2015 Federal Committee on Statistical Methodology Research Conference

Linearization Variance Estimators for Mixed-mode Survey Data when Response Indicators are Modeled as Discrete-time Survival

A. Demnati

Independent Consultant

Ottawa, Canada, Abdellatif_Demnati@msn.com

Abstract

Collecting information from sampled units over the Internet or by mail is much more cost–efficient than conducting interviews. These methods make self–enumeration an attractive data–collection method for surveys and censuses. Despite the benefits associated with self–enumeration data collection—in particular Internet-based data collection— self–enumeration can produce low response rates compared to interviews. To increase response rates, non–respondents are subject to a mixed mode of follow–up treatments, which influence the resulting probability of response, to encourage them to participate. Because response occurrence is intrinsically conditional, we preliminary record response occurrence in discrete intervals, and we then characterize the probability of response by a discrete time hazard. This approach facilitates examining when a response is most likely to occur and how the probability of responding varies over both time and follow–up treatments. We use regression analysis to investigate the effect of mixed–mode on the response probability. Factors and interactions are commonly treated in regression analyses, and have important implications for the interpretation of statistical models. The nonresponse bias can be avoided by multiplying the sampling weight of respondents by the inverse of an estimate of the response probability. Estimators and associated variance estimators of model parameters as well as of parameters of interest are studied. We take into account correlation over time for the same unit in variance estimation. The problem of optimal resources allocation within stages of the survey design is also investigated.

Key Words: Event history analysis; Longitudinal data; Maximum likelihood; Optimal resources allocation; Partially classified units.

1. Introduction

Mixing modes of follow–up and data collection offers the possibility of offsetting the disadvantages of one mode with the advantages of another. For example, recognizing that the Internet, unlike mail, offers the ability to move data capture and editing closer to the respondent, many statistical agencies are now offering electronic questionnaires as a voluntary option to both improve quality of statistical processes and reduce survey costs. This potential increase in survey quality in combination with the fact that collecting information from sampled units over the Internet or by mail is much more cost–effective than conducting interviews makes self–enumeration an attractive data–collection method for surveys and censuses. Although there are benefits associated with self–enumeration, in particular Internet–based surveys, as well as an expected wider application of this approach in future, self–enumeration brings particular difficulties to surveys and censuses. Observed values of typical variable of interest y might depend on the variable y_m associated with mode m of data collection, $m=1,...,M$, where M is the number of modes of data collection under consideration for a given survey. In principle, each unit k of the finite population P of size N has all responses, i.e., a response $y_{m;k}$ that would have resulted if it had chosen mode m. Since each unit receives or chooses only one mode, only one response is observed. If the variable of interest is defined uniquely and independently from each mode, then $y_{m;k}$ represents the value the unit k believes is the correct answer to y, resulting from the medium of mode m in which the question is presented to the unit. We assume that census parameter $\mathbf{\Theta}_N(y)$ associated with the variable of interest y is defined as solution to an estimating equation (EE) of the form

$$\mathbf{S}(\mathbf{\Theta}) = \sum_k \mathbf{s}(y_k;\mathbf{\Theta}) - \mathbf{v}(\mathbf{\Theta}) = \mathbf{0}\,, \tag{1.1}$$

where $\sum_k$ is the sum over all the population units, the known function $\mathbf{s}(y_k;\mathbf{\Theta})$ is a q_Θ –dimensional vector–valued function of y_k and the know function $\mathbf{v}(\mathbf{\Theta})$ allows for explicitly defined parameters. For linear and logistic regression models, $\mathbf{s}(y_k;\mathbf{\Theta}) = \boldsymbol{\chi}_k(y_k - \mu_k(\boldsymbol{\chi}_k^T\mathbf{\Theta}))$ and $\mathbf{v}(\mathbf{\Theta}) = \mathbf{0}$, where $\mu(\boldsymbol{\chi}^T\mathbf{\Theta}) = E_y(y)$, $\boldsymbol{\chi} = (\chi_1,...,\chi_{q_\Theta})^T$ is a $q_\Theta \times 1$ vector of explanatory variables, $\mathbf{\Theta} = (\Theta_1,...,\Theta_{q_\Theta})^T$ is the $q_\Theta \times 1$ vector of model parameter and E_y denotes model expectation. For the special case of the finite population total $Y = \sum_k y_k$, $\mathbf{s}(y_k;\mathbf{\Theta}) = y_k$, $\mathbf{s}(\mathbf{\Theta}) = \Theta_N$ and $\Theta_N = Y$.

Similarly, the m^{th} EE of the vector parameter $\mathbf{\Theta}_{m;N}(y_m)$ $\mathbf{\Theta}_m$ associated with mode m is of the form

$$\mathbf{S}(\mathbf{\Theta}_m) = \sum_k \mathbf{s}(y_{m;k};\mathbf{\Theta}_m) - \mathbf{v}(\mathbf{\Theta}_m) = \mathbf{0}\,,\ m=1,...,M\,. \tag{1.2}$$

The solution to (1.2) constitutes both the census vector parameter $\mathbf{\Theta}_{m;N}(y_m)$ and an estimator of the model parameter $\mathbf{\Theta}_m$.

One of the main objectives of the mixed–mode of data collection is to influence the unit to get its cooperation, regardless of its preference for data–collection mode. The overall response indicator for unit k for the combined modes can be defined as $r_k = 1 - \prod_{m=1}^M (1 - r_{m;k}) = \sum_{m=1}^M r_{m;k}$ and the overall probability of response can be represented in the mixture form $\xi_k = \sum_{m=1}^M \phi_{m;k}^{(dc)} \xi_{m;k}$, where $r_{m;k}$ is the response indicator by mode m, $\phi_{m;k}^{(dc)}$ is the probability that respondent k uses mode m, $\xi_{m;k} = E_r(r_{m;k})$ is the associated response probability, E_r denotes expectation with respect to the response mechanism and the superscript (dc) stands for data collection. If only mode m is assigned to unit k then $\phi_{m;k}^{(dc)} = 1$ and $\phi_{i;k}^{(dc)} = 0$ for $i \neq m$. If the mixed–mode of data collection can increase the overall response rates, we will, of course, be pleased to quantify and examine the contribution of each mode on the response probability. In reality, self–enumeration can produce low response rates in comparison to interviews. To gain non–respondents' cooperation and therefore maximize survey quality, each non–respondent is assigned to a follow–up strategy, where each strategy consists of a mixed–mode of predefined follow–up treatments. Different costs are associated with different follow–up treatments. For example, face–to–face follow–up is more expensive than telephone follow–up. Currently, in some business surveys, to reduce the global cost of data collection, follow–up for nonresponse is performed on only a portion of non–respondents. These units are often identified in a deterministic way based, for example, on their expected contribution to the estimate. In addition, since a significant number of units are never followed up for nonresponse, the final response rate can be very low. The nonresponse bias can be avoided by multiplying the sampling weight of respondents by the inverse of the response probability. Since the response probability is unknown, estimated probability is used. As noted by Rosenbaum (1987) and others, estimators using the estimated response probability can be more efficient than estimators using the true response probability.

Given the above issues, one question should first be of particular interest to statistical agencies: How should both the response probabilities under mixed–mode and the influence of the follow–up treatments on the resulting probability of response be modelled? Other relevant questions include the following: If one factor of the mixed–mode is improved, what will be the effect on the performance of the response mechanism? How can we estimate the response probability due to a particular mixed–mode factor of interest with the presence of the other mixed–mode factor? As intensive follow–up is expensive, a follow–up strategy is needed to make wise use of global resources compared to the quality of the estimates. Since a follow–up treatment could produce estimates with better quality, the strategy should consist of allocating non–respondents to different treatments while controlling for data collection costs. In an attempt to discuss some of these and other issues, we first characterized the response probability by a discrete–time hazard (Demnati, 2014) and we used regression analysis to investigate the effect of mixed–mode on the response probability. We estimated the regression (or nuisance) parameter using the EM algorithm (Hartley, 1958; Dempster et al., 1977) to then estimate the parameter of interest. Our present work below is organized as follows: in Section 2, we briefly review the discrete–time hazard approach to the analysis of response indicators; in Section 3, estimators of the regression parameter are derived using Newton–Raphson iterative method, and estimators of the parameters of interest under mixed–mode surveys are studied; and, in Section 4, linearization variance estimators are studied and optimal resources allocation within stages of the survey design is determined.

2. Modeling Response Indicators as Discrete–time Hazard

2.1 Discrete–time Hazard

Consider a homogeneous sample of units, each at risk of experiencing a single target event, − responding. The target event is nonrepeatable. To record response occurrence in discrete intervals, we divided continuous time of the entire data collection period into a sequence of continuous time periods: 1, 2, and so on. Suppose the duration of data collection is made up of I time periods. Let t represent the discrete random variable that indicates the time period i when the response occurs for a randomly selected unit from the sample. Then each unit k is observed until some period I_k, with $I_k \le I$. Observation of the unit could be discontinued for two reasons: 1) the unit responds; or 2) the survey ends. In the first case, $t = I_k$. In the second case, it is only known that $t > I$. Units with $t > I$ are right–censored—it is unknown whether they respond. Because response occurrence is intrinsically conditional, we characterize t 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, $h_{ki}(\boldsymbol{x}_k, \boldsymbol{\beta})$, h_{ki} 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:

$$h_{ki} = \Pr(t = i \mid t \geq i), \tag{2.1}$$

where $\boldsymbol{x}_k$ refers to both time–invariant and time–varying explanatory variables and $\boldsymbol{\beta}$ is the unknown $q_r \times 1$ vector parameter to be estimated. For unit with $t = i$, the probability of obtaining a response at time period i could be expressed in terms of the hazard as

$$\Pr(t = i) = h_{ki} \prod_{j=1}^{i-1}(1 - h_{kj}). \tag{2.2}$$

For units with $t > i$, the probability of obtaining a response can be expressed as

$$\Pr(t > i) = \prod_{j=1}^{i}(1 - h_{kj}). \tag{2.3}$$

We assume that every unit in the sample lives through each successive discrete time period until the unit responds or is censored by the end of data collection. The use of mixed–mode modifies the expression for the hazard function in (2.1) as $h_{ki} = \sum_{m=1}^{M} \phi_{m;k}^{(dc)} h_{ki|m}$, where $h_{ki|m}$ is the discrete–hazard function for mode m. The marginal probability of obtaining a response after I time periods is given by

$$\xi_k = 1 - \prod_{i=1}^{I}(1 - h_{ki}) = \sum_{i=1}^{I} \xi_k^{(i)}, \tag{2.4}$$

where $\xi_k^{(i)} = h_{ki} \prod_{j=1}^{i-1}(1 - h_{kj})$. It is easily seen from (2.4) that ξ_k increases (or stays the same) as the level of effort increases, where the level of effort is seen in terms of follow–up treatments and data–collection periods. This suggests that costs and benefits of increasing the level of effort should be explored given that, in some circumstances, there a number of follow up treatments made with a high percentage of cost expanded to get values from a few non-respondents.

2.2 Influence of Follow–up on Response Probability

We expressed the inverse link function of the hazard rate as a function of explanatory variables $\boldsymbol{x}_{ki}$ and a vector parameter $\boldsymbol{\beta}$ to be estimated. For units under self–enumeration data collection without any follow–up, the inverse link form of the hazard–rate is expressed as

$$g^{-1}(h_{ki}) = \eta(\boldsymbol{x}_{ki}^{(0)}, \boldsymbol{\beta}^{(0)}), \tag{2.5}$$

for known function $\eta(.)$, where $\boldsymbol{x}_{ki}^{(0)}$ is the vector of explanatory variables for self–enumeration without any follow–up, $\boldsymbol{\beta}^{(0)}$ is the associated unknown vector parameter to be estimated, $\boldsymbol{x}_{ki} = \boldsymbol{x}_{ki}^{(0)}$, $\boldsymbol{\beta} = \boldsymbol{\beta}^{(0)}$ and $g(.)$ is a link function—although the link function is generally used to transform (or to link) the conditional mean to the linear predictor $\boldsymbol{x}_{ki}^{T}\boldsymbol{\beta}$. For example, $g(a) = a$ with $\eta(\boldsymbol{x}_{ki}, \boldsymbol{\beta}) = \boldsymbol{x}_{ki}^{T}\boldsymbol{\beta}$ gives a linear regression model and $g(a) = \exp(a)/\{1 + \exp(a)\}$ with $\eta(\boldsymbol{x}_{ki}, \boldsymbol{\beta}) = \boldsymbol{x}_{ki}^{T}\boldsymbol{\beta}$ gives a logistic regression model for binary responses r_{ki}, where r_{ki} is a sequence of response indicators defined for each unit k whose values are defined as $r_{ki} = 1$ if the unit does respond in period i and $r_{ki} = 0$ if the unit does not respond in period i.

Additional influences on response probability can be investigated by adding further predictors to the initial discrete–time hazard model. For instance, the following model differs from the model in (2.5) by the inclusion of the time–variant follow–up predictor $\gamma_{ki}^{(1)}\boldsymbol{x}_{ki}^{(1)}$, the influence of which is captured by the parameter $\boldsymbol{\beta}^{(1)}$:

$$g^{-1}(h_{ki}) = \eta(\boldsymbol{x}_{ki}^{(0)}, \boldsymbol{\beta}^{(0)}; \gamma_{ki}^{(1)}\boldsymbol{x}_{ki}^{(1)}, \boldsymbol{\beta}^{(1)}) , \tag{2.6}$$

where the value of $\gamma_{ki}^{((1)}$ is set to 1 if the first follow–up treatment is started, or set to 0 if this is not the case, with $\boldsymbol{x}_{ki} = (\boldsymbol{x}_{ki}^{(0)T}, \gamma_{ki}^{(1)}\boldsymbol{x}_{ki}^{(1)T})^T$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}^{(0)T}, \boldsymbol{\beta}^{(1)T})^T$. Note that (2.6) can be used to define different slopes and intercepts, in which case the parameter $\boldsymbol{\beta}^{(1)}$ reflects the changes in the intercepts and in the slopes associated with changing from self–enumeration only to self–enumeration followed by the first follow–up treatment. For example, in the specification $\eta(\boldsymbol{x}_{ki}, \boldsymbol{\beta}) = \boldsymbol{x}_{ki}^{(0)T}\boldsymbol{\beta}^{(0)} + \boldsymbol{x}_{ki}^{(1)T}\boldsymbol{\beta}^{(1)}$, $i = 1, \ldots, I$, with $\boldsymbol{x}_{ki}^{(0)T}\boldsymbol{\beta}^{(0)} = \alpha_{0i}^{(0)} + x_{ki}\alpha_{1i}^{(0)}$ and $\boldsymbol{x}_{ki}^{(1)T}\boldsymbol{\beta}^{(1)} = \gamma_{ki}^{(1)}(\alpha_{0i}^{(1)} + x_{ki}\alpha_{1i}^{(1)})$, the regression parameters $\alpha_{0i}^{(0)}$, $\alpha_{1i}^{(0)}$ and the values x_{ki} represent respectively the intercept, the slope and the predictor associated with self–enumeration data collection in time period i. We have $\boldsymbol{x}_{ki}^{(0)} = (D_{k1}^{(0)}, ..., D_{kI}^{(0)}, x_{k1}^{(0)}, ..., x_{kI}^{(0)})^T$ and $\boldsymbol{\beta}^{(0)} = (\alpha_{01}^{(0)}, ..., \alpha_{0I}^{(0)}, \alpha_{11}^{(0)}, ..., \alpha_{1I}^{(0)})^T$, where $D_{ki}^{(0)} = 1$, $x_{ki}^{(0)} = x_{ki}$, $D_{kj}^{(0)} = 0$ and $x_{kj}^{(0)} = 0$ for $j \neq i$. The vector predictor follow–up is given by $\boldsymbol{x}_{ki}^{(1)} = (D_{k1}^{(1)}, ..., D_{kI}^{(1)}, x_{k1}^{(1)}, ..., x_{kI}^{(1)})^T$ and the changes due to the follow–up in the intercepts and slopes are reflected by vector parameter $\boldsymbol{\beta}^{(1)} = (\alpha_{01}^{(1)}, ..., \alpha_{0I}^{(1)}, \alpha_{11}^{(1)}, ..., \alpha_{1I}^{(1)})^T$, where $D_{ki}^{(1)} = 1$, $x_{ki}^{(1)} = x_{ki}$, $D_{kj}^{(1)} = 0$ and $x_{kj}^{(1)} = 0$ for $j \neq i$. To increase response rates, non–respondents are subject to intensive multiple follow–ups by telephone or other treatments to encourage them to participate. A follow–up treatment can take the form of mailed reminders, emailed reminders, telephone calls or in–person interviews. The follow–up process through treatments is conducted using data collection calendars with a specific strategy for each sampled unit. In case of $1 + \mathrm{T}$ follow–up treatments, the inverse link form of the hazard–rate can be expressed as $g^{-1}(h_{ki}) = \eta(\boldsymbol{x}_{ki}, \boldsymbol{\beta})$, where $\boldsymbol{x}_{ki} = (\boldsymbol{x}_{ki}^{(0)T}, \gamma_{ki}^{(1)}\boldsymbol{x}_{ki}^{(1)T}, ..., \gamma_{ki}^{(\mathrm{T})}\boldsymbol{x}_{ki}^{(\mathrm{T})T})^T$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}^{(0)T}, \boldsymbol{\beta}^{(1)T}, ..., \boldsymbol{\beta}^{(\mathrm{T})T})^T$.

Consider the case of $\mathrm{T} = 1$ where T_1 consists of intensive follow–up and T_0 consists of sending the questionnaire, and suppose for simplicity the case in which the response outcome is instant. After collecting the response from self–enumeration respondents, follow–up is performed in a deterministic way—non–respondents with $u_k \geq c_u$ are assigned to treatment T_1, where c_u is a predetermined constant and u is an auxiliary variable with values available for all sampled units. Suppose all units under T_1 responded, while the other units have still not responded. We have $\xi_k = h_{k1} + (1 - h_{k1})1 = 1$ for unit k with $u_k \geq c_u$ and $\xi_k = h_{k1} + (1 - h_{k1})0 = h_{k1}$ for units with $u_k < c_u$. This highlights the significant effect of follow-up on the probabilities of response.

3. Estimation Under Mixed–mode Survey

3.1 Mixed–mode Indicator Variables

Suppose we have S follow–up strategies, and define a vector of follow–up strategy indicator variables as $l_{s;k}^{(f)}=1$ if unit k is assigned to strategy s, and $l_{s;k}^{(f)}=0$ if not, where $\boldsymbol{l}_k^{(f)}=(l_{1;k}^{(f)},...,l_{S;k}^{(f)})^T$ are realizations of independent distributed variables according to a multinomial distribution, $Mult_S(1,\boldsymbol{\varphi}_k^{(f)})$, $\boldsymbol{\varphi}_k^{(f)}=(\phi_{1;k}^{(f)},...,\phi_{S;k}^{(f)})^T=E_f(\boldsymbol{l}_k^{(f)})$ is the vector of probabilities with $\sum_{s=1}^{S}\phi_{s;k}^{(f)}=1$, 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

$$\log(\phi_{s;k}^{(f)}/\phi_{S;k}^{(f)})=\boldsymbol{v}_{f;k}^{T}\boldsymbol{\lambda}_s^{(f)},\ s=1,...,S-1,$$

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

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

and for $s=1,...,S-1$

$$\phi_{s;k}^{(f)}=\phi_{S;k}^{(f)}\exp(\boldsymbol{v}_{f;k}^{T}\boldsymbol{\lambda}_s^{(f)}).$$

Similarly, let's define a vector of data collection mode indicator variables as $l_{m;k}^{(dc)}=1$ if unit k uses mode m, and $l_{m;k}^{(dc)}=0$ if not, where $\boldsymbol{l}_k^{(dc)}=(l_{1;k}^{(dc)},...,l_{M;k}^{(dc)})^T$ are realizations of independent distributed random variables according to a multinomial distribution, $Mult_M(1,\boldsymbol{\varphi}_k^{(dc)})$, $\boldsymbol{\varphi}_k=(\phi_{1;k}^{(dc)},...,\phi_{M;k}^{(dc)})^T=E_{dc}(\boldsymbol{l}_k^{(dc)})$ is the vector of data collection mode probabilities with $\sum_{m=1}^{M}\phi_{m;k}^{(dc)}=1$, and E_{dc} denotes expectation with respect to the data collection model. The mode of data collection for unit k is characterized by the matrix $\boldsymbol{\Phi}_k^{(dc)}$ which consists of the conditional probability $\phi_{m|s;k}^{(dc)}$ of using mode m given that unit is assigned to follow–up strategy s defined by its components

$$\phi_{m|s;k}^{(dc)}=\Pr(l_{m;k}^{(dc)}=1\,|\,l_{s;k}^{(f)}=1).$$

For example, for a 2×2 case ($S=M=2$) it depends on two probabilities: $\phi_{1|1;k}^{(dc)}=\Pr(l_{1;k}^{(dc)}=1\,|\,l_{1;k}^{(f)}=1)$ and $\phi_{1|2;k}^{(dc)}=\Pr(l_{1;k}^{(dc)}=1\,|\,l_{2;k}^{(f)}=1)$. Hence

$$\boldsymbol{\Phi}_k^{(dc)}=\begin{pmatrix}\phi_{1|1;k}^{(dc)} & 1-\phi_{1|1;k}^{(dc)}\\ \phi_{1|2;k}^{(dc)} & 1-\phi_{1|2;k}^{(dc)}\end{pmatrix}.$$

The marginal distribution of $l_{m;k}^{(dc)}$ is:

$$\phi_{m;k}^{(dc)}=\Pr(l_{m;k}^{(dc)}=1)=\textstyle\sum_{s=1}^{S}\phi_{s;k}^{(f)}\phi_{m|s;k}^{(dc)},\ m=1,...,M,$$

where $\sum_{m=1}^{M}\phi_{m|s;k}^{(dc)}=1$. 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

$$\log(\phi_{m|s;k}^{(dc)}/\phi_{M|s;k}^{(dc)})=\boldsymbol{v}_{dc;k}^{T}\boldsymbol{\lambda}_{m|s}^{(dc)},\ m=1,...,M-1,$$

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

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

and for $m = 1,..., M-1$
$$\phi_{m|s;k}^{(dc)} = \phi_{M|s;k}^{(dc)} \exp(\boldsymbol{v}_{dc;k}^T \boldsymbol{\lambda}_{m|s}^{(dc)}) .$$

3.2 Estimation of Regression Parameter

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. Consequently the likelihood for the census observed data may be decomposed as

$$L_k(\boldsymbol{\Gamma}) = \prod_{s=1}^{S} \{\phi_{s;k}^{(f)} L_{R|s;k} L_{M|s;k}\}^{I_{m;k}^{(f)}} , \tag{3.1}$$

with
$$L_{R|s;k} = \{\prod_{m=1}^{M} [\phi_{m|s;k}^{(dc)} f_{m|s}(t_k)]^{I_{m|s;k}^{(dc)}}\}^{\delta_k} \text{ and } L_{M|s;k} = \{\sum_{m=1}^{M} \phi_{m|s;k}^{(dc)} f_{m|s}(t_k)\}^{(1-\delta_k)} ,$$

where $\boldsymbol{\Gamma} = (\lambda^{(f)T}, \lambda^{(dc)T}, \boldsymbol{\beta}^T)^T$, $f_{m|s}(t_k)$ is $f(t_k)$ for mode m of data collection under follow-up strategy s,

$$f(t_k) = \Pr(t = I_k)^{\delta_k} \Pr(t > I_k)^{1-\delta_k} , \tag{3.2}$$

$\delta_k = 1$ if unit k is uncensored and $\delta_k = 0$ if unit k is censored. When unit k is censored, either unit k will respond at some future time period $t > I$ or the unit will never respond. The function $\boldsymbol{\Gamma}_N$ such that $\boldsymbol{\Gamma}_N = \arg\max_{\boldsymbol{\Gamma}} \sum_k \ell_k(\boldsymbol{\Gamma})$ is the maximum likelihood estimator (MLE) of $\boldsymbol{\Gamma}$, where $\ell_k(\boldsymbol{\Gamma}) = \log L_k(\boldsymbol{\Gamma})$. Under certain conditions the MLE is taken to be the solution of the system

$$\mathbf{S}(\boldsymbol{\Gamma}) = \sum_k \mathbf{s}_k(\boldsymbol{\Gamma}) = \mathbf{0} , \tag{3.3}$$

with $E\{\mathbf{s}_k(\boldsymbol{\Gamma})\} = \mathbf{0}$, where $\mathbf{s}_k(\boldsymbol{\Gamma}) = \partial \ell_k(\boldsymbol{\Gamma}) / \partial \boldsymbol{\Gamma}$ is the score function. For general sampling design with known positive inclusion probabilities, π_k, an design–unbiased estimator of the EE defined by (3.3) is given by

$$\hat{\mathbf{S}}(\boldsymbol{\Gamma}) = \sum_k d_k(\wp) \mathbf{s}_k(\boldsymbol{\Gamma}) = \mathbf{0} , \tag{3.4}$$

where $d_k(\wp) = 1_k(\wp) / \pi_k$ are the design weights with $1_k(\wp) = 1(k \in \wp)$ is the sample $\wp$ membership indicator variable for unit k, $1(\mathit{condition})$ is the truth function, i.e., $1(\mathit{condition}) = 1$ if the condition is true and $1(\mathit{condition}) = 0$ if not, $\pi_k = E_p(1_k(\wp))$, and E_p denotes expectation with respect to the sampling design. Starting with a guessed value, $\boldsymbol{\Gamma}_0$, then for $b = 1,2,...$ updates are made using

$$\boldsymbol{\Gamma}_b = \boldsymbol{\Gamma}_{b-1} + \{\hat{\mathbf{J}}_{\boldsymbol{\Gamma}}(\boldsymbol{\Gamma}_{b-1})\}^{-1} \hat{\mathbf{S}}(\boldsymbol{\Gamma}_{b-1}) ,$$

where $\hat{\mathbf{J}}_{\boldsymbol{\Gamma}}(\boldsymbol{\Gamma}) = -\partial \hat{\mathbf{S}}^T(\boldsymbol{\Gamma}) / \partial \boldsymbol{\Gamma}$. The solution obtained by a Newton–Raphson–type iterative method gives the estimator $\hat{\boldsymbol{\Gamma}}$ of $\boldsymbol{\Gamma}$.

3.3 Estimation of the Parameters of Interest

An estimator $\hat{\boldsymbol{\Theta}}$ of the parameter $\boldsymbol{\Theta}_N(y)$ associated with EE defined by (1.1) is defined as the solution of the weighted EE

$$\hat{\mathbf{S}}(\boldsymbol{\Theta}) = \sum_k \hat{\varpi}_k \mathbf{s}(y_k; \boldsymbol{\Theta}) - \mathbf{v}(\boldsymbol{\Theta}) = \mathbf{0} . \tag{3.5}$$

A first set of weights is given by

$$\hat{\varpi}_k = d_k(\wp) r_k / \hat{\xi}_k \equiv \hat{\varpi}_k^{(com)}\,, \tag{3.5a}$$

with
$$\xi_k = \sum_{s=1}^{S} \phi_{s;k}^{(f)} \sum_{m=1}^{M} \phi_{m|s;k}^{(dc)} \xi_{m|s;k}\,,$$

where $\hat{\xi}_k = \xi_k(\hat{\boldsymbol{\Gamma}})$. The solution of (3.5) using the set of weights given by (3.5a) is denoted by $\hat{\boldsymbol{\Theta}}^{(com)}$. The above set of weights which combined the follow–up strategies and modes of data collection may not be preferred when follow–up strategies contribute toward explaining units behaviour. An alternative set of weights with separate strategies and modes is given by

$$\hat{\varpi}_k = d_k(\wp) \sum_{s=1}^{S} l_{s;k}^{(f)} \sum_{m=1}^{M} l_{m|s;k}^{(dc)} (r_{m|s;k} / \hat{\xi}_{m|s;k}) \equiv \hat{\varpi}_k^{(sep)}\,. \tag{3.5b}$$

The solution of (3.5) using the set of weights given by (3.5b) is denoted by $\hat{\boldsymbol{\Theta}}^{(sep)}$.

Since respondents in one mode cannot be considered a simple random subsample of the sample, we defined an estimator $\hat{\boldsymbol{\Theta}}_m$ of the parameter of interest $\boldsymbol{\Theta}_{m;N}(y_m)$ associated with EE (1.2) as the solution to the weighted EE

$$\hat{\mathbf{S}}(\boldsymbol{\Theta}_m) = \sum_k \hat{\varpi}_{m;k} \mathbf{s}(y_k; \boldsymbol{\Theta}_m) - \mathbf{v}(\boldsymbol{\Theta}_m) = \mathbf{0}\,. \tag{3.6}$$

Again, we consider two sets of weights. The combined set

$$\hat{\varpi}_{m;k} = d_k(\wp)(r_k / \hat{\xi}_k)(l_{m;k}^{(dc)} / \hat{\phi}_{m;k}^{(dc)}) \equiv \hat{\varpi}_{m;k}^{(com)}\,, \tag{3.6a}$$

and the separate set

$$\hat{\varpi}_{m;k} = d_k(\wp) \sum_{s=1}^{S} l_{s;k}^{(f)} \{ l_{m|s;k}^{(dc)} r_{m|s;k} / (\hat{\xi}_{m|s;k} \hat{\phi}_{m|s;k}^{(dc)}) \} \equiv \hat{\varpi}_{m;k}^{(sep)}\,, \tag{3.6b}$$

where $\hat{\phi}_{m;k}^{(dc)} = \sum_{s=1}^{S} \hat{\phi}_{s;k}^{(f)} \hat{\phi}_{m|s;k}^{(dc)}$. The solution of (3.6) using the set of weights given by (3.6a) is denoted by $\hat{\boldsymbol{\Theta}}_m^{(com)}$, while the solution of (3.6) using the set of weights given by (3.6b) is denoted by $\hat{\boldsymbol{\Theta}}_m^{(sep)}$.

4. Linearization Variance Estimators and Optimal Resources Allocation

4.1 Derivation of the Variance

We use the linearization method of Demnati and Rao (2004, 2010) to derive variances and variance estimators. We first give a brief account of the Demnati–Rao (DR) approach. Let $\boldsymbol{d}_k = (\boldsymbol{d}_{1;k}^T, \boldsymbol{d}_{2;k}^T, ..., \boldsymbol{d}_{g;k}^T)^T$ be a $G \times 1$ vector of random weights and $\boldsymbol{u}_k = (\boldsymbol{u}_{1;k}, ..., \boldsymbol{u}_{g;k})$ be a $p \times G$ matrix of constants for $k = 1, ..., N$. Let $\hat{\boldsymbol{U}} = \sum_k \boldsymbol{u}_k \boldsymbol{d}_k$ be a linear combination and, using an operator notation, let $\mathbf{V}(\boldsymbol{u})$ and $\vartheta(\boldsymbol{u})$ denote respectively the variance of $\hat{\boldsymbol{U}}$ and its variance estimator. DR expressed an estimator $\hat{\boldsymbol{\theta}}$ and its induced parameter $\boldsymbol{\theta} = E(\hat{\boldsymbol{\theta}})$ as $\hat{\boldsymbol{\theta}} = f(\mathbf{A}_{\boldsymbol{d}})$ and $\boldsymbol{\theta} = f(\mathbf{A}_{\mu})$, where $\mathbf{A}_{\boldsymbol{d}}$ is a $G \times N$ matrix with k^{th} column $\boldsymbol{d}_k$, $\mathbf{A}_{\boldsymbol{u}}$ is a $G \times N$ matrix with k^{th} column $\boldsymbol{\mu}_k = E(\boldsymbol{d}_k)$ and E denotes expectation under random processes involved. The DR linearization variance and variance estimator of $\hat{\boldsymbol{\theta}} = f(\mathbf{A}_{\boldsymbol{d}})$ are simply given by $\mathrm{V}_{DR}(\hat{\boldsymbol{\theta}}) = \mathbf{V}(\tilde{\boldsymbol{z}})$ and $\vartheta_{DR}(\hat{\boldsymbol{\theta}}) = \vartheta(\boldsymbol{z})$ respectively, where $\mathrm{V}(\tilde{\boldsymbol{z}})$ is obtained from $\mathrm{V}(\boldsymbol{u})$ by replacing $\boldsymbol{u}_k$ by $\tilde{\boldsymbol{z}}_k = \partial f(\mathbf{A}_{\boldsymbol{b}}) / \partial \boldsymbol{b}_k^T |_{\mathbf{A}_{\boldsymbol{b}} = \mathbf{A}_{\mu}}$, and $\vartheta(\boldsymbol{z})$ is obtained from $\vartheta(\boldsymbol{u})$

by replacing $\boldsymbol{u}_k$ by $\boldsymbol{z}_k = \partial f(\mathbf{A}_{\boldsymbol{b}}) / \partial \boldsymbol{b}_k^T |_{\mathbf{A}_{\boldsymbol{b}} = \mathbf{A}_{\boldsymbol{d}}}$, where $\mathbf{A}_{\boldsymbol{b}}$ is a $G \times N$ matrix of arbitrary real numbers with k^{th} column $\boldsymbol{b}_k = (\boldsymbol{b}_{1;k}^T, \boldsymbol{b}_{2;k}^T, ..., \boldsymbol{b}_{g;k}^T)^T$. Seeking clarity the rest of my work below considers only the case where values of the variable of interest y are fixed. We now consider the derivation of the variance of a first compact form given by

$$\hat{\boldsymbol{U}} = \sum_k \boldsymbol{u}_k \boldsymbol{d}_k \, , \tag{4.1}$$

with $$g = 2 \, , \ \boldsymbol{d}_{1;k} = d_k(\wp) \nabla_k \ \text{ and } \ \boldsymbol{d}_{2;k} = d_k(\wp) \mathbf{s}_k(\boldsymbol{\Gamma}) \, ,$$

where ∇_k is a dichotomous variable with expectation $E_\nabla(\nabla_k) = \overline{\nabla}_k$, and $\boldsymbol{u}_k = (\boldsymbol{u}_{1;k}, \boldsymbol{u}_{2;k})$. We may decompose the variance of $\hat{\mathbf{U}}$ as

$$Var(\hat{\mathbf{U}}) = E_p Var_\nabla(\hat{\mathbf{U}}) + Var_p E_\nabla(\hat{\mathbf{U}}) \equiv \mathbf{V}_\nabla + \mathbf{V}_p \, . \tag{4.2}$$

Under independent mechanism on ∇_k, the first component $\mathbf{V}_\nabla = E_p Var_\nabla(\hat{\mathbf{U}})$ of $Var(\hat{\mathbf{U}})$ given by (4.2) is given by

$$\mathbf{V}_\nabla = \sum_k (1/\pi_k) \boldsymbol{u}_k \begin{pmatrix} \overline{\nabla}_k (1 - \overline{\nabla}_k) & E\{\nabla_k \mathbf{s}_k^T(\boldsymbol{\Gamma})\} \\ E\{\mathbf{s}_k(\boldsymbol{\Gamma}) \nabla_k\} & E\{\mathbf{s}_k(\boldsymbol{\Gamma}) \mathbf{s}_k^T(\boldsymbol{\Gamma})\} \end{pmatrix} \boldsymbol{u}_k^T \, , \tag{4.3a}$$

provided $E\{\mathbf{s}_k(\boldsymbol{\Gamma})\} = \mathbf{0}$.

The second component $\mathbf{V}_p = Var_p\{\sum_k \boldsymbol{u}_{1;k} d_k(\wp) \overline{\nabla}_k\}$ of $Var(\hat{\mathbf{U}})$ given by (4.2) is given by

$$\mathbf{V}_p = \sum_k \sum_l \omega_{kl}^{-1} (1 - \omega_{kl}) \boldsymbol{u}_{1;k} \overline{\nabla}_k \overline{\nabla}_l \boldsymbol{u}_{1;l}^T \, , \tag{4.3b}$$

where $\omega_{kl} = \pi_{kl}^{-1} \pi_k \pi_l$, $\omega_{kk} = \omega_k = \pi_k$ and $\pi_{kl} = E_p\{1_k(\wp) 1_l(\wp)\}$.

The sum of (4.3a) and (4.3b) constitutes $\mathbf{V}(\boldsymbol{u}) = \mathbf{V}_\nabla + \mathbf{V}_p$, the variance of $\hat{\mathbf{U}}$ under (4.1).

It follows from (3.4), (3.5) with (3.5a) and (3.6) with (3.6a) that the compact form given by (4.1) can be used to derive the variance of $\hat{\boldsymbol{\Gamma}}$, $\hat{\boldsymbol{\Theta}}^{(com)}$ and $\hat{\boldsymbol{\Theta}}_m^{(com)}$ respectively. In the first case of $\hat{\boldsymbol{\Gamma}}$, $\boldsymbol{u}_{1;k} = \mathbf{0}$ and $\tilde{\boldsymbol{z}}_{2;k} = \{\mathbf{J}_{\boldsymbol{\Gamma}_N}(\boldsymbol{\Gamma}_N)\}^{-1}$ where $\boldsymbol{\Gamma}_N$ is the solution to the census EE $\mathbf{S}(\boldsymbol{\Gamma}) = \sum_k \mathbf{s}_k(\boldsymbol{\Gamma}) = \mathbf{0}$, and $\mathbf{J}_{\mathbf{b}_*}(\mathbf{a}_*) = -\partial \mathbf{S}^T(\mathbf{a}) / \partial \mathbf{b}$ evaluated at $(\mathbf{a}, \mathbf{b}) = (\mathbf{a}_*, \mathbf{b}_*)$; in the second case of $\hat{\boldsymbol{\Theta}}^{(com)}$, $\nabla_k = r_k$, $\overline{\nabla}_k = \xi_k$, $\tilde{\boldsymbol{z}}_{1;k} = \mathbf{s}(y_k; \boldsymbol{\Theta}_N) / \xi_k(\boldsymbol{\Gamma}_N)$, and $\tilde{\boldsymbol{z}}_{2;k} = -\mathbf{J}_{\boldsymbol{\Gamma}_N}(\boldsymbol{\Theta}_N)\{\mathbf{J}_{\boldsymbol{\Gamma}_N}(\boldsymbol{\Gamma}_N)\}^{-1}$; while in the third case of $\hat{\boldsymbol{\Theta}}_m^{(com)}$, $\nabla_k = r_k l_{m;k}^{(dc)}$, $\overline{\nabla}_k = \xi_k \phi_{m;k}^{(dc)}$, $\tilde{\boldsymbol{z}}_{1;k} = \mathbf{s}(y_k; \boldsymbol{\Theta}_{m;N}) / \{\xi_k(\boldsymbol{\Gamma}_N) \phi_{m;k}^{(dc)}(\boldsymbol{\Gamma}_N)\}$, and $\tilde{\boldsymbol{z}}_{2;k} = -\mathbf{J}_{\boldsymbol{\Gamma}_N}(\boldsymbol{\Theta}_N)\{\mathbf{J}_{\boldsymbol{\Gamma}_N}(\boldsymbol{\Gamma}_N)\}^{-1}$. In the first case $G = q_{\boldsymbol{\Gamma}}$, while in the second and third cases $G = 1 + q_{\boldsymbol{\Gamma}}$ where $q_{\boldsymbol{\Gamma}} = q_{\boldsymbol{f}} + q_{\boldsymbol{dc}} + q_{\boldsymbol{r}}$.

We now consider a second compact form given by

$$\hat{\boldsymbol{U}} = \sum_k \boldsymbol{u}_k \boldsymbol{d}_k \, , \tag{4.4}$$

with $$g = 2 \, , \ \boldsymbol{d}_{1;k} = d_k(\wp)(J_{1;k} \nabla_{1;k}, ..., J_{B;k} \nabla_{B;k})^T \ \text{ and } \ \boldsymbol{d}_{2;k} = d_k(\wp) \mathbf{s}_k(\boldsymbol{\Gamma}) \, ,$$

where $\boldsymbol{J}_k = (J_{1;k},...,J_{B;k})^T$ is a vector of random variables distributed according to a multinomial distribution $\boldsymbol{J}_k \sim Mult_B(1, \boldsymbol{p}_{J;k})^T$ with $\boldsymbol{p}_{J;k} = E(\boldsymbol{J}_k)$, $\nabla_{b;k}$ is a dichotomous variable with expectation $E_\nabla(\nabla_{b;k}) = \overline{\nabla}_{b;k}$, and $\boldsymbol{u}_{1;k} = (\boldsymbol{u}_{1;k}^{(1)},...,\boldsymbol{u}_{1;k}^{(B)})$ .We may decompose the variance of $\hat{\mathbf{U}}$ as (4.2).

Under independent mechanism on $\nabla_{b;k}$, the first component $\mathbf{V}_\nabla = E_p Var_\nabla(\hat{\mathbf{U}})$ of $Var(\hat{\mathbf{U}})$ given by (4.2) under (4.4) is given by

$$\mathbf{V}_\nabla = \sum_k (1/\pi_k)\begin{pmatrix} \sum_b \boldsymbol{u}_{1;k}^{(b)} p_{b;k}\overline{\nabla}_{b;k}\{\boldsymbol{u}_{1;k}^{(b)T} - \bar{\boldsymbol{s}}_{1;k}^T\} & E\{\boldsymbol{s}_{1;k}\mathbf{s}_k^T(\boldsymbol{\Gamma})\}\boldsymbol{u}_{2;k}^T \\ \boldsymbol{u}_{2;k}E\{\mathbf{s}_k(\boldsymbol{\Gamma})\boldsymbol{s}_{1;k}^T\} & \boldsymbol{u}_{2;k}E\{\mathbf{s}_k(\boldsymbol{\Gamma})\mathbf{s}_k^T(\boldsymbol{\Gamma})\}\boldsymbol{u}_{2;k}^T \end{pmatrix}, \tag{4.5a}$$

provided $E\{\mathbf{s}_k(\boldsymbol{\Gamma})\} = \mathbf{0}$, where $\boldsymbol{s}_{1;k} = \sum_b \boldsymbol{u}_{1;k}^{(b)} J_{b;k}\nabla_{b;k}$ and $\bar{\boldsymbol{s}}_{1;k} = \sum_b \boldsymbol{u}_{1;k}^{(b)} p_{b;k}\overline{\nabla}_{b;k}$.

The second component $\mathbf{V}_p = Var_p\{\sum_k d_k(\wp)\bar{\boldsymbol{s}}_{1;k}\}$ of $Var(\hat{\mathbf{U}})$ given by (4.2) under (4.4) is given by

$$\mathbf{V}_p = \sum_k \sum_l \omega_{kl}^{-1}(1-\omega_{kl})\bar{\boldsymbol{s}}_{1;k}\bar{\boldsymbol{s}}_{1;l}^T. \tag{4.5b}$$

The sum of (4.5a) and (4.5b) constitutes $\mathbf{V}(\boldsymbol{u}) = \mathbf{V}_\nabla + \mathbf{V}_p$, the variance of $\hat{\mathbf{U}}$ under (4.4).

It follows from (3.5) with (3.5b) and from (3.6) with (3.6b) that the compact form given by (4.4) can be used to derive the variance of $\hat{\boldsymbol{\Theta}}^{(sep)}$ and $\hat{\boldsymbol{\Theta}}_m^{(sep)}$ respectively. In the first case of $\hat{\boldsymbol{\Theta}}^{(sep)}$, $B = SM$, $J_{sm;k} = l_{s;k}^{(f)} l_{m|s;k}^{(dc)}$, $\nabla_{sm;k} = r_{m|s;k}$, $\overline{\nabla}_{sm;k} = \xi_{m|s;k}$, $\tilde{\boldsymbol{z}}_{1;k}^{(s)} = \mathbf{s}(y_k;\boldsymbol{\Theta}_N)/\xi_{s;k}(\boldsymbol{\Gamma}_N)$, and $\tilde{\boldsymbol{z}}_{2;k} = -\mathbf{J}_{\boldsymbol{\Gamma}_N}(\boldsymbol{\Theta}_N)\{\mathbf{J}_{\boldsymbol{\Gamma}_N}(\boldsymbol{\Gamma}_N)\}^{-1}$; while in the second of $\hat{\boldsymbol{\Theta}}_m^{(sep)}$, $B = S$, $J_{s;k} = l_{s;k}^{(f)}$, $\nabla_{s;k} = r_{s;k} l_{m|s;k}^{(dc)}$, $\overline{\nabla}_{s;k} = \phi_{s;k}^{(f)}\phi_{m|s;k}^{(dc)}$, $\tilde{\boldsymbol{z}}_{1;k}^{(s)} = \mathbf{s}(y_k;\boldsymbol{\Theta}_{m;N})/\{\xi_{m;k}(\boldsymbol{\Gamma}_N)\phi_{m;k}(\boldsymbol{\Gamma}_N)\}$, and $\tilde{\boldsymbol{z}}_{2;k} = -\mathbf{J}_{\boldsymbol{\Gamma}_N}(\boldsymbol{\Theta}_N)\{\mathbf{J}_{\boldsymbol{\Gamma}_N}(\boldsymbol{\Gamma}_N)\}^{-1}$.

4.2 Variance Estimation

We first consider the estimation of the variance of the compact form given by (4.1). Under independent mechanism on ∇_k, an estimator of the first component of $Var(\hat{\mathbf{U}})$ given by (4.2) is given by

$$\vartheta_\nabla = \sum_k d_k^2(\wp)\boldsymbol{u}_k \begin{pmatrix} \nabla_k(1-\hat{\overline{\nabla}}_k) & \nabla_k \mathbf{s}_k^T(\hat{\boldsymbol{\Gamma}}) \\ \mathbf{s}_k(\hat{\boldsymbol{\Gamma}})\nabla_k & \mathbf{s}_k(\hat{\boldsymbol{\Gamma}})\mathbf{s}_k^T(\hat{\boldsymbol{\Gamma}}) \end{pmatrix}\boldsymbol{u}_k^T, \tag{4.6a}$$

provided $E\{\mathbf{s}_k(\boldsymbol{\Gamma})\} = \mathbf{0}$, where $\hat{\overline{\nabla}}_k$ is an estimator of $\overline{\nabla}_k$.

If we use the HT variance estimator for arbitrary designs, then an estimator of the second component of $Var(\hat{\mathbf{U}})$ given by (4.2) is given by

$$\vartheta_p = \sum_k d_k^2(\wp)(1-\omega_k)\nabla_k\hat{\overline{\nabla}}_k\boldsymbol{u}_{1;k}\boldsymbol{u}_{1;k}^T + \sum_k\sum_{l\neq k} d_k(\wp)d_l(s)(1-\omega_{kl})\nabla_k\nabla_l\boldsymbol{u}_{1;k}\boldsymbol{u}_{1;l}^T. \tag{4.6b}$$

The sum of the two terms (4.6a) and (4.6b) constitutes $\vartheta(\boldsymbol{u}) = \vartheta_p + \vartheta_\nabla$, our estimator of the variance of $\hat{\mathbf{U}}$ under (4.1).

In the first case of $\hat{\boldsymbol{\Gamma}}$, $\boldsymbol{z}_{2;k} = \{\hat{\mathbf{J}}_{\hat{\boldsymbol{\Gamma}}}(\hat{\boldsymbol{\Gamma}})\}^{-1}$ where $\hat{\mathbf{J}}_{\hat{\mathbf{b}}}(\hat{\mathbf{a}}) = -\partial \hat{\mathbf{S}}^{T}(\mathbf{a}) / \partial \mathbf{b}$ evaluated at $(\mathbf{a}, \mathbf{b}) = (\hat{\mathbf{a}}, \hat{\mathbf{b}})$; in the second case of $\hat{\boldsymbol{\Theta}}^{(com)}$, $\boldsymbol{z}_{1;k} = \mathbf{s}(y_k; \hat{\boldsymbol{\Theta}}) / \hat{\xi}_k$, and $\boldsymbol{z}_{2;k} = -\hat{\mathbf{J}}_{\hat{\boldsymbol{\Gamma}}}(\hat{\boldsymbol{\Theta}})\{\hat{\mathbf{J}}_{\hat{\boldsymbol{\Gamma}}}(\hat{\boldsymbol{\Gamma}})\}^{-1}$; while in the third case of $\hat{\boldsymbol{\Theta}}_m^{(com)}$, $\boldsymbol{z}_{1;k} = \mathbf{s}(y_k; \hat{\boldsymbol{\Theta}}_m) / (\hat{\xi}_k \phi_{m;k}^{(dc)})$, and $\boldsymbol{z}_{2;k} = -\hat{\mathbf{J}}_{\hat{\boldsymbol{\Gamma}}}(\hat{\boldsymbol{\Theta}})\{\hat{\mathbf{J}}_{\hat{\boldsymbol{\Gamma}}}(\hat{\boldsymbol{\Gamma}})\}^{-1}$.

We now turn to the estimation of the variance of the compact form given by (4.4). Under independent mechanism on $\nabla_{b;k}$, an estimator of the first component of $Var(\hat{\mathbf{U}})$ given by (4.2) is given by

$$\vartheta_{\nabla} = \sum_k d_k^2(s)) \begin{pmatrix} \sum_b \boldsymbol{u}_{1;k}^{(b)} J_{b;k} \nabla_{b;k} \{\boldsymbol{u}_{1;k}^{(b)T} - \hat{\hat{\boldsymbol{s}}}_{1;k}^{T}\} & \boldsymbol{s}_{1;k} \mathbf{s}_k^T(\hat{\boldsymbol{\Gamma}})\} \boldsymbol{u}_{2;k}^T \\ \boldsymbol{u}_{2;k} \mathbf{s}_k(\hat{\boldsymbol{\Gamma}}) \boldsymbol{s}_{1;k}^T & \boldsymbol{u}_{2;k} \mathbf{s}_k(\hat{\boldsymbol{\Gamma}}) \mathbf{s}_k^T(\hat{\boldsymbol{\Gamma}}) \boldsymbol{u}_{2;k}^T \end{pmatrix}, \tag{4.7a}$$

provided $E\{\mathbf{s}_k(\boldsymbol{\Gamma})\} = \mathbf{0}$, where $\hat{\hat{\boldsymbol{s}}}_{1;k} = \sum_b \boldsymbol{u}_{1;k}^{(b)} \hat{p}_{b;k} \hat{\overline{\nabla}}_{b;k}$, $\hat{\overline{\nabla}}_{b;k}$ is an estimator of $\overline{\nabla}_k$, and $\hat{p}_{b;k}$ is an estimator of $p_{b;k}$.

If we use the HT variance estimator for arbitrary designs, then an estimator of the second component of $Var(\hat{\mathbf{U}})$ given by (4.2) is given by

$$\vartheta_p = \sum_k d_k^2(\wp)(1-\omega_k) \boldsymbol{s}_{1;k} \hat{\hat{\boldsymbol{s}}}_{1;k}^T + \sum_k \sum_{l \neq k} d_k(\wp) d_l(\wp)(1-\omega_{kl}) \boldsymbol{s}_{1;k} \boldsymbol{s}_{1;l}^T . \tag{4.7b}$$

The sum of the two terms (4.7a) and (4.7b) constitutes $\vartheta(\boldsymbol{u}) = \vartheta_p + \vartheta_{\nabla}$, an estimator of the variance of $\hat{\mathbf{U}}$ under (4.4).

In the first case of $\hat{\boldsymbol{\Theta}}^{(sep)}$, $\boldsymbol{z}_{1;k}^{(s)} = \mathbf{s}(y_k; \hat{\boldsymbol{\Theta}}) / \hat{\xi}_{m|s;k}$, and $\boldsymbol{z}_{2;k} = -\hat{\mathbf{J}}_{\hat{\boldsymbol{\Gamma}}}(\hat{\boldsymbol{\Theta}})\{\hat{\mathbf{J}}_{\hat{\boldsymbol{\Gamma}}}(\hat{\boldsymbol{\Gamma}})\}^{-1}$; while in the second of $\hat{\boldsymbol{\Theta}}_m^{(sep)}$, $\boldsymbol{z}_{1;k}^{(s)} = \mathbf{s}(y_k; \hat{\boldsymbol{\Theta}}_m) / (\hat{\xi}_{m|s;k} \hat{\phi}_{m|s;k}^{(dc)})$, and $\boldsymbol{z}_{2;k} = -\hat{\mathbf{J}}_{\hat{\boldsymbol{\Gamma}}}(\hat{\boldsymbol{\Theta}})\{\hat{\mathbf{J}}_{\hat{\boldsymbol{\Gamma}}}(\hat{\boldsymbol{\Gamma}})\}^{-1}$.

4.3 Optimal Resources Allocation

To reduce nonresponse bias, units are subject to intensive follow–up activities to obtain their cooperation. As this intensive follow–up is extensive, optimal resources allocation within sampling, follow–up, and data collection stages of the survey design is needed to make wise use of resources compared to the quality of the estimates. The portion of survey design that incorporates sampling, follow–up and data collection consists of: a) determining the probability of selection for each unit; b) determining the allocation probability of each follow–up strategy to each unit; c) selecting the sample; d) assigning a follow–up strategy to each unit to get their cooperation; e) collecting data from respondents; f) integrating results from sampling, follow–up and data collection activities into the estimation step of the ultimate parameter of interest; and, g) evaluating the impact of sampling errors as well as errors due to data collection and nonresponse on the ultimate estimator. A question that is of great importance is how a portion of resources that is allocated to a given stage of the survey design influences the estimator of the ultimate parameter of interest? Or simply, what is the optimal portion of global resources that should be allocated to a given stage of the survey design in comparison to the other stages? In general, a survey design is a large and complex program such that the ultimate estimator response to resources changes in each stage is not transparent. As there is no general measure that would capture all information on the impact of resources allocation on the ultimate estimator, the survey producer tends to combine various measures to get a broader effect and interactions between different stages. To allocate optimally resources within stages of the survey design, we determine for each unit simultaneously: a) the probability of selection in the sample; and b) the probability of allocation of each follow–up strategy among a set of strategies for each unit. This is done either by minimizing the variance for a given survey global cost; or in the case of multiple parameters of interest by minimizing the global cost subject to constraints on the variances

The global cost of sampling, follow–up, and data collection activities can maybe expressed as

$$C = C^{(p)} + C^{(f)} + C^{(dc)} . \tag{2.2}$$

where $C^{(p)} = \sum_k 1_k(\wp) c_k^{(p)}$ is the component associated with sampling cost, $C^{(f)} = \sum_k 1_k(\wp) \sum_{s=1}^{S} l_{s;k}^{(f)} c_{s;k}^{(f)}$ is the component associated with follow–up cost, and $C^{(dc)} = \sum_k 1_k(\wp) \sum_{s=1}^{S} l_{s;k}^{(f)} \sum_m l_{m|s;k}^{(dc)} r_{m|s;k} c_{m;k}^{(dc)}$ is the component associated with data collection cost. Here, $c_k^{(p)}$ is the sampling cost for unit k, $c_{s;k}^{(f)}$ denotes the cost associated with follow–up strategy s ($s = 1,..., S$) for unit k, and $c_{m;k}^{(dc)}$ denote the data collection cost associated with mode m ($m = 1,..., M$) for unit k. Since the above global cost is random, we consider its expectation given by

$$E(C) = \overline{C}^{(p)} + \overline{C}^{(f)} + \overline{C}^{(dc)} \equiv \overline{C} , \tag{4.8}$$

where $\overline{C}^{(p)} = \sum_k \pi_k c_k^{(p)}$, $\overline{C}^{(f)} = \sum_k \pi_k \sum_{s=1}^{S} \phi_{s;k}^{(f)} c_{s;k}^{(f)}$, and $\overline{C}^{(dc)} = \sum_k \pi_k \sum_{s=1}^{S} \phi_{s;k}^{(f)} \sum_{m=1}^{M} \phi_{m|s;k}^{(dc)} \xi_{m|s;k} c_{m;k}^{(dc)}$.

To create a design in the case of one parameter of interest, we determine the optimal π_k and $\phi_{s;k}^{(f)}$ ($k = 1,...N; s = 1,...S$) such that the variance, $Var(\hat{\Theta})$, is minimized subject to constraints on the expected global cost $\overline{C} \le C_{\max}$, where $C_{\max}$ is the survey global cost limit.

In the case of $\Lambda(>1)$ parameters of interest, we determine the optimal π_k and $\phi_{s;k}^{(f)}$ such that the expected global cost given by (4.8) is minimized subject to constraints on Λ variances:

$$Var(\hat{\Theta}_\kappa) \le \mathrm{V}_\kappa , \ \kappa = 1,..., \Lambda ,$$

where V_κ are specified tolerances, and $Var(\hat{\Theta}_\kappa)$ is the variance of the estimator $\hat{\Theta}$ for the κ^{th} parameter of interest $\kappa = 1,..., \Lambda$. For example, one could specify an upper limits, $\Im_\kappa$, on the coefficient of variation of $\hat{\Theta}_\kappa$ so that $\mathrm{V}_\kappa = \{\Im_\kappa E(\hat{\Theta}_\kappa)\}^2$.

Concluding Remarks

We introduced discrete–time hazard to the analysis of response indicators in surveys and censuses. The proposed approach facilitates examination of the shape of the hazard function. Since inspection of the shape of the hazard function indicates when a response is most likely to occur, and how the probability varies over both time and follow–up treatments, the description of the shapes of hazard function have an important role to play in survey quality and cost. We also used regression analysis to investigate the effect of mixed–mode on the response probability. Estimators of model parameters as well as estimators of parameters of interest are given, and associated variance estimators are studied under mixed–mode surveys. Finally, optimal resources allocation within stages of the mixed–mode survey design is determined to make wise use of global survey resources in terms of the quality of the ultimate estimate.

Currently, we are studying the issue of reducing side effect of prior information on survey design. 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. 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. This work will be presented at the 2016 Joint Statistical Meeting.

References

Allison, P. D. (1982). "Discrete–time methods for the analysis of event histories." S. Leinhardt (ed.). *Sociological Methodology*. San Francisco: Jossey–Bass. pp. 61–98.

Demnati, A. (2014). "Modeling self–enumeration and follow–up response indicators as discrete-time survival." *Proceeding of Statistics Canada Symposium*.

Demnati, A. and J.N.K. Rao. (2004). "Linearization variance estimators for survey data (with discussion)." *Survey Methodology*, **30**, 17-34.

Demnati, A. and J.N.K. Rao. (2010). "Linearization variance estimators for model parameters from complex survey data." *Survey Methodology*, **36**, 193-199.

Dempster, A. P., N.M. Laird and D.B. Rubin. (1977). "Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion)." *Journal of the Royal Statistical Society B.* 39, pp. 1–38.

Hansen, M. H. and W. N. Hurwitz. (1946). "The problem of nonresponse in sample survey." *Journal of the American Statistical Association*, 41, 517-529.

Hartley, H. O. (1958). "Maximum likelihood estimation from incomplete data." *Biometrics*. 4, pp. 174–194.

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