Welcome to the world of non-ignorable nonresponse: Review of the existing methods

Jae-Kwang Kim

Department of Statistics, Iowa State University

NISS Workshop
November 12, 2015

Call for papers: Special Issue on Nonignorable Nonresponse

- Special Issue at *Statistica Sinica*.
- Submission deadline: Feb 28th, 2016.
- Will be handled by Guest Editor Jae-kwang Kim (ISU) and Co-Editor Zhiliang Ying (Columbia U.)

Introduction

- (X, Y): random variable, y is subject to missingness
- Response indicator function

$$\delta_i = \begin{cases} 1 & \text{if } y_i \text{ is observed} \\ 0 & \text{otherwise.} \end{cases}$$

- Nonignorable nonresponse

$$f(y \mid \mathbf{x}) \neq f(y \mid \mathbf{x}, \delta = 1).$$

- In general,

$$f(y \mid \mathbf{x}, \delta = 1) = \frac{P(\delta = 1 \mid \mathbf{x}, y)}{P(\delta = 1 \mid \mathbf{x})} f(y \mid \mathbf{x}).$$

Thus, $P(\delta = 1 \mid \mathbf{x}, y) \neq P(\delta = 1 \mid \mathbf{x})$ implies nonignorable nonresponse.

Observed likelihood

- $f(y \mid \mathbf{x}; \theta)$: model of y on $\mathbf{x}$
- $g(\delta \mid \mathbf{x}, y; \phi)$: model of δ on $(\mathbf{x}, y)$
- Observed likelihood

$$
\begin{aligned}
L_{obs}(\theta, \phi) &= \prod_{\delta_i = 1} f\left(y_i \mid \mathbf{x}_i; \theta\right) g\left(\delta_i \mid \mathbf{x}_i, y_i; \phi\right) \\
&\quad \times \prod_{\delta_i = 0} \int f\left(y_i \mid \mathbf{x}_i; \theta\right) g\left(\delta_i \mid \mathbf{x}_i, y_i; \phi\right) dy_i
\end{aligned}
$$

- Under what conditions are the parameters identifiable (or estimable)?

Lemma

Suppose that we can decompose the covariate vector $\mathbf{x} = (\mathbf{u}, \mathbf{z})$ such that

$$g(\delta|y, \mathbf{x}) = g(\delta|y, \mathbf{u}) \tag{1}$$

and, for any given $\mathbf{u}$, there exist $z_{\mathbf{u},1}$ and $z_{\mathbf{u},2}$ such that

$$f(y|\mathbf{u}, \mathbf{z} = z_{\mathbf{u},1}) \neq f(y|\mathbf{u}, \mathbf{z} = z_{\mathbf{u},2}). \tag{2}$$

Under some other minor conditions, all the parameters in f and g are identifiable.

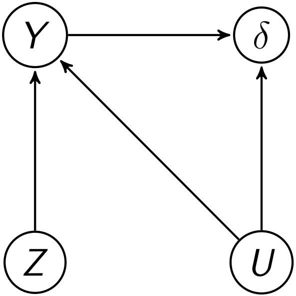
Remark

- Condition (1) means

$$\delta \perp \mathbf{z} \mid y, \mathbf{u}.$$

- That is, given $(y, \mathbf{u})$, $\mathbf{z}$ does not help in explaining δ.

Figure: A DAG for understanding nonresponse instrumental variable Z



- We may call $\mathbf{z}$ the nonresponse instrument variable.
- Rigorous theory developed by Wang et al. (2014).

Remark

- MCAR (Missing Completely at random): $P(\boldsymbol{\delta} \mid \mathbf{y})$ does not depend on $\mathbf{y}$.
- MAR (Missing at random): $P(\boldsymbol{\delta} \mid \mathbf{y}) = P(\boldsymbol{\delta} \mid \mathbf{y}_{obs})$
- NMAR (Not Missing at random): $P(\boldsymbol{\delta} \mid \mathbf{y}) \neq P(\boldsymbol{\delta} \mid \mathbf{y}_{obs})$
- Thus, MCAR is a special case of MAR.

Parameter estimation under nonresponse instrument variable

- Full likelihood-based ML estimation
- Generalized method of moment (GMM) approach (Section 6.3 of KS)
- Conditional likelihood approach (Section 6.2 of KS)
- Pseudo likelihood approach (Section 6.4 of KS)
- Exponential tilting method (Section 6.5 of KS)
- Latent variable approach (Section 6.6 of KS)

Reference

Kim, J.K. and Shao, J. (2013). "Statistical Methods for Handling Incomplete Data", Chapman & Hall / CRC.

Full likelihood-based ML estimation

- Wish to find $\hat{\eta} = (\hat{\theta}, \hat{\phi})$, that maximizes the observed likelihood

$$
\begin{aligned}
L_{obs}(\eta) &= \prod_{\delta_i = 1} f(y_i \mid \mathbf{x}_i; \theta)\, g(\delta_i \mid \mathbf{x}_i, y_i; \phi) \\
&\quad \times \prod_{\delta_i = 0} \int f(y_i \mid \mathbf{x}_i; \theta)\, g(\delta_i \mid \mathbf{x}_i, y_i; \phi)\, dy_i
\end{aligned}
$$

- Mean score theorem: Under some regularity conditions, finding the MLE by maximizing the observed likelihood is equivalent to finding the solution to

$$\bar{S}(\eta) \equiv E\{S(\eta) \mid \mathbf{y}_{obs}, \boldsymbol{\delta}; \eta\} = 0,$$

where $\mathbf{y}_{obs}$ is the observed data. The conditional expectation of the score function is called mean score function.

EM algorithm

- Interested in finding $\hat{\eta}$ that maximizes $L_{obs}(\eta)$. The MLE can be obtained by solving $S_{obs}(\eta) = 0$, which is equivalent to solving $\bar{S}(\eta) = 0$ by the mean score theorem.
- EM algorithm provides an alternative method of solving $\bar{S}(\eta) = 0$ by writing

$$\bar{S}(\eta) = E\left\{ S(\eta) \mid \mathbf{y}_{obs}, \boldsymbol{\delta}; \eta \right\}$$

and using the following iterative method:

$$\hat{\eta}^{(t+1)} \leftarrow \text{ solve } E\left\{ S(\eta) \mid \mathbf{y}_{obs}, \boldsymbol{\delta}; \hat{\eta}^{(t)} \right\} = 0.$$

EM algorithm

Definition

Let $\eta^{(t)}$ be the current value of the parameter estimate of η. The EM algorithm can be defined as iteratively carrying out the following E-step and M-steps:

- E-step: Compute

$$Q\left(\eta \mid \eta^{(t)}\right) = E\left\{\ln f\left(\mathbf{y}, \boldsymbol{\delta}; \eta\right) \mid \mathbf{y}_{\rm obs}, \boldsymbol{\delta}, \eta^{(t)}\right\}$$

- M-step: Find $\eta^{(t+1)}$ that maximizes $Q(\eta \mid \eta^{(t)})$ w.r.t. η.

Monte Carlo EM

Motivation: Monte Carlo samples in the EM algorithm can be used as imputed values.

Monte Carlo EM

1. In the EM algorithm defined by
   - [E-step] Compute

$$Q\left(\eta \mid \eta^{(t)}\right) = E\left\{\ln f\left(\mathbf{y}, \boldsymbol{\delta}; \eta\right) \mid \mathbf{y}_{\rm obs}, \boldsymbol{\delta}; \eta^{(t)}\right\}$$

   - [M-step] Find $\eta^{(t+1)}$ that maximizes $Q\left(\eta \mid \eta^{(t)}\right)$,

   E-step is computationally cumbersome because it involves integral.
2. Wei and Tanner (1990): In the E-step, first draw

$$\mathbf{y}_{mis}^{*(1)}, \cdots, \mathbf{y}_{mis}^{*(m)} \sim f\left(\mathbf{y}_{mis} \mid \mathbf{y}_{\rm obs}, \boldsymbol{\delta}; \eta^{(t)}\right)$$

   and approximate

$$Q\left(\eta \mid \eta^{(t)}\right) \cong \frac{1}{m}\sum_{j=1}^{m} \ln f\left(\mathbf{y}_{obs}, \mathbf{y}_{mis}^{*(j)}, \boldsymbol{\delta}; \eta\right).$$

Monte Carlo EM

Example 1

- Suppose that
$$y_i \sim f\left(y_i \mid x_i; \theta\right)$$
Assume that x_i is always observed but we observe y_i only when $\delta_i = 1$ where $\delta_i \sim Bernoulli\left[\pi_i\left(\phi\right)\right]$ and
$$\pi_i\left(\phi\right) = \frac{\exp\left(\phi_0 + \phi_1 x_i + \phi_2 y_i\right)}{1 + \exp\left(\phi_0 + \phi_1 x_i + \phi_2 y_i\right)}.$$
- To implement the MCEM method, in the E-step, we need to generate samples from
$$f(y_i \mid x_i, \delta_i = 0; \hat{\theta}, \hat{\phi}) = \frac{f(y_i \mid x_i; \hat{\theta})\{1 - \pi_i(\hat{\phi})\}}{\int f(y_i \mid x_i; \hat{\theta})\{1 - \pi_i(\hat{\phi})\} dy_i}.$$

Monte Carlo EM

Example 1 (Cont'd)

- We can use the following rejection method to generate samples from $f(y_i \mid x_i, \delta_i = 0; \hat{\theta}, \hat{\phi})$:
  1. Generate y_i^* from $f(y_i \mid x_i; \hat{\theta})$.
  2. Using y_i^*, compute

$$\pi_i^*(\hat{\phi}) = \frac{\exp(\hat{\phi}_0 + \hat{\phi}_1 x_i + \hat{\phi}_2 y_i^*)}{1 + \exp(\hat{\phi}_0 + \hat{\phi}_1 x_i + \hat{\phi}_2 y_i^*)}.$$

     Accept y_i^* with probability $1 - \pi_i^*(\hat{\phi})$.
  3. If y_i^* is not accepted, then goto Step 1.

Monte Carlo EM

Example 1 (Cont'd)

- Using the m imputed values of y_i, denoted by $y_i^{*(1)}, \cdots, y_i^{*(m)}$, and the M-step can be implemented by solving

$$\sum_{i=1}^{n}\sum_{j=1}^{m} S\left(\theta; x_i, y_i^{*(j)}\right) = 0$$

and

$$\sum_{i=1}^{n}\sum_{j=1}^{m} \left\{\delta_i - \pi(\phi; x_i, y_i^{*(j)})\right\}\left(1, x_i, y_i^{*(j)}\right) = 0,$$

where $S(\theta; x_i, y_i) = \partial \log f(y_i \mid x_i; \theta)/\partial\theta$.

Remark

- Identifiability condition is needed to guarantee the convergence of EM sequence.
- The fully parametric model approach is known to be sensitive to the failure of model assumptions: Little (1985), Kenward and Molenberghs (1988).
- Sensitivity analysis is often recommended: Scharfstein et al. (1999).

Partial Likelihood approach

- A classical likelihood-based approach for parameter estimation under non ignorable nonresponse is to maximize $L_{obs}(\theta, \phi)$ with respect to (θ, ϕ), where

$$L_{obs}(\theta, \phi) = \prod_{\delta_i=1} f(y_i \mid \mathbf{x}_i; \theta) g(\delta_i \mid \mathbf{x}_i, y_i; \phi) \times \prod_{\delta_i=0} \int f(y_i \mid \mathbf{x}_i; \theta) g(\delta_i \mid \mathbf{x}_i, y_i; \phi) dy_i$$

- Such approach can be called full likelihood-based approach because it uses full information available in the observed data.
- On the other hand, partial likelihood-based approach (or conditional likelihood approach) uses a subset of the sample.

Conditional Likelihood approach

Idea

- Since
$$f(y \mid \mathbf{x})g(\delta \mid \mathbf{x}, y) = f_1(y \mid \mathbf{x}, \delta)g_1(\delta \mid \mathbf{x}),$$
for some f_1 and g_1, we can write
$$\begin{aligned} L_{obs}(\theta) &= \prod_{\delta_i=1} f_1\left(y_i \mid \mathbf{x}_i, \delta_i = 1\right) g_1\left(\delta_i \mid \mathbf{x}_i\right) \\ &\quad \times \prod_{\delta_i=0} \int f_1\left(y_i \mid \mathbf{x}_i, \delta_i = 0\right) g_1\left(\delta_i \mid \mathbf{x}_i\right) dy_i \\ &= \prod_{\delta_i=1} f_1\left(y_i \mid \mathbf{x}_i, \delta_i = 1\right) \times \prod_{i=1}^{n} g_1\left(\delta_i \mid \mathbf{x}_i\right). \end{aligned}$$

- The conditional likelihood is defined to be the first component:
$$L_c(\theta) = \prod_{\delta_i=1} f_1\left(y_i \mid \mathbf{x}_i, \delta_i = 1\right) = \prod_{\delta_i=1} \frac{f(y_i \mid \mathbf{x}_i; \theta)\pi(\mathbf{x}_i, y_i)}{\int f(y \mid \mathbf{x}_i; \theta)\pi(\mathbf{x}_i, y)dy},$$
where $\pi(\mathbf{x}, y_i) = Pr(\delta_i = 1 \mid x_i, y_i)$. Popular in biased sampling literature.

Pseudo Likelihood approach

Idea

- Consider bivariate (x_i, y_i) with density $f(y \mid x; \theta)h(x)$ where y_i are subject to missingness.
- We are interested in estimating θ.
- Suppose that $Pr(\delta = 1 \mid x, y)$ depends only on y. (i.e. x is nonresponse instrument)
- Note that $f(x \mid y, \delta) = f(x \mid y)$.
- Thus, we can consider the following conditional likelihood

$$L_c(\theta) = \prod_{\delta_i=1} f(x_i \mid y_i, \delta_i = 1) = \prod_{\delta_i=1} f(x_i \mid y_i).$$

- We can consider maximizing the pseudo likelihood

$$L_p(\theta) = \prod_{\delta_i=1} \frac{f(y_i \mid x_i; \theta)\hat{h}(x_i)}{\int f(y_i \mid x; \theta)\hat{h}(x)dx},$$

where $\hat{h}(x)$ is a consistent estimator of the marginal density of x.

Pseudo Likelihood approach

Idea

- We may use the empirical density in $\hat{h}(x)$. That is, $\hat{h}(x) = 1/n$ if $x = x_i$. In this case,

$$L_c(\theta) = \prod_{\delta_i = 1} \frac{f(y_i \mid x_i; \theta)}{\sum_{k=1}^{n} f(y_i \mid x_k; \theta)}.$$

- We can extend the idea to the case of $\mathbf{x} = (\mathbf{u}, \mathbf{z})$ where $\mathbf{z}$ is a nonresponse instrument. In this case, the conditional likelihood becomes

$$\prod_{i:\delta_i = 1} p(\mathbf{z}_i \mid y_i, \mathbf{u}_i) = \prod_{i:\delta_i = 1} \frac{f(y_i \mid \mathbf{u}_i, \mathbf{z}_i; \theta) p(\mathbf{z}_i | \mathbf{u}_i)}{\int f(y_i \mid \mathbf{u}_i, \mathbf{z}; \theta) p(\mathbf{z} | \mathbf{u}_i) d\mathbf{z}}. \tag{3}$$

Pseudo Likelihood approach

- Let $\hat{p}(\mathbf{z}|\mathbf{u})$ be an estimated conditional probability density of $\mathbf{z}$ given $\mathbf{u}$. Substituting this estimate into the likelihood in (3), we obtain the following pseudo likelihood:

$$\prod_{i:\delta_i=1} \frac{f(y_i \mid \mathbf{u}_i, \mathbf{z}_i; \theta)\hat{p}(\mathbf{z}_i|\mathbf{u}_i)}{\int f(y_i \mid \mathbf{u}_i, \mathbf{z}; \theta)\hat{p}(\mathbf{z}|\mathbf{u}_i)d\mathbf{z}}. \tag{4}$$

- The pseudo maximum likelihood estimator (PMLE) of θ, denoted by $\hat{\theta}_p$, can be obtained by solving

$$S_p(\theta; \hat{\alpha}) \equiv \sum_{\delta_i=1} [S(\theta; \mathbf{x}_i, y_i) - E\{S(\theta; \mathbf{u}_i, \mathbf{z}, y_i) \mid y_i, \mathbf{u}_i; \theta, \hat{\alpha}\}] = 0$$

for θ, where $S(\theta; \mathbf{x}, y) = S(\theta; \mathbf{u}, \mathbf{z}, y) = \partial \log f(y \mid \mathbf{x}; \theta)/\partial\theta$ and

$$E\{S(\theta; \mathbf{u}_i, \mathbf{z}, y_i) \mid y_i, \mathbf{u}_i; \theta, \hat{\alpha}\} = \frac{\int S(\theta; \mathbf{u}_i, \mathbf{z}, y_i) f(y_i \mid \mathbf{u}_i, \mathbf{z}; \theta) p(\mathbf{z} \mid \mathbf{u}_i; \hat{\alpha}) d\mathbf{z}}{\int f(y_i \mid \mathbf{u}_i, \mathbf{z}; \theta) p(\mathbf{z} \mid \mathbf{u}_i; \hat{\alpha}) d\mathbf{z}}.$$

Pseudo Likelihood approach

- The Fisher-scoring method for obtaining the PMLE is given by

$$\hat{\theta}_p^{(t+1)} = \hat{\theta}_p^{(t)} + \left\{ \mathcal{I}_p \left(\hat{\theta}^{(t)}, \hat{\alpha} \right) \right\}^{-1} S_p(\hat{\theta}^{(t)}, \hat{\alpha})$$

where

$$\mathcal{I}_p (\theta, \hat{\alpha}) = \sum_{\delta_i = 1} \left[E\{S(\theta; \mathbf{u}_i, \mathbf{z}, y_i)^{\otimes 2} \mid y_i, \mathbf{u}_i; \theta, \hat{\alpha}\} - E\{S(\theta; \mathbf{u}_i, \mathbf{z}, y_i) \mid y_i, \mathbf{u}_i; \theta, \hat{\alpha}\}^{\otimes 2} \right].$$

- First considered by Tang et al. (2003) and further developed by Zhao and Shao (2015).

Basic setup

- (X, Y): random variable
- θ: Defined by solving
$$E\{U(\theta; X, Y)\} = 0.$$
- y_i is subject to missingness
$$\delta_i = \begin{cases} 1 & \text{if } y_i \text{ responds} \\ 0 & \text{if } y_i \text{ is missing.} \end{cases}$$
- Want to find w_i such that the solution $\hat{\theta}_w$ to
$$\sum_{i=1}^{n} \delta_i w_i U(\theta; x_i, y_i) = 0$$
is consistent for θ.

Basic Setup

- Result 1: The choice of
$$w_i = \frac{1}{E(\delta_i \mid x_i, y_i)} \tag{5}$$
makes the resulting estimator $\hat{\theta}_w$ consistent.
- Result 2: If $\delta_i \sim \text{Bernoulli}(\pi_i)$, then using $w_i = 1/\pi_i$ also makes the resulting estimator consistent, but it is less efficient than $\hat{\theta}_w$ using w_i in (5).

Parameter estimation : GMM method

- Because **z** is a nonresponse instrumental variable, we may assume

$$P(\delta = 1 \mid \mathbf{x}, y) = \pi(\phi_0 + \phi_1 \mathbf{u} + \phi_2 y)$$

for some (ϕ_0, ϕ_1, ϕ_2).

- Kott and Chang (2010): Construct a set of estimating equations such as

$$\sum_{i=1}^{n} \left\{ \frac{\delta_i}{\pi(\phi_0 + \phi_1 \mathbf{u}_i + \phi_2 y_i)} - 1 \right\} (1, \mathbf{u}_i, \mathbf{z}_i) = 0$$

that are unbiased to zero.

- May have overidentified situation: Use the generalized method of moments (GMM).

GMM method

Example 2

- Suppose that we are interested in estimating the parameters in the regression model
$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i \tag{6}$$
where $E(e_i \mid \mathbf{x}_i) = 0$.
- Assume that y_i is subject to missingness and assume that
$$P(\delta_i = 1 \mid x_{1i}, x_{i2}, y_i) = \frac{\exp(\phi_0 + \phi_1 x_{1i} + \phi_2 y_i)}{1 + \exp(\phi_0 + \phi_1 x_{1i} + \phi_2 y_i)}.$$
Thus, x_{2i} is the nonresponse instrument variable in this setup.

Example 2 (Cont'd)

- A consistent estimator of ϕ can be obtained by solving

$$\hat{U}_2(\phi) \equiv \sum_{i=1}^{n} \left\{ \frac{\delta}{\pi(\phi; x_{1i}, y_i)} - 1 \right\} (1, x_{1i}, x_{2i}) = (0, 0, 0). \tag{7}$$

Roughly speaking, the solution to (7) exists almost surely if $E\{\partial \hat{U}_2(\phi)/\partial \phi\}$ is of full rank in the neighborhood of the true value of ϕ. If x_2 is vector, then (7) is overidentified and the solution to (7) does not exist. In the case, the GMM algorithm can be used.

- Finding the solution to $\hat{U}_2(\phi) = 0$ can be obtained by finding the minimizer of $Q(\phi) = \hat{U}_2(\phi)' \hat{U}_2(\phi)$ or $Q_W(\phi) = \hat{U}_2(\phi)' W \hat{U}_2(\phi)$ where $W = \{V(\hat{U}_2)\}^{-1}$.

Example 2 (Cont'd)

- Once the solution $\hat{\phi}$ to (7) is obtained, then a consistent estimator of $\beta = (\beta_0, \beta_1, \beta_2)$ can be obtained by solving

$$\hat{U}_1(\beta, \hat{\phi}) \equiv \sum_{i=1}^{n} \frac{\delta_i}{\hat{\pi}_i} \{y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i}\} (1, x_{1i}, x_{2i}) = (0, 0, 0) \quad (8)$$

for β.

Asymptotic Properties

- The asymptotic variance of $\hat{\beta}$ obtained from (8) with $\hat{\phi}$ computed from the GMM can be obtained by

$$V(\hat{\theta}) \cong \left(\Gamma_a' \Sigma_a^{-1} \Gamma_a\right)^{-1}$$

  where

$$\begin{aligned} \Gamma_a &= E\{\partial \hat{U}(\theta)/\partial\theta\} \\ \Sigma_a &= V(\hat{U}) \\ \hat{U} &= (\hat{U}_1', \hat{U}_2')' \end{aligned}$$

  and $\theta = (\beta, \phi)$.

- Rigorous theory developed by Wang et al. (2014).

Exponential tilting method

Motivation

- Parameter θ defined by $E\{U(\theta; X, Y)\} = 0$.
- We are interested in estimating θ from an expected estimating equation:

$$\sum_{i=1}^{n} [\delta_i U(\theta; \mathbf{x}_i, y_i) + (1 - \delta_i) E\{U(\theta; \mathbf{x}_i, Y) \mid \mathbf{x}_i, \delta_i = 0\}] = 0. \tag{9}$$

- The conditional expectation in (9) can be evaluated by using

$$f(y|\mathbf{x}, \delta = 0) = f(y|\mathbf{x}) \frac{P(\delta = 0|\mathbf{x}, y)}{E\{P(\delta = 0|\mathbf{x}, y)|\mathbf{x}\}} \tag{10}$$

which requires correct specification of $f(y \mid \mathbf{x}; \theta)$. Known to be sensitive to the choice of $f(y \mid \mathbf{x}; \theta)$.

Exponential tilting method

Idea

Instead of specifying a parametric model for $f(y \mid \mathbf{x})$, consider specifying a parametric model for $f(y \mid \mathbf{x}, \delta = 1)$, denoted by $f_1(y \mid \mathbf{x})$. In this case,

$$f_0\left(y_i \mid \mathbf{x}_i\right) = f_1\left(y_i \mid \mathbf{x}_i\right) \times \frac{O\left(\mathbf{x}_i, y_i\right)}{E\left\{O\left(\mathbf{x}_i, Y_i\right) \mid \mathbf{x}_i, \delta_i = 1\right\}}, \tag{11}$$

where $f_\delta\left(y_i \mid \mathbf{x}_i\right) = f\left(y_i \mid \mathbf{x}_i, \delta_i = \delta\right)$ and

$$O\left(\mathbf{x}_i, y_i\right) = \frac{Pr\left(\delta_i = 0 \mid \mathbf{x}_i, y_i\right)}{Pr\left(\delta_i = 1 \mid \mathbf{x}_i, y_i\right)} \tag{12}$$

is the conditional odds of nonresponse.

Remark

- If the response probability follows from a logistic regression model

$$\pi(\mathbf{x}_i, y_i) \equiv Pr(\delta_i = 1 \mid \mathbf{x}_i, y_i) = \frac{\exp\{g(\mathbf{x}_i) + \phi y_i\}}{1 + \exp\{g(\mathbf{x}_i) + \phi y_i\}}, \tag{13}$$

  where $g(\mathbf{x})$ is completely unspecified, the expression (11) can be simplified to

$$f_0(y_i \mid \mathbf{x}_i) = f_1(y_i \mid \mathbf{x}_i) \times \frac{\exp(\gamma y_i)}{E\{\exp(\gamma Y) \mid \mathbf{x}_i, \delta_i = 1\}}, \tag{14}$$

  where $\gamma = -\phi$ and $f_1(y \mid \mathbf{x})$ is the conditional density of y given $\mathbf{x}$ and $\delta = 1$.
- Model (14) states that the density for the nonrespondents is an exponential tilting of the density for the respondents. The parameter γ is the tilting parameter that determines the amount of departure from the ignorability of the response mechanism. If $\gamma = 0$, the the response mechanism is ignorable and $f_0(y|\mathbf{x}) = f_1(y|\mathbf{x})$.

Estimation of tilting parameter

- Sverchkov (2008) considered direct maximization of the observed likelihood for ϕ: Given a parametric model for $f_1(y \mid \mathbf{x})$ and $\pi(\mathbf{x}, y; \phi)$, find $\hat{\phi}$ that maximizes

$$l_{obs}(\phi) = \sum_{i=1}^{n} \delta_i \log \pi(\mathbf{x}_i, y_i; \phi) + \sum_{i=1}^{n} (1 - \delta_i) \log \int \{1 - \pi(\mathbf{x}_i, y; \phi)\} \hat{f}_1(y \mid \mathbf{x}_i) dy.$$

- Riddles et al. (2015) proposed an alternative computational tool that avoids computing the above integration using an EM-type algorithm.
- Semiparametric extension (Morikawa et al., 2015): Use a nonparametric density for $f_1(y \mid \mathbf{x})$.

A Toy Example: Categorical Data (All dichotomous)

Example (SRS, $n = 10$)

ID	Weight	x_1	x_2	y
1	0.1	1	0	1
2	0.1	1	1	1
3	0.1	0	1	M
4	0.1	1	0	0
5	0.1	0	1	1
6	0.1	1	0	M
7	0.1	0	1	M
8	0.1	1	0	0
9	0.1	0	0	0
10	0.1	1	1	0

M: Missing

A Toy Example (Cont'd)

Assume $P(\delta = 1 \mid x_1, x_2, y) = \pi(x_1, y)$

ID	Weight	x_1	x_2	y
1	0.1	1	0	1
2	0.1	1	1	1
3	$0.1 \cdot w_{3,0}$	0	1	0
	$0.1 \cdot w_{3,1}$	0	1	1
4	0.1	1	0	0
5	0.1	0	1	1

$$
\begin{aligned}
w_{3,j} &= \hat{P}(Y = j \mid X_1 = 0, X_2 = 1, \delta = 0) \\
&\propto \hat{P}(Y = j \mid X_1 = 0, X_2 = 1, \delta = 1)\frac{1 - \hat{\pi}(0, j)}{\hat{\pi}(0, j)}
\end{aligned}
$$

with

$$w_{3,0} + w_{3,1} = 1$$

A Toy Example (Cont'd)

ID	Weight	x_1	x_2	y
6	$0.1 \cdot w_{6,0}$	1	0	0
	$0.1 \cdot w_{6,1}$	1	0	1
7	$0.1 \cdot w_{7,0}$	0	1	0
	$0.1 \cdot w_{7,1}$	0	1	1
8	0.1	1	0	0
9	0.1	0	0	0
10	0.1	1	1	0

$$
\begin{aligned}
w_{6,j} &\propto \hat{P}(Y = j \mid X_1 = 1, X_2 = 0, \delta = 1)\frac{1-\hat{\pi}(1,j)}{\hat{\pi}(1,j)} \\
w_{7,j} &\propto \hat{P}(Y = j \mid X_1 = 0, X_2 = 1, \delta = 1)\frac{1-\hat{\pi}(0,j)}{\hat{\pi}(0,j)}
\end{aligned}
$$

with

$$
w_{6,0} + w_{6,1} = w_{7,0} + w_{7,1} = 1.
$$

Example (Cont'd)

- E-step: Compute the conditional probability using the estimated response probability $\hat{\pi}_{ab}$.
- M-step: Update the response probability using the fractional weights. For fully nonparametric model,

$$\hat{\pi}_{ab} = \frac{\sum_{\delta_i=1} I(x_{1i} = a, y_i = b)}{\sum_{\delta_i=1} I(x_{1i} = a, y_i = b) + \sum_{\delta_i=0} \sum_{j=0}^{1} w_{i,j} I(x_{1i} = a, y_{ij}^* = b)}$$

- The solution from the proposed method is $\hat{\pi}_{11} = 1$, $\hat{\pi}_{10} = 3/4$, $\hat{\pi}_{01} = 1/3$, $\hat{\pi}_{00} = 1$.

A Toy Example (Cont'd)

Example (Cont'd)

- The method can be viewed as a fractional imputation method of Kim (2011).
- On the other hand, GMM method is more close to nonresponse weighting adjustment.

A Toy Example (Cont'd)

Example GMM method

ID	Wgt 1	Wgt2	x_1	x_2	y
1	0.1	$0.1\hat{\pi}_{11}^{-1}$	1	0	1
2	0.1	$0.1\hat{\pi}_{11}^{-1}$	1	1	1
3	0.1	0.0	0	1	M
4	0.1	$0.1\hat{\pi}_{10}^{-1}$	1	0	0
5	0.1	$0.1\hat{\pi}_{01}^{-1}$	0	1	1
6	0.1	0.0	1	0	M
7	0.1	0.0	0	1	M
8	0.1	$0.1\hat{\pi}_{10}^{-1}$	1	0	0
9	0.1	$0.1\hat{\pi}_{00}^{-1}$	0	0	0
10	0.1	$0.1\hat{\pi}_{10}^{-1}$	1	1	0

M: Missing

A Toy Example (Cont'd)

- GMM method: Calibration equation

$$\sum_i \frac{\delta_i}{\hat{\pi}_i} I(x_{1i} = a, x_{2i} = b) = \sum_i I(x_{1i} = a, x_{2i} = b).$$

1. $X_1 = 1, X_2 = 1$: $\hat{\pi}_{11}^{-1} + \hat{\pi}_{10}^{-1} = 2$
2. $X_1 = 1, X_2 = 0$: $\hat{\pi}_{11}^{-1} + \hat{\pi}_{10}^{-1} + \hat{\pi}_{10}^{-1} = 4$
3. $X_1 = 0, X_2 = 1$: $\hat{\pi}_{01}^{-1} = 3$
4. $X_1 = 0, X_2 = 0$: $\hat{\pi}_{00}^{-1} = 1$.

- The solution of GMM method does not exist.

Concluding remarks

- Uses a model for the response probability.
- Parameter estimation for response model can be implemented using the idea of maximum likelihood method.
- Instrumental variable needed for identifiability of the response model.
- Likelihood-based approach vs GMM approach
- Less tools for model diagnostics or model validation

The end

REFERENCES

Kenward, M. G. and G. Molenberghs (1988). Likelihood based frequentist inference when data are missing at random. *Statistical Science 13*, 236–247.

Kim, J. K. (2011). Parametric fractional imputation for missing data analysis. *Biometrika 98*, 119–132.

Kim, J. K. and J. Shao (2013). *Statistical Methods for Handling Incomplete Data*. Chapman & Hall / CRC.

Kott, P. S. and T. Chang (2010). Using calibration weighting to adjust for nonignorable unit nonresponse. *Journal of the American Statistical Association 105*, 1265–1275.

Little, R. (1985). A note about models for selectivity bias. *Econometrica 53*, 1469–1474.

Morikawa, K., J. K. Kim, and Y. Kano (2015). Semiparametric inference under nonignorable nonresponse. Submitted.

Riddles, M. K., J. K. Kim, and J. Im (2015). Propensity score adjustment for nonignorable nonresponse. *Journal of Survey Statistics and Methodology*. Accepted for publication.

Scharfstein, D., A. Rotnizky, and J. M. Robins (1999). Adjusting for nonignorable dropout using semi-parametric models. *Journal of the American Statistical Association 94*, 1096–1146.

Sverchkov, M. (2008). A new approach to estimation of response probabilities when missing data are not missing at random. In *Proc. Survey Res. Meth. Sect.*, Washington, DC, pp. 867–874. American Statistical Association.

Tang, G., R. J. A. Little, and T. E. Raghunathan (2003). Analysis of multivariate missing data with nonignorable nonresponse. *Biometrika 90*, 747–764.

Wang, S., J. Shao, and J. K. Kim (2014). Identifiability and estimation in problems with nonignorable nonresponse. *Statistica Sinica 24*, 1097 – 1116.

Wei, G. C. and M. A. Tanner (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association 85*, 699–704.

Zhao, J. and J. Shao (2015). Semiparametric pseudo likelihoods in generalized linear models with nonignorable missing data. *Journal of the American Statistical Association*. In press.