Responsive Design – Side Effect Reduction of Prior Information on Survey Design

A. Demnati
Independent Researcher, Ottawa, Canada, Abdellatif_Demnati@msn.com

Abstract

It is difficult to design a survey because prior information on response rates and the like is likely generated from a different random process than the target one governing the survey to be designed, and the impact on the properties of the estimator can be significant. We are concerned with reducing the side effect of error-prone prior information on the precision of the estimator of the parameter of interest. Nowadays, computer-assisted data collection methods provide an instant variety of observations on the target random process governing the survey under consideration. These data and paradata enable the survey producer to make decisions regarding the need for methodology-process revision, which involves consideration of both a model that represents how the target information relates to the prior information and the design that describes how the observations are obtained. We think of the prior and target information as a random process that has a joint distribution with some probability function. Then at each phase of data collection, after receiving the information that the target random process has taken specific values, we update the joint probability distribution, to revise the design specification in the course of the data collection period. A coefficient of reliability for a survey as a whole set of processes as well for a single process is also discussed.

Key Words: Adaptive design; Misclassification; Optimal resources allocation; Paradata; Partially observed units; Two sources of information.

1. Introduction

Survey or census studies start with a collection of distinct units of interest; termed population. There are multiple random variables attached to each unit, as each holds their own individual characteristics and attitudes. Each particular study targets a small subset of these random variables and involves a questionnaire to collect the data of sampled respondents in order to draw inferences at the estimation stage about the parameter of interest. Measurements on these variables of interest are intended to be collected during data collection stage from each selected unit.

At the planning stage of a survey the question of resources determination and allocation within stages of the survey design is a difficult and critical one. Survey developers must justify resources to be used, and the survey producer should review the justification to ensure the survey produces results within resources, quality and timing constraints. Efficiency is a very important issue because inefficient allocation may lead to: a) imprecise results; and b) waste of time, resources, and money. To determine optimally : i) the duration of the survey; and ii) the amount of resources and their allocation within stages of the survey design, design pre-specification requires the following: a) specification of the parameter of interest, and the associated estimator to be used; b) specification of the desired precision or the global cost; c) specification of the cost function; d) specification of the precision function; e) obtaining prior information, from the sampling frame, administrative files, or from previous surveys, needed to compute unknown quantities in formulas for both precision and cost functions; and, f) optimization of some utility function – that perhaps involves both precision and cost functions.

Suppose previous surveys suggest that the conditional probability of responding $h^{(rq)}$ (in a time period) is constant over time, where the superscript " rq " stands for "response to questionnaire". When the conditional response probability $h^{(rq)}$ is constant over time, then the marginal response probability over I time periods is given by $\xi^{(rq)}_{\mathrm{I};k} = 1-(1-h^{(rq)})^{\mathrm{I}}$. To reach a marginal probability of response close to 1 under constant conditional response probability, it will take around 17 time periods when $h^{(rq)} = .5$, and over 100 time periods when $h^{(rq)} = .1$. Since collecting data over such a long data collection period is time consuming, costly and the results may vary from one time period to another, the way survey sampling handles the problem of capturing information from, or estimating parameters with respect to, finite population generated from such random processes is as follows: a) selecting a random sample of units from the population; and b) increasing level of efforts in term of follow-up treatments to improve units cooperation. Sampling is based on the idea that, within a certain margin of error, one can infer something about the parameter of interest from a small sample, as long as the sample is chosen at random. Efficient follow-up requires information on the error-free target random process governing the survey under consideration. It is difficult to pre-specify the design for certain surveys because prior information is likely generated from a different random process than the one under consideration. A naive approach simplifies the problem under the assumption that resources should be big enough to have good estimates. However, often a survey has limited budget and timing, and those in turn, in combination with the resources allocation used within stages of the survey design based on prior information, determine the achievable precision. Nowadays, computer-assisted data collection methods provide an instant variety of observations about the target random process that can be used to revise survey design during the course of its data collection. Although, previous survey designs are mostly done deterministically using prior information, there is a wide spread need for responsive design where the design is revised during the data collection period. The intent of such revision is to reduce errors attached to design pre-specification on prior information grounds. Objectives of responsive design are formulated in Groves and Heeringa (2006). Starting with an expected design based on prior information, then, cumulative collected data can be used to: 1) update information used for design specification; and, 2) revise, if necessary, specification of the design at each phase of data collection. Estimates with a certain margin of error can be obtained using few time periods, since prior information provides information about the joint probability distribution.

Design pre-specification is a special case of measurement error which refers here to the case where the prior (or error-prone) information, say χ, is not necessarily identical to the target (or error-free) information, say ψ, of the processes underlying the finite population. We assume that the assessment of error in χ can be carried out on the basis of observations on ψ. We also assume that the error-prone information χ has a potential bias $\boldsymbol{b}$ when used to estimate ψ and that the error-free information ψ has no error. Thus, the assessment of errors allows quantification of such bias. Under two random processes, we are interested in the error-free random variable ψ, knowing its probability density function, the probability density function of another random variable χ, together with the joint probability density function with vector parameter denoted by λ. We assume that census parameter λ_N associated with the vector model parameter λ is defined as solution to an estimating equation (EE) of the form $\mathbf{S}(\psi,\chi;\lambda)=\sum_k \mathbf{s}(\psi_k,\chi_k;\lambda)=\mathbf{0}$, where $\sum_k$ is the sum over all the population units. It is assumed that the sampling frame has no coverage bias. It is also assumed that values of the error-prone variable are available for all units in the population, while values of the error-free variable are unknown but observable.

Once an estimate of λ, ψ_k, or of the parameter of interest is obtained, the question follows; what is the reliability of this estimate? In a general sense, reliability of an estimate refers to the degree to which the estimate is free from error and therefore truly measures the quantity which it is intended to measure. When reliability measures are available at all various stages of the survey process, they can serve as performance measures. Such measures enable the survey producer to make decisions regarding the need for methodology-process modification. As there is no general reliability measure that would capture all information on the impact of each stage of the survey design on the ultimate estimate, the survey producer tends to combine various measures to get a broader effect and interactions between different factors. A key step in defining reliability was the introduction of an error criterion that measures, in a probabilistic sense, the error between the desired quantity θ and an estimate $\hat{\theta}$ of it. Possible sources of error in surveys include sampling frame, sampling scheme, nonresponse, measurement, disclosure-avoidance, etc. A criterion which is commonly used in judging the performance of an estimator $\hat{\theta}$ of a quantity θ is its Mean Square Error (MSE) defined by $MSE(\hat{\theta})=E\{(\hat{\theta}-\theta)^2\}$, where E denotes total expectation under random processes involved. We can also interpret this via the MSE decomposition. For any random variable z, we have $E(z^2)=E\{[z-E(z)]^2\}+\{E(z)\}^2$. Applying this to $z=\hat{\theta}-\theta$ we get

$$E\{(\hat{\theta}-\theta)^2\}=E\{[(\hat{\theta}-\theta)-E(\hat{\theta}-\theta)]^2\}+\{E(\hat{\theta}-\theta)\}^2 . \tag{1.1}$$

The first term of (1.1) is the variance of $\hat{\theta}-\theta$. It is the error of the estimator due to the random processes involved. The second term of (1.1) is the square of the bias of $\hat{\theta}$, the best one can do is make this zero.

In an attempt to discuss side effect reduction of prior information on the ultimate estimator, our work below is organized as follows: in Section 2 stochastic models underlying nonresponse follow-up strategy, mode of data collection, and response behaviour are presented; in Section 3, basic ingredients for designing a survey are defined; in Section 4, design pre-specification and revision in the course of data collection is studied; in Section 5, estimator of the regression parameter as well as estimator of the parameter of interest are derived; in Section 6, surveys requiring units contactibility and eligibility identification are briefly discussed; and, in Section 7, a reliability coefficient is discussed.

2. Stochastic Models

To revise the survey design in discrete intervals, we divided the continuous time of the entire data collection period into a sequence of continuous time periods: 1, 2, and so on, and let's $\mathrm{I}_{\min}$ denote the minimum length of data collection period to obtain full responses. Suppose the survey limited length of duration of data collection is made up of $\mathrm{P}_{\max}$ phases, the p^{th} being of size n_{p} time periods, so that the limited duration of data collection is made up of $\mathrm{I}_{\max} = \sum_{\mathrm{p}=1}^{\mathrm{P}_{\max}} n_{\mathrm{p}}$ time periods, with $\mathrm{I}_{\max} < \mathrm{I}_{\min}$. So we shall be dealing with $N \times \mathrm{P}_{\max}$ rectangular array of phases of data collection, where N is the size of the finite population. In this Section, we consider stochastic models for nonresponse follow-up strategy, mode of data collection, and response behaviour. Such models are based on the paradigm of a random experiment modeled by a probability measure on an underlying sample space; i.e., an experiment whose outcome cannot be predicted with certainty, before the experiment is run.

2.1 Nonresponse Follow-up Strategy Model

Suppose we have S nonresponse follow-up strategies, and define a vector of strategy indicator variables as $J_{\mathrm{s}|\mathrm{p};k}^{(f)} = 1$ if unit k is assigned to strategy s at phase p, and $J_{\mathrm{s}|\mathrm{p};k}^{(f)} = 0$ if not, where $\boldsymbol{J}_{|\mathrm{p};k}^{(f)} = (J_{1|\mathrm{p};k}^{(f)},\ldots,J_{S|\mathrm{p};k}^{(f)})^T$ are realizations of independent distributed variables according to a multinomial distribution, $Mult_S(1,\boldsymbol{\varphi}_{|\mathrm{p};k}^{(f)})$, $\boldsymbol{\varphi}_{|\mathrm{p};k}^{(f)} = (\phi_{1|\mathrm{p};k}^{(f)},\ldots,\phi_{S|\mathrm{p};k}^{(f)})^T = E_f(\boldsymbol{J}_{|\mathrm{p};k}^{(f)})$ is the vector of strategy probabilities with $\sum_{s=1}^{S}\phi_{\mathrm{s}|\mathrm{p};k}^{(f)} = 1$, E_f denotes expectation under the strategy allocation model, and the superscript "f" stands for "follow-up". We consider the S^{th} strategy as an omitted or reference strategy. For the multinomial logistic regression model, logits of the first $S-1$ strategies are constructed with the reference strategy in the denominator

$$\log(\phi_{\mathrm{s}|\mathrm{p};k}^{(f)} / \phi_{S|\mathrm{p};k}^{(f)}) = \boldsymbol{v}_{f|\mathrm{p};k}^T \boldsymbol{\eta}_{|\mathrm{p}} ,\ s = 1,\ldots,S-1 ,$$

where $\boldsymbol{v}_{f|\mathrm{p};k}$ is the $q_{f|\mathrm{p}}^{(1)} \times 1$ vector of explanatory variables and $\boldsymbol{\eta}_{|\mathrm{p}} = (\boldsymbol{\eta}_{1|\mathrm{p}}^T,\ldots,\boldsymbol{\eta}_{(S-1)|\mathrm{p}}^T)^T$ is the $q_{f|\mathrm{p}} = q_{f|\mathrm{p}}^{(1)}(S-1) \times 1$ unknown vector parameter to be estimated. It follows that the S conditional probabilities given the vector of explanatory variables are

$$\phi_{S|\mathrm{p};k}^{(f)} = \{1 + \textstyle\sum_{s=1}^{S-1} \exp(\boldsymbol{v}_{f|\mathrm{p};k}^T \boldsymbol{\eta}_{\mathrm{s}|\mathrm{p}})\}^{-1} ,$$

and for $s = 1,\ldots,S-1$

$$\phi_{\mathrm{s}|\mathrm{p};k}^{(f)} = \phi_{S|\mathrm{p};k}^{(f)} \exp(\boldsymbol{v}_{f|\mathrm{p};k}^T \boldsymbol{\eta}_{\mathrm{s}|\mathrm{p}}) .$$

Let's ${}^{(t)}c_{\mathrm{s}|\mathrm{p};k}^{(f)}$ and ${}^{(e)}c_{\mathrm{s}|\mathrm{p};k}^{(f)}$ denote respectively the error-free and error-prone cost associated with follow-up strategy s ($s = 1,\ldots,S$) for unit k during phase p, where the superscripts "t" and "e" stand for "error-free" and "error-prone" respectively.

2.2 Data Collection Models

Similarly, suppose we have M modes of data collection, and define a vector of error-free data collection mode indicator variables as ${}^{(t)}J_{m;k}^{(dc)} = 1$ if unit k uses mode m, and ${}^{(t)}J_{m;k}^{(dc)} = 0$ if not, where ${}^{(t)}\boldsymbol{J}_k^{(dc)} = ({}^{(t)}J_{1;k}^{(dc)},\ldots,{}^{(t)}J_{M;k}^{(dc)})^T$ are realizations of independent distributed random variables according to a multinomial distribution, $Mult_M(1,{}^{(t)}\boldsymbol{\varphi}_k^{(dc)})$, ${}^{(t)}\boldsymbol{\varphi}_k^{(dc)} = ({}^{(t)}\phi_{1;k}^{(dc)},\ldots,{}^{(t)}\phi_{M;k}^{(dc)})^T = E_{dc}({}^{(t)}\boldsymbol{J}_k^{(dc)})$ is the vector of data collection mode probabilities with $\sum_{m=1}^{M}{}^{(t)}\phi_{m;k}^{(dc)} = 1$, and E_{dc} denotes expectation with respect to the data collection model. We consider the M^{th} mode as a reference mode. For the multinomial logistic regression model, logits of the first $M-1$ modes are constructed with the reference mode in the denominator

$$\log({}^{(t)}\phi_{m;k}^{(dc)} / {}^{(t)}\phi_{M;k}^{(dc)}) = \boldsymbol{v}_{dc;k}^T\, {}^{(t)}\boldsymbol{\alpha}_m ,\ m = 1,\ldots,M-1 ,$$

where $\boldsymbol{v}_{dc;k}$ is the $q_{dc}^{(1)} \times 1$ vector of explanatory variables and ${}^{(t)}\boldsymbol{\alpha} = ({}^{(t)}\boldsymbol{\alpha}_1^T,\ldots,{}^{(t)}\boldsymbol{\alpha}_{M-1}^T)^T$ is the ${}^{(t)}q_{dc} = q_{dc}^{(1)}(M-1) \times 1$ unknown vector parameter to be estimated. It follows that the M conditional probabilities given the vector of explanatory variables are

$${}^{(t)}\phi_{M;k}^{(dc)} = \{1 + \textstyle\sum_{m=1}^{M-1} \exp(\boldsymbol{v}_{dc;k}^T\, {}^{(t)}\boldsymbol{\alpha}_m)\}^{-1} ,$$

and for $m = 1,\ldots,M-1$

$${}^{(t)}\phi_{m;k}^{(dc)} = {}^{(t)}\phi_{M;k}^{(dc)} \exp(\boldsymbol{v}_{dc;k}^T\, {}^{(t)}\boldsymbol{\alpha}_m) .$$

Similarly, the vector of error-prone data collection mode indicator variables ${}^{(e)}\boldsymbol{J}_k^{(dc)} = ({}^{(e)}J_{1;k}^{(dc)},\ldots,{}^{(e)}J_{M;k}^{(dc)})^T$ are realizations of independent distributed random variables according to a multinomial distribution, $Mult_M(1,{}^{(e)}\boldsymbol{\varphi}_k)$, ${}^{(e)}\boldsymbol{\varphi}_k^{(dc)} = ({}^{(e)}\phi_{1;k}^{(dc)},\ldots,{}^{(e)}\phi_{M;k}^{(dc)})^T = E_{dc}({}^{(e)}\boldsymbol{J}_k^{(dc)})$ is the vector of probabilities with $\sum_{m=1}^{M}{}^{(e)}\phi_{m;k}^{(dc)} = 1$. The error-prone data collection mode for unit k is characterized by the matrix ${}^{(e|t)}\mathbf{\Phi}_{k}^{(dc)}$ which consists of the conditional probability ${}^{(e|t)}\phi_{i|j;k}^{(dc)}$ of assigning mode i given that unit k should have been assigned mode j defined by its components

$$ {}^{(e|t)}\phi_{i|j;k}^{(dc)} = \Pr({}^{(e)}J_{i;k}^{(dc)} = 1|{}^{(t)}J_{j;k}^{(dc)} = 1) . $$

The marginal distribution of ${}^{(e)}J_{i;k}^{(dc)}$ is:

$$ {}^{(e)}\phi_{i;k}^{(dc)} = \Pr({}^{(e)}J_{i;k}^{(dc)} = 1) = \sum_{j=1}^{M} {}^{(e|t)}\phi_{i|j;k}^{(dc)}\,{}^{(t)}\phi_{j;k}^{(dc)} ,\ i = 1,\dots, M , $$

where $\sum_{i=1}^{M} {}^{(e|t)}\phi_{i|j;k}^{(dc)} = 1$. We consider the M^{th} mode as the reference mode. For the multinomial logistic regression model, logits of the first $M-1$ conditional modes are constructed with the reference mode in the denominator

$$ \log({}^{(e|t)}\phi_{i|j;k}^{(dc)} / {}^{(e|t)}\phi_{M|j;k}^{(dc)}) = {}^{(|j)}\boldsymbol{v}_{dc;k}^{T}\, {}^{(|j)}\boldsymbol{\alpha} ,\ i = 1,\dots, M-1 , $$

where ${}^{(|j)}\boldsymbol{v}_{dc;k}$ is the ${}^{(|j)}q \times 1$ vector of explanatory variables, ${}^{(|j)}\boldsymbol{\alpha} = ({}^{(1|j)}\boldsymbol{\alpha}^{T},\dots,{}^{(M-1|j)}\boldsymbol{\alpha}^{T})^{T}$ and ${}^{(e)}\boldsymbol{\alpha} = ({}^{(|1)}\boldsymbol{\alpha}^{T},\dots,{}^{(|M)}\boldsymbol{\alpha}^{T})^{T}$ is the ${}^{(e)}q_{dc} \times 1$ unknown vector parameter to be estimated, with ${}^{(e)}q_{dc} = \sum_{j=1}^{M} {}^{(|j)}q(M-1)$. It follows that the M conditional probabilities of each mode given mode j and the vector of explanatory variables are

$$ {}^{(e|t)}\phi_{M|j;k}^{(dc)} = \{1 + \sum_{m=1}^{M-1} \exp({}^{(|j)}\boldsymbol{v}_{dc;k}^{T}\, {}^{(|j)}\boldsymbol{\alpha}_{m})\}^{-1} , $$

and for $m = 1,\dots, M-1$

$$ {}^{(e|t)}\phi_{m|j;k}^{(dc)} = {}^{(e|t)}\phi_{M|j;k}^{(dc)} \exp({}^{(|j)}\boldsymbol{v}_{dc;k}^{T}\, {}^{(|j)}\boldsymbol{\alpha}_{m}) . $$

Let's ${}^{(t)}c_{m;k}^{(dc)}$ and ${}^{(e)}c_{m;k}^{(dc)}$ denote respectively the error-free and error-prone cost associated with data collection mode m ($m = 1,\dots, M$) for unit k.

2.3 Response Models

Let ${}^{(a)}t^{(rq)}$ represent the discrete random variable that indicates the time period i when the response occurs under random process a with $a \in \{e, t, e|t\}$ for a randomly selected unit from the sample, where e, t, and $e|t$ are used to indicate error-prone, error-free and conditional random processes respectively. After P phases (or equivalently $\mathrm{I}_{\mathrm{P}} = \sum_{\mathrm{p}=1}^{\mathrm{P}} n_{\mathrm{p}}$ time periods) of data collection, each unit k is observed from the $e_{k}^{(dc)}$ time period until the period ${}^{(a)}\mathrm{I}_{\mathrm{P};k}$, with ${}^{(a)}\mathrm{I}_{\mathrm{P};k} \le \mathrm{I}_{\mathrm{P}}$, where $e_{k}^{(dc)}$ denotes the entry time period for unit k into data collection window. Usually $e_{k}^{(dc)}$ equals 0. Observation of the unit could be discontinued for two reasons: 1) the unit responds; or 2) the phase P of data collection ends. In the first case, ${}^{(a)}t_{k}^{(rq)} = {}^{(a)}\mathrm{I}_{\mathrm{P};k}$. In the second case, it is only know that ${}^{(a)}t_{k}^{(rq)} > \mathrm{I}_{\mathrm{P}}$. Units with ${}^{(a)}t_{k}^{(rq)} > \mathrm{I}_{\mathrm{P}}$ are right-censored – it is unknown when they respond. Note that ${}^{(a)}t^{(rq)}$ is defined only when the unit will respond eventually using appropriate follow-up strategy. Since censoring is planned and observation is terminated at the end of data collection, the censuring mechanism is noninformative (Lagakos, 1979) in the sense that the act of censoring imparts no information about the response mechanism. The overall response indicator from the $e_{k}^{(dc)}$ to the I_{P} time period, with $e_{k}^{(dc)} < \mathrm{I}_{\mathrm{P}}$, is given by ${}^{(a)}r_{\mathrm{P};k}^{(rq)} = 1 - \prod_{i=e_{k}^{(dc)}+1}^{\mathrm{I}_{\mathrm{P};k}} (1 - {}^{(a)}r_{ki}^{(rq)}) = \sum_{i=e_{k}^{(dc)}+1}^{\mathrm{I}_{\mathrm{P};k}} {}^{(a)}r_{ki}^{(rq)}$, where ${}^{(a)}r_{ki}^{(rq)}$ is a sequence of response indicators defined for each unit k whose values are defined as ${}^{(a)}r_{ki}^{(rq)} = 1$ if the unit does respond in period i, and ${}^{(a)}r_{ki}^{(rq)} = 0$ if the unit does not respond in period i. Because response occurrence is intrinsically conditional, Demnati (2015) characterized ${}^{(a)}t^{(rq)}$ by its conditional probability function – the distribution of the probability that a response will occur in each time period given that it has not already occurred in a previous time period – known as the discrete-time hazard function. Discrete-time hazard, ${}^{(a)}h_{ki}^{(rq)}({}^{(a)}\boldsymbol{v}_{rq;k}, {}^{(a)}\boldsymbol{\beta})$, ${}^{(a)}h_{ki}^{(rq)}$ for short, is defined as the conditional probability that unit k will respond in time period i, given that the unit did not respond prior to i:

$$ {}^{(a)}h_{ki}^{(rq)} = \Pr({}^{(a)}t_{k}^{(rq)} = i|{}^{(a)}t_{k}^{(rq)} \ge i) , $$

where ${}^{(a)}\boldsymbol{v}_{rq;k}$ refers to both time-invariant and time-varying explanatory variables and ${}^{(a)}\boldsymbol{\beta}$ is the unknown ${}^{(a)}q_{rq} \times 1$ vector parameter to be estimated. For unit with ${}^{(a)}t_{k}^{(rq)} = i$, the probability of obtaining a response at time period i could be expressed in terms of the hazard as

$$ \Pr({}^{(a)}t_{k}^{(rq)} = i) = {}^{(a)}h_{ki}^{(rq)} \prod_{j=e_{k}^{(dc)}+1}^{i-1} (1 - {}^{(a)}h_{kj}^{(rq)}) . $$

For units with ${}^{(a)}t_{k}^{(rq)} > i$, the probability of obtaining a response can be expressed as

$$ \Pr({}^{(a)}t_{k}^{(rq)} > i) = \prod_{j=e_{k}^{(dc)}+1}^{i} (1 - {}^{(a)}h_{kj}^{(rq)}) . $$

After P phases of data collection, we have

$$ g({}^{(a)}t_{k}^{(rq)}) = \{\Pr({}^{(a)}t_{k}^{(rq)} = {}^{(a)}\mathrm{I}_{\mathrm{P};k})\}^{{}^{(a)}\delta_{k}^{(dc)}} \{\Pr({}^{(a)}t_{k}^{(rq)} > {}^{(a)}\mathrm{I}_{\mathrm{P};k})\}^{1-{}^{(a)}\delta_{k}^{(dc)}} \equiv {}^{(a)}g_{\mathrm{I}_{\mathrm{P}};k}^{(rq)} , $$

where ${}^{(a)}\delta_k^{(dc)} = 1$ if unit k is uncensored (responds) and ${}^{(a)}\delta_k^{(dc)} = 0$ if unit k is censored under process a. When unit k is censored, it is unknown when the unit responds. The joint distribution of $({}^{(e)}t_k^{(rq)}, {}^{(t)}t_k^{(rq)})$ is characterized by the matrix ${}^{(e|t)}\mathbf{G}_k^{(rq)}$ defined by its components

$${}^{(e|t)}g_{i|j;k}^{(rq)} = \Pr({}^{(e)}t_k^{(rq)} = i|{}^{(t)}t_k^{(rq)} = j)\,,\ i, j = e_k^{(dc)} + 1, \ldots, \mathrm{I_P}\,.$$

The marginal distribution of ${}^{(e)}t_k^{(rq)}$ is:

$${}^{(e)}g_{\mathrm{I_P};k}^{(rq)} = \textstyle\sum_{j=e_k^{(dc)}+1}^{\mathrm{I_P}} {}^{(e|t)}g_{\mathrm{I_P};k|j;k}^{(rq)}\ {}^{(t)}g_{j;k}^{(rq)}\,,\ \mathrm{I}_{\mathrm{P};k} = e_k^{(dc)} + 1, \ldots, \mathrm{I_P}\,.$$

The marginal probability of obtaining a response from the $e_k^{(dc)}$ to the $\mathrm{I_P}$ time period is given by

$${}^{(a)}\xi_{\mathrm{P};k}^{(rq)} = 1 - \textstyle\prod_{i=e_k^{(dc)}+1}^{\mathrm{I_P}} (1 - {}^{(a)}h_{ki}^{(rq)}) = \textstyle\sum_{i=e_k^{(dc)}+1}^{\mathrm{I_P}} \Pr({}^{(a)}t_{ki}^{(rq)} = i)\ .$$

2.4 Modeling Sample Selection Probability

For sampling, we parameterize the probability of selection as $\pi_k = \{lb + ub \times \exp(\boldsymbol{v}_{\wp;k}^T \varphi)\}/\{1 + \exp(\boldsymbol{v}_{\wp;k}^T \varphi)\}$, where lb and ub are respectively the lower and upper bounds with $0 < lb < ub \le 1$, $\boldsymbol{v}_{\wp;k}$ is the $q_\wp \times 1$ vector of explanatory variables and φ is the $q_\wp \times 1$ unknown vector parameter to be determined. Unlike the customary probability of selection $\pi_{c;k} = np_k$, where n is the expected sample size and p_k is a measure of size, this expression for π_k fulfill the two criterions: : a) $0 < lb \le \pi_k \le ub \le 1$; and b) $\pi_k \to ub$ as $n \to N$. It is common practice to set $(lb, ub) = (0,1)$.

3. Basic Ingredients for Designing a Survey

In this Section, we elaborate on each step of the technical steps mentioned in the introduction. These steps are required in designing a survey. We consider examples of parameter of interest and an associated estimator, a precision function, a cost function, and an utility function.

3.1 Parameter of Interest

We assume that census parameter $\boldsymbol{\Theta}_N(\boldsymbol{\psi})$ associated with the variable of interest $\boldsymbol{\psi}$ is defined as solution to an EE of the form

$$\mathbf{S}(\boldsymbol{\psi};\boldsymbol{\Theta}) = \textstyle\sum_k \mathbf{s}(\boldsymbol{\psi}_k;\boldsymbol{\Theta}) - \mathbf{v}(\boldsymbol{\Theta}) = \mathbf{0}\,, \qquad (3.1)$$

where the known function $\mathbf{s}(\boldsymbol{\psi}_k;\boldsymbol{\Theta})$ is a q_Θ-dimensional vector-valued function of $\boldsymbol{\psi}_k$ and the known function $\mathbf{v}(\boldsymbol{\Theta})$ allows for explicitly defined parameters. For linear and logistic regression models, $\mathbf{s}(\boldsymbol{\psi}_k;\boldsymbol{\Theta}) = \boldsymbol{x}_k(y_k - \mu_k(\boldsymbol{x}_k^T\boldsymbol{\Theta}))$ and $\mathbf{v}(\boldsymbol{\Theta}) = \mathbf{0}$, where $\mu_k(\boldsymbol{x}^T\boldsymbol{\Theta}) = E_y(y)$, $\boldsymbol{x} = (x_1, \ldots, x_{q_\Theta})^T$ is a $q_\Theta \times 1$ vector of explanatory variables, $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_{q_\Theta})^T$ is the $q_\Theta \times 1$ vector of model parameter and E_y denotes model expectation on the variable of interest y. For the special case of the finite population total $Y = \sum_k y_k$, $\mathbf{s}(\boldsymbol{\psi}_k;\boldsymbol{\Theta}) = y_k$, $\mathbf{v}(\boldsymbol{\Theta}) = \Theta_N(\boldsymbol{\psi})$, and $\Theta_N(\boldsymbol{\psi}) = Y$. The finite population parameter $\boldsymbol{\Theta}_N(\boldsymbol{\psi})$, obtained as the solution of (3.1), under the assumed ideal situation which consists of census case with complete response and without any measurement error, plays the role of a "gold standard".

3.2 Estimator of the Parameter of Interest

Suppose that the response probability, $\xi_{\mathrm{P};k}^{(rq)} = E_r(r_{\mathrm{P};k}^{(rq)})$, during P phases of data collection, is known for every unit in the population, where $r_{\mathrm{P};k}^{(rq)}$ is the overall response indicator for unit k, and E_r denotes expectation under the response mechanism. For general sampling design with known positive inclusion probabilities, π_k, a design-response unbiased estimator of the EE defined by (3.1) is given by

$$\breve{\mathbf{S}}(\boldsymbol{\psi};\boldsymbol{\Theta}) = \textstyle\sum_k d_k(\wp)(r_{\mathrm{P};k}^{(rq)}/\xi_{\mathrm{P};k}^{(rq)})\mathbf{s}(\boldsymbol{\psi}_k;\boldsymbol{\Theta}) - \mathbf{v}(\boldsymbol{\Theta}) = \mathbf{0}\,, \qquad (3.2)$$

where $d_k(\wp) = 1_k(\wp)/\pi_k$ are the design weights, $1_k(\wp) = 1(k \in \wp)$ is the sample $\wp$ membership indicator variable for unit k, $1(condition\ )$ is the truth function, i.e., $1(condition\ ) = 1$ if the $condition$ is true and $1(condition\ ) = 0$ if not, $\pi_k = E_\wp\{1_k(\wp)\}$ is the sample $\wp$ inclusion probability for unit k, and $E_\wp$ denotes expectation with respect to the sampling design. The solution obtained by a Newton-Raphson-type iterative method gives the estimator $\breve{\boldsymbol{\Theta}}_{\mathrm{P}}(\boldsymbol{\psi})$ of $\boldsymbol{\Theta}_N(\boldsymbol{\psi})$.

3.3 Derivation of the Variance Function

We assume that P phases of data collection are completed, with $1 \le \mathrm{P} \le \mathrm{P}_{\max}$, and we consider for illustration the simple estimator $\breve{\mathbf{\Theta}}_{\mathrm{P};k}(\boldsymbol{\psi})$ of $\mathbf{\Theta}_N(\boldsymbol{\psi})$ solution to (3.2). We first consider the derivation of the variance of a compact form given by

$$\hat{\mathrm{U}} = \textstyle\sum_k \boldsymbol{u}_k d_k(\wp) r_{\mathrm{P};k}^{(rq)} , \tag{3.3}$$

where $\boldsymbol{u}_k$ is a vector of constants. We may decompose the variance of $\hat{\mathrm{U}}$ as

$$Var(\hat{\mathbf{U}}) = E_{\wp} Var_r(\hat{\mathbf{U}}) + Var_{\wp} E_r(\hat{\mathbf{U}}) \equiv \mathbf{V}_r + \mathbf{V}_{\wp} = \mathbf{V}(\boldsymbol{u}) , \tag{3.4}$$

where $Var_{\wp}$ and Var_r denote variance with respect to sampling design and response mechanism respectively.
Under independent mechanism on $r_{\mathrm{P};k}$, the first component $\mathbf{V}_r = E_{\wp} Var_r(\hat{\mathrm{U}})$ of $Var(\hat{\mathrm{U}})$ given by (3.4) is given by

$$\mathbf{V}_r = \textstyle\sum_k (1/\pi_k) \boldsymbol{u}_k \xi_{\mathrm{P};k}^{(rq)} (1 - \xi_{\mathrm{P};k}^{(rq)}) \boldsymbol{u}_k^T . \tag{3.4a}$$

The second component $\mathbf{V}_{\wp} = Var_{\wp}\{\sum_k \boldsymbol{u}_k d_k(\wp) \xi_{\mathrm{P};k}^{(rq)}\}$ of $Var(\hat{\mathrm{U}})$ given by (3.4) is given by

$$\mathbf{V}_{\wp} = \textstyle\sum_k \sum_l \omega_{kl}^{-1} (1 - \omega_{kl}) \boldsymbol{u}_k \xi_{\mathrm{P};k}^{(rq)} \xi_{\mathrm{P};l}^{(rq)} \boldsymbol{u}_l^T , \tag{3.4b}$$

where $\omega_{kl} = \pi_{kl}^{-1} \pi_k \pi_l$, $\omega_{kk} = \omega_k = \pi_k$ and $\pi_{kl} = E_{\wp}\{1_k(\wp) 1_l(\wp)\}$. The sum of (3.4a) and (3.4b) constitutes $\mathbf{V}(\boldsymbol{u}) = \mathbf{V}_r + \mathbf{V}_{\wp}$, the variance of $\hat{\mathrm{U}}$ given by (3.3).

It follows that the compact form given by (3.3) can be used to derive the linearization variance $Var_L\{\breve{\mathbf{\Theta}}_{\mathrm{P}}(\boldsymbol{\psi})\}$ of $\breve{\mathbf{\Theta}}_{\mathrm{P}}(\boldsymbol{\psi})$ using $\boldsymbol{u}_k = \{\mathbf{J}(\mathbf{\Theta}_N(\boldsymbol{\psi}))\}^{-1} \mathbf{s}(\boldsymbol{\psi}_k; \mathbf{\Theta}_N(\boldsymbol{\psi})) / \xi_{\mathrm{P};k}^{(rq)}$, where $\mathbf{J}(\mathbf{\Theta}) = -\partial \mathbf{S}^T(\boldsymbol{\psi}; \mathbf{\Theta}) / \partial \mathbf{\Theta}$. It is easily shown that, under stratified simple random sample and Poisson sampling, we can express the linearization variance $Var_L\{\breve{\mathbf{\Theta}}_{\mathrm{P}}(\boldsymbol{\psi})\}$ of $\breve{\mathbf{\Theta}}_{\mathrm{P}}(\boldsymbol{\psi})$ in the separate form as

$$Var_L\{\breve{\mathbf{\Theta}}_{\mathrm{P}}(\boldsymbol{\psi})\} = \boldsymbol{v}_0(\boldsymbol{\psi}) + \textstyle\sum_k \boldsymbol{v}_{\wp;k}(\boldsymbol{\psi}) / \pi_k + \sum_k \boldsymbol{v}_{\wp r;k}(\boldsymbol{\psi}) / \{\pi_k \xi_{\mathrm{P};k}^{(rq)}\} ,$$

where $\boldsymbol{v}_0(\boldsymbol{\psi})$, $\boldsymbol{v}_{\wp;k}(\boldsymbol{\psi})$, and $\boldsymbol{v}_{\wp r;k}(\boldsymbol{\psi})$ are functions independents of π_k and $\xi_{\mathrm{P};k}^{(rq)}$. We denote in operator notation the estimator $\breve{\mathbf{\Theta}}_{\mathrm{P}}(\boldsymbol{\psi})$ and its linearization variance $Var_L\{\breve{\mathbf{\Theta}}_{\mathrm{P}}(\boldsymbol{\psi})\}$ by $\breve{\mathbf{\Theta}}\{\boldsymbol{\lambda}, \mathbf{A}(\boldsymbol{\psi})\}$ and $Var_L\{\breve{\mathbf{\Theta}}\{\boldsymbol{\lambda}, \mathbf{A}(\boldsymbol{\psi})\}\}$ respectively, where $\mathbf{A}(\boldsymbol{\psi})$ is an N-column matrix with k^{th} column $\boldsymbol{\psi}_k$.

3.4 Specification of the Cost Function

We may decompose the global cost over P phases of data collection as

$$C(\mathrm{P}) = \textstyle\sum_{\mathrm{p}=1}^{\mathrm{P}} C_{\mathrm{p}} , \tag{3.5}$$

with $1 \le \mathrm{P} \le \mathrm{P}_{\max}$. The P components of the global cost are

$$C_1 = c_1 + C_1^{(\wp)} + C_1^{(f)} + C_1^{(dc)} ,$$

and for $\mathrm{p} = 2, \ldots, \mathrm{P}$

$$C_{\mathrm{p}} = c_{\mathrm{p}} + C_{\mathrm{p}}^{(f)} + C_{\mathrm{p}}^{(dc)} ,$$

where $C_1^{(\wp)} = \sum_k 1_k(\wp) c_k^{(\wp)}$ is the component associated with sampling cost, $C_{\mathrm{p}}^{(f)} = \sum_k 1_k(\wp)(1 - r_{\mathrm{p-1};k}^{(rq)}) \sum_{s=1}^{S} J_{s|\mathrm{p};k}^{(f)} c_{s|\mathrm{p};k}^{(f)}$ is the component associated with nonresponse follow-up cost for phase p, with $r_{0;k}^{(rq)} = 0$, and $C_{\mathrm{p}}^{(dc)} = \sum_k 1_k(\wp)(1 - r_{\mathrm{p-1};k}^{(rq)}) \sum_{s=1}^{S} J_{s|\mathrm{p};k}^{(f)} \sum_{m=1}^{M} J_{m|s;k}^{(dc)} r_{\mathrm{p};k}^{(rq)} c_{m|\mathrm{p};k}^{(dc)}$ is the component associated with data collection cost for phase p. Here c_{p} is the fixed cost for phase p, and $c_k^{(\wp)}$ is the sampling cost for unit k.

3.5 Specification of the Utility Function

Suppose guessed values $\boldsymbol{\lambda}_{\mathrm{p}}$ and $\mathbf{A}(\boldsymbol{\psi}_{\mathrm{p}})$ are available for $\boldsymbol{\lambda}$ and for $\mathbf{A}(\boldsymbol{\psi})$ respectively. To create a design in the case of one parameter of interest, we minimize the linearization variance, $Var_L\{\breve{\Theta}\{\boldsymbol{\lambda}_{\mathrm{p}}, \mathbf{A}(\boldsymbol{\psi}_{\mathrm{p}})\}\}$ subject to constraints on the duration of the data collection and on the conditional expected global cost: $1 \le \mathrm{P} \le \mathrm{P}_{\max}$ and $C\overline{C}\{\boldsymbol{\lambda}_{\mathrm{p}}, \mathbf{A}(\boldsymbol{\psi}_{\mathrm{p}})\} \le C_{\max}$, where $C\overline{C}\{\boldsymbol{\lambda}_{\mathrm{p}}, \mathbf{A}(\boldsymbol{\psi}_{\mathrm{p}})\}$ denotes the conditional expectation of the cost function $C(\mathrm{P})$ given by (3.5), and $C_{\max}$ is the survey global cost limit.

In the case of $\Lambda (>1)$ parameters of interest, let Θ_{κ} denote the parameter of interest κ, $\kappa = 1, \ldots, \Lambda$. To create a design, we optimize the conditional expected global cost given by $C\overline{C}\{\boldsymbol{\lambda}_{\mathrm{p}}, \mathbf{A}(\boldsymbol{\psi}_{\mathrm{p}})\}$ subject to constraints on the duration of data collection and on Λ variances:

$$1 \le \mathrm{P} \le \mathrm{P}_{\max} ,$$

and $$Var_L\{\breve{\Theta}_\kappa\{\boldsymbol{\lambda}_{\mathrm{p}},\mathbf{A}(\boldsymbol{\psi}_{\mathrm{p}})\}\}\leq \mathrm{V}_{\mathrm{p};\kappa},\ \ \kappa=1,...,\Lambda\ ,$$

where $\mathrm{V}_{\mathrm{p};\kappa}$ are specified tolerances, and $Var_L\{\breve{\Theta}_\kappa\{\boldsymbol{\lambda}_{\mathrm{p}},\mathbf{A}(\boldsymbol{\psi}_{\mathrm{p}})\}\}$ is the linearization variance of the estimator $\breve{\Theta}_\kappa\{\boldsymbol{\lambda}_{\mathrm{p}},\mathbf{A}(\boldsymbol{\psi}_{\mathrm{p}})\}$ for the κ^{th} parameter of interest $\kappa=1,...,\Lambda$. For example, one could specify an upper limit, $\Im_{\mathrm{p};\kappa}$, on the coefficient of variation of $\breve{\Theta}_\kappa\{\boldsymbol{\lambda}_{\mathrm{p}},\mathbf{A}(\boldsymbol{\psi}_{\mathrm{p}})\}$ so that $\mathrm{V}_{\mathrm{p};\kappa}=\{\Im_{\mathrm{p};\kappa}\mathrm{E}\{\breve{\Theta}_\kappa\{\boldsymbol{\lambda}_{\mathrm{p}},\mathbf{A}(\boldsymbol{\psi}_{\mathrm{p}})\}\}\}^2$. One may repeat the optimization process with different value of $\Im_{\mathrm{p};\kappa}$ $\kappa=1,...,\Lambda$ to obtain the desired minimum cost.

4. Design Pre-specification and Revision

The complete data for sampled unit k is given by $(\boldsymbol{\psi}_k^T,\boldsymbol{\chi}_k^T)^T$, where $\boldsymbol{\psi}_k={}^{(t)}\varpi_k$, $\boldsymbol{\chi}_k={}^{(e)}\varpi_k$, ${}^{(\varepsilon)}\varpi_k=({}^{(\varepsilon)}\mathrm{I}_k,{}^{(\varepsilon)}\delta_k^{(dc)},{}^{(\varepsilon)}\boldsymbol{J}_k^{(dc)T},{}^{(\varepsilon)}\boldsymbol{y}_k^T,{}^{(\varepsilon)}\boldsymbol{v}_k^T,{}^{(\varepsilon)}\boldsymbol{c}_k^T)^T$, ${}^{(\varepsilon)}\boldsymbol{c}_k=({}^{(\varepsilon)}\boldsymbol{c}_k^{(f)T},{}^{(\varepsilon)}\boldsymbol{c}_k^{(dc)T})^T$, and $\varepsilon\in\{t,e\}$. The vectors parameter associated with the marginal distributions of ${}^{(\varepsilon)}\varpi_k$ is ${}^{(\varepsilon)}\boldsymbol{\lambda}=({}^{(\varepsilon)}\boldsymbol{\alpha}^T,{}^{(\varepsilon)}\boldsymbol{\beta}^T,{}^{(\varepsilon)}\boldsymbol{\gamma}_y^T,{}^{(\varepsilon)}\boldsymbol{\gamma}_c^T)^T$, where ${}^{(\varepsilon)}\boldsymbol{\gamma}_y$ is the vector parameter associated with the vector of variables of interest ${}^{(\varepsilon)}\boldsymbol{y}$, and ${}^{(\varepsilon)}\boldsymbol{\gamma}_c$ is the vector parameter associated with the vector cost ${}^{(\varepsilon)}\boldsymbol{c}$. We set ${}^{(t)}\boldsymbol{\lambda}_0={}^{(e)}\boldsymbol{\lambda}$ and $\boldsymbol{\psi}_{0;k}=\boldsymbol{\chi}_k$ for design pre-specification. In this Section, we discuss survey design pre-specification, and the Observation-Revision-Optimization steps for design revision in the course of its data collection period.

4.1 Pre-specification of the Survey Design

Suppose guessed values are available for design pre-specification. Since the above global cost given by (3.5) under the models comes from what we think of as random, we consider its expectation under the sampling design and models for strategy, mode and response behaviour, given by

$$E\{C(\mathrm{P})\}=\textstyle\sum_{\mathrm{p}=1}^{\mathrm{P}}\overline{C}_{\mathrm{p}}\equiv C\overline{C}\{\boldsymbol{\lambda},\mathbf{A}(\boldsymbol{\psi})\}\ . \tag{4.1}$$

The P components of the expected global cost are

$$\overline{C}_1=c_1+\overline{C}_1^{(\wp)}+\overline{C}_1^{(f)}+\overline{C}_1^{(dc)}\ ,$$

and for $\mathrm{p}=2,...,\mathrm{P}$ $$\overline{C}_{\mathrm{p}}=c_{\mathrm{p}}+\overline{C}_{\mathrm{p}}^{(f)}+\overline{C}_{\mathrm{p}}^{(dc)}\ ,$$

where $\overline{C}_1^{(\wp)}=\sum_k\pi_kc_k^{(\wp)}$, $\overline{C}_{\mathrm{p}}^{(f)}=\sum_k\pi_k(1-\xi_{\mathrm{p}\text{-}1;k}^{(rq)})\sum_{s=1}^S\phi_{s|\mathrm{p};k}^{(f)}c_{s|\mathrm{p};k}^{(f)}$, and $\overline{C}_{\mathrm{p}}^{(dc)}=\sum_k\pi_k(1-\xi_{\mathrm{p}\text{-}1;k}^{(rq)})\sum_{s=1}^S\phi_{s|\mathrm{p};k}^{(f)}\sum_{m=1}^M\phi_{m|s;k}^{(dc)}\xi_{\mathrm{p};k}^{(rq)}c_{m|\mathrm{p};k}^{(dc)}$, with $\xi_{0;k}^{(rq)}=0$.

To create a pre-design using guessed values $\boldsymbol{\lambda}_0$, and $\mathbf{A}(\boldsymbol{\psi}_0)$, the conditional expected cost $C\overline{C}\{\boldsymbol{\lambda}_0,\mathbf{A}(\boldsymbol{\psi}_0)\}$, and the variance $Var_L\{\breve{\boldsymbol{\Theta}}\{\boldsymbol{\lambda}_0,\mathbf{A}(\boldsymbol{\psi}_0)\}\}$, we determine the optimal P, φ, and $\boldsymbol{\eta}$ by optimizing the utility function. We denote the solution by φ, P_0 and $\boldsymbol{\eta}_0$, then we draw the sample $\wp$ using φ.

4.2 Revision of a Design in the Course of its Progress

Our method basically consists of a series of optimizations using additional information on the error-free target random process. After optimization using additional information from a certain phase of data collection is completed, the result indicates if the active design should be revised. We do not, therefore, use a fixed design, although an expected design is always specified in the survey design. Starting with an expected pre-specified design, then for $\mathrm{p}=1,2,...$ design update is made using the Observation-Revision-Optimization (ORO) steps:

- ➢ **Observation Step:** Obtain next phase p of observations on the error-free process.
- ➢ **Revision Step:**
  - **Maximization**: Update $\boldsymbol{\lambda}_{\mathrm{p}-1}$ to get $\boldsymbol{\lambda}_{\mathrm{p}}$ using $\mathbf{D}_{\mathrm{p}}$ and the update step given in Subsection 4.3, where $\mathbf{D}_{\mathrm{p}}$ denotes all observed information until the end of phase p of data collection.
  - **Imputation**: Impute missing values of each component ψ of $\boldsymbol{\psi}$ to get $\psi_{\mathrm{p};k}=E_\psi(\psi_k\,|\,\mathbf{D}_{\mathrm{p}},\boldsymbol{\lambda}_{\mathrm{p}})$, where E_ψ denotes expectation with respect to the random process governing the component ψ. Note that $\psi_{\mathrm{p};k}=\psi_k$ when item ψ_k is observed.
  - **Conditional Expectation of the Cost Function:** Compute the conditional expectation of the global cost to get $C\overline{C}\{\boldsymbol{\lambda}_{\mathrm{p}},\mathbf{A}(\boldsymbol{\psi}_{\mathrm{p}})\}$
  - **Conditional Expectation of the Precision Function:** Compute the conditional expectation of the variance to get $Var_L\{\breve{\boldsymbol{\Theta}}\{\boldsymbol{\lambda}_{\mathrm{p}},\mathbf{A}(\boldsymbol{\psi}_{\mathrm{p}})\}\}$.
- ➢ **Optimization Step:**

- Determine the optimal P, and $\boldsymbol{\eta}$ conditional on φ, $\boldsymbol{J}_{|1;k}^{(f)},\ldots,\boldsymbol{J}_{|\mathrm{p}\text{-}1;k}^{(f)}$, $r_{\mathrm{p}\text{-}1;k}^{(rq)}$, $\boldsymbol{\lambda}_{\mathrm{p}}$, and $\mathbf{A}(\boldsymbol{\psi}_{\mathrm{p}})$ under the usual constraints and the following additional constraint $\mathrm{p} \le \mathrm{P} \le \mathrm{P}_{\max}$, using $C\overline{C}\{\boldsymbol{\lambda}_{\mathrm{p}},\mathbf{A}(\boldsymbol{\psi}_{\mathrm{p}})\}$ and $Var_L\{\breve{\Theta}_\kappa\{\boldsymbol{\lambda}_{\mathrm{p}},\mathbf{A}(\boldsymbol{\psi}_{\mathrm{p}})\}\}$. The solution is denoted by P_{p}, and $\boldsymbol{\eta}_{\mathrm{p}}$.

The three steps are repeated until $\mathrm{p} = \mathrm{P}_{\mathrm{p}}$.

4.3 Revision – Maximization Step

When the duration of data collection period is taken into account, the likelihood function of the joint distribution under census data is defined for unit k as

$$L_{\mathrm{I}_{\min};k}^{(\mathrm{I}_{\min})}(\boldsymbol{\lambda}) = f_{\mathrm{I}_{\min}}^{(\mathrm{I}_{\min})}(\boldsymbol{\psi}_k,\boldsymbol{\chi}_k)\,, \tag{4.1}$$

where the subscript I in $f_{\mathrm{I}}^{(\mathrm{I}_{\min})}(\zeta)$ denotes that ζ is observed during the interval $[0,\mathrm{I}]$. Hence when data collection period is taken into account, the census case means that $\mathrm{I} = \mathrm{I}_{\min}$ and $d_k(\wp) = 1_k(\wp) = \pi_k = 1$. To simplify our notation we drop the superscript $\mathrm{I}_{\min}$, and write (4.1) as $L_{\mathrm{I}_{\min};k}(\lambda) = f_{\mathrm{I}_{\min}}(\boldsymbol{\psi}_k,\chi_k)$. After observing I_{p} time periods of data collection, the joint observations on χ, and ψ are known for respondents during the I_{p} time periods, while only observations on χ are known for nonrespondents during the rest of the periods. Consequently, we decompose the likelihood of observed data for unit k in two parts

$$L_{\mathrm{I}_{\mathrm{p}};k}(\boldsymbol{\lambda}) = f_{\mathrm{I}_{\max}}(\boldsymbol{\chi}_k)\, f_{\mathrm{I}_{\mathrm{p}}}(\boldsymbol{\psi}_k \mid \boldsymbol{\chi}_k)\,, \tag{4.2}$$

in which case the log-likelihood is given by

$$\ell_{\mathrm{I}_{\mathrm{p}};k}(\boldsymbol{\lambda}) = \log f_{\mathrm{I}_{\max}}(\boldsymbol{\chi}_k) + \log f_{\mathrm{I}_{\mathrm{p}}}(\boldsymbol{\psi}_k \mid \boldsymbol{\chi}_k)\,, \tag{4.3}$$

where $f_{\mathrm{I}_{\max}}(\boldsymbol{\chi}_k) = \int f_{\mathrm{I}_{\max}}(\boldsymbol{\chi}_k,\boldsymbol{\psi}_k)d\boldsymbol{\psi}_k$, and $f_{\mathrm{I}_{\mathrm{p}}}(\boldsymbol{\psi}_k \mid \boldsymbol{\chi}_k) = f_{\mathrm{I}_{\mathrm{p}}}(\boldsymbol{\chi}_k,\boldsymbol{\psi}_k)/\, f_{\mathrm{I}_{\mathrm{p}}}(\boldsymbol{\chi}_k)$. Note that $f_{\mathrm{I}_{\max}}(\boldsymbol{\chi}_k)\, f_{\mathrm{I}_{\mathrm{p}}}(\boldsymbol{\psi}_k \mid \boldsymbol{\chi}_k) \to f_{\mathrm{I}_{\max}}(\boldsymbol{\psi}_k,\boldsymbol{\chi}_k)$ as $\mathrm{I}_{\mathrm{p}} \to \mathrm{I}_{\max}$. In arriving to (4.2), we decomposed χ in two parts: $\boldsymbol{\chi} = (\boldsymbol{\chi}_{[e_k^{(dc)},\mathrm{I}_{\mathrm{p}}]}^T, \boldsymbol{\chi}_{[\mathrm{I}_{\mathrm{p}}+1,\mathrm{I}_{\max}]}^T)^T$, where $\boldsymbol{\chi}_{[e_k^{(dc)},\mathrm{I}_{\mathrm{p}}]}$ denotes observation during the interval $[e_k^{(dc)},\mathrm{I}_{\mathrm{p}}]$, while $\boldsymbol{\chi}_{[\mathrm{I}_{\mathrm{p}}+1,\mathrm{I}_{\max}]}$ denotes observation during the interval $[\mathrm{I}_{\mathrm{p}}+1,\mathrm{I}_{\max}]$. The joint distribution is given by

$$f(\boldsymbol{\chi},\boldsymbol{\psi}) = f(\boldsymbol{\chi}_{[e_k^{(dc)},\mathrm{I}_{\mathrm{p}}]},\boldsymbol{\chi}_{[\mathrm{I}_{\mathrm{p}}+1,\mathrm{I}_{\max}]},\boldsymbol{\psi}) = f(\boldsymbol{\chi}_{[e_k^{(dc)},\mathrm{I}_{\mathrm{p}}]})\, f(\boldsymbol{\chi}_{[\mathrm{I}_{\mathrm{p}}+1,\mathrm{I}_{\max}]} \mid \boldsymbol{\chi}_{[e_k^{(dc)},\mathrm{I}_{\mathrm{p}}]})\, f(\boldsymbol{\psi} \mid \boldsymbol{\chi}_{[e_k^{(dc)},\mathrm{I}_{\mathrm{p}}]}) = f(\boldsymbol{\chi})\, f(\boldsymbol{\psi} \mid \boldsymbol{\chi}_{[e_k^{(dc)},\mathrm{I}_{\mathrm{p}}]})\,.$$

Taking the derivatives of (4.3) and adjusting for unequal probability of selection, we get the weighted EE

$$\hat{\mathbf{S}}_{\mathrm{I}_{\mathrm{p}}}(\boldsymbol{\psi},\boldsymbol{\chi};\boldsymbol{\lambda}) = \textstyle\sum_k d_k(\wp)\{\mathbf{s}_{\mathrm{I}_{\max}}(\boldsymbol{\chi}_k;\boldsymbol{\lambda}) + \mathbf{s}_{\mathrm{I}_{\mathrm{p}}}(\boldsymbol{\psi}_k;\boldsymbol{\lambda} \mid \boldsymbol{\chi}_k)\} = \mathbf{0}\,, \tag{4.4}$$

where $\mathbf{s}_{\mathrm{I}_{\max}}(\boldsymbol{\chi}_k;\boldsymbol{\lambda}) = \partial \log f_{\mathrm{I}_{\max}}(\boldsymbol{\chi}_k)/\partial\boldsymbol{\lambda}$ and $\mathbf{s}_{\mathrm{I}_{\mathrm{p}}}(\boldsymbol{\psi}_k;\boldsymbol{\lambda} \mid \boldsymbol{\chi}_k) = \partial \log f_{\mathrm{I}_{\mathrm{p}}}(\boldsymbol{\psi}_k \mid \boldsymbol{\chi}_k)/\partial\boldsymbol{\lambda}$.

The estimator $\hat{\boldsymbol{\lambda}}_{\mathrm{p}}$ of $\boldsymbol{\lambda}$ is obtained using the following update step.

Update Step of $\boldsymbol{\lambda}$: Starting with a guessed value, $\boldsymbol{\lambda}^{(0)} = \boldsymbol{\lambda}_{\mathrm{p}\text{-}1}$, then for $b = 1,2,\ldots$ updates are made using

$$\boldsymbol{\lambda}^{(b)} = \boldsymbol{\lambda}^{(b-1)} + \{\hat{\mathbf{J}}_{\mathrm{I}_{\mathrm{p}}}(\boldsymbol{\lambda}^{(b-1)})\}^{-1}\hat{\mathbf{S}}_{\mathrm{I}_{\mathrm{p}}}(\boldsymbol{\psi},\boldsymbol{\chi};\boldsymbol{\lambda}^{(b-1)})\,,$$

where $\hat{\mathbf{J}}_{\mathrm{I}_{\mathrm{p}}}(\boldsymbol{\lambda}) = -\partial\hat{\mathbf{S}}_{\mathrm{I}_{\mathrm{p}}}^T(\boldsymbol{\psi},\boldsymbol{\chi};\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}$. The solution to (4.4) gives the estimator $\hat{\boldsymbol{\lambda}}_{\mathrm{p}}$ of $\boldsymbol{\lambda}$.

To update only the vector parameter $\boldsymbol{\beta}$ associated with the response mechanism after observing p phases of data collection, we set $\boldsymbol{\chi}_k = {}^{(e)}t_k^{(rq)}$, $\boldsymbol{\psi}_k = {}^{(t)}t_k^{(rq)}$, $\boldsymbol{\lambda} = \boldsymbol{\beta}$,

$$f_{\mathrm{I}}(\boldsymbol{\chi}_k) = \textstyle\prod_{i=1}^{\mathrm{I}}\{\sum_{j=1}^{\mathrm{I}} {}^{(e|t)}g_{i|j;k}^{(rq)}\, {}^{(t)}g_{j;k}^{(rq)}\}^{{}^{(e)}t_{i;k}^{(rq)}}\,,$$

and

$$f_{\mathrm{I}_{\mathrm{p}}}(\boldsymbol{\psi}_k,\boldsymbol{\chi}_k) = \textstyle\prod_{i=1}^{\mathrm{I}_{\mathrm{p}}}\prod_{j=1}^{\mathrm{I}_{\mathrm{p}}}\{{}^{(e|t)}g_{i|j;k}^{(rq)}\, {}^{(t)}g_{j;k}^{(rq)}\}^{{}^{(e)}t_{i;k}^{(rq)}\,{}^{(t)}t_{j;k}^{(rq)}}\,,$$

where ${}^{(a)}t_{i;k}^{(rq)} = 1({}^{(a)}t_k^{(rq)} = i)$ for $i = 1,\ldots,\mathrm{I}_{\max}$.

After observing the first phase of data collection, the joint distribution is given by $f(\boldsymbol{\psi},\boldsymbol{\chi}) = f_{\mathrm{I}_{\max}}(\boldsymbol{\chi})\, f_{\mathrm{I}_1}(\boldsymbol{\psi} \mid \boldsymbol{\chi})$, and the conditional distribution of ψ given χ is given by

$$f_{\mathrm{I}_1}(\boldsymbol{\psi} \mid \boldsymbol{\chi}) = \frac{f_{\mathrm{I}_1}(\boldsymbol{\chi} \mid \boldsymbol{\psi})}{f_{\mathrm{I}_1}(\boldsymbol{\chi})}\, f_{\mathrm{I}_1}(\boldsymbol{\psi})\,,$$

where the factor $\{f_{\mathrm{I}_1}(\boldsymbol{\chi} \mid \boldsymbol{\psi})/\, f_{\mathrm{I}_1}(\boldsymbol{\chi})\}$ represents the impact of the error-prone information χ on the distribution of the target error-free information ψ. After observing two phases of data collection, we may decompose the conditional distribution as

$$f_{\mathrm{I}_2}(\boldsymbol{\psi} \mid \boldsymbol{\chi}) = f_{\mathrm{I}_1}(\boldsymbol{\psi} \mid \boldsymbol{\chi}) \times \frac{f(\boldsymbol{\chi}_{[\mathrm{I}_1+1,\mathrm{I}_2]} \mid \boldsymbol{\psi}_{[e_k^{dc)}+1,\mathrm{I}_1]}, \boldsymbol{\psi}_{[\mathrm{I}_1+1,\mathrm{I}_2]})}{f(\boldsymbol{\chi}_{[\mathrm{I}_1+1,\mathrm{I}_2]} \mid \boldsymbol{\psi}_{[e_k^{dc)}+1,\mathrm{I}_1]})} f(\boldsymbol{\psi}_{[\mathrm{I}_1+1,\mathrm{I}_2]} \mid \boldsymbol{\psi}_{[e_k^{dc)}+1,\mathrm{I}_1]}) ,$$

where the factor $\{f(\boldsymbol{\chi}_{[\mathrm{I}_1+1,\mathrm{I}_2]} \mid \boldsymbol{\psi}_{[e_k^{dc)}+1,\mathrm{I}_1]}, \boldsymbol{\psi}_{[\mathrm{I}_1+1,\mathrm{I}_2]})\}/\{f(\boldsymbol{\chi}_{[\mathrm{I}_1+1,\mathrm{I}_2]} \mid \boldsymbol{\psi}_{[e_k^{dc)}+1,\mathrm{I}_1]})\}$ represents the partial impact of the information $\boldsymbol{\chi}_{[\mathrm{I}_1+1,\mathrm{I}_2]}$ during phase 2 on the conditional distribution of $\boldsymbol{\psi}_{[\mathrm{I}_1+1,\mathrm{I}_2]}$ during phase 2 given observation $\boldsymbol{\psi}_{[e_k^{dc)}+1,\mathrm{I}_1]}$ during phase 1, provided $e_k^{dc)} < \mathrm{I}_1$.

4.4 Revision – Imputation Step

Our interest here is in estimating the variable ψ by an estimator θ using the well known mean square approach. One may estimate the unknown quantity ψ as follows: a) set the lost function to the square error $SE(\psi,\theta) = (\psi - \theta)^2$; then, b) pick estimate θ to minimize $E_\psi\{SE(\theta,\psi)\}$. If all of the available information on ψ is summarised in its distribution $f(\psi)$, then we need to solve $\min_\theta E_\psi\{SE(\theta,\psi)\} = \min_\theta \int (\psi-\theta)^2 f(\psi) d\psi$. Differentiating with respect to θ gives the optimal estimate $\theta = E_\psi(\psi)$. Thus, the case of no additional information the minimum MSE estimator is simply the expected value $\theta = E_\psi(\psi)$. When we have additional information in the form of observed information χ that is related somehow to ψ . We could use that information to get a better estimation than its mean. A simple model assumes that $\chi = \psi + b$, where b is a random variable that represents error. Suppose that ψ and b are independent, ψ has a normal distribution with mean μ_ψ and variance σ_ψ^2, and b has a normal distribution with mean 0 and variance σ_b^2 . Then the conditional expectation of ψ given χ is $E_\psi(\psi \mid \chi) = \mu_\psi + \sigma_\psi^2(\chi - \mu_\psi)/(\sigma_\psi^2 + \sigma_b^2)$, and the conditional variance of ψ given χ is given by $Var_\psi(\psi \mid \chi) = \sigma_\psi^2\{1 - \sigma_\psi^2/(\sigma_\psi^2 + \sigma_b^2)\}$. When the variance of the error is smaller, ψ becomes closer to χ , i.e. $Var_\psi(\psi \mid \chi)$ becomes closer to 0. Hence, the conditional variance of ψ given χ decreases as χ becomes closer to ψ . More generally, if the form of the joint distribution $f(\chi,\psi)$ is known, then our estimator of ψ given χ is of course some function of χ , say $\theta(\chi)$ and our aim now is to minimize $\min_{\theta(\chi)} E_\psi[SE\{\psi, \theta(\chi) \mid \chi\}] = \min_{\theta(\chi)} \int \{\psi - \theta(\chi)\}^2 f(\psi \mid \chi) d\psi$. Exactly the same calculations as in the case of no additional information then show that $\theta(\chi) = E_\psi(\psi \mid \chi)$, the conditional expectation of ψ given χ . The updated probability distribution of ψ is the conditional probability distribution of ψ given χ . If ψ and χ are independent, then all conditional expectation of ψ are independent of χ , and coincide with the unconditional expectation $E_\psi(\psi)$.

4.5 Revision – Cost Conditional Expectation Step

Consider for simplicity the case of $\mathrm{P} = 2$. In this case, the global cost given by (3.5) reduces to $C(\mathrm{P}) = C_1 + C_2$. After observing the first phase of data collection, the first part $C_1 = c_1 + C_1^{(\wp)} + C_1^{(f)} + C_1^{(dc)}$ is known, while the second part $C_2 = c_2 + C_2^{(f)} + C_2^{(dc)}$ is partially unknown; and the resulting conditional expected cost is given by

$$C\overline{C}\{\boldsymbol{\lambda}_1, \mathbf{A}(\boldsymbol{\psi}_1)\} = C_1 + C\overline{C}_2 ,$$

with
$$C\overline{C}_2 = c_2(\boldsymbol{\lambda}_1) + C\overline{C}_2^{(f)} + C\overline{C}_2^{(dc)} ,$$

where $C\overline{C}_2^{(f)} = \sum_k 1_k(\wp)(1 - r_{1;k}^{(rq)}) \sum_{s=1}^S \phi_{s|2;k}^{(f)}(\boldsymbol{\lambda}_1) c_{s|2;k}^{(f)}(\mathbf{A}(\boldsymbol{\psi}_1))$, and $C\overline{C}_2^{(dc)} = \sum_k 1_k(\wp)(1 - r_{1;k}^{(rq)}) \sum_{s=1}^S \phi_{s|2;k}^{(f)}(\boldsymbol{\lambda}_1) \sum_{m=1}^M \phi_{m|s;k}^{(dc)}(\boldsymbol{\lambda}_1) \xi_{2;k}^{(rq)}(\boldsymbol{\lambda}_1) c_{m|2;k}^{(dc)}(\mathbf{A}(\boldsymbol{\psi}_1))$. Note that the remaining cost is given by $C_{\max} - C_1$.

5. Estimation

5.1 Estimation of the Regression Parameter

Once data collection is completed, observed values of the target information in combination with prior information values are used in the EE given by (4.4) at the estimation stage to get estimate of regression parameter. However it may happens that prior information is ignored in the estimation stage, and only observed values of the target information are used. In this case, the likelihood for unit k is given by

$$L_{\mathrm{P};k}(\boldsymbol{\lambda}) = f_{\mathrm{P}}(\boldsymbol{\psi}_k) ,$$

Taking the derivatives of the log-likelihood, and adjusting for sampling unequal probabilities, we get the weighted EE

$$\hat{\mathbf{S}}_{\mathrm{P}}(\boldsymbol{\psi};\boldsymbol{\lambda}) = \sum_k d_k(\wp)\mathbf{s}(\boldsymbol{\psi}_k;\boldsymbol{\lambda}) = \mathbf{0} , \qquad (5.1)$$

where $\mathbf{s}(\boldsymbol{\psi}_k;\boldsymbol{\lambda}) = \partial \log f_{\mathrm{P}}(\boldsymbol{\psi}_k)/\partial\boldsymbol{\lambda}$. Starting with a guessed value, $\boldsymbol{\lambda}^{(0)}$, then for $b = 1,2,...$ updates are made using

$$\boldsymbol{\lambda}^{(b)} = \boldsymbol{\lambda}^{(b-1)} + \{\hat{\mathbf{J}}_{\mathrm{P}}(\boldsymbol{\psi};\boldsymbol{\lambda}^{(b-1)})\}^{-1}\hat{\mathbf{S}}_{\mathrm{P}}(\boldsymbol{\psi};\boldsymbol{\lambda}^{(b-1)}) ,$$

where $\hat{\mathbf{J}}_{\mathrm{P}}(\boldsymbol{\psi};\boldsymbol{\lambda}) = -\partial\hat{\mathbf{S}}_{\mathrm{P}}^{T}(\boldsymbol{\psi};\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}$. The solution of (5.1) obtained by a Newton-Raphson-type iterative method gives the estimator $\hat{\boldsymbol{\lambda}}$ of ${}^{(t)}\boldsymbol{\lambda}$. Let's consider our case which consists on models for strategy, mode and response behaviour. Since, the mode of data collection is known for a unit that responds at any time of data collection, while it is unknown for a unit that is censored, the likelihood for unit k may be decomposed as

$$L_{\mathrm{P};k}(\boldsymbol{\lambda}) = f(\boldsymbol{J}_k^{(f)})L_{R|\boldsymbol{J}_k^{(f)};k}L_{M|\boldsymbol{J}_k^{(f)};k} ,$$

with $$L_{R|\boldsymbol{J}_k^{(f)};k} = \{\textstyle\prod_{m=1}^{M}[\phi^{(dc)}_{m|\boldsymbol{J}_k^{(f)};k}\, f_{m|\boldsymbol{J}_k^{(f)}}(t_k^{(rq)})]^{J^{(dc)}_{m|s;k}}\}^{\delta_k^{(dc)}} \text{ and } L_{M|\boldsymbol{J}_k^{(f)};k} = \{\textstyle\sum_{m=1}^{M}\phi^{(dc)}_{m|\boldsymbol{J}_k^{(f)};k}\, f_{m|\boldsymbol{J}_k^{(f)}}(t_k^{(rq)})\}^{(1-\delta_k^{(dc)})} ,$$

where $\boldsymbol{\lambda} = (\boldsymbol{\eta}^T, \boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)^T$, $\boldsymbol{J}_k^{(f)} = (\boldsymbol{J}_{|1;k}^{(f)T}, \ldots, \boldsymbol{J}_{|\mathrm{P};k}^{(f)T})^T$ is the stochastic process representing the evolution of nonresponse follow-up strategies over data collection phases, $\boldsymbol{J}_{|\mathrm{p};k}^{(f)}$ indicates which follow-up strategy is assigned to unit k at phase p, and $f_{m|\boldsymbol{J}_k^{(f)}}(t_k^{(rq)})$ is $f(t_k^{(rq)})$ for mode m of data collection under follow-up stochastic process $\boldsymbol{J}_k^{(f)}$. It remains to specify the joint probability distribution of the random process $\boldsymbol{J}_k^{(f)} = (\boldsymbol{J}_{|1;k}^{(f)T}, \ldots, \boldsymbol{J}_{|\mathrm{P};k}^{(f)T})^T$. It maybe informative in our context to use Markov chains model with S states. Markov chains model is characterized by – the probability of the next follow-up strategy depends only on the current follow-up strategy and not on the sequence of follow-up strategies that proceed it – the matrix which consists of the conditional probability defined by its components

$$\Pr(J_{i|\mathrm{p}+1;k}^{(f)} = 1 \mid J_{j|\mathrm{p};k}^{(f)} = 1) ,\ i, j = 1, \ldots, S \text{ , and } \mathrm{p} = 1, \ldots, \mathrm{P} - 1 .$$

An excellent source of information on Marcov Chains estimation is the volume by Brémaud (1998).

5.2 Estimation of the Parameter of Interest

Since the probability of response is unknown, estimated response probability $\hat{\xi}_{\mathrm{P};k} = \xi_{\mathrm{P};k}(\hat{\boldsymbol{\lambda}})$ is used in the EE given by (3.2) to get

$$\hat{\mathbf{S}}(\boldsymbol{\psi};\boldsymbol{\Theta}) = \textstyle\sum_k d_k(\wp)(r_{\mathrm{P};k}^{(rq)} / \hat{\xi}_{\mathrm{P};k}^{(rq)})\mathbf{s}(\boldsymbol{\psi}_k;\boldsymbol{\Theta}) - \mathbf{v}(\boldsymbol{\Theta}) = \mathbf{0} , \qquad (5.2)$$

and the solution gives the estimator $\hat{\boldsymbol{\Theta}}_{\mathrm{P}}(\boldsymbol{\psi})$ of $\boldsymbol{\Theta}_N(\boldsymbol{\psi})$. As noted by Rosenbaum (1987) and others, estimator $\hat{\boldsymbol{\Theta}}_{\mathrm{P}}(\boldsymbol{\psi})$ using the estimated response probability can be more efficient than estimator $\breve{\boldsymbol{\Theta}}_{\mathrm{P}}(\boldsymbol{\psi})$ using the true response probability.

If prior information values are used, then the likelihood for unit k in given by

$$L_{\mathrm{P};k}(\boldsymbol{\Theta}) = \{f_{\mathrm{I}_{\max}}(\boldsymbol{\chi}_k)\}^{1-1_k(\wp)r_{\mathrm{P};k}}\, f_{\mathrm{I}_{\mathrm{P}}}(\boldsymbol{\psi}_k, \boldsymbol{\chi}_k)^{1_k(\wp)r_{\mathrm{P};k}} ,$$

in which case the weighed EE is given by

$$\hat{\mathbf{S}}_{\mathrm{P}}(\boldsymbol{\psi}, \boldsymbol{\chi};\boldsymbol{\Theta}) = \textstyle\sum_k \mathbf{s}_{\mathrm{I}_{\max}}(\boldsymbol{\chi}_k;\boldsymbol{\Theta}) + \sum_k d_k(\wp)(r_{\mathrm{P};k} / \hat{\xi}_{\mathrm{P};k})\mathbf{s}_{\mathrm{I}_{\mathrm{P}}}(\boldsymbol{\psi}_k;\boldsymbol{\Theta} \mid \boldsymbol{\chi}_k) = \mathbf{0} ,$$

where $$\mathbf{s}_{\mathrm{I}_{\max}}(\boldsymbol{\chi}_k;\boldsymbol{\Theta}) = \partial \log f_{\mathrm{I}_{\max}}(\boldsymbol{\chi}_k)/\partial\boldsymbol{\Theta} \text{ and } \mathbf{s}_{\mathrm{I}_{\mathrm{P}}}(\boldsymbol{\psi}_k;\boldsymbol{\Theta} \mid \boldsymbol{\chi}_k) = \partial \log f_{\mathrm{I}_{\mathrm{P}}}(\boldsymbol{\psi}_k \mid \boldsymbol{\chi}_k)/\partial\boldsymbol{\Theta} .$$

For the simple model defined in Section 4.4, where the vectors $(\chi_k, \psi_k)^T$ are realizations of independent distributed random variables according to a bivariate Normal distribution,

$$\begin{pmatrix} \chi_k \\ \psi_k \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_\psi \\ \mu_\psi \end{pmatrix}, \begin{pmatrix} \sigma_\psi^2 + \sigma_b^2 & \sigma_\psi^2 \\ \sigma_\psi^2 & \sigma_\psi^2 \end{pmatrix}\right),$$

we have $\mathbf{s}_{\mathrm{I}_{\max}}(\boldsymbol{\chi}_k;\boldsymbol{\Theta}) = (\chi_k - \mu_\psi)\rho_{\psi\chi}^2 / \sigma_\psi^2$, and $\mathbf{s}_{\mathrm{I}_{\mathrm{P}}}(\boldsymbol{\psi}_k;\boldsymbol{\Theta} \mid \boldsymbol{\chi}_k) = \{\psi_k - \mu_\psi - \rho_{\psi\chi}^2(\chi_k - \mu_\psi)\}/\{(1 - \rho_{\psi\chi}^2)\sigma_\psi^2\}$, provided $\rho_{\psi\chi}^2 < 1$, where $\rho_{\psi\chi}^2 = \sigma_\psi^2/(\sigma_\psi^2 + \sigma_b^2)$. If $\rho_{\psi\chi} = 0$, then the EE reduces to $\hat{\mathbf{S}}_{\mathrm{P}}(\boldsymbol{\psi}, \boldsymbol{\chi};\boldsymbol{\Theta}) = \sum_k d_k(\wp)(r_{\mathrm{P};k}^{(rq)} / \hat{\xi}_{\mathrm{P};k}^{(rq)})(\psi_k - \mu_\psi)/\sigma_\psi^2 = 0$.

6. Quick Look at Surveys Requiring Contactibility and Eligibility Identification

The process of determining the contactibility and the eligibility status of each sampled unit is another stage of the survey design where the side effect of prior information can be reduced during its phase of follow-up to establish a first contact. In the following, we sketch the necessitate ingredients.

6.1 Contactibility Models

Let ${}^{(a)}t^{(fc)}$ represent the discrete random variable that indicates the time period i when the first contact occurs under random process a, where the superscript "fc" stands for "first contact". The process of follow-up to reach unit k at specified time periods is conducted until some period ${}^{(a)}\mathrm{I}_k^{(fc)}$, with ${}^{(a)}\mathrm{I}_k^{(fc)} \le \mathrm{I}_{\max}^{(fc)}$, where $\mathrm{I}_{\max}^{(fc)}$ is the maximum time periods for follow-up to get a first contact, and $\mathrm{I}_{\max}^{(fc)} < \mathrm{I}_{\max}$. The process of reaching a first contact could be discontinued for two reasons: 1) the unit is contacted for the first time; or 2) the period to reach sampled unit ends. In the first case, ${}^{(a)}t_k^{(fc)} = {}^{(a)}\mathrm{I}_k^{(fc)}$. In the second case, it is only known that ${}^{(a)}t_k^{(fc)} > \mathrm{I}_{\max}^{(fc)}$. Units with ${}^{(a)}t_k^{(fc)} > \mathrm{I}_{\max}^{(fc)}$ are right-censored – it is unknown whether they are contactable during survey data collection period. The overall contactibility indicator over $\mathrm{I}_k^{(fc)}$ time periods is given by ${}^{(a)}r_k^{(fc)} = 1 - \prod_{i=1}^{\mathrm{I}_k^{(fc)}} (1 - {}^{(a)}r_{ki}^{(fc)})^{z_{ki}^{(fc)}}$, where ${}^{(a)}r_{ki}^{(fc)}$ is a sequence of contactibility indicators defined for each unit k whose values are defined as ${}^{(a)}r_{ki}^{(fc)} = 1$ if the unit is contacted in time period i, and ${}^{(a)}r_{ki}^{(fc)} = 0$ if not, and ${}^{(a)}z_{ki}^{(fc)}$ denote the 0/1 variable indicating whether a unit k is to be contacted at time period i. Because contactibility occurrence is intrinsically conditional, we characterized ${}^{(a)}t^{(fc)}$ by its discrete-time hazard function

$$ {}^{(a)}h_{ki}^{(fc)} = \Pr({}^{(a)}t_k^{(fc)} = i \mid {}^{(a)}t_k^{(fc)} \ge i), $$

where ${}^{(a)}h_{ki}^{(fc)} = {}^{(a)}h_{ki}^{(fc)}({}^{(a)}\boldsymbol{v}_{fc;k}, {}^{(a)}\boldsymbol{\beta}^{(fc)})$, ${}^{(a)}\boldsymbol{v}_{fc;k}$ refers to both time-invariant and time-varying explanatory variables and ${}^{(a)}\boldsymbol{\beta}^{(fc)}$ is the unknown ${}^{(a)}q_{fc} \times 1$ vector parameter to be estimated. For unit with ${}^{(a)}t_k^{(fc)} = i$, the probability of obtaining a first contact at time period i could be expressed in terms of the hazard as

$$ \Pr({}^{(a)}t_k^{(fc)} = i) = {}^{(a)}z_{ki}^{(fc)}\,{}^{(a)}h_{ki}^{(fc)} \prod_{j=1}^{i-1} (1 - {}^{(a)}h_{kj}^{(fc)})^{{}^{(a)}z_{ki}^{(fc)}} . \tag{6.1} $$

For units with ${}^{(a)}t_k^{(fc)} > i$, the probability of obtaining a response can be expressed as

$$ \Pr({}^{(a)}t_k^{(fc)} > i) = \prod_{j=1}^{i} (1 - {}^{(a)}h_{kj}^{(fc)})^{{}^{(a)}z_{kj}^{(fc)}} . $$

We have

$$ g({}^{(a)}t_k^{(fc)} = {}^{(a)}\mathrm{I}_k^{(fc)}) = \{\Pr({}^{(a)}t_k^{(fc)} = {}^{(a)}\mathrm{I}_k^{(fc)})\}^{{}^{(a)}\delta_k^{(fc)}} \{\Pr({}^{(a)}t_k^{(fc)} > {}^{(a)}\mathrm{I}_k^{(fc)})\}^{1 - {}^{(a)}\delta_k^{(fc)}} , $$

where ${}^{(a)}\delta_k^{(fc)} = 1$ if unit k is uncensored (contacted) and ${}^{(a)}\delta_k^{(fc)} = 0$ if unit k is censored under process a. When unit k is censored, either unit k will be reachable at some future time period ${}^{(a)}t_k^{(fc)} > \mathrm{I}_{\max}^{(fc)}$ or the unit will not be contactable during survey period. The joint distribution of $({}^{(e)}t_k^{(fc)}, {}^{(t)}t_k^{(fc)})$ is characterized by the matrix ${}^{(e|t)}\mathbf{G}_k^{(fc)}$ defined by its components

$$ {}^{(e|t)}g_{i|j;k}^{(fc)} = \Pr({}^{(e)}t_k^{(fc)} = i \mid {}^{(t)}t_k^{(fc)} = j) , \; i, j = 1, \ldots, \mathrm{I}_{\max}^{(fc)} . $$

The marginal probability of obtaining a first contact after $\mathrm{I}_{\max}^{(fc)}$ time periods is given by

$$ \Pr({}^{(a)}r_k^{(fc)} = 1) = 1 - \prod_{i=1}^{\mathrm{I}_{\max}^{(fc)}} (1 - {}^{(a)}h_{ki}^{(fc)})^{{}^{(a)}z_{ki}} . $$

6.2 Parameter of Interest and Associated Estimator

We now define the census parameter for eligible subpopulation as the solution to

$$ \mathbf{S}(\boldsymbol{\psi};\boldsymbol{\Theta}) = \sum_k J_k^{(el)} \mathbf{s}(\boldsymbol{\psi}_k;\boldsymbol{\Theta}) - \mathbf{v}(\boldsymbol{\Theta}) = \mathbf{0} , \tag{6.2} $$

where $J_k^{(el)}$ denotes the eligibility indicator for unit k, i.e., $J_k^{(el)} = 1$ if unit k is eligible for the given survey and $J_k^{(el)} = 0$ if unit k is not eligible. Suppose that the probability, $\xi_k = E(\Delta_k)$, of getting a response to questionnaire during P phases of follow-ups to both establish a contact and get response to questionnaire is known for every eligible unit in the population, where $\Delta_k = r_k^{(fc)} r_k^{(rq)}$, $r_k^{(fc)}$ is the overall contactibility indicator, and $r_k^{(rq)}$ is the overall response indicator for eligible unit k. A design-response unbiased estimator of the EE defined by (6.2) is given by

$$ \breve{\mathbf{S}}(\boldsymbol{\psi};\boldsymbol{\Theta}) = \sum_k J_k^{(el)} d_k(\wp)(\Delta_k / \xi_k)\mathbf{s}(\boldsymbol{\psi}_k;\boldsymbol{\Theta}) - \mathbf{v}(\boldsymbol{\Theta}) = \mathbf{0} , \tag{6.3} $$

where the probability of getting response to questionnaire for eligible unit k is given by

$$ \xi_k = \sum_{i=1}^{\mathrm{I}_{\max}^{(fc)}} \Pr(t_k^{(fc)} = i)\{1 - \prod_{j=i+1}^{\mathrm{I}_{max}} (1 - h_{kj}^{(rq)})\} , $$

and $\Pr(t_k^{(fc)} = i)$ is given by (6.1).

The variance of the estimator obtained as the solution to (6.3) can be derived along the lines of Section 3.3 by replacing $r_{\mathrm{P};k}^{(rq)}$ and $\xi_{\mathrm{P};k}^{(rq)}$ by Δ_k and ξ_k respectively.

6.3 Specification of the Cost Function

We may decompose the global cost over P phases of follow-ups and data collection as

$$ C(\mathrm{P}) = \sum_{\mathrm{p}=1}^{\mathrm{P}} C_{\mathrm{p}} , $$

with $1 \le \mathrm{P} \le \mathrm{P}_{\max}$. The P components of the global cost are

$$C_1 = c_1 + C_1^{(\wp)} + C_1^{(fc)} + C_1^{(f)} + C_1^{(dc)},$$

and for $\mathrm{p} = 2,\ldots,\mathrm{P}$

$$C_\mathrm{p} = c_\mathrm{p} + C_\mathrm{p}^{(fc)} + C_\mathrm{p}^{(f)} + C_\mathrm{p}^{(dc)},$$

where the extra component associated with the establishment of a first contact is $C_\mathrm{p}^{(fc)} = \sum_k 1_k(\wp)(1 - r_{\mathrm{p\text{-}1};k}^{(fc)}) \sum_{i=\mathrm{I}_{\mathrm{p}-1}+1}^{\mathrm{I}_\mathrm{p}} (1 - r_{i\text{-}1;k}^{(fc)}) z_{ki}^{(fc)} \{c_{ki}^{(ffc)} + r_{ki}^{(fc)} c_{ki}^{(rfc)}\}$ for $1 \le \mathrm{p} \le \mathrm{P}_{\max}^{(fc)}$ and $C_\mathrm{p}^{(fc)} = 0$ for $\mathrm{P}_{\max}^{(fc)} < \mathrm{p} \le \mathrm{P}_{\max}$, with $r_{0;k}^{(fc)} = 0$ and $\mathrm{I}_0 = 0$, $c_{ki}^{(ffc)}$ is the cost associated with follow-up to get a first contact at time period i, and $c_{ki}^{(rfc)}$ is the cost associated with first contact data collection at time period i. Then the two remaining components of the global cost are adjusted as follows: $C_\mathrm{p}^{(f)} = \sum_k 1_k(\wp) r_{\mathrm{p\text{-}1};k}^{(fc)} J_k^{(el)} (1 - r_{\mathrm{p\text{-}1};k}^{(rq)}) \sum_{s=1}^{S} J_{s|\mathrm{p};k}^{(f)} c_{s|\mathrm{p};k}^{(f)}$, and $C_\mathrm{p}^{(dc)} = \sum_k 1_k(\wp) r_{\mathrm{p\text{-}1};k}^{(fc)} J_k^{(el)} (1 - r_{\mathrm{p\text{-}1};k}^{(rq)}) \sum_{s=1}^{S} J_{s|\mathrm{p};k}^{(f)} \sum_{m=1}^{M} J_{m|s;k}^{(dc)} r_{\mathrm{p};k}^{(rq)} c_{m|\mathrm{p};k}^{(dc)}$ for $1 \le \mathrm{p} \le \mathrm{P}_{\max}$.

6.4 Estimation of the Regression Parameter

The eligibility indicator $J_k^{(el)}$, with $J_k^{(el)} \in \{0,1\}$, is known for a unit that have been contacted during any time of the contactibility period, while it is unknown for a unit that is censored, where the superscript "el" stands for "eligible". Consequently the likelihood for the census observed data associated with (4.1) may be decomposed as

$$L_k^{(all)} = f(z_k) L_{R|z_k;k}^{(all)} L_{M|z_k;k}^{(all)},$$

with $L_{R|z_k;k}^{(all)} = \{\prod_{i=0}^{1} \phi_{i;k}^{(el)} f(t_k^{(fc)} \mid J_{i;k}^{(el)}) L_k(\lambda \mid J_{i;k}^{(el)})]^{J_{i;k}^{(fc)}}\}^{\delta_k^{(fc)}}$ and $L_{M|z_k;k}^{(all)} = \{\sum_{i=0}^{1} [\phi_{i;k}^{(el)} f(t_k^{(fc)} \mid J_{i;k}^{(el)})]^{J_{i;k}^{(fc)}}\}^{1-\delta_k^{(fc)}}$,

where $J_{i;k}^{(el)} = 1(J_k^{(el)} = i)$, $\phi_{i;k}^{(el)} = \Pr(J_k^{(el)} = i)$, z_k is the stochastic process representing the evolution of follow-up to get a first contact, $L_k(\lambda \mid J_{i;k}^{(el)}) = L_k(\lambda)$ for illegible units, $L_k(\lambda \mid J_{i;k}^{(el)}) = 1$ for ineligible units, and $L_k(\lambda)$ is given by (4.1). In likelihood components that incorporate a contactibility piece, appropriate conditioning would be assumed.

7. Coefficient of Reliability

To start our discussion on the reliability coefficient, suppose interest is in estimating the variable ψ by an estimator θ rather than evaluating the reliability of an estimator. In the case of no additional information, the minimum MSE of $\theta = E(\psi)$ is the variance of ψ, namely $Var(\psi) = E\{[\psi - E(\psi)]^2\}$; while in the case in which we have additional information, the min MSE of $\theta(\chi) = E(\psi \mid \chi)$ is also the variance, $Var(\psi \mid \chi) = E\{[\psi - E(\psi \mid \chi)]^2 \mid \chi\}$, but of the conditional density $f(\psi \mid \chi)$. This conditional variance characterizes the spread of ψ about its conditional expectation $E(\psi \mid \chi)$ for a given value of χ. If χ and ψ are independent, then $Var(\psi \mid \chi) = Var(\psi)$. The remaining relative error (or the missed information) of ψ based on the knowledge of χ is given by $Var(\psi \mid \chi)/Var(\psi)$; so that the proportion of knowledge (or the attained information) about ψ obtained after observing χ constitutes our coefficient of reliability given by

$$\mathrm{K}\{\psi;\chi\} = 1 - \frac{Var(\psi \mid \chi)}{Var(\psi)}. \tag{7.1}$$

If $Var(\psi \mid \chi) = Var(\psi)$ then $\mathrm{K}\{\psi;\chi\} = 0$, and if $Var(\psi \mid \chi) = 0$ then $\mathrm{K}\{\psi;\chi\} = 1$.

Suppose that ψ and χ have a correlation coefficient $\rho_{\psi\chi}$, ψ has a normal distribution with mean μ_ψ and variance σ_ψ^2, and χ has a normal distribution with mean μ_χ and variance σ_χ^2. Then the conditional distribution has

$$E(\psi \mid \chi) = \mu_\psi + \rho_{\psi\chi} \sigma_\psi (\chi - \mu_\chi)/\sigma_\chi, \tag{7.2}$$

and

$$Var(\psi \mid \chi) = \sigma_\psi^2 (1 - \rho_{\psi\chi}^2).$$

As shown by Goldberger (1962), the linear estimator given by (7.2) is the best linear unbiased predictor of ψ under the general linear model. In this case under the normality assumption, the coefficient of reliability of ψ based on the knowledge χ given by (7.1) reduces to

$$\mathrm{K}\{\psi;\chi\} = \rho_{\psi\chi}^2 = \left(\frac{Cov(\psi,\chi)}{\sigma_\psi \sigma_\chi}\right)^2. \tag{7.3}$$

Tenenbein (1970) introduced the square of the correlation coefficient given by (7.3) as a measure of reliability between the error-prone and error-free classification variables to measure the strength of the relationship between the true and fallible classifications; i.e., it measures how well the true classification can be predicted from the fallible classification on a given sampling unit. Expression (7.3) gives a convenient way to derive the coefficient of reliability: It is reasonable in practice to replace conditional variance, which depends on the joint distribution, with correlation which can be calculated more easily. That being said, conditional independence is more meaningful and preferable than zero-correlation.

For the survey case, where the estimator $\hat{\Theta}_{\mathrm{P}}(\boldsymbol{\psi})$ of our gold standard $\Theta_N(\boldsymbol{\psi})$ is obtained as the solution of (5.2), the coefficient of reliability under the normality assumption is

$$\mathrm{K}\{\Theta_N(\boldsymbol{\psi});\hat{\Theta}_{\mathrm{P}}(\boldsymbol{\psi})\} = \rho^2_{\Theta_N(\boldsymbol{\psi})\hat{\Theta}_{\mathrm{P}}(\boldsymbol{\psi})} = \frac{\{Cov(\Theta_N(\boldsymbol{\psi}),\hat{\Theta}_{\mathrm{P}}(\boldsymbol{\psi}))\}^2}{Var(\Theta_N(\boldsymbol{\psi}))Var(\hat{\Theta}_{\mathrm{P}}(\boldsymbol{\psi}))}\ .$$

Suppose $\Theta_N(\boldsymbol{\psi}) = \sum_k y_k$ and $\hat{\Theta}_{\mathrm{P}}(\boldsymbol{\psi}) = \sum_k F_k y_k$, where F_k is the frame membership indicator for unit k. We have under independent observations, $Var(\Theta_N(\boldsymbol{\psi})) = N\sigma^2_y$, $Var(\hat{\Theta}_{\mathrm{P}}(\boldsymbol{\psi})) = N_F\sigma^2_y$, $Cov(\Theta_N(\boldsymbol{\psi}),\hat{\Theta}_{\mathrm{P}}(\boldsymbol{\psi})) = N_F\sigma^2_y$, and $\mathrm{K}\{\Theta_N(\boldsymbol{\psi});\hat{\Theta}_{\mathrm{P}}(\boldsymbol{\psi})\} = N_F/N$, where $N_F = \sum_k F_k$ and σ^2_y denotes the variance of the variable y. If $\hat{\Theta}_{\mathrm{P}}(\boldsymbol{\psi}) = \sum_k F_k v_k$ is instead used to estimate $\Theta_N(\boldsymbol{\psi})$, then $\mathrm{K}\{\Theta_N(\boldsymbol{\psi});\hat{\Theta}_{\mathrm{P}}(\boldsymbol{\psi})\} = \rho^2_{yv}(N_F/N)$, where ρ_{yv} denotes the correlation coefficient between y and v. If $\hat{\Theta}_{\mathrm{P}}(\boldsymbol{\psi}) = \sum_k d_k(\wp)F_k v_k$ under simple random sampling is used to estimate $\Theta_N(\boldsymbol{\psi})$, then $\mathrm{K}\{\Theta_N(\boldsymbol{\psi});\hat{\Theta}_{\mathrm{P}}(\boldsymbol{\psi})\} = \rho^2_{yv}(n_F/N)$, where $Var(\hat{\Theta}_{\mathrm{P}}(\boldsymbol{\psi})) = N_F\sigma^2_v + N_F\sigma^2_v(N_F - n_F)/n_F$, σ^2_v denotes the variance of the variable v, and n_F is the sample size. Finally, if $\hat{\Theta}_{\mathrm{P}}(\boldsymbol{\psi}) = \sum_k d_k(\wp)F_k v_k$ under Bernoulli sampling is used to estimate $\Theta_N(\boldsymbol{\psi})$, then $\mathrm{K}\{\Theta_N(\boldsymbol{\psi});\hat{\Theta}_{\mathrm{P}}(\boldsymbol{\psi})\} = \rho^2_{yv}(n_F/N)\{1+(1-\pi_F)(\mu_v/\sigma_v)^2\}^{-1}$, where $\pi_F = n_F/N_F$, n_F is the expected sample size, and μ_v is the mean of the variable v.

Suppose now a third set of information $\aleph$ was available previously to χ, with $E(\aleph) = \mu_\aleph$, $Var(\aleph) = \sigma^2_\aleph$, $\rho(\psi,\aleph) = \rho_{\psi\aleph}$, and $\rho(\chi,\aleph) = \rho_{2\aleph}$. Then, the conditional variance-covariance matrix of (ψ,χ) given $\aleph$ under the normality assumption is given by

$$\Sigma_{|\aleph} = \begin{bmatrix} \sigma^2_\psi(1-\rho^2_{\psi\aleph}) & \sigma_\psi\sigma_\chi(\rho_{\psi\chi} - \rho_{\psi\aleph}\rho_{\chi\aleph}) \\ \sigma_\psi\sigma_\chi(\rho_{\psi\chi} - \rho_{\psi\aleph}\rho_{\chi\aleph}) & \sigma^2_\chi(1-\rho^2_{\chi\aleph}) \end{bmatrix},$$

and the resulting partial coefficient of reliability after removing the effect of $\aleph$ from each variable is

$$\mathrm{K}\{\psi;\chi\,|\,\aleph\} = \frac{\{\rho_{\psi\chi} - \rho_{\psi\aleph}\rho_{\chi\aleph}\}^2}{(1-\rho^2_{\psi\aleph})(1-\rho^2_{\chi\aleph})}\ ,$$

provided $|\,\rho_{\psi\aleph}| < 1$ and $|\,\rho_{\chi\aleph}| < 1$.

The quality of an estimate is usually characterized by MSE. However, coefficients of reliability for each survey process as well as for a survey as a whole set of processes when supplied with MSE enhance the quality of information on a) the survey results; b) the comparisons between surveys; and c) the contribution of the given survey as addition to prior information.

Concluding Remarks

We formulated an optimization problem for designing a survey, and we identified steps for its revision in the course of the data collection period. We considered the error-prone and error-free information as a random variable with a joint distribution with some probability function. Then, we updated the joint probability distribution after observing some of realizations of the error-free random process at each phase of data collection, to revise the design specification in the course of the data collection period. The proposed approach makes full use of error-prone information while requiring only few observations from the error-free and expensive random process. Since revision of a design indicates when a design is nearly "optimal", and how the error-free random process varies from the error-prone random process, the revision of the design has an important role to play in survey quality and cost. In the case of insufficient sample size on prior information grounds, our approach can be extended to increase the sample in the course of data collection. Details are omitted to focus on the response mechanism. A reliability coefficient for a survey as a whole set of processes, as well as for a single process, is also discussed. Such a coefficient when supplied with MSE enhances information on a) the survey results; b) the comparisons between surveys; and c) the contribution of the given survey as addition to prior information.

Acknowledgement

I thank Mary Thompson Distinguished Professor from the University of Waterloo for many helpful comments made during the development of this manuscript.

References

Brémaud, P. 1998. *Markov chains: Gibbs fields, Monte Carlo simulation, and Queues*. Springer.

Demnati, A. 2015. Linearization variance estimators for mixed-mode survey data when response indicators are modeled as discrete-time survival. In *Proceeding of Federal Committee on Statistical Methodology Research Conference.*

Goldberger, A.S. 1962. Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association*, **57**, pp. 369-375.

Groves, R.M. and S.G. Heeringa. 2006. Responsive design for household surveys: tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society*, Series A, **169**, pp. 439-457.

Lagakos, S.W. 1979. General right censoring and its impact on the analysis of survival data. *Biometrics*, **35**, pp. 139–156.

Rosenbaum, P. R. 1987. Model-based direct adjustment. *Journal of the American Statistical Association*, **82**, pp. 387–394.

Tenenbein, A. 1970. A double sampling scheme for estimating from binomial data with misclassifications. *Journal of the American Statistical Association*, **331**, pp. 1350–1361.