density (18). Define the $N_k \times N_k$ matrix $\bar{Q}^{(k)}$ by its entries

$$\bar{Q}_{t_1, t_2}^{(k)} \equiv N_k \cdot P\left(u_1 \in \left(\frac{t_1 - 1}{N_k}, \frac{t_1}{N_k}\right], u_2 \in \left(\frac{t_2 - 1}{N_k}, \frac{t_2}{N_k}\right] \mid \lceil N_k u_1 \rceil \neq \lceil N_k u_2 \rceil\right) \quad (19)$$

where $t_1, t_2 \in \{1, \dots, N_k\}$, and $\lceil x \rceil$ denotes the smallest integer $\geq x$. It is clear from the definition (19) that $\bar{Q}^{(k)}$ is doubly stochastic and symmetric, with 0's on its diagonal.

Our next goal is to show that, when the random pair $\mathbf{Y}_k^s \equiv (Y_{j_1,k}, Y_{j_2,k})$ is selected within the population cluster k according to the sampling plan of Definition 1, the logarithm of the joint density of \mathbf{Y}_k^s is close to the log of the limiting joint density as N_k gets large. (Note that, although the random index pair (j_1, j_2) depends on k , we suppress that dependence in the notation.) By definition, the joint density of \mathbf{Y}_k^s is the mixture with weights $N_k^{-1} \bar{Q}_{t_1, t_2}^{(k)}$ of the joint densities of the order statistics $(Y_{(t_1),k}, Y_{(t_2),k})$, over all $1 \leq t_1, t_2 \leq N_k$. Another way of expressing this is that the joint density of \mathbf{Y}_k^s is the same as the joint density of $(a_k + \eta_2^{1/2} z_{(\lceil N_k u_1 \rceil)}, a_k + \eta_2^{1/2} z_{(\lceil N_k u_2 \rceil)})$, where $z_{(i)}$ are the order-statistics for a set of iid $\mathcal{N}(0, 1)$ variables $\{z_i\}_{i=1}^{N_k}$, conditionally given that $\lceil N_k u_1 \rceil \neq \lceil N_k u_2 \rceil$. As $N_k \rightarrow \infty$, it is easy to see that the density of the distributional limit $\mathbf{Y}_* \equiv (a_k + \eta_2^{1/2} \Phi^{-1}(u_1), a_k + \eta_2^{1/2} \Phi^{-1}(u_2))$ of \mathbf{Y}_k^s is the bivariate-normal found in Prop. 1, with means 0, variances $\eta_1 + \eta_2$, and correlation $(\eta_1 + \rho\eta_2)/(\eta_1 + \eta_2)$. An important tool for quantifying the closeness in probability of the random vectors \mathbf{Y}_k^s and \mathbf{Y}_* is the Bahadur-Kiefer Representation of sample quantiles (Encycl. of Math. 2020), that says in particular (e.g., as a corollary of the more sophisticated result given in formula (1.27) of Deheuvels and Mason 1990) that

$$\max_{i \leq N_k} |\Phi(z_{(i)}) - i/N_k| = O_P(N_k^{-1/2} \log N_k) \quad \text{as } N_k \rightarrow \infty$$

Without supplying further details of proof, which will be given elsewhere, we state next a Proposition expressing the mathematical property of survey-sampling loglikelihood ratios that implies, by way of the general criterion (11), the asymptotic nonidentifiability of the parameters η_1, η_2, ρ under the model (17).

Proposition 2 Assume (8) and (17), with random sampling of $n_k = 2$ elements done within large finite population clusters $k = 1, \dots, M$ according to Definition 1 with the matrices $Q^{(k)} \equiv \bar{Q}^{(k)}$ given by (19). Assume that m clusters k are sampled SRS. Then the first-order inclusion weights are all $\omega_k = M/m$, $w_{j|k} = N_k/2$. Assume further that the cluster-sizes N_k are such that

$$\liminf_{m, N \rightarrow \infty} \min_k N_k/m > 0 \quad \text{as} \quad m, N \rightarrow \infty$$

Then the logarithm of the likelihood ratio for the data $\{(Y_{j_1, k}, Y_{j_2, k})\}_{k \in \mathcal{S}_G}$ based on the models with parameters (η_1, η_2, ρ) respectively equal to $(\sigma_1^2, \sigma_2^2, \rho)$ and to $(\sigma_1^2 + \rho^2 \sigma_2^2, \sigma_2^2(1 - \rho^2), 0)$, is of order $O_P(1)$ for both models, as $m, N \rightarrow \infty$.

Propositions 1 and 2 complete our formal justifications that informative within-cluster sampling in general prevents the possibility of model and design consistent estimation of mixed-model superpopulation parameters from survey data and single-inclusion weights. However, the constructions given here also include many other informative within-cluster designs for which estimation of variance components η_1, η_2 may be possible but extremely imprecise due to considerations of Fisher information and the large number of nuisance parameters associated with unrestricted informative sampling. The point here is that the matrices $Q^{(k)}$ used in the informative Q-sample designs of Definition 1 may depend on parameters (like p in Example 1) that are different from one population cluster to another, but that do not affect the constant single-inclusion weights within clusters. These parameters related to joint inclusion probabilities, or their empirical distribution over many clusters, would be nuisance parameters affecting the joint cluster-level joint distributions of sampled units. Even if consistent semiparametric estimation of superpopulation parameters (η_1, η_2) were possible in such complicated Q-sampling designs, the precision of the estimates would be made much weaker by the nuisance parameters.

5. Discussion

The primary focus of this paper has been to present an impossibility result for general consistent estimation of two-level model parameters based on complex survey data with single-inclusion weights. The biased-sampling construction of the paper should not be construed as recommendation for such a sampling design, which is studied for theoretical purposes only since it requires knowledge of the order statistics of all outcome variable values $Y_{j,k}$ in the sampled population clusters. The reason for interest in such examples is that actual within-cluster sampling-and-response patterns may resemble such biased sampling patterns and defeat the possibility of parametric mixed-model inference in two-level clustered survey data.

The consequence of nonidentifiability results like those presented here is that, when a finite population exhibits cluster structure at a level not accounted for in a complex survey design, no purely design-based method of analysis based exclusively on first-order inclusion weights can protect against features of the design that might differentially encourage or exclude simultaneous sampling of units from the same cluster. When clusters *are* designed into the survey, but analyses are to be made at unit level, the design and allowed nonresponse patterns affecting members of the same cluster should be controlled as strictly as possible.

Disclaimer

This paper is released to inform interested parties of ongoing research and to encourage discussion of it. The views expressed are the author's and not the Census Bureau's.

REFERENCES

- Asparouhov, T. (2006), "General Multi-level Modeling with Sampling Weights," *Communications in Statistics – Theory and Methods*, 35, 439-460.
- Beaumont, J. (2008), "A New Approach to Weighting and Inference in Sample Surveys," *Biometrika*, 95, 539-553.
- Binder, D. (1983), "On the Variance of Asymptotically Normal Estimators from Complex Surveys," *International Statistical Review*, 51, 279-292.
- Deheuvels, P. and Mason, D. (1990), "Bahadur-Kiefer-type Processes," *Annals of Probability*, 18, 669-697.
- Deville, J. and Särndal, C.-E. (1992), "Calibration Estimators in Survey Sampling," *Journal of the American Statistical Association*, 87, 376-382.
- Encyclopedia of Mathematics (2020), "Bahadur Representation," accessed 9/24/2020, https://encyclopediaofmath.org/wiki/Bahadur_representation.
- Fuller, W.(2009), *Sampling Statistics*, Wiley.
- Kim, J.-K. and Shao, J. (2014), *Statistical Methods for Handling Incomplete Data*, CRC Press, Taylor & Francis.
- Korn, E. and Graubard, B. (2003), "Estimating Variance Components by using Survey Data," *Journal of the Royal Statistical Society, Ser. B*, 65, 175-190.
- Krewski, D. and Rao, J. (1981), "Inference from Stratified Samples: Properties of the Linearization, Jackknife and Balanced Repeated Replication Methods," *Annals of Statistics*, 5, 1010-1019.
- Lindsay, B. (1988), "Composite Likelihood Methods," *Contemporary Mathematics*, 80, 220-239.
- Lohr, S. (2009), *Sampling: Design and Analysis*, 2nd ed., Brooks-Cole.
- Magee, L. (1998), "Improving Survey-weighted Least-squares Regression," *Journal of the Royal Statistical Society, Ser. B*, 60, 115-126.
- Pfeffermann, D. and Sverchkov, M. (2009), "Inference under Informative Sampling," pp. 455-487 in: *Sample Surveys: Inference and Analysis, Handbook of Statistics*, vol. 29B, North-Holland.
- Pfeffermann, D., Skinner, C., Goldstein, H., Holmes, D. and Rasbash, J. (1998), "Weighting for Unequal Selection Probabilities in Multilevel Models (with discussion), *Journal of the Royal Statistical Society, Ser. B*, 60, 23-40.
- Rabe-Hesketh, S. and Skrondal, A. (2006), "Multilevel Modeling of Complex Survey Data," *Journal of the Royal Statistical Society, Ser. A*, 169, 805-827.
- Rao, J. and Scott, A. (1983), "The Analysis of Categorical Data from Complex Sample Surveys: Chi-squared Tests for Goodness of Fit and Independence in Two-way Tables," *Journal of the American Statistical Association*, 76, 221-230.
- Rao, J., Verret, F. and Hidiroglou, M. (2013), "A Weighted Estimating Equations Approach to Inference for Two-level Models from Survey Data," *Survey Methodology*, 39, 263-282.
- Rubin-Bleuer, S. and Kratina, I. (2006), "On the Two-phase Framework for Joint Model and Design-based Inference," *Annals of Statistics*, 33, 2789-2810.
- Särndal, C.-E., Swenssen, J. and Wretman, J. (1992), *Model-Assisted Survey Sampling*, Springer.
- Savitsky, T. and Williams, M. (2019), "Bayesian Mixed Models under Informative Sampling," arXiv:1904.07680
- Savitsky, T. and Williams, M. (2020), "Pseudo-Bayesian Estimation of One-way ANOVA Model in Complex Surveys," preprint.
- Slud, E. (2020), "Model-Assisted Estimation of Mixed-Effect Model Parameters in Complex Surveys," *Statistical Research Report Series RRS 2020-05*, US Census Bureau, <https://www.census.gov/library/working-papers/2020/adrm/RRS2020-05.html>
- Tillé, Y. (2006), *Sampling Algorithms*, Springer.
- Tsiatis, A. (2006), *Semiparametric Theory and Missing Data*, Springer.
- Yi, G., Rao, J. and Li, H. (2016), "A Weighted Composite Likelihood Approach for Analysis of Survey Data under Two-level Models," *Statistica Sinica*, 26, 569-587.
- Zheng, H. and Little, R. (1993), "Penalized Spline Model-based Estimation of the Finite Populations Total from Probability-Proportional-to-Size Samples," *Journal of Official Statistics*, 19, 99-117.